Analysis of Curves as Data

by

Pat Slettehaugh

A thesis submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Science in the Department of Statistics.

Chapel Hill
1991

Approved by:

____

Advisor

____

Reader

____

Reader

PAT MARIE SLETTEHAUGH. Analysis of Curves as Data (Under the direction of J. Stephen Marron.)

ABSTRACT

Bandwidth choice plays a crucial role in the area of kernel density estimation. Because many choices are available, it is of interest to know any similarities or differences between them. In this thesis I use principal component analysis to investigate the variability in kernel density estimates which are based on different bandwidths.

I found that bandwidths which over smooth yield more random kernel density estimates, while bandwidths which under smooth yield kernel density estimates which vary in a systematic way. In comparing ISE and MISE, I found that the difference was not systematic, and in comparing IAE and ISE differences were due to under vs over smoothing effects.

ACKNOWLEDGEMENTS

First of all, I would like to thank my advisor, J. Stephen Marron, for his help and encouragement throughout this project. I would also like to thank my committee members, Norman Johnson and Willem van Zwet, who provided helpful comments concerning this thesis. In addition, they were very considerate with me as well as extremely flexible with their time.

I also thank the secretaries and members of the department who have assisted me in numerous ways, and have more or less been wonderful. Thanks also go to T.A.W., who has listened to me and often provided me with encouragement and motivation.

Finally, I would like to thank my friend and classmate, Amy Grady, who has been there with me through it all. She knows what she's done for me; without her, I might never have gotten this far.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

IAE	integrated absolute error
ISE	integrated squared error
LSCV	least squares cross-validation
MISE	mean integrated squared error
MISE / 2	(mean integrated squared error) / 2
MISE * 2	(mean integrated squared error) * 2
PC	principal component
PCA	principal component analysis
PC(j)	principal component j
PMPI	Park-Marron plug-in
SJPI	Sheather Jones plug-in

Chapter 1
Introduction

Curves are often the observed responses in experiments. The motivation for this work was our interest in the effects different fixed bandwidths and bandwidth selectors have on kernel density estimation. Kernel density estimation is a very useful tool for examining an unknown population's distribution structure; it can show structure that may be very difficult to see using classical methods. The bandwidth choice, or smoothing parameter, has a crucial effect on the kernel density estimation. The main idea of this paper is to study the variability in these random curves using principal component analysis (PCA) so that we can see any similarities or differences between the various bandwidths. In order to do so, we "digitized" the curves, and proceeded to treat them as vectors of data. By "digitized" we mean that instead of examining the kernel density estimates as curves, we examined their values at a certain number of fixed, equally spaced points. By keeping the space between points small, we are able to retain much of the original information on our curves. The idea of studying variability with PCA as applied to digitized curves has been proposed and previously investigated by others, including: Jones and Rice (1990), Rice and Silverman(1991), Besse and Ramsay(1986), Servain and Legler(1986), Ramsay(1982), and Winant *et al*(1975).

The goal of kernel density estimation is to estimate a probability density, f(x), using a random sample $X_1, \ldots, X_n$ from f. The kernel density estimator is given by:

$$\hat{f}_h = \frac{1}{n}\sum_{i=1}^{n} K_h\,(x - X_i)$$

where $K_h(\cdot) = \dfrac{K(\cdot/h)}{h}$ and K is the kernel. The K used here is always the standard normal density. The scale parameter h is called the bandwidth or smoothing parameter, and plays a crucial role in how the estimator performs (Silverman (1986)). As an example, Figure 1.1 shows 20 kernel density estimates based on 20 independent samples of size n=100 from the Gaussian distribution, using the fixed bandwidth which minimizes the error criterion MISE (mean integrated squared error). MISE represents the h to minimize:

$$\mathrm{MISE}(h) = \mathbb{E}\int[\hat{f}_h - f]^2\,dx.$$

The solid curve is the true density, and the dotted curves are the linearly interpolated versions of the digitized curves. The vertical dotted lines show the ordinates of the "digitized" points.

In this case, we used 61 points, which means we are able to study the kernel density estimates at the values -3.0, -2.9, -2.8, . . . , 2.9, 3.0. By linearly connecting the values at these points, we get a good idea as to how the kernel density estimate curves actually look.

In Chapter 2 we will discuss the PCA ideas and notation used throughout this paper. Chapter 3 will examine several simple examples (test cases) which we will use to build our intuition about the PCA. Chapter 4 will discuss the effects of under smoothing and over smoothing on density estimation. In Chapter 5 we will compare and contrast different optimal bandwidths. Finally, our last chapter will focus on the differences among different bandwidth selectors.

Chapter 2

Principal Component Analysis

PCA is a tool used for "reducing the dimensionality" of a dataset of high dimensional vectors. The goal of the analysis is to explain as much of the total variation in the data as possible with only a few factors (PC's). These factors explain the variability in a simple, additive way. The first principal component, PC(1), is the weighted linear combination of the p variables which accounts for the largest amount of total variation in the data. The second principal component, PC(2), is the weighted linear combination which accounts for the next largest amount of variability in the data; PC(2) is also orthogonal to PC(1) in $\mathbb{R}^p$. And similarly we have PC(3), PC(4), ... , PC(p), all orthogonal to each other in $\mathbb{R}^p$.

For our purposes, we will refer to our data as:

$$\underset{\sim}{X} = (\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots, \underset{\sim}{X}_m)$$

where m is the number of datasets (curves) we are studying, and each

$$\underset{\sim}{X}_j = (x_{1,j}, x_{2,j}, \ldots, x_{p,j})' \quad j = 1, \ldots, m,$$

where p is the number of variables each data set contains. (In our case, p is the number of points which we have digitized the curve into.) When carrying out our PCA, we used data which was corrected for the mean; ie we looked at $(\underset{\sim}{X} - \overline{\underset{\sim}{X}}\,\underset{\sim}{1}_m)$ instead of just $\underset{\sim}{X}$, where $\overline{\underset{\sim}{X}}_{(p \times 1)}$ is the component wise mean of the data, and $\underset{\sim}{1}_m$ is the $(1 \times m)$ vector of 1's. We define this mean as:

$$\overline{\underset{\sim}{X}} = (\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_p)' = \frac{1}{m} \underset{\sim}{X}\,\underset{\sim}{1}_m{}',$$

where

$$\overline{x}_i = \frac{1}{m} \sum_{j=1}^{m} x_{i,j} \quad i = 1, \ldots, p.$$

The mean, $\overline{X}$, of the kernel density estimates in Figure 1.1 looks the standard normal density. (In particular, the data we studied was based on m=100 datasets digitized into p=61 points.) Let $\mathbf{S}_{(p \times p)}$ denote the sample covariance matrix for $\underset{\sim}{X}$:

$$\mathbf{S} = \frac{1}{m-1} (\underset{\sim}{X} - \overline{\underset{\sim}{X}}\,\underset{\sim}{1}_m)(\underset{\sim}{X} - \overline{\underset{\sim}{X}}\,\underset{\sim}{1}_m)',$$

where the i , i' th entry is

$$\frac{1}{m-1} \sum_{j=1}^{m} (x_{i,j} - \overline{x}_i)(x_{i',j} - \overline{x}_{i'})$$

By the Spectral Decomposition Theorem, there exists a $p \times p$ orthogonal matrix Γ and a diagonal matrix $\Lambda = \mathrm{diag}[\lambda_1, \ldots, \lambda_p]$ (where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$), such that $\Gamma' \mathbb{S} \Gamma = \Lambda$. (We call $\lambda_1, \ldots, \lambda_p$ the eigenvalues.) The principal component transformation is then defined by

$$\underset{\sim}{Y} = \Gamma' (\underset{\sim}{X} - \bar{\underset{\sim}{X}} \underset{\sim}{1}_m),$$

and the ith component of $\underset{\sim}{Y} = (\underset{\sim}{Y}_1, \ldots, \underset{\sim}{Y}_p)'$ is called the ith principal component of $\underset{\sim}{X}$. We also define

$$\bar{\underset{\sim}{Y}} = (\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_p)' = \frac{1}{m} \underset{\sim}{Y} \underset{\sim}{1}_m{}'$$

(where

$$\bar{y}_i = \frac{1}{m} \sum_{j=1}^{m} y_{i,j} \quad i = 1, \ldots, p),$$

and

$$\mathrm{cov}(\underset{\sim}{Y}) = \frac{1}{m-1} (\underset{\sim}{Y} - \bar{\underset{\sim}{Y}} \underset{\sim}{1}_m)(\underset{\sim}{Y} - \bar{\underset{\sim}{Y}} \underset{\sim}{1}_m)' = \frac{1}{m-1} \underset{\sim}{Y} \underset{\sim}{Y}' = \Gamma' \mathbb{S} \Gamma = \Lambda,$$

(where the i, i' the entry is of the form

$$\frac{1}{m-1} \sum_{j=1}^{m} (y_{i,j} - \bar{y}_i)(y_{i',j} - \bar{y}_{i'}) = \frac{1}{m-1} \sum_{j=1}^{m} y_{i,j}\, y_{i',j} \;)$$

which yields the properties:

1) $\bar{\underset{\sim}{Y}} = \underset{\sim}{0}$,
2) $\mathrm{cov}(\underset{\sim}{Y}_i, \underset{\sim}{Y}_i') = \mathrm{var}(\underset{\sim}{Y}_i) = \lambda_i$,
3) $\mathrm{cov}(\underset{\sim}{Y}_i, \underset{\sim}{Y}_j') = 0, \; i \neq j$,
4) $\mathrm{var}(\underset{\sim}{Y}_1) \geq \ldots \geq \mathrm{var}(\underset{\sim}{Y}_p)$,
5) $\sum_{i=1}^{p} \mathrm{var}(\underset{\sim}{Y}_i) = \sum_{i=1}^{p} \lambda_i = \mathrm{tr}(\mathbb{S}) = \mathrm{var}(\underset{\sim}{X})$.

From 3) we know that the PC's are orthogonal to each other; 4) says that the first PC has the largest variance; 5) says the total variation of the original data is preserved, and that the ith component

accounts for $\frac{\lambda_i}{\sum_{i=1}^{p} \lambda_i}$ of the total variation in the original data. Hence, we know that λ_1 is the amount

of total variability explained by PC(1), with $\frac{\lambda_1}{\sum_{i=1}^{p} \lambda_i}$ the proportion of total variability explained by

PC(1). Similarly λ_2 is the amount of total variability explained by PC(2), with $\frac{\lambda_2}{\sum_{i=1}^{p} \lambda_i}$ as the

proportion of total variability explained by PC(2). The other eigenvalues correspond accordingly with the other PC's. Since the matrix Γ is orthogonal, the weights or coefficients (also called eigenvectors) in our PC transformation, denoted by

$$\underset{\sim}{\Gamma} = (\underset{\sim}{A}_1, \underset{\sim}{A}_2, \ldots, \underset{\sim}{A}_p),$$

with each

$$\underset{\sim}{A}_j = (a_{1,j}, a_{2,j}, \ldots, a_{p,j})' \quad j = 1, \ldots, p,$$

satisfy the property

$$\sum_{k=1}^{p} (a_{k,j})^2 = (\underset{\sim}{A}_j)' \underset{\sim}{A}_j = 1 \quad \text{for all j} = 1, \ldots, p.$$

These eigenvectors are "unit vectors" in the direction of the PC's. $\underset{\sim}{A}_1$ is a "unit vector" in the direction of PC(1); $\underset{\sim}{A}_2$ is in the direction of PC(2); . . . ; $\underset{\sim}{A}_p$ is in the direction of PC(p). The graphs in this paper will show these direction "unit vectors", not the PC's themselves. In our analysis we will be interested in various sums of squares, so we define them as follows:

$$SS_0 = \sum_{j=1}^{m} (\underset{\sim}{X}_j)' \underset{\sim}{X}_j \ ;$$

$$SS_1 = \sum_{j=1}^{m} (\underset{\sim}{X}_j - \underset{\sim}{\bar{X}})'(\underset{\sim}{X}_j - \underset{\sim}{\bar{X}})$$

$$= \sum_{i=1}^{p} \lambda_i \ ;$$

$$SS_2 = \sum_{j=1}^{m} [\underset{\sim}{X}_j - \underset{\sim}{\bar{X}} - PC(1)]'[\underset{\sim}{X}_j - \underset{\sim}{\bar{X}} - PC(1)] = \sum_{j=1}^{m} [\underset{\sim}{X}_j - \underset{\sim}{\bar{X}} - \underset{\sim}{A}_1'(\underset{\sim}{X}_j - \underset{\sim}{\bar{X}})]'[\underset{\sim}{X}_j - \underset{\sim}{\bar{X}} - \underset{\sim}{A}_1'(\underset{\sim}{X}_j - \underset{\sim}{\bar{X}})]$$

$$= \sum_{i=2}^{p} \lambda_i \ ;$$

$$SS_3 = \sum_{j=1}^{m} [\underset{\sim}{X}_j - \underset{\sim}{\bar{X}} - PC(1) - PC(2)]' [\underset{\sim}{X}_j - \underset{\sim}{\bar{X}} - PC(1) - PC(2)]$$

$$= \sum_{j=1}^{m} [\underset{\sim}{X}_j - \underset{\sim}{\bar{X}} - \underset{\sim}{A}_1'(\underset{\sim}{X}_j - \underset{\sim}{\bar{X}}) - \underset{\sim}{A}_2'(\underset{\sim}{X}_j - \underset{\sim}{\bar{X}})]' [\underset{\sim}{X}_j - \underset{\sim}{\bar{X}} - \underset{\sim}{A}_1'(\underset{\sim}{X}_j - \underset{\sim}{\bar{X}}) - \underset{\sim}{A}_2'(\underset{\sim}{X}_j - \underset{\sim}{\bar{X}})]$$

$$= \sum_{i=3}^{p} \lambda_i \ ;$$

$$SS_4 = \sum_{j=1}^{m} [\underset{\sim}{X}_j - \underset{\sim}{\bar{X}} - PC(1) - PC(2) - PC(3)]'[\underset{\sim}{X}_j - \underset{\sim}{\bar{X}} - PC(1) - PC(2) - PC(3)]$$

$$= \sum_{j=1}^{m} [\underset{\sim}{X}_j - \underset{\sim}{\bar{X}} - \underset{\sim}{A}_1'(\underset{\sim}{X}_j - \underset{\sim}{\bar{X}}) - \underset{\sim}{A}_2'(\underset{\sim}{X}_j - \underset{\sim}{\bar{X}}) - \underset{\sim}{A}_3'(\underset{\sim}{X}_j - \underset{\sim}{\bar{X}})]'$$

$$\times \ [\underset{\sim}{X}_j - \underset{\sim}{\bar{X}} - \underset{\sim}{A}_1'(\underset{\sim}{X}_j - \underset{\sim}{\bar{X}}) - \underset{\sim}{A}_2'(\underset{\sim}{X}_j - \underset{\sim}{\bar{X}}) - \underset{\sim}{A}_3'(\underset{\sim}{X}_j - \underset{\sim}{\bar{X}})]$$

$$= \sum_{i=4}^{p} \lambda_i \ ;$$

$$SS_5 = \sum_{j=1}^{m} [\underset{\sim}{X}_j - \overline{\underset{\sim}{X}} - PC(1) - PC(2) - PC(3) - PC(4)]'[\underset{\sim}{X}_j - \overline{\underset{\sim}{X}} - PC(1) - PC(2) - PC(3) - PC(4)]$$

$$= \sum_{j=1}^{m} [\underset{\sim}{X}_j - \overline{\underset{\sim}{X}} - \underset{\sim}{A}_1'(\underset{\sim}{X}_j - \overline{\underset{\sim}{X}}) - \underset{\sim}{A}_2'(\underset{\sim}{X}_j - \overline{\underset{\sim}{X}}) - \underset{\sim}{A}_3'(\underset{\sim}{X}_j - \overline{\underset{\sim}{X}}) - \underset{\sim}{A}_4'(\underset{\sim}{X}_j - \overline{\underset{\sim}{X}})]'$$

$$\times \ [\underset{\sim}{X}_j - \overline{\underset{\sim}{X}} - \underset{\sim}{A}_1'(\underset{\sim}{X}_j - \overline{\underset{\sim}{X}}) - \underset{\sim}{A}_2'(\underset{\sim}{X}_j - \overline{\underset{\sim}{X}}) - \underset{\sim}{A}_3'(\underset{\sim}{X}_j - \overline{\underset{\sim}{X}}) - \underset{\sim}{A}_4'(\underset{\sim}{X}_j - \overline{\underset{\sim}{X}})]$$

$$= \sum_{i=5}^{p} \lambda_i \ .$$

Chapter 3

Test Cases

In order to better understand the random perturbations in general curves, we examined numerous carefully selected test cases. In each case we generated a family of curves corresponding to "small" random perturbations and carried out Taylor expansions in the size of the perturbations. This gives useful interpretation because the eigenvectors show up as the coefficients in the expansion.

Suppose each $\underset{\sim}{X}_j$, j = 1, . . . , d , has the form:

$$\underset{\sim}{X} = \underset{\sim}{a} + (\sigma Z)\, \underset{\sim}{b} + (\sigma Z)^2\, \underset{\sim}{c} + (\sigma Z)^3\, \underset{\sim}{d} + \ldots$$

where $\underset{\sim}{a}$, $\underset{\sim}{b}$, $\underset{\sim}{c}$, $\underset{\sim}{d}$, . . . are fixed; Z is a standard normal random variable, different realizations of which yield the family of curves; and σ is a constant which is "small". The idea here is to look at this function as $\sigma \to 0$.

Upon examining $\underset{\sim}{X}$ at the first level, we have:

$$\underset{\sim}{X} = \underset{\sim}{a} + \text{"small noise"}.$$

In other words, $\underset{\sim}{a}$ is representing most of the "mean", $\underset{\sim}{\overline{X}}$. In this paper, we rescale the mean so it is a

"unit vector" , and denote this rescaled mean by $\underset{\sim}{A}_0$ (like the eigenvectors, $\underset{\sim}{A}_0$ is a "unit vector" which gives the direction of the mean). Thus, we have:

$$\underset{\sim}{A}_0 = \frac{\underset{\sim}{\overline{X}}}{\sqrt{\underset{\sim}{\overline{X}}'\underset{\sim}{\overline{X}}}}, \text{ and } \underset{\sim}{A}_0{}' \underset{\sim}{A}_0 = 1.$$

At the second level we want to examine structure in the data beyond the mean, so we look at:

$$(\underset{\sim}{X} - \underset{\sim}{a}) = (\sigma Z)\, \underset{\sim}{b} + \text{"small noise"}.$$

The coefficient here is $\underset{\sim}{b}$, which gives us the direction of the first principal component. In other words, the direction of maximum change is $\underset{\sim}{b}$. We rescale $\underset{\sim}{b}$ so that it is a "unit vector" like $\underset{\sim}{A}_1$; we denote this "unit vector" by

$$\underset{\sim}{b}(u) = \frac{\underset{\sim}{b}}{\sqrt{\underset{\sim}{b}'\underset{\sim}{b}}}.$$

From the covariance matrix we get λ_1, the amount of total variability explained by PC(1).

At the third level we examine the remaining structure in the population, the dominant part of which is:

$$(\underset{\sim}{X} - \underset{\sim}{a} - (\sigma Z)\, \underset{\sim}{b}\,) = (\sigma Z)^2\, \underset{\sim}{c} + \text{"small noise"}.$$

Here the coefficient is $\underset{\sim}{c}$, which will determine the direction of PC(2). We rescale $\underset{\sim}{c}$ so that it is a "unit vector",

$$\underset{\sim}{c}\,(u) = \frac{\underset{\sim}{c}}{\sqrt{\underset{\sim}{c}\,'\,\underset{\sim}{c}}},$$

and we get λ_2 from the covariance matrix. Here the orthogonality issue needs to be addressed. It is sometimes the case that the Taylor expansion coefficients are nearly, but not quite, orthogonal to each other. However, we know that the PC's must be orthogonal. In order to satisfy this requirement, the Gram-Schmidt Orthogonalization Process is used. The result of this process is

$$\underset{\sim}{A}_2 = \frac{\underset{\sim}{c} - (\underset{\sim}{c} \cdot \underset{\sim}{A}_1)\, \underset{\sim}{A}_1}{\sqrt{[\underset{\sim}{c} - (\underset{\sim}{c} \cdot \underset{\sim}{A}_1)\, \underset{\sim}{A}_1]'\, [\underset{\sim}{c} - (\underset{\sim}{c} \cdot \underset{\sim}{A}_1)\, \underset{\sim}{A}_1]}}.$$

Note that $\underset{\sim}{A}_2$ is a "unit vector" in the direction of PC(2).

At the fourth level we have:

$$(\underset{\sim}{X} - \underset{\sim}{a} - (\sigma Z)\, \underset{\sim}{b} - (\sigma Z)^2\, \underset{\sim}{c}\,) = (\sigma Z)^3\, \underset{\sim}{d} + \text{"small noise"}.$$

Here we find that PC(3) is in the direction of $\underset{\sim}{d}$. As before, $\underset{\sim}{d}$ is rescaled so that it is a "unit vector",

$$\underset{\sim}{d}\,(u) = \frac{\underset{\sim}{d}}{\sqrt{\underset{\sim}{d}\,'\,\underset{\sim}{d}}},$$

and the corresponding eigenvalue is λ_3. Again, the Gram-Schmidt Orthogonalization Process is used to ensure that the PC's are all orthogonal to each other. Similarly, we can find the direction of all remaining PC's.

In all test cases we calculated and graphed $\underset{\sim}{A}_0$ - $\underset{\sim}{A}_5$ vs each corresponding "unit vector" based on the Taylor expansion coefficient. On the graphs we refer to this "unit vector" as the "Raw Coeff" curve (for raw coefficient curve). This "Raw Coeff" curve does not always look like the corresponding eigenvector. To explain the discrepancy, we need to account for orthogonality. The "Theory" curve on the graphs refers to the orthogonalized version of the "Raw Coeff" curve. Note that for all graphs showing the direction of PC(1) the "Raw Coeff" curve and the "Theory" curve are the same thing, so we merely label it "Theory". This is because we use the first Taylor expansion coefficient as our starting point; we make all the other coefficients orthogonal to this one. Due to the large amount of graphs we produced, we will show only a sample of them, and merely comment on the others.

3.1 Test Case 1

The first test case we examined was on curves of the form:

$$f_{\sigma,z}(x) = \phi(x+\sigma Z)\ ,$$

where $\sigma = .15$, Z is as defined above, and ϕ is the standard normal density (See Figure 3.1.1). Note that the effect of adding σZ inside ϕ is a "horizontal shift" in the test curves. Carrying out the Taylor expansion on this function gives:

$$f_{\sigma,z}(x) = \phi(x) + (\sigma Z)\phi'(x) + \frac{(\sigma Z)^2}{2}\phi''(x) + \frac{(\sigma Z)^3}{6}\phi^{(3)}(x) + \frac{(\sigma Z)^4}{24}\phi^{(4)}(x) + \ldots$$

Examining the first level we have:

$$f_{\sigma,z}(x) = \phi(x) + R,$$

where $R = O_p(\sigma)$, and has nearly mean 0. This suggests that PC(0), the "average direction", should be in the direction of $\phi(x)$. We found that this was indeed the case. $\underset{\sim}{A}_0$ was ≥ 0, looked bell-shaped, was symmetric, and coincided almost exactly with the curve $\dfrac{\phi(x)}{\sqrt{\phi(x)'\phi(x)}}$.

At the second level we have:

$$[f_{\sigma,z}(x) - \phi(x)] = (\sigma Z)\phi'(x) + O_p(\sigma^2).$$

The coefficient here is $\phi'(x)$, which suggests that PC(1) may go in this direction. After rescaling $\phi'(x)$ so that it becomes a "unit vector", Figure 3.1.2 shows that this is the case. To the right of the graph we find that PC(1) accounts for approximately 98% of the total variability present. Notice that $\underset{\sim}{A}_1$ is not exactly $\phi'(x)$, probably due to the random variability from the simulation.

At the third level we have:

$$[f_{\sigma,z}(x) - \phi(x) - (\sigma Z)\phi'(x)] = \frac{(\sigma Z)^2}{2}\phi''(x) + O_p(\sigma^3).$$

In this situation, the dot product $\underset{\sim}{b}' \cdot \underset{\sim}{c}$ is a Riemann sum for $m\int \phi'(x)\phi''(x)dx = 0$ (by integration by parts), which would imply that $\phi'(x)$ and $\phi''(x)$ are orthogonal to each other. Since this integral is 0, the Taylor expansion coefficients are nearly orthogonal to each other. Here we find that PC(2) should be in the direction of $\phi''(x)$, and we find this to be true (see Figure 3.1.3). Notice in Table 3.1.1 that together PC(1) and PC(2) explain nearly all of the variability present. This can also be seen by examining the sums of squares in Table 3.1.2. Again, the direction of PC(2) (as shown by $\underset{\sim}{A}_2$) and $\phi''(x)$ differ mostly due to Monte Carlo variability.

At the fourth level we have:

$$[f_{\sigma,z}(x) - \phi(x) - (\sigma Z)\phi(x) - \frac{(\sigma Z)^2}{2}\phi''(x)] = \frac{(\sigma Z)^3}{6}\phi^{(3)}(x) + O_p(\sigma^4).$$

Figure 3.1.4 shows us that PC(3) is in the direction of $\phi^{(3)}(x)$, as it should be. However, here $\underset{\sim}{A}_3$ and the "Raw Coeff" curve have departed from each other in a significant manner. As was mentioned prior to this section, this is due to a lack of orthogonality. Using the Gram-

Schmidt process we get the orthogonalized version, rescale it so it is a "unit vector", and find that it is nearly identical to $\underset{\sim}{A}_3$. It is shown as the "Theory" curve on the graph.

To get the orthogonalized version, we subtracted off the component which is in the direction of $\underset{\sim}{A}_1$. Thus, we want to find:

$$\underset{\sim}{d} - (\underset{\sim}{d} \cdot \underset{\sim}{A}_1)\, \underset{\sim}{A}_1$$

which, since $\underset{\sim}{d}$ in this case is $\phi^{(3)}(x)$ and $\underset{\sim}{A}_1$ is $\dfrac{\phi'(x)}{\sqrt{\phi'(x) \cdot \phi'(x)}}$, is the same as

$$\phi^{(3)}(x) - [\phi^{(3)}(x) \cdot \frac{\phi'(x)}{\sqrt{\phi'(x) \cdot \phi'(x)}}]\, \frac{\phi'(x)}{\sqrt{\phi'(x) \cdot \phi'(x)}} = \phi^{(3)}(x) - [\, \frac{\phi^{(3)}(x) \cdot \phi'(x)}{\phi'(x) \cdot \phi'(x)} \,]\, \phi'(x).$$

Here the dot product $\phi^{(3)}(x) \cdot \phi'(x)$ is a Riemann sum for

$m \int \phi^{(3)}(x)\, \phi'(x)\, dx = \frac{-3m}{2^3\sqrt{\pi}}$ (Corollary 4.6 Aldershof *et al* (1991)),

and the dot product $\phi'(x) \cdot \phi'(x)$ is a Riemann sum for

$m \int \phi'(x) \cdot \phi'(x)\, dx = \frac{m}{2^{\,2}\sqrt{\pi}}$ (Corollary 4.7 Aldershof *et al* (1991)).

Now our vector of interest can be rewritten as

$$\phi^{(3)}(x) - [\, \tfrac{-3}{2} \,]\, \phi'(x) = -(x^3 - 3x)\, \phi(x) - [\, \tfrac{3}{2} \,]\, x\, \phi(x) = (\, \tfrac{3}{2} x - x^3)\, \phi(x)\,.$$

We rescale this vector so that it is a "unit vector", and find that our "Theory" curve is:

$$\frac{(\, \frac{3}{2} x - x^3)\, \phi(x)}{\sqrt{[(\, \frac{3}{2} x - x^3)\, \phi(x)]'\, [(\, \frac{3}{2} x - x^3)\, \phi(x)]}}\,.$$

At the fifth level we have:

$$[f_{\sigma,z}(x) - \phi(x) - (\sigma Z)\phi'(x) - \frac{(\sigma Z)^2}{2}\, \phi''(x) - \frac{(\sigma Z)^3}{6}\, \phi^{(3)}(x)] = \frac{(\sigma Z)^4}{24}\, \phi^{(4)}(x) + \mathcal{O}_p(\sigma^5).$$

Here the difference between the "Raw Coeff" curve and the eigenvector, $\underset{\sim}{A}_4$, is even more pronounced than we found at the fourth level. As was discussed just above, we need to use the Gram-Schmidt orthogonalization process so that we can find our "Theory" curve.

Table 3.1.1

Percent Variability Explained

	PC(1)	PC(2)	PC(3)	PC(4)
INDIVIDUAL	98.35765	1.63682	0.00551	0.00002
CUMULATIVE	98.35765	99.99447	99.99998	100.00000

Table 3.1.2

Sums of Squares

SS_0	SS_1	SS_2	SS_3	SS_4	SS_5
282.08781	3.28132	0.05389	0.00018	0.00000	0.00000

The initial sum of squares, SS_0, is quite large; however, it reduces rather quickly when the mean is taken out. From this we can see that the mean explains a lot of the overall structure in the dataset of curves. Recall, however, that we conducted our PCA on the data ($\underset{\sim}{X}$ - $\overline{\underset{\sim}{X}}\,\underset{\sim}{1}_m$), so we in fact should consider the initial variability to be SS_1. Once the mean is subtracted off from the data, most of the remaining variability is explained by PC(1). This can be seen both by the drastic reduction in the sum of squares from SS_1 to SS_2, and in the percent variability explained by PC(1), about 98%, as found in Table 3.1.1. PC(2) helps explain some variability, but it is minimal when compared to PC(1). PC(3)-PC(4) also help explain a percentage of the variability, but they are quite insignificant compared to PC(1), or even to PC(2). The sums of squares are nearly zero even before these last two PC's are subtracted off. The horizontal shift in the test curves of Figure 3.1.1 produces simultaneous vertical changes at our collection of "digitized" points. When the curves are shifted horizontally from right to left (analogously, we could have looked from left to right), the effects on the "digitized" points are as follows: in the right tail there is a small downward change; in the left tail there is a small upward change; at the peak there is a small downward change; between the peak and the right tail the change is downward, but of a greater magnitude than in either the tail or peak; a similar, but upward, change is found between the peak and the left tail. The magnitude of these changes is largest near the points of greatest slope (inflection points), and smallest near the points of little to no slope (the tails and peak). PC(1) is large because it measures the type of variability just described. PC(2) also contributes, but we did not analyze the type of variability it is explaining.

3.2 Test Case 2

This case was based on curves of the form:

$$f_{\sigma,z}(x) = \phi(x) + (\sigma Z),$$

where $\sigma = .02$. From Figure 3.2.1 we can see that the effect of adding σZ is a "vertical shift" in the data (ie the perturbations in these curves are in a vertical direction). This is a simpler example than test case 1 because the Taylor expansion has exactly one nonzero term. We made the picture, and found that $\underset{\sim}{A}_0$ is bell shaped, and falls almost exactly along the "unit vector", $\frac{\phi(x)}{\sqrt{\phi(x)'\phi(x)}}$. Figure 3.2.2 shows that PC(1) is in the direction of a constant vector (since the coefficient of σZ is the constant vector $\underset{\sim}{1}$). There are no other coefficients; all remaining PC's are just roundoff error. For a typical example, see Figure 3.2.3.

Table 3.2.1

Percent Variability Explained

	PC(1)	PC(2)	PC(3)	PC(4)
INDIVIDUAL	100.00000	0.00000	0.00000	0.00000
CUMULATIVE	100.00000	100.00000	100.00000	100.00000

Table 3.2.2

Sums of Squares

SS_0	SS_1	SS_2	SS_3	SS_4	SS_5
289.85644	2.56801	0.00000	0.00000	0.00000	0.00000

Once again SS_0 is quite large, but subtracting the mean significantly reduces it. Here our initial variability, SS_1, is less than it was in test case 1. We find that PC(1) explains more or less everything: it explains 100% of the total variability, and, when subtracted from the data along with the mean, accounts for all the sum of squares. This situation will be common whenever the Taylor expansion has exactly one nonzero term, as will be seen in further test cases. Here we examined the effects (on the "digitized" points) of shifting the test curves vertically from top to bottom. Visually it appears that the magnitude of change is greatest in the peak and tails, and least at the points of inflection. However, this is not the case. At each "digitized" point there is a constant change in the vertical direction as the test curves are shifted.

3.3 Test Case 3

Curves in this case were of the form:

$$f_{\sigma,z}(x) = \phi(x) + (\sigma Z)x,$$

where $\sigma = .03$. A linear change has been brought to the curve, as can be seen in Figure 3.3.1. Again, this is a simpler case where only a one term Taylor expansion is needed. At the first level we find that PC(0) is in the direction of $\phi(x)$. The curves are nearly the same, except

$\underset{\sim}{A}_0$ slightly underestimates the "Raw Coeff" vector at one end, and slightly overestimates it at the other. At the second level we find that PC(1) is in the direction of a line since the coefficient is x. Figure 3.3.2 shows that this is the case. As in the previous test case, all higher order PC's are merely roundoff error.

Table 3.3.1

Percent Variability Explained

	PC(1)	PC(2)	PC(3)	PC(4)
INDIVIDUAL	100.00000	0.00000	0.00000	0.00000
CUMULATIVE	100.00000	100.00000	100.00000	100.00000

Table 3.3.2

Sums of Squares

SS_0	SS_1	SS_2	SS_3	SS_4	SS_5
300.28638	17.91184	0.00000	0.00000	0.00000	0.00000

As in the previous case, we find that PC(1) does all the explaining once the mean is removed. SS_1 here is much larger than it was in either of the first two test cases, and is due to the increased variability found in the tails of the test curves (see Figure 3.3.1). The effects of this linear change on the"digitized" points as the curves are rocked from right to left are: a downward change which decreases in magnitude as we move from the right tail to the peak, and an upward change which increases in magnitude as we move from the peak to the left tail.

3.4 Test Case 4

In this case,

$$f_{\sigma,z}(x) = \phi(x) + (\sigma Z)\phi(x),$$

with $\sigma = .05$. The effect here is that the curves have been vertically scaled (see Figure 3.4.1). Once again the Taylor expansion has exactly one nonzero term. $\underset{\sim}{A}_0$ looks as it should; it is impossible to distinguish between it and the "Raw Coeff" curve. PC(1) is in the direction of $\phi(x)$, as shown in Figure 3.4.2; these two curves are also impossible to distinguish between. All higher order PC's are again roundoff error.

Table 3.4.1

Percent Variability Explained

	PC(1)	PC(2)	PC(3)	PC(4)
INDIVIDUAL	100.00000	0.00000	0.00000	0.00000
CUMULATIVE	100.00000	100.00000	100.00000	100.00000

Table 3.4.2

Sums of Squares

SS_0	SS_1	SS_2	SS_3	SS_4	SS_5
286.48974	0.74222	0.00000	0.00000	0.00000	0.00000

The pattern from test cases 2 and 3 continues here. PC(1) accounts for all the variability present once the mean is removed. Here SS_1 is quite small, reflecting the fact that the curves in Figure 3.4.1 are not very variable; they are tight at both ends, and vary most in the middle. The effects at the "digitized" points of shifting the curves vertically from top to bottom are all downward changes. At the peak the change has the greatest magnitude; between the peak and the tails the magnitude of change tapers off, until there is virtually no change in the tails themselves. Once again, PC(1) does an excellent job of capturing the up and down variability of the curves.

3.5 Test Case 5

Curves are of the form:

$$f_{\sigma,z}(x) = \phi(x) + (\sigma Z)\, x\phi(x),$$

where $\sigma = .2$. In Figure 3.5.1 we can see the asymmetry that has been introduced. $\underset{\sim}{A}_0$ looks as it should, but with a slight overestimation of the "Raw Coeff" curve from x = 0 to x = 2, and a slight underestimation from x = -2 to x = 0. PC(1) is in the direction of $x\phi(x)$; it is hard to tell $\underset{\sim}{A}_1$ and the "Theory" curve apart (see Figure 3.5.2). All higher PC's are roundoff error. Notice the similarity between this $\underset{\sim}{A}_1$ and $\underset{\sim}{A}_1$ from the first test case (Figure 3.1.2). This is due to the fact that $\phi'(x) = -x\phi(x)$.

Table 3.5.1

Percent Variability Explained

	PC(1)	PC(2)	PC(3)	PC(4)
INDIVIDUAL	100.00000	0.00000	0.00000	0.00000
CUMULATIVE	100.00000	100.00000	100.00000	100.00000

Table 3.5.2

Sums of Squares

SS_0	SS_1	SS_2	SS_3	SS_4	SS_5
288.12046	5.93594	0.00000	0.00000	0.00000	0.00000

Again there is only one nonzero term in the expansion, so we find that PC(1) explains all variability beyond the mean. The initial sum of squares is fairly large with respect to SS_1's of the previous test cases. This reflects the greater variability present in the test curves. As these curves are rolled from right to left, we again have upward and downward changes, which explains why PC(1) captures the variability. In the left tail there is a fairly small downward change. This downward change increases in magnitude as we move left toward the inflection point, followed by a decrease in magnitude until we reach the peak. At the peak there is no downward change. A symmetric, but upward change occurs to the left of the peak.

3.6 Test Case 6

In this case we have:

$$f_{\sigma,z}(x) = \phi(x + (\sigma Z)x),$$

with $\sigma = .1$. Figure 3.6.1 illustrates this horizontal scale change. Carrying out the Taylor expansion yields:

$$f_{\sigma,z}(x) = \phi(x) + (\sigma Z)x\phi'(x) + \frac{(\sigma Z)^2}{2}x^2\phi''(x) + \frac{(\sigma Z)^3}{6}x^3\phi^{(3)}(x) + \ldots$$

$\underset{\sim}{A}_0$ looks as it should; PC(1) is in the direction of $x\phi'(x)$ ($= -x^2\phi(x)$), and is only slightly off in the tails (see Figure 3.6.2). All higher order PC's follow in the directions of their corresponding Taylor expansion coefficients, as we saw in test case 1.

Table 3.6.1

Percent Variability Explained

	PC(1)	PC(2)	PC(3)	PC(4)
INDIVIDUAL	98.85374	1.13854	0.00770	0.00003
CUMULATIVE	98.85374	99.99228	99.99998	100.00000

Table 3.6.2

Sums of Squares

SS_0	SS_1	SS_2	SS_3	SS_4	SS_5
281.38867	2.17457	0.02493	0.00017	0.00000	0.00000

As in test case 1, PC(1) no longer explains all the variability, but comes very close. SS_0 is large, but reduces significantly once the mean is removed. The sum of squares also reduce significantly after PC(1) is removed, and again after PC(2) is removed. PC(1) is again dominant because most of the changes in the test curves are in an up and down direction as the curves expand and contract. Examining the curves from greatest expansion to greatest contraction, we find a relatively small downward change at the "digitized" points in the tails. The magnitude of this downward change increases quickly until we reach inflection points, and is followed by a slower decrease in magnitude until we reach the peak, where there is no change. PC(2) helps to explain most of the remaining variability.

3.7 Test Case 7

Here the curves are of the form:

$$f_{\sigma,z}(x) = \phi(x + (\sigma Z)x)(1 + (\sigma Z)),$$

where $\sigma = .05$. The curves have been scaled inside and out so that they are more variable in the tails and in the peak, as is seen in Figure 3.7.1. Note that each realization of $f_{\sigma,z}(x)$ is a probability density. Taylor expansion gives:

$$f_{\sigma,z}(x) = \phi(x) + (\sigma Z)(1 - x^2)\phi'(x) + \frac{(\sigma Z)^2}{2} x^2 (x^2 - 3)\phi''(x) + \ldots$$

$\underset{\sim}{A}_0$ looks as it should; PC(1) is almost exactly in the direction of $(1-x^2)\phi(x)$ (see Figure 3.7.2). The other PC's follow in the direction of their corresponding coefficients.

Table 3.7.1

Percent Variability Explained

	PC(1)	PC(2)	PC(3)	PC(4)
INDIVIDUAL	99.70625	0.29315	0.00060	0.00000
CUMULATIVE	99.70625	99.99940	100.00000	100.00000

Table 3.7.2

Sums of Squares

SS_0	SS_1	SS_2	SS_3	SS_4	SS_5
283.91162	0.54990	0.00162	0.00000	0.00000	0.00000

SS_0 is drastically reduced by subtracting the mean, as is SS_1 by subtracting PC(1). Here PC(1) is even more dominant than in the previous case. Again, most of the changes at the "digitized" points are in the up and down motions we have discussed before. If we examine the curves from highest peak to lowest peak (a sort of "twanging" motion), we find a downward change between x=1 and x=-1 (zero points), and an upward change everywhere else. The magnitude of downward change is greatest at the peak, and decreases to zero at the zero points. The magnitude of upward change increases from zero at the zero points, and is followed by a decrease at the tails. The magnitude of the largest downward change is much greater than the magnitude of largest upward change. PC(2) helps explain most of the remaining variability.

3.8 Test Case 8

Curves are of the form:

$$f_{\sigma,z}(x) = \phi(x)\,(1 - \frac{(\sigma Z)}{4}) + (\sigma Z)\, x^2\, \phi(x) = \phi(x) + (\sigma Z)\,(x^2 - \tfrac{1}{4})\,\phi(x),$$

where σ = .1. Figure 3.8.1 models the kurtosis in the kernel density estimates, with fixed points at $x = \pm\ \frac{1}{2}$. One term in the Taylor expansion is enough. PC(0) is in the direction of $\phi(x)$; PC(1) is in the direction of $(x^2 - \frac{1}{4})\ \phi(x)$. It corresponds almost exactly with the "Theory" curve (see Figure 3.8.2). All higher order PC's are roundoff error.

Table 3.8.1

Percent Variability Explained

	PC(1)	PC(2)	PC(3)	PC(4)
INDIVIDUAL	100.00000	0.00000	0.00000	0.00000
CUMULATIVE	100.00000	100.00000	100.00000	100.00000

Table 3.8.2

Sums of Squares

SS_0	SS_1	SS_2	SS_3	SS_4	SS_5
285.60358	1.66525	0.00000	0.00000	0.00000	0.00000

The pattern which we found in test case 2 continues: PC(1) accounts for all the variability, and reduces the sum of squares to zero. SS_1 is neither large nor small; it reflects the symmetric variability found in the tails, as well as at the peak in the middle. We can examine the curves in the same manner as was described for test case 7. However, here the maximum downward change is smaller than the maximum upward change. The points of no change (zero points) are now at $x=\frac{1}{2}$ and $x=-\frac{1}{2}$.

3.9 Test Case 9

Curves in this case are of the form:

$$f_{\sigma,z}(x) = \phi(x) + (\sigma Z)\,(x^2 - \tfrac{1}{4})\,(x^2 - 3)\phi(x),$$

with $\sigma = .07$. Again, the test curves show the effects of kurtosis, but now with four fixed points (see Figure 3.9.1). There is exactly one nonzero term, and we find that $\underset{\sim}{A}_0$ looks as it should, while PC(1) is precisely in the direction of $(x^2 - \frac{1}{4})\,(x^2 - 3)\phi(x)$ (see Figure 3.9.2). All higher PC's are roundoff error.

Table 3.9.1

Percent Variability Explained

	PC(1)	PC(2)	PC(3)	PC(4)
INDIVIDUAL	100.00000	0.00000	0.00000	0.00000
CUMULATIVE	100.00000	100.00000	100.00000	100.00000

Table 3.9.2

Sums of Squares

SS_0	SS_1	SS_2	SS_3	SS_4	SS_5
283.59622	2.11625	0.00000	0.00000	0.00000	0.00000

The initial sum of squares, SS_1, reflects the symmetric variability found in the kernel density estimates. The changes at the "digitized" points are in the vertical direction, so PC(1) explains it all. Again there is a "twanging" motion of the curves, now with four zero points. If we examine the curves from highest peak to lowest peak we find: from the right tail to the first zero point there is a downward change; from first zero point to second zero point there is an upward change; from second to third zero points (the peak) there is a downward change; from third to fourth there is an upward change; and finally from fourth to left tail there is a downward change. More generally, if the curve is low on one interval (between zero points), it is high on the next one. The magnitude of change is greatest halfway between zero points.

3.10 Test Case 10

Here curves are of the form:

$$f_{\sigma,z}\,(x) = \phi(x) + (\sigma Z)\,\sin(\frac{1.7\ x}{\sigma}),$$

with $\sigma = .03$. Figure 3.10.1 shows the shape of these curves, where we find 11 fixed points. Only a one term Taylor expansion is needed, and we find that PC(0) is in the direction of $\phi(x)$, while $\underset{\sim}{A}_1$ is almost exactly the "unit vector" based on $(\sigma Z)\,\sin(\frac{1.7\ x}{\sigma})$ (see Figure 3.10.2). All higher order PC's are roundoff error.

Table 3.10.1

Percent Variability Explained

	PC(1)	PC(2)	PC(3)	PC(4)
INDIVIDUAL	100.00000	0.00000	0.00000	0.00000
CUMULATIVE	100.00000	100.00000	100.00000	100.00000

Table 3.10.2

Sums of Squares

SS_0	SS_1	SS_2	SS_3	SS_4	SS_5
285.03278	2.89649	0.00000	0.00000	0.00000	0.00000

SS_1 is relatively large, as compared to other test cases, reflecting the increased variability found about the 11 zero points. As in the last test case, if a curve is low on one interval, it is high on the next interval; the magnitude of change is greatest halfway between the zero points. The change is all in upward or downward motions.

3.11 Test Case 11

The curves are of the form:

$$f_{\sigma,z}(x) = \phi(x) + (\sigma)\sin\left(\frac{x}{2\sigma(1+\sigma Z)}\right),$$

where $\sigma = .05$. Figure 3.11.1 shows the effects of having random frequency in the sine function. The frequency is very high near the tails, and decreases towards the middle. Taylor expansion yields:

$$f_{\sigma,z}(x) = \phi(x) + (\sigma)\sin\left(\frac{x}{2\sigma}\right) - (\sigma Z)\,\frac{x}{2}\cos\left(\frac{x}{2\sigma}\right) + (\sigma Z)^2\,\frac{x}{2}\cos\left(\frac{x}{2\sigma}\right) - \ \ldots$$

We find that PC(0) is in the direction of $\phi(x) + (\sigma)\sin(\frac{x}{2\sigma})$. The graph shows $\underset{\sim}{A}_0$ basically following the "unit vector" of $\phi(x)$, but in the form of a sine function, so that it repeatedly crosses the "Raw Coeff" curve. PC(1) is in the direction of $-(\sigma Z)\,\frac{x}{2}\cos(\frac{x}{2\sigma})$, as can be seen in Figure 3.11.2. Note that here there is a greater discrepancy between the coefficient and the PC than we have had before. This problem is partly due to the randomness of the sine frequency, but mostly due to the orthogonality issue, which we did not analyze further here.

Table 3.11.1

Percent Variability Explained

	PC(1)	PC(2)	PC(3)	PC(4)
INDIVIDUAL	56.20493	40.00832	3.44752	0.33379
CUMULATIVE	56.20493	96.21325	99.66077	99.99456

Table 3.11.2

Sums of Squares

SS_0	SS_1	SS_2	SS_3	SS_4	SS_5
289.74474	3.37628	1.47864	0.12785	0.01145	0.00018

SS_0 is quite large, but is greatly reduced when we subtract the mean. We find that when we also subtract PC(1), only 56% of the variability is explained. So we know that although there is variability in the vertical direction, there is also a good deal of variability in some other direction(s). PC(2) significantly adds to the percent variability explained, and also reduces the sum of squares quite a bit. PC(3) also contributes a sizable amount. Looking at Figure 3.11.1, we can see that there is indeed a good deal of variability in many directions.

3.12 Test Case 12

The last case we looked at was curves of the form:

$$f_{\sigma,z}(x) = \phi(x) + \text{White Noise}$$

($\sigma = .02$). In all 11 cases discussed previously we dealt with very systematic changes in the test curve. Here we are dealing with the opposite extreme: these curves are very wiggly in a very random way, as can be seen in Figure 3.12.1. $\underset{\sim}{A}_0$ looks as it should, but is not quite as smooth as the "Raw Coeff" curve. The remaining PC's can do little to explain anything; they merely reflect the noise. This can be seen by examining Figures 3.12.2 - 3.12.3 (note the extremely low percentage of variability explained). The higher order PC's look more or less "the same".

Table 3.12.1

Percent Variability Explained

	PC(1)	PC(2)	PC(3)	PC(4)
INDIVIDUAL	5.05969	4.84842	4.43223	4.23117
CUMULATIVE	5.05969	9.90811	14.34034	18.57151

Table 3.12.2

Sums of Squares

SS_0	SS_1	SS_2	SS_3	SS_4	SS_5
285.03775	2.38840	2.26756	2.15176	2.04590	1.94484

Although the curves do not vary in a systematic way here, SS_0 and SS_1 are no larger than in any of the other test cases we have looked at. This is due to the fact that these estimates fit fairly tightly around the underlying curve, $\phi(x)$. PC(1) attempts to pick up the variability in the vertical direction, but we find there is little, if any, here. The randomness is the dominating feature here, as can be seen both by the extremely low percentage of variability explained, and by the nearly nondecreasing sums of squares.

Chapter 4
Density Estimation:
Under versus Over Smoothing

A question which naturally arises in kernel density estimation is the effect the bandwidth size has on the estimate. More specifically, what are the effects of using fixed bandwidths for several datasets, where the fixed bandwidths are: a) too large (over smoothing), b) optimal (ie h to minimize MISE), and c) too small (under smoothing). To investigate this, we studied kernel density estimates of four distributions using the bandwidths: (h to minimize MISE)*2, h to minimize MISE, and (h to minimize MISE)/2. We refer to these bandwidths as MISE*2, MISE, and MISE/2. Three of these will be discussed below; results from analysis of the fourth distribution coincide with the results found from analysis of the first three. We will no longer be comparing the PC's to Taylor expansions; rather, we will compare them to the corresponding derivatives of the density we are trying to estimate. We compare the PC's to the theoretical derivatives because these looked like the derivatives of f which appeared in test cases 1, 7, 8, and 10 .

4.1 Gaussian

The first example we looked at was kernel density estimates from the Gaussian distribution. We further broke this distribution down into two cases: one based on an 0original sample size of 100, the other based on an original sample size of 1000.

4.1.A n=100

Upon examining Figures 4.1.1 - 4.1.3, it is easy to see the effects the three bandwidths have on the estimates. MISE*2 yields overly smooth estimates which cause us to lose features of the true density (ie the peak is too low), while MISE/2 produces too wiggly (undersmoothed) estimates, which cause us to gain many spurious features. The MISE based estimates fall somewhere in the middle in terms of smoothness. We found that the mean, PC(0), was oversmoothed for MISE*2, as was clearly expected from Figure 4.1.1: it was a good deal too high at the tails, and too low at the peak in the middle; $\underset{\sim}{A}_0$ was only a little biased for MISE and less so for MISE/2: slightly high at the tails, and slightly low at the peak. At the first level, the under and over smoothing effects can better be seen (see Figures 4.1.4 - 4.1.6). PC(1) for MISE*2 explains the largest percent of variability, followed by MISE, and then MISE/2. This is due to the fact that a much lower percentage of the total variability for MISE*2 was random noise. These same smoothing effects are carried out in the higher order PC's as well. All three have the same general shape with regards to corresponding PC's, but we find that the over smoothness in MISE*2 causes it's $\underset{\sim}{A}_1$ to be closer to $f^{(1)}$ than the other two are. $\underset{\sim}{A}_1$ for MISE*2 is fairly close to PC(1) in test case 1, where the curves have more or less been horizontally shifted. MISE falls somewhere between test case 1 and test cases 7 or 8 in terms of similarity of $\underset{\sim}{A}_1$'s; here the data has more of a vertical shift. $\underset{\sim}{A}_1$ for MISE/2 looks a lot like $\underset{\sim}{A}_1$ in test case 10 in terms of the sharpness of the curves. There are a lot of wiggles in the estimates, but it is not nearly as bad as the white noise example (test case 12).

The percent of variability explained by the first five PC's can be found in Table 4.1.1. Overall, MISE*2 explains the most with the first few PC's, again followed by MISE, then MISE/2. This is sensible in view of the different levels of randomness apparent in Figures 4.1.1 - 4.1.3. Most of the variability is captured in the first few PC's for MISE*2; for MISE/2 there is still a considerable percentage of variability left to be explained even after considering PC(5) (for similar reasons as were discussed in the white noise test case). These results are also reflected in Table 4.1.2, where we list the sum of squares. MISE*2 has the smallest sum of squares, and MISE/2 has the largest, which makes sense after examining the original curves. The estimates for MISE/2 are visually a lot more variable than are those for MISE*2 or even MISE.

Table 4.1.1

Percent Variability Explained

	PC(1)	PC(2)	PC(3)	PC(4)	PC(5)
MISE * 2	62	32	5	1	0
CUMULATIVE	62	94	99	100	100
MISE	44	34	9	7	3
CUMULATIVE	44	78	87	94	97
MISE / 2	30	25	14	9	6
CUMULATIVE	30	55	69	78	84

Table 4.1.2

Sums of Squares

	SS_0	SS_1	SS_2	SS_3	SS_4	SS_5	SS_6
MISE * 2	2215.151	1.173	0.451	0.076	0.020	0.003	0.001
MISE	266.386	4.208	2.367	0.938	0.540	0.253	0.128
MISE / 2	290.541	10.324	7.256	4.693	3.299	2.349	1.767

Note that the total sums of squares are really not all that different from each other; SS_0 for MISE/2 is only larger than SS_0 for MISE*2 by a factor of about 1.35. However, once we get to the higher levels, this factor has increased dramatically. At SS_4 the same factor is now nearly 165. This is because the variability is really being picked up well by the first few PC's in MISE*2, while there is a larger component of noise in MISE/2, perhaps not so different from that of test case 12.

4.1.B n=1000

Here we find that our kernel density estimates are "tighter" than they were for n = 100, most noticeably in MISE/2 (see Figures 4.1.7 - 4.1.9). $\underset{\sim}{A}_0$ is again oversmoothed for MISE*2, although less so than it was for n=100. $\underset{\sim}{A}_0$ for MISE is just a little biased, and for MISE/2 $\underset{\sim}{A}_0$ is almost exactly $f(0) = \phi(x)$. $\underset{\sim}{A}_1$ for MISE*2 is fairly close to it's derivative; for MISE $\underset{\sim}{A}_1$ is too sharp; for MISE/2 $\underset{\sim}{A}_1$ is much too sharp and wiggly (see Figures 4.1.10 - 4.1.12). This pattern follows in the higher order PC's, getting progressively worse. MISE/2 is even more similar to test case 12 here. As in the case where n=100, there is the pattern of MISE*2 explaining the highest percent of variability both in PC(1) and overall, followed by MISE and then MISE/2. Also, MISE*2 has the smallest sum of squares, MISE/2 has the largest. However, the sum of squares are larger for n=1000 than for n=100 at all levels, while the total amount of variability explained is smaller than for n=100.

Table 4.1.3

Percent Variability Explained

	PC(1)	PC(2)	PC(3)	PC(4)	PC(5)
MISE * 2	48	36	8	5	2
CUMULATIVE	62	94	99	100	100
MISE	36	25	15	8	5
CUMULATIVE	36	61	76	84	89
MISE / 2	24	15	14	11	6
CUMULATIVE	24	39	53	64	70

Table 4.1.4

Sums of Squares

	SS_0	SS_1	SS_2	SS_3	SS_4	SS_5	SS_6
MISE * 2	255.708	0.290	0.152	0.047	0.024	0.009	0.004
MISE	282.252	0.840	0.537	0.326	0.202	0.137	0.097
MISE / 2	291.109	1.858	1.407	1.120	0.851	0.654	0.537

Here the disparity between the sums of squares for MISE*2 and MISE/2 is not nearly as great as it was for n=100. At SS_0 the factor is only a little greater than 1, while at level 4 the factor has only increased to approximately 35. This is due to the greater amount of noise found in the MISE/2 curves as opposed to the MISE*2 curves.

4.2 Bimodal

For reference to this distribution, see Marron and Wand (1991), distribution #6.

4.2.A n=100

The same patterns continue here as they did for the Gaussian distribution (see Figure 4.2.1). MISE*2 basically smooths out the bimodal part so that the kernel density estimates look unimodal. The $\underset{\sim}{A}_j$'s in all three cases are visually similar to the corresponding $\underset{\sim}{A}_j$'s in test case 1. The main differences are that here it takes more PC's to explain approximately the same amount of variability as it did above, and that the corresponding sums of squares are smaller. What is interesting about this case is that although the $\underset{\sim}{A}_j$'s are not close to their corresponding derivatives for PC(1) - PC(3), they come back and are fairly accurate for PC(4) and PC(5).

Table 4.2.1

Percent Variability Explained

	PC(1)	PC(2)	PC(3)	PC(4)	PC(5)
MISE * 2	65	24	9	2	1
CUMULATIVE	62	89	98	100	101
MISE	39	24	16	9	5
CUMULATIVE	39	63	79	88	93
MISE / 2	21	20	13	12	9
CUMULATIVE	21	41	54	66	75

Table 4.2.2

Sums of Squares

	SS_0	SS_1	SS_2	SS_3	SS_4	SS_5	SS_6
MISE * 2	197.985	1.715	0.608	0.193	0.046	0.012	0.003
MISE	229.255	5.101	3.133	1.890	1.060	0.580	0.328
MISE / 2	249.454	12.352	9.708	7.259	5.643	4.174	3.109

At SS_0 the factor of difference between MISE*2 and MISE/2 is only 1.26, but at level 4 it jumps significantly to nearly 123, and at level 9 it is about 1036. Again, most of the variability is picked up in the MISE*2 case, but MISE/2 still has a lot of noise.

4.2.B n=1000

We carried out the same analysis here, and found no new lessons beyond what was found in section 4.1.B as far as comparing n=100 to n=1000, and in section 4.1.A as far as this distribution is concerned.

4.3 Strongly Skewed

For reference to this distribution, see Marron and Wand (1991), distribution #3.

The pattern continues here: the first few PC's for MISE*2 explains the most, MISE/2 the least; less overall variability is explained for n=1000 than for n=100, while the sums of squares are greater for n=1000. Here quite a bit less of the variability is explained as compared to the first two distributions. MISE/2 appears to be mainly noise. See Figure 4.3.1 to get a general idea as to the shape of the true density and the kernel density estimates. MISE/2 provided the best fit to the mean. None of the three $\underset{\sim}{A}_1$'s resemble $\underset{\sim}{A}_1$ from any test case, but the higher order PC's are becoming similar to those of test case 12. This makes sense since the randomness of the kernel density estimates is much like that found in test case 12 (see Figure 3.12.1), although those test curves are a little sharper than the estimates found here.

Table 4.3.1

n=100

Percent Variability Explained

	PC(1)	PC(2)	PC(3)	PC(4)	PC(5)
MISE * 2	27	22	13	9	7
CUMULATIVE	27	49	62	71	78
MISE	20	16	11	8	7
CUMULATIVE	20	36	47	55	62
MISE / 2	16	11	8	7	6
CUMULATIVE	16	27	35	42	48

Table 4.3.2

n=100

Sums of Squares

	SS_0	SS_1	SS_2	SS_3	SS_4	SS_5	SS_6
MISE * 2	448.310	10.770	7.848	5.428	4.070	3.062	2.286
MISE	525.708	23.963	19.058	15.145	12.462	10.441	8.709
MISE / 2	664.958	57.149	47.983	41.966	37.146	33.010	29.470

None of the three bandwidths was significantly better than the others. Even at SS_6, the sums of squares for MISE*2 and MISE/2 differ in magnitude by less than 13. For this distribution, the PC's did not appear to be sensitive to the over and under smoothing effects. This is also similar to what happened in test case 12. Although the percentages and sums of squares are not as low here as are those found in test case 12, we can see similarities between it and our three bandwidths. The percent explained is not very high for any of these three bandwidths, and though the sums of squares are decreasing, they are doing so very slowly. Similar results held in test case 12.

Again, we carried out the same analysis for n=1000, but found no new lessons.

Chapter 5

"Optimal" Bandwidths

We now turn our focus to a comparison of various kernel density estimate "optimal" bandwidths. In particular, we studied MISE, ISE (integrated squared error), and IAE (integrated absolute error). MISE was defined on page 1; ISE represents the h to minimize:

$$\mathrm{ISE}(h) = \int (\hat{f}_h - f)^2 \, dx;$$

IAE represents the h to minimize:

$$\mathrm{IAE}(h) = \int | \hat{f}_h - f | \, dx.$$

To investigate any differences or similarities, we again examined kernel density estimates of four distributions (Gaussian, bimodal, strongly skewed, smooth comb), using each of the three bandwidths. As before, we further broke each distribution down into two cases: one based on an original sample size of n=100, the other on n=1000. In our study, we found that the results for the four distributions were very similar, and that the results for n=100 and n=1000 within each distribution were virtually identical. Because of this, we will restrict our discussion to the bimodal, n=100 case, and merely comment on any differences found in the other three distributions.

We found the three sets of kernel density estimates to be visually very close to each other; to get an idea of the shape, see Figure 5.1. The directions of the means were virtually indistinguishable from each other; all over smoothed near the peaks and valley (see Figure 5.2). The PC(1)'s are also very much in the same direction; however, they do not seem to pick up the bimodality of the curves, as they are very similar to $\underset{\sim}{A}_1$ from test case 1 (the test curves were unimodal) and they do not correspond well with the derivative, $f^{(1)}$, (compare Figures 5.3 and 3.1.2). A few differences are more apparent in the directions of PC(2) and PC(3). ISE and IAE are very visually close, while MISE is similar, but with not quite as pronounced peaks and valleys. They are not very close to their derivatives. Once we reach $\underset{\sim}{A}_4$ and $\underset{\sim}{A}_5$, the three bandwidths are again almost exactly the same. In addition, they are fairly close to the corresponding derivatives (see Figures 5.4 and 5.5). For n=1000 similar results hold, although the kernel density estimates fit tighter around the true density. All three bandwidths are again very similar, with MISE following a little closer to the derivatives.

MISE had the largest bandwidth, followed by ISE and then IAE. However, all explained virtually the same percentage of variability, and had very similar sums of squares.

For n=1000, the sums of squares increased and the percent variability explained decreased. Despite evidence of an important difference between MISE and ISE (see theorem 2.1 of Hall and Marron (1987)), this difference is not systematic in any sense which can be observed through our PCA. Hence, this difference may not be so important as previously believed.

Table 5.1

Bimodal

n=100

	average bandwidth	% variability PC(1)→PC(5)	SS_0	SS_1
MISE	.3854	93	229.255	5.101
ISE	.3831	92	228.729	4.727
IAE	.3682	91	229.408	4.937

The Gaussian distribution findings are nearly the same as those just discussed. MISE again explained the most, followed by ISE and then IAE (see Table 5.2). The mean for MISE was closest to the true density; ISE and IAE slightly underestimated the peak. The PC directions were virtually the same in all three cases, and appeared much like those of test case 1. In this distribution, MISE was the best "optimal" bandwidth; again, however, it was not significantly better than the other two.

Table 5.2

Gaussian

n=100

	average bandwidth	% variability PC(1)→PC(5)	SS_0	SS_1
MISE	.4455	97	266.386	4.208
ISE	.4309	94	266.124	3.604
IAE	.4101	92	268.275	3.958

In the strongly skewed distribution, there were more important differences. The kernel density estimates for ISE and IAE all fell below the true density at it's peak; for MISE this was not true. The IAE bandwidth was about twice that of MISE or ISE, which were about the same (see Table 5.3). Correspondingly, the first few PC's of IAE explained the most variability, and had the smallest SS_0. (Recall our results from Chapter 4: the larger the bandwidth, the larger the percent variability explained, and the smaller the sums of squares.) ISE and MISE were about the same in these two areas. The PC directions were pretty much the same for all three bandwidths. However, none of the three did a very good job of explaining the variability. Even for IAE, there was still about 24% of the variability left unexplained after the first five PC's. Our belief here is that the main difference between IAE and ISE is due only to the under vs over smoothing effect.

Table 5.3

Strongly Skewed

n=100

	average bandwidth	% variability PC(1)→PC(5)	SS_0	SS_1
IAE	.1795	76	444.348	13.780
ISE	.0848	61	521.545	22.673
MISE	.0827	62	525.708	23.963

For the smooth comb distribution (reference: Marron and Wand (1991), distribution #14), the kernel density estimates and PC directions were virtually the same for the three bandwidths. IAE had the largest bandwidth, but was not significantly better than ISE or MISE. Again, none of the three did a very good job of explaining the variability (see Table 5.4).

Table 5.4

Smooth Comb

n=100

	average bandwidth	% variability PC(1)→PC(5)	SS_0	SS_1
IAE	.1447	60	258.714	16.281
MISE	.1404	61	258.121	15.669
ISE	.1398	60	259.343	16.063

In general, we found that for smooth, symmetric densities a large percentage of variability was explained by all three bandwidths. MISE explained the most, followed by ISE and then IAE. For the densities which were neither symmetric nor very smooth, a fairly low percentage of variability was explained by any of the three bandwidths. Though not very good itself, IAE explained more than the other two in these situations.

We did not observe a difference between MISE and ISE that was systematic in a sense that is well measured by PCA. While there has been considerable discussion of which is the more "reasonable" criterion, (see Hall and Marron (1991), Mammen (1988), and Jones (1991)), this provides some evidence that the difference is of a "pure noise" type, so there may not be a large practical difference between these assessment methods, although MISE may be preferred on grounds of less random noise. Devroye and Györfi (1985) point out important differences in IAE and ISE. While these were nearly the same in many of our examples, there was a major difference in the strongly skewed case. However, in terms of shapes of the resulting estimates, this difference showed up only as an under vs over smoothing effect.

Chapter 6

Data-Driven Bandwidth Selectors

Now we will study the differences between various data-driven bandwidth selectors, as opposed to the "optimal" bandwidths we compared in Chapter 5. Most of these selectors can be thought of as an attempt to estimate MISE. LSCV (least squares cross-validation) was proposed by Rudemo (1982) and Bowman (1984), and represents the h to minimize:

$$LSCV(h) = \int \hat{f}_h^2 \, dx - \frac{2}{n} \sum_{j=1}^{n} \hat{f}_{j,h}(X_j) ,$$

where $\hat{f}_{j,h}$ denotes the leave-one-out kernel estimator constructed from the data with X_j deleted. It has been found that LSCV suffers from a great deal of sample variability (Hall and Marron (1987)). PIP0 was proposed by Silverman (1987), equation 3.28, and represents

$$h = \left(\tfrac{4}{3}\right)^{\frac{1}{5}} \hat{\sigma} \, n^{-\left(\frac{1}{5}\right)} ,$$

where $\hat{\sigma}$ is the sample standard deviation. PIP0 notoriously over smooths the kernel density estimates when the underlying distribution is far from normal. The PMPI (Park-Marron plug-in) selector was proposed by Park and Marron (1990), and appears in a much more complicated form. It makes use of the following definitions:

$R(g) = \int g(x)^2 dx$, for any function $g(x)$;

$\sigma^2_g = \int x^2 \, g(x) \, dx$, the variance of the g distribution (mean-zero probability density $g(x)$;

$\hat{f}_a$ is a kernel density estimate, with bandwidth now represented by a [allowed to be different from h because estimation of this integral is a different smoothing problem from estimation of f(x)];

$\hat{R}_{f''}(a) = R(\hat{f}_a'') - n^{-1}a^{-5} \, R(K'')$

K*K denotes the convolution $K*K(x) = \int K(x-t) \, K(t)dt$;

$C_3(K) = \{18R(K^{(4)}) \, \sigma^8_K / \sigma^4_{K*K} \, R(K)^2\}^{\frac{1}{13}}$;

$C_4(g_1) = \{R(g_1) \, R(g''_1)^2 \, / \, R(g_1^{(3)})^2\}^{\frac{1}{13}}$, where g_1 is the normal density;

$a_{\hat{\lambda}}(h) = C_3(K)C_4(g_1)\lambda^{\frac{3}{13}} \, h^{\frac{10}{13}}$;

λ is the scale of f;

$a_\lambda(h) = C_3(K) \, C_4(g_1) \, \lambda^{\frac{3}{13}} \, h^{\frac{10}{13}}$.

PMPI represents the h which is the root (when it exists, the largest if there are more than one) over the range [$\underline{B}\, n^{\frac{-1}{5}}$, $\overline{B}\, n^{\frac{-1}{5}}$] (for some constants $0 < \underline{B} < \overline{B}$) of the equation

$$h = \{R(K)/\sigma^4{}_K\ \widehat{R}_{f''}(a_{\widehat{\lambda}}(h))\}^{\frac{1}{5}}\ n^{\frac{-1}{5}}\ ,$$

where $\widehat{\lambda}$ is a good (ie $n^{\frac{1}{2}}$ consistent) estimator of λ. The SJPI (Sheather Jones plug-in) selector was proposed by Sheather and Jones (1991), and is often thought of as a slight improvement of PMPI. It differs from PMPI only slightly, defining

$\widehat{R}_{f''}(a) = R(\widehat{f}_a{}'')$, and not as above.

As in the previous chapter, we will focus our discussion on the bimodal, n=100 distribution, and merely comment on the other three distributions. [Note: for the other three distributions, we have data but not the graphs of $\underset{\sim}{A}_j$ vs $f^{(j)}$, so we can not make visual comparisons of the PC's.]

The kernel density estimates for PIP0 are slightly oversmoothed at the two peaks (see Figure 6.1); for PMPI and SJPI they appear virtually the same (see Figure 6.2); for LSCV they are a lot noisier at the peaks (see Figure 6.3). The means for all four bandwidths were pretty similar in that they all oversmoothed the true density; PIP0 was least accurate because it oversmoothed the most, while LSCV was most accurate, and the other two appeared about the same (see Figure 6.4). All four again had similar PC(1) directions, although LSCV's was noisier than the rest. The kernel density estimates seemed to "miss" the bimodality of the true density, as the PC(1)'s corresponded more to the direction of PC(1) found in test case 1, which had a unimodal density (compare Figures 6.5 and 3.1.2). The PC(2) directions were again about the same: PIP0 is still smoothest, LSCV is still noisiest, and the other two are about the same, falling somewhere in between (see Figure 6.6). All the $\underset{\sim}{A}_2$'s were closer in relation to $f^{(2)}$ than the $\underset{\sim}{A}_1$'s were to $f^{(1)}$. The relation gets better for the directions of PC(3) - PC(5) (see Figures 6.7 - 6.9). Due to the greater amount of over smoothing, PIP0 followed closer to test case 1 than did the others. LSCV remained the noisiest, and the other two were about the same. Their directions were most similar, and closest to the corresponding derivative for PC(5)

Corresponding to these visual observations, we found that PIP0 explained the most variability, LSCV explained the least, and the other two explained about the same amount (see Table 6.1). Although LSCV ranks second in terms of bandwidth size, and it's sums of squares are relatively close to those of the other selectors, it explains a considerable amount less of the variability than they do. PIP0 obviously oversmoothed things-it does not appear to feel the peaks, as is evidenced by the PC directions looking like those of the unimodal distribution in test case 1. SJPI is not better than PMPI in terms of percent variability explained, as would be expected if indeed SJPI is an improvement over PMPI. For n=1000, the same results hold, although less overall variability is explained, and the sums of squares are larger for all four) bandwidth selectors. Here, however, LSCV dropped to fourth in terms of average bandwidth size, which would coincide with Hall and Marron finding (1987) that this selector suffers from a great deal of sample variability. It is closer to the two bandwidths SJPI and PMPI in terms of variability explained than it was for n=100. PIP0 stands out more from the others, and explains much more of the variability.

SJPI is the best estimate of MISE in terms of bandwidth size, percent variability explained, and sums of squares (see Table 5.1), although PMPI is quite close, and PIP0 is not too far off. This is reflected throughout the graphs of the $\underset{\sim}{A}_j$'s vs the $f^{(j)}$'s. SJPI corresponds closest with MISE, with PMPI pretty close behind. PIP0 shows signs of more over smoothing than MISE, while LSCV is not really even close. For a visual impression, we can compare Figures: 6.1 - 6.3 with 5.1, 6.4 and 5.2, 6.5 and 5.3, 6.8 and 5.4, and 6.9 and 5.5.

Table 6.1

Bimodal

n=100

	average bandwidth	% variability PC(1)—PC(5)	SS_0	SS_1
PIP0	.5059	98	218.904	3.540
PMPI	.4763	95	223.523	5.140
LSCV	.4024	82	232.497	8.370
SJPI	.3822	92	231.912	6.658

For the Gaussian distribution, all the kernel density estimates look basically the same, both for n=100 and n=1000. Here PMPI appears slightly better than the others; LSCV again explains the least (see Table 6.2). For n=1000 there is less variability explained, and larger sums of squares. LSCV becomes more separated from the others, while the other four become more similar amongst themselves. PMPI explains more than SJPI. The results listed in Table 6.2 are all are very close to the results found using the bandwidth MISE (Table 5.2). LSCV has the closest average bandwidth, but explains less variability than MISE did. We can conclude that all provided good estimates of MISE.

Table 6.2

Gaussian

n=100

	average bandwidth	% variability PC(1)→PC(5)	SS_0	SS_1
PMPI	.4812	98	263.157	4.803
LSCV	.4450	90	267.949	6.950
PIP0	.4212	97	269.294	4.924
SJPI	.4023	95	271.667	6.049

In the strongly skewed distribution, PIP0 is over smoothed (see Figure 6.10); PMPI and SJPI are also oversmoothed, but to a lesser degree (see Figure 6.11); LSCV does the best job of capturing the peak (see Figure 6.12). PIP0 again explains the most, and LSCV the least (see Table 6.3). The other two are about the same, with PMPI again explaining more than SJPI. For n=1000 we found the unique situation that a selector explained more of the variability than it did for n=100: LSCV went from 66% to 73%, as well as moving up in the rank of bandwidth size, from fifth to second. This did not happen anywhere else, nor to any other selector. We did not investigate why. All the average bandwidths are larger than the MISE one (see Table 5.3), and all explain more of the variability than MISE did. In this case, only LSCV was close to estimating MISE.

Table 6.3

Strongly Skewed

n=100

	average bandwidth	% variability PC(1)—PC(5)	SS_0	SS_1
PIP0	.4332	97	303.820	3.508
PMPI	.1761	82	445.411	12.874
SJPI	.1569	79	459.837	14.041
LSCV	.0932	66	539.515	33.8548

PIP0 is over smoothed a lot in the smooth comb distribution (see Figure 6.13); PMPI and SJPI are so to a lesser degree (see Figure 6.14); and LSCV provides pretty good estimates (see Figure 6.15). PMPI again explains more than SJPI does. PIP0 explains the most, and LSCV explains the least (see Table 6.4). Again all average bandwidths were larger than the MISE one (see Table 5.4); all except LSCV explain a significant amount of the variability, but have smaller sums of squares. LSCV came closest to estimating MISE here.

Table 6.4

Smooth Comb

n=100

	average bandwidth	% variability PC(1)—PC(5)	SS_0	SS_1
PIP0	.6925	100	169.698	2.423
PMPI	.3037	87	213.273	7.303
SJPI	.2638	81	222.125	8.378
LSCV	.1509	59	256.000	17.484

Except for the Gaussian distribution (where all except LSCV were virtually the same), PIP0 comes out on top and LSCV comes out on bottom in terms of percent variability explained. SJPI and PMPI are more or less interchangeable. For the smooth, symmetric densities, all except LSCV are about the same. The differences among the three "groups" are magnified as the density becomes less symmetric and smooth.

References

Aldershof, B., Marron, J. S., Park, B. U., and Wand, M. P. (1991), Facts about the normal density, unpublished manuscript.

Arnold, Steven F, (1981), *The Theory of Linear Models and Multivariate Analysis*, New York: John Wiley & Sons.

Besse, P. and Ramsay, J.O. (1986), Principal components analysis of sampled functions, *Psychometrika*, **51**, 285-311.

Bowman, A. (1984), An alternative method of cross-validation for the simulations of density estimates, *Biometrika*, **71**, 353-360.

Chatfield, Christopher and Collins, Alexander J. (1980), *Introduction to Multivariate Analysis*, London: Chapman and Hall.

Devore, Jay and Peck, Roxy (1986), *Statistics: The Exploration and Analysis of Data*, St. Paul: West Publishing Co.

Devroye, L. and Györfi, L. (1985), *Nonparametric Density Estimation: The L_1 View*, New York: Wiley & Sons.

Everitt, B.S. (1978), *Graphical Techniques for Multivariate Data*, London: Heinemann Educational Books.

Gnanadesikan, R. (1977), *Methods for Statistical Data Analysis of Multivariate Observations*, New York: John Wiley & Sons.

Hall, P., and Marron, S. J. (1987), Extent to which least-squares cross-validation minimises integrated square error in nonparametric density estimation, *Probability Theory and Related Fields*, **74**, 567-581.

Hall, P. and Marron, J. S. (1991), Lower bounds for bandwidth selection in density estimation, *Probability Theory and Related Fields*, to appear.

Harris, Richard J. (1975), *A Primer of Multivariate Statistics*, New York: Academic Press.

Jones, M.C. (1991), The roles of ISE and MISE in density estimation, unpublished manuscript.

Jones, M.C. and Rice, John A. (1990), Displaying the important features of large collections of similar curves, *The American Statistician*, 1992, to appear.

Kleinbaum, David G., Kupper, Lawrence L., and Muller, Keith E. (1988), *Applied Regression Analysis and Other Multivariate Methods*, Kent Publishing Company.

Mammen, E. (1988), A short note on optimal bandwidth selection for kernel estimators, *Statist. Prob. Letters 9,* 23-25.

Marron, J.S. (1988a), Automatic smoothing parameter selection: A survey, *Empirical Economics,* **13**, 187-208.

Marron, J.S. and Wand, M.P. (1991), Exact mean integrated squared error, unpublished manuscript.

Morrison, Donald F. (1976), *Multivariate Statistical Methods, 2nd ed,* New York: McGraw -Hill Book Company.

Park, B.U. and Marron, J.S. (1990), Comparison of data driven bandwidth selectors, *JASA,* **85**, 66-72.

Press, S. James. (1972), *Applied Multivariate Analysis,* New York: Holt, Rinehart, and Winston, Inc.

Ramsay, J.O. (1982), When the data are functions, *Psychometrika,* **47**, 379-396.

Rice, John A. and Silverman, B. W. (1991), Estimating the mean and covariance structure nonparametrically when the data are curves, *J.R. Statist. Soc. B.,* **53**, No. 1, 233-243.

Rudemo, M. (1982), Empirical choice of histograms and kernel density estimators, *Scandinavian Journal of Statistics,* **9**, 65-78.

Servain, J. and Legler, D.M. (1986), Empirical orthogonal function analyses of tropical atlantic sea surface temperature and wind stress: 1964-1979, *J. Geophys. Res.*, **91**, 14181-14191.

Sheather, S. J., and Jones, M. C. (1991), A reliable data-based bandwidth selection method for kernel density estimation, *J. Roy. Statist. Soc. Ser. B*, **53**, to appear.

Silverman, B.W. (1986), *Density estimation for statistics and data analysis*, London: Chapman & Hall.

Winant, C.D., Inman, D.L. and Nordstrom, C.E. (1975), Description of seasonal beach changes using empirical eigenfunctions, *J. Geophys. Res.*, **80**, 1979-1986.

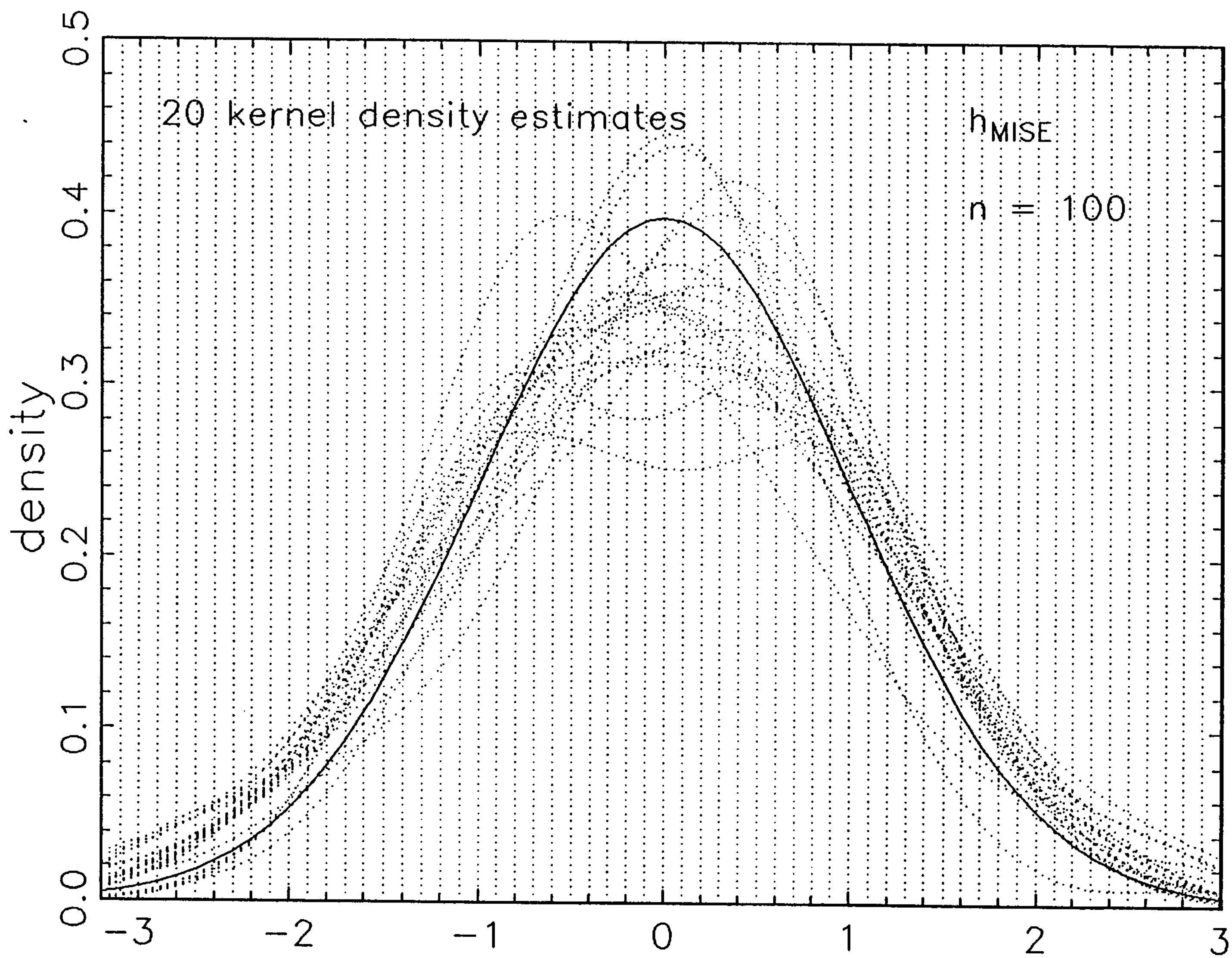


Figure 1.1 Gaussian distribution

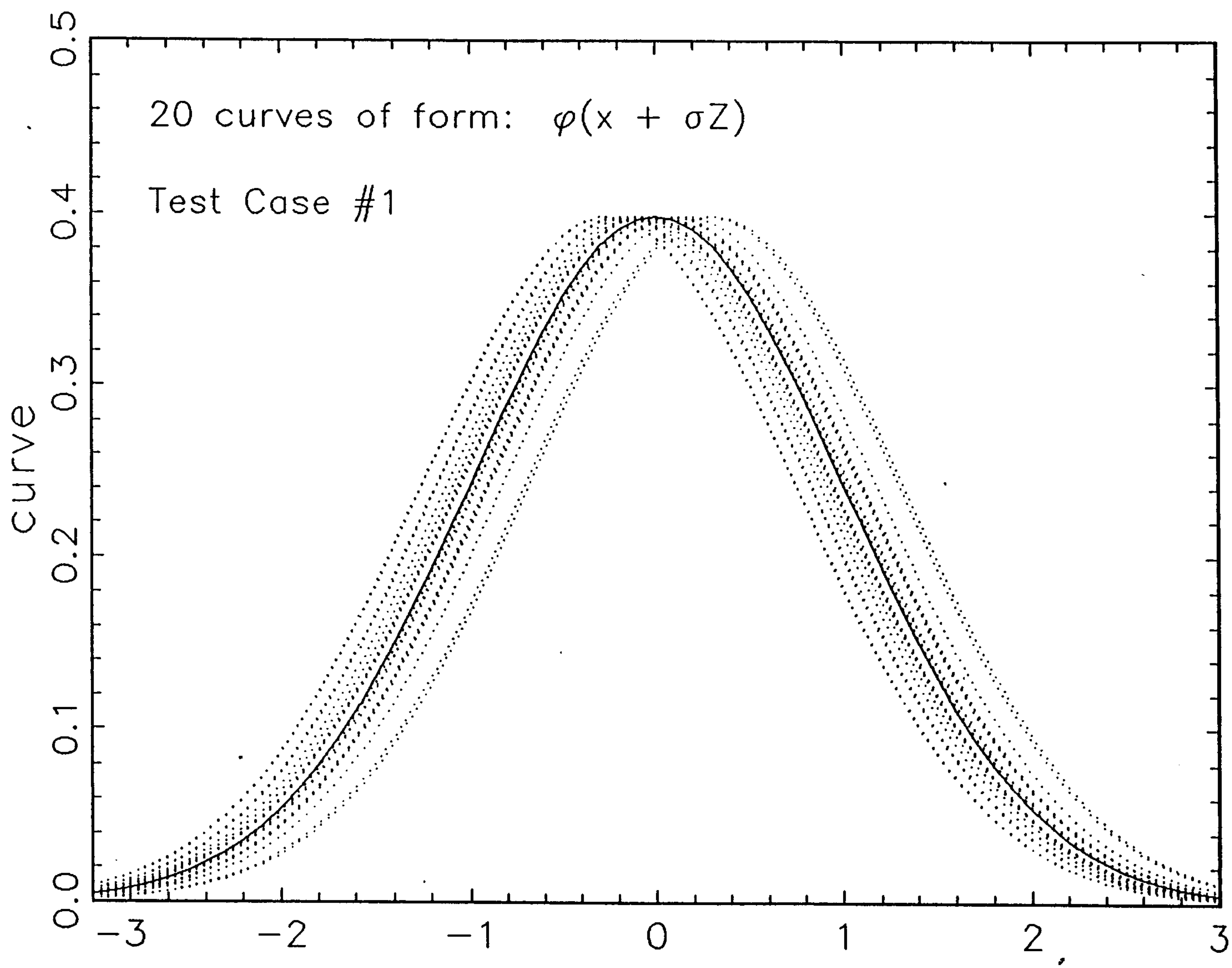

20 curves of form: φ(x + σZ)
Test Case #1
curve
0.0
0.1
0.2
0.3
0.4
0.5
−3
−2
−1
0
1
2
3
x

Figure 3.1.1

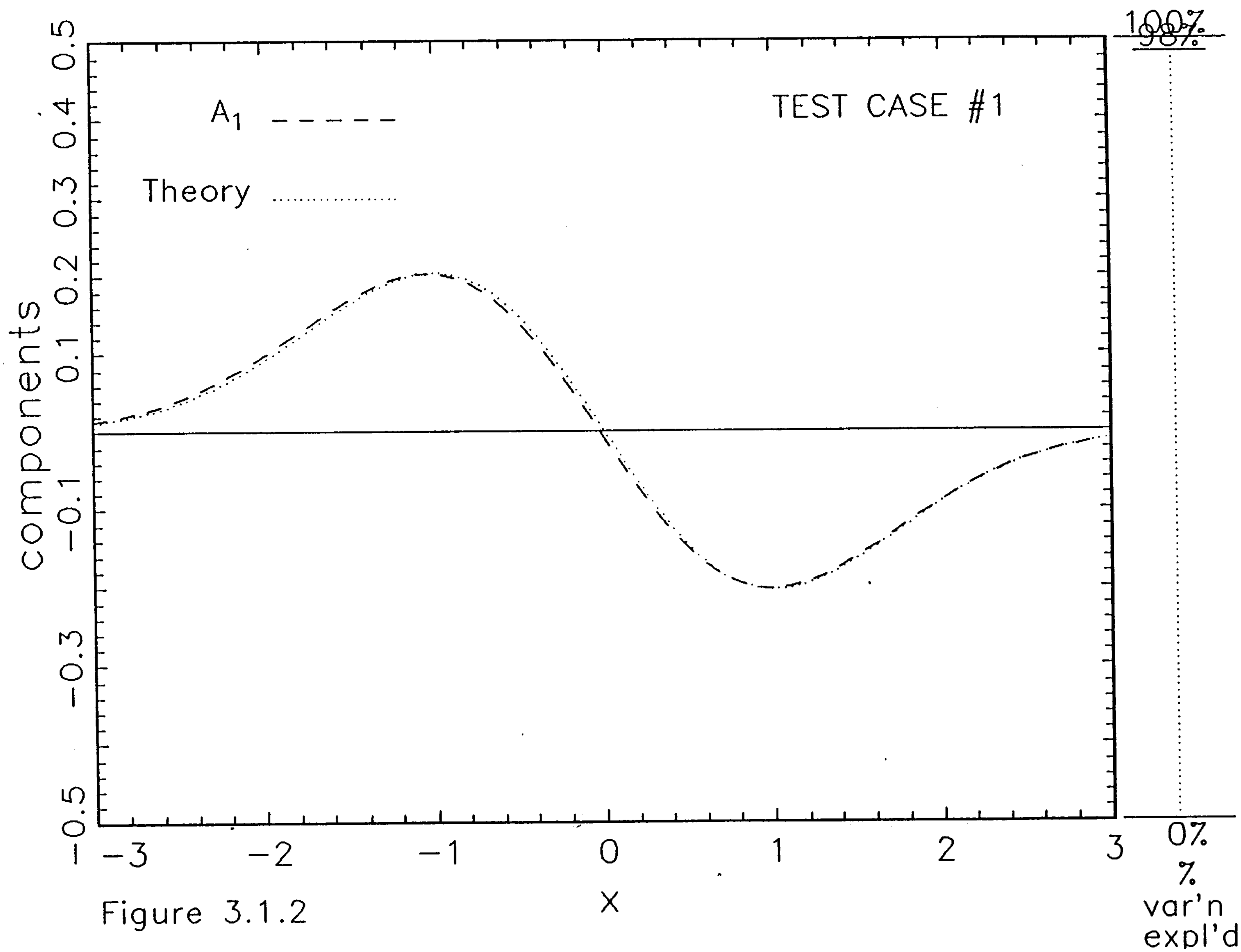

TEST CASE #1
A1
Theory
components
0.5
0.4
0.3
0.2
0.1
−0.1
−0.3
−0.5
−3
−2
−1
0
1
2
3
X
100%
98%
0%
%
var'n
expl'd

Figure 3.1.2

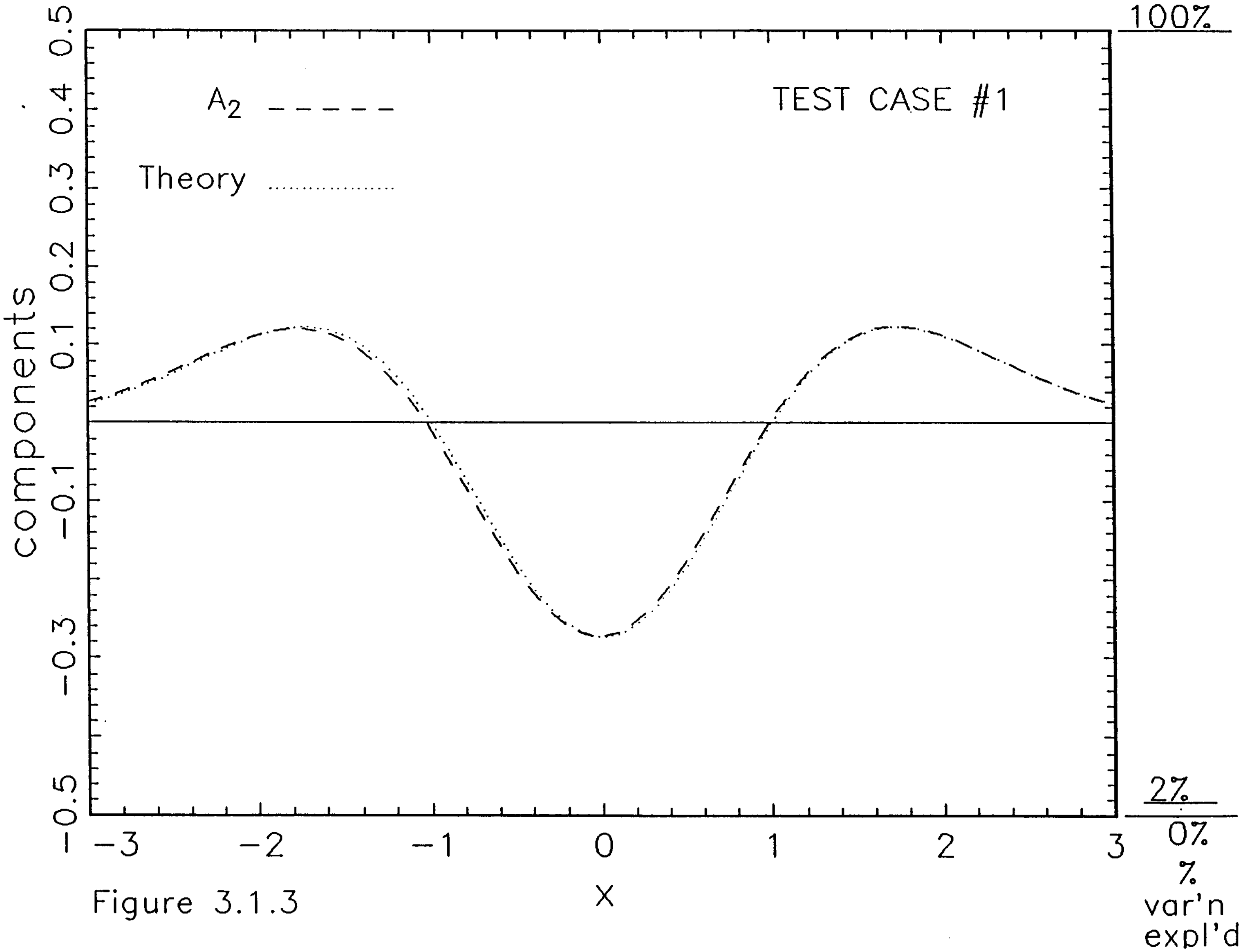
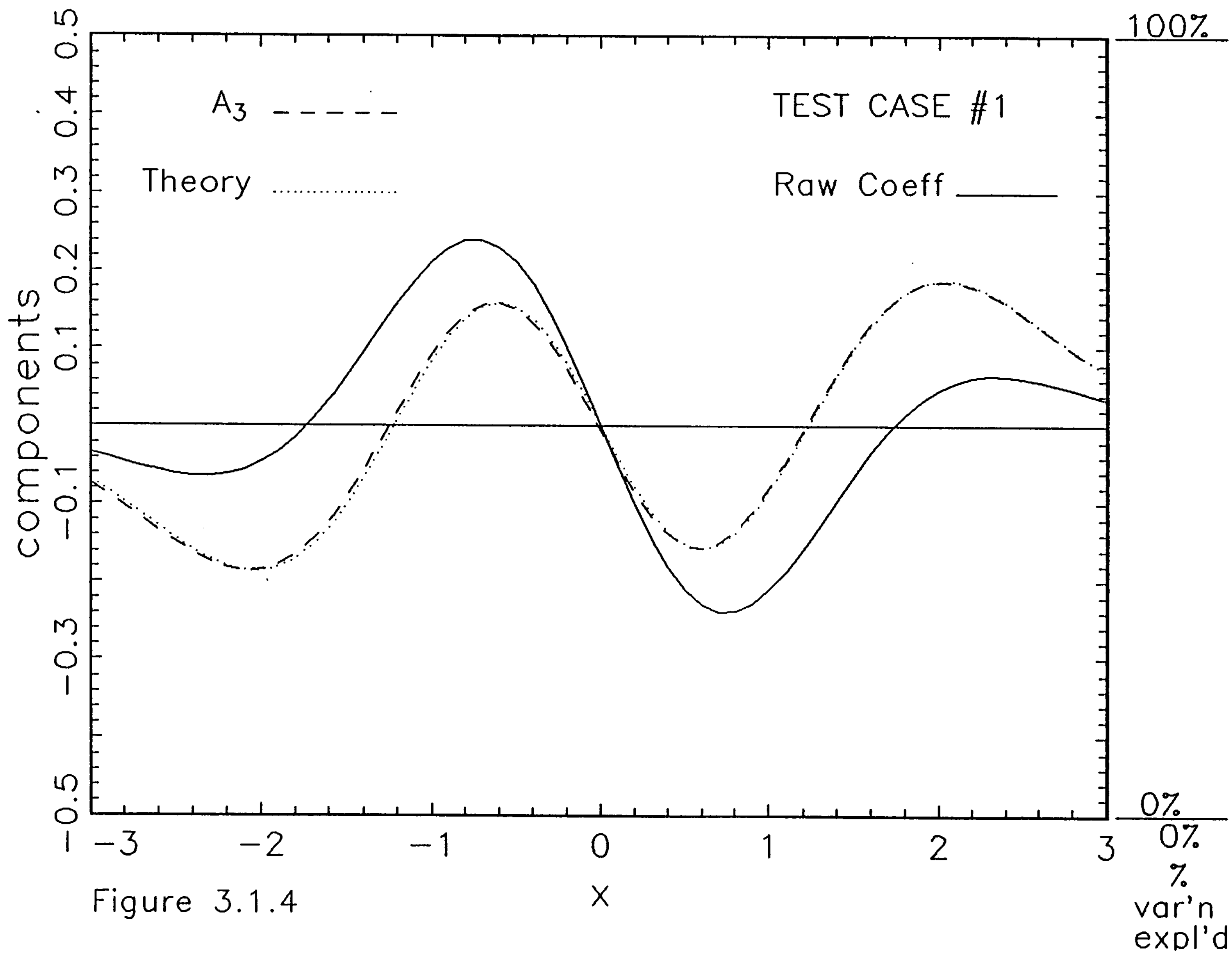
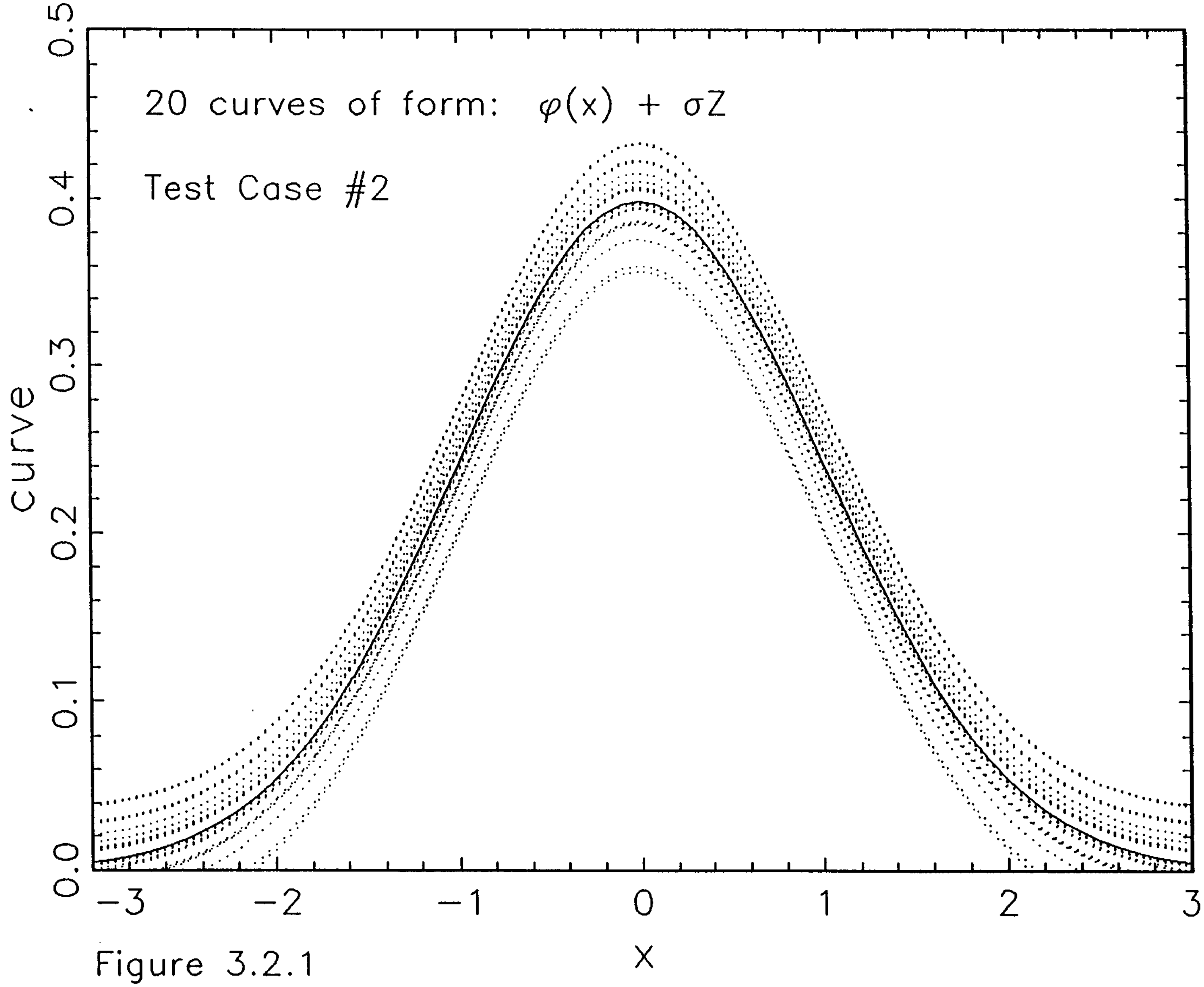

components
0.5
0.4
0.3
0.2
0.1
-0.1
-0.3
-0.5
A_2
Theory
TEST CASE #1
-3
-2
-1
0
1
2
3
X
100%
2%
0%
%
var'n
expl'd

Figure 3.1.3



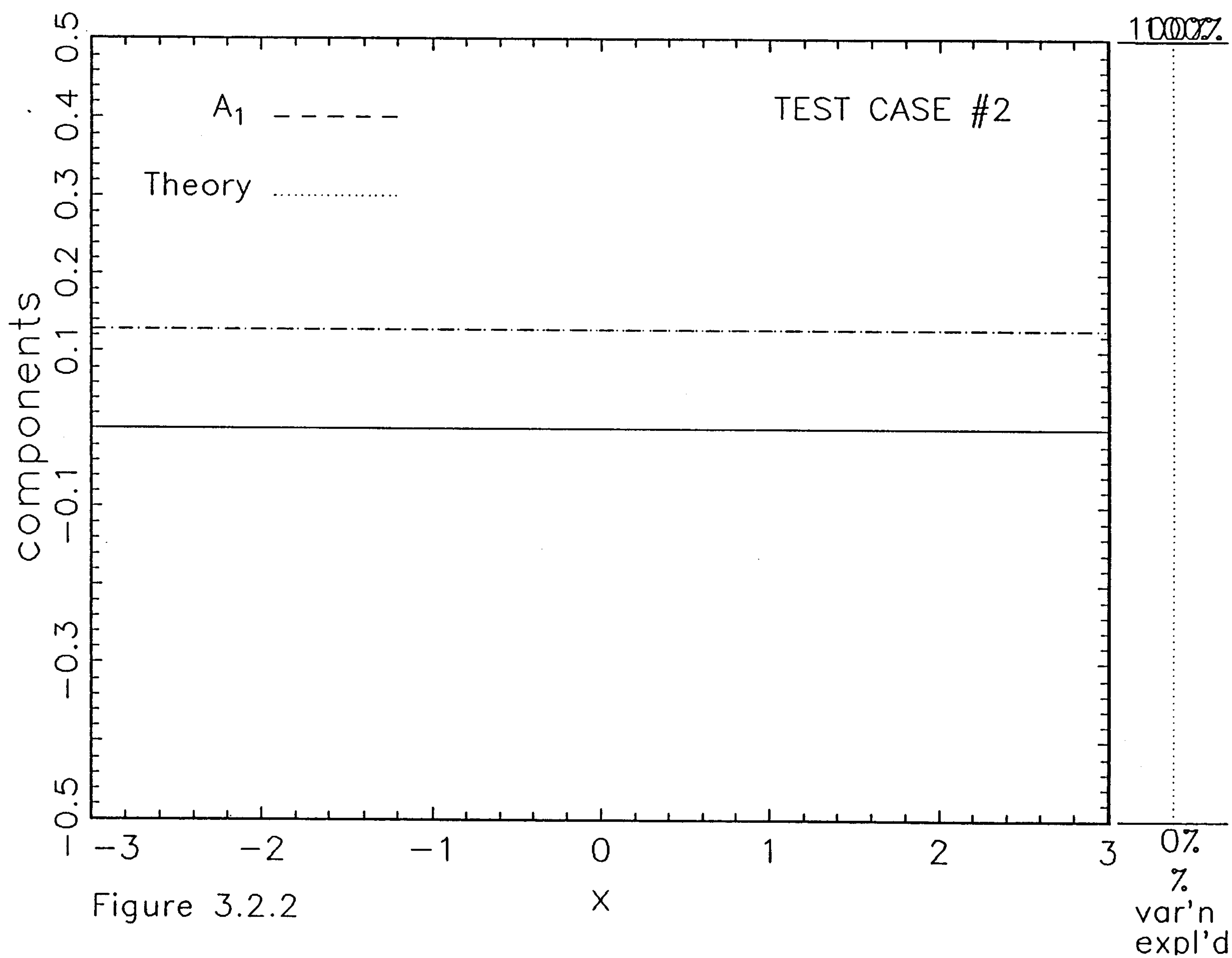
A3
Theory
TEST CASE #1
Raw Coeff
components
0.5
0.4
0.3
0.2
0.1
−0.1
−0.3
−0.5
−3
−2
−1
0
1
2
3
X
100%
0%
0%
%
var'n
expl'd

Figure 3.1.4



20 curves of form: $\varphi(x) + \sigma Z$
Test Case #2
curve
0.0
0.1
0.2
0.3
0.4
0.5
−3
−2
−1
0
1
2
3
x

Figure 3.2.1



TEST CASE #2
A_1
Theory
components
0.5
0.4
0.3
0.2
0.1
−0.1
−0.3
−0.5
−3
−2
−1
0
1
2
3
X
100%
0%
%
var'n
expl'd

Figure 3.2.2

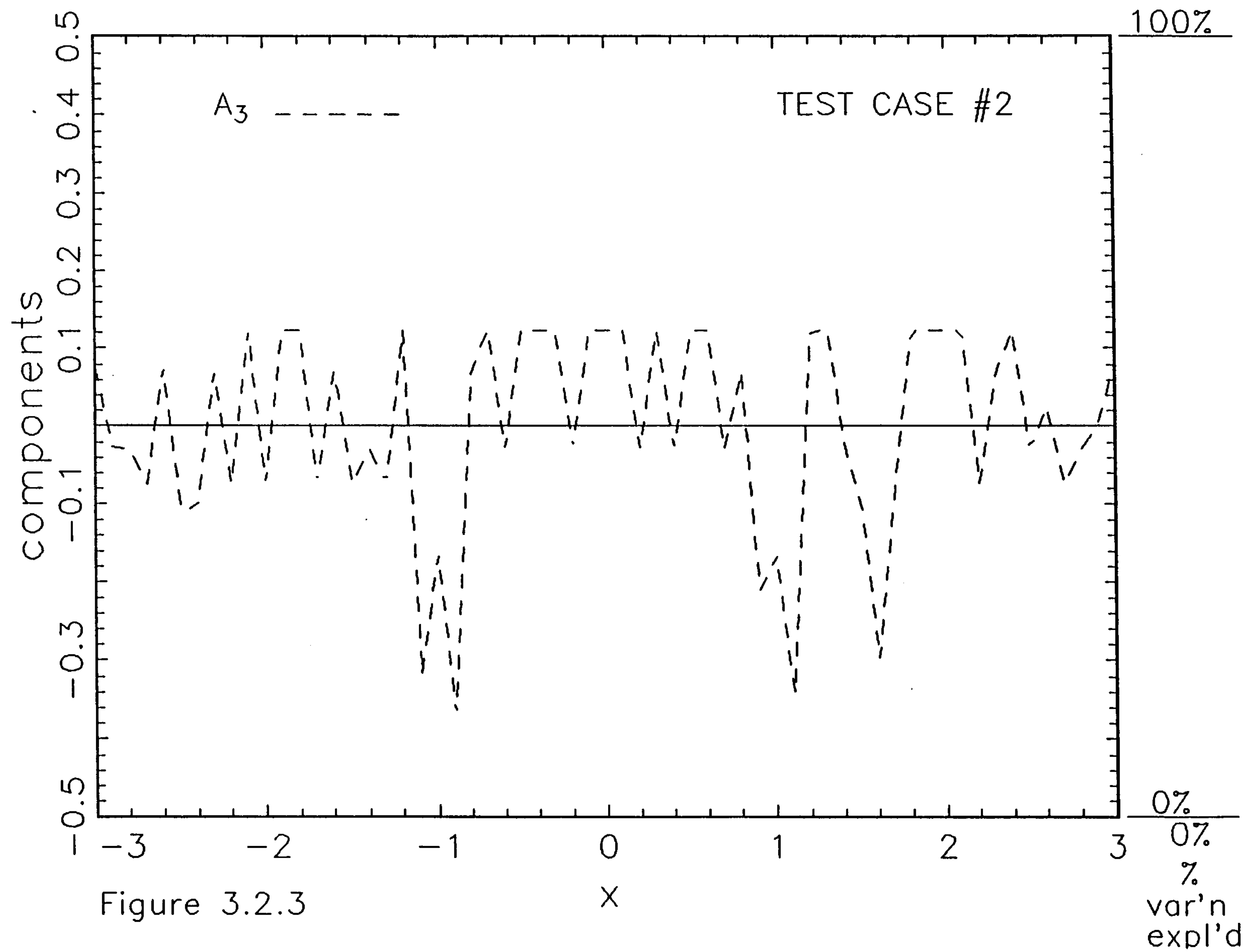

TEST CASE #2
A_3
components
0.5
0.4
0.3
0.2
0.1
−0.1
−0.3
−0.5
−3
−2
−1
0
1
2
3
x
100%
0%
0%
%
var'n
expl'd

Figure 3.2.3

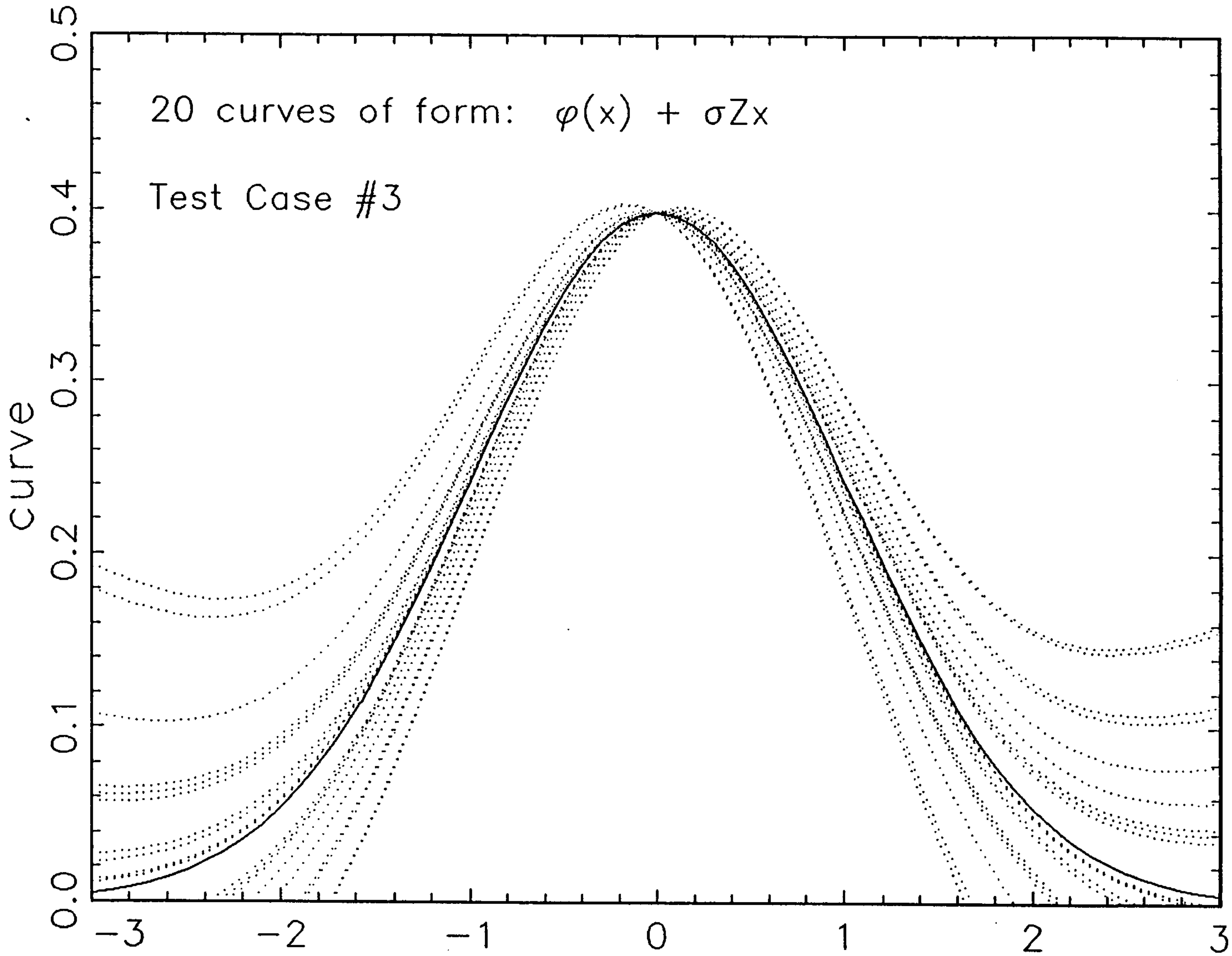
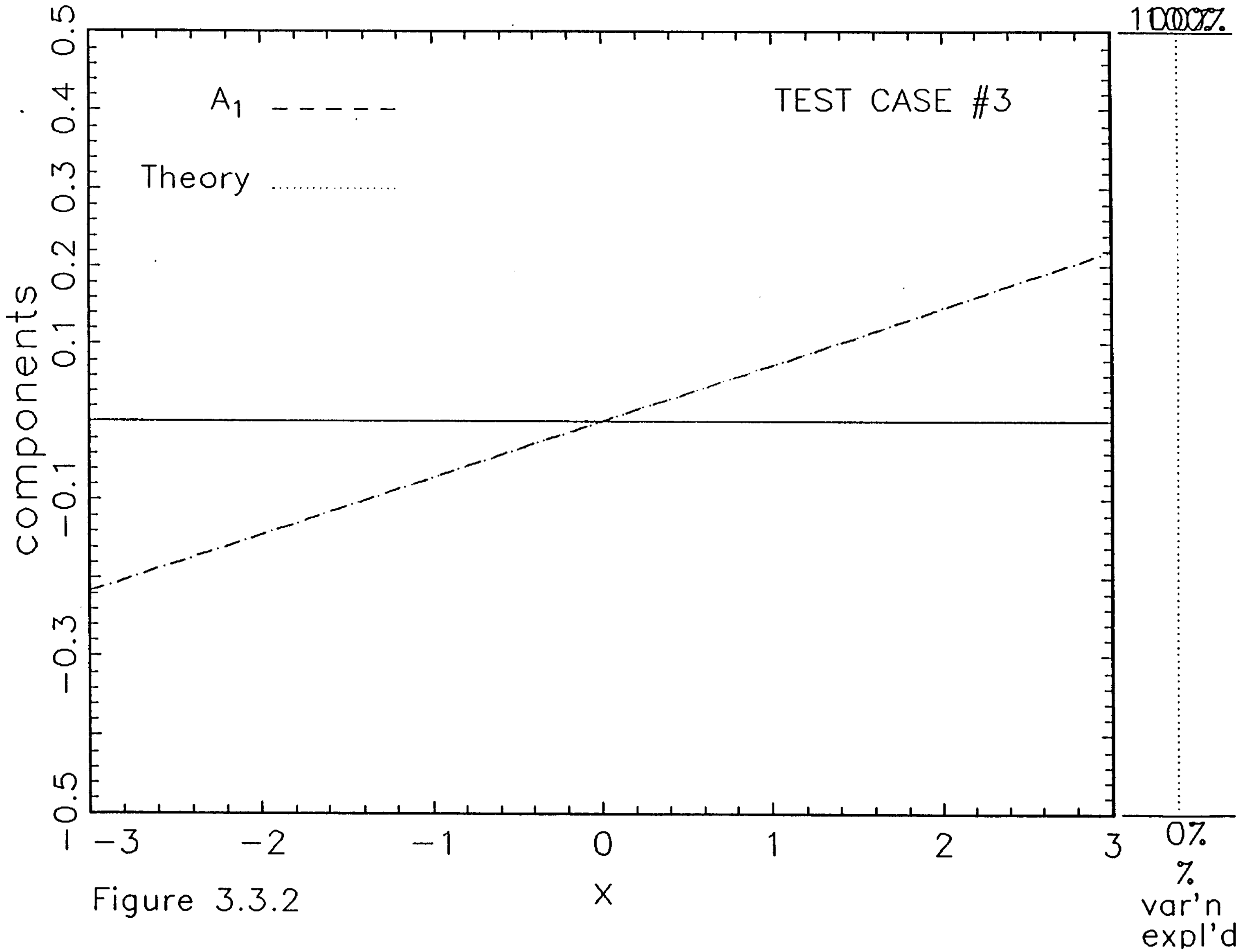
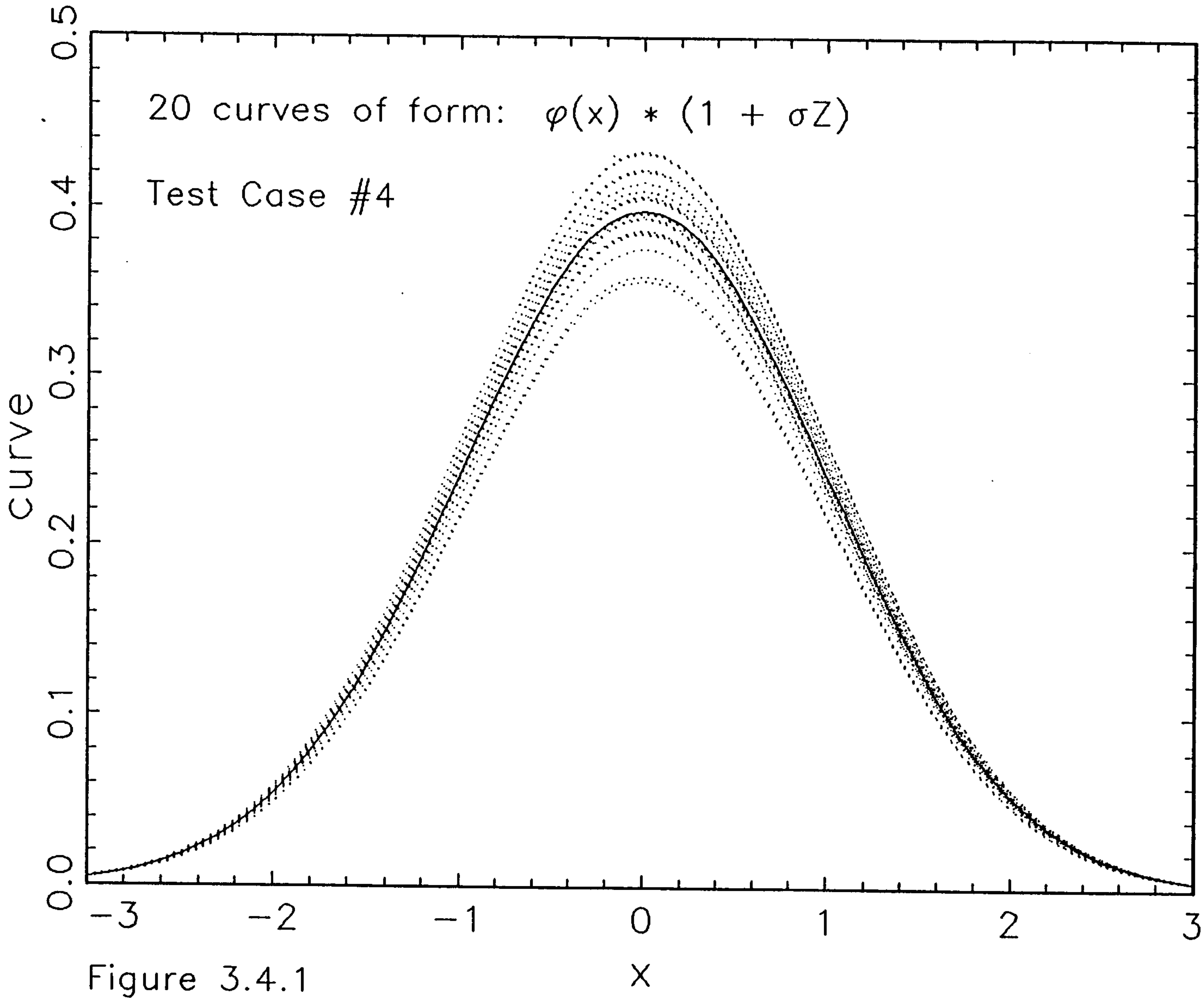

20 curves of form: φ(x) + σZx
Test Case #3
curve
0.0
0.1
0.2
0.3
0.4
0.5
−3
−2
−1
0
1
2
3
x

Figure 3.3.1



components
0.5
0.4
0.3
0.2
0.1
−0.1
−0.3
−0.5
A_1
Theory
TEST CASE #3
−3
−2
−1
0
1
2
3
X
100%
0%
%
var'n
expl'd

Figure 3.3.2


20 curves of form: $\varphi(x) * (1 + \sigma Z)$
Test Case #4
curve
0.0
0.1
0.2
0.3
0.4
0.5
−3
−2
−1
0
1
2
3
X
Figure 3.4.1

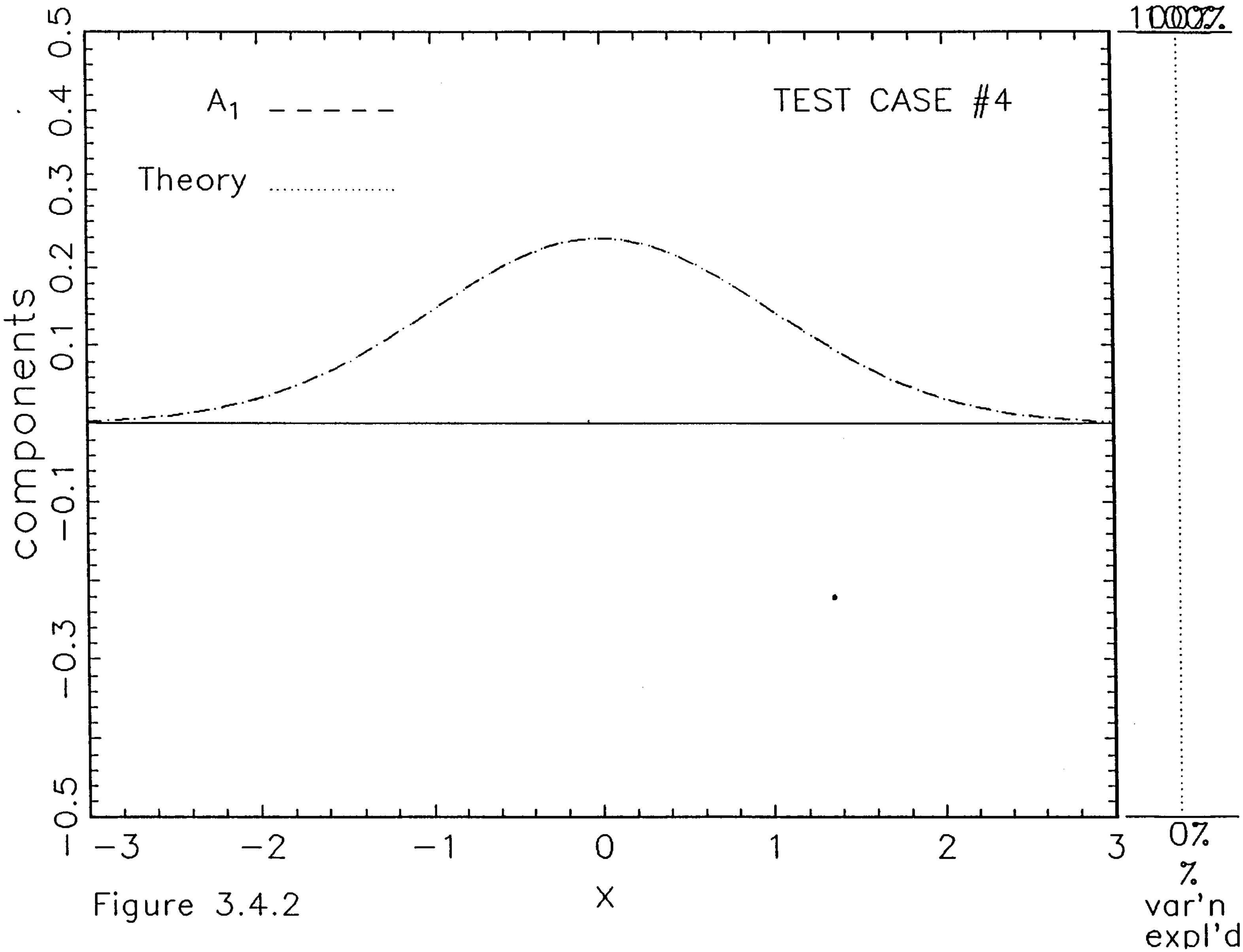
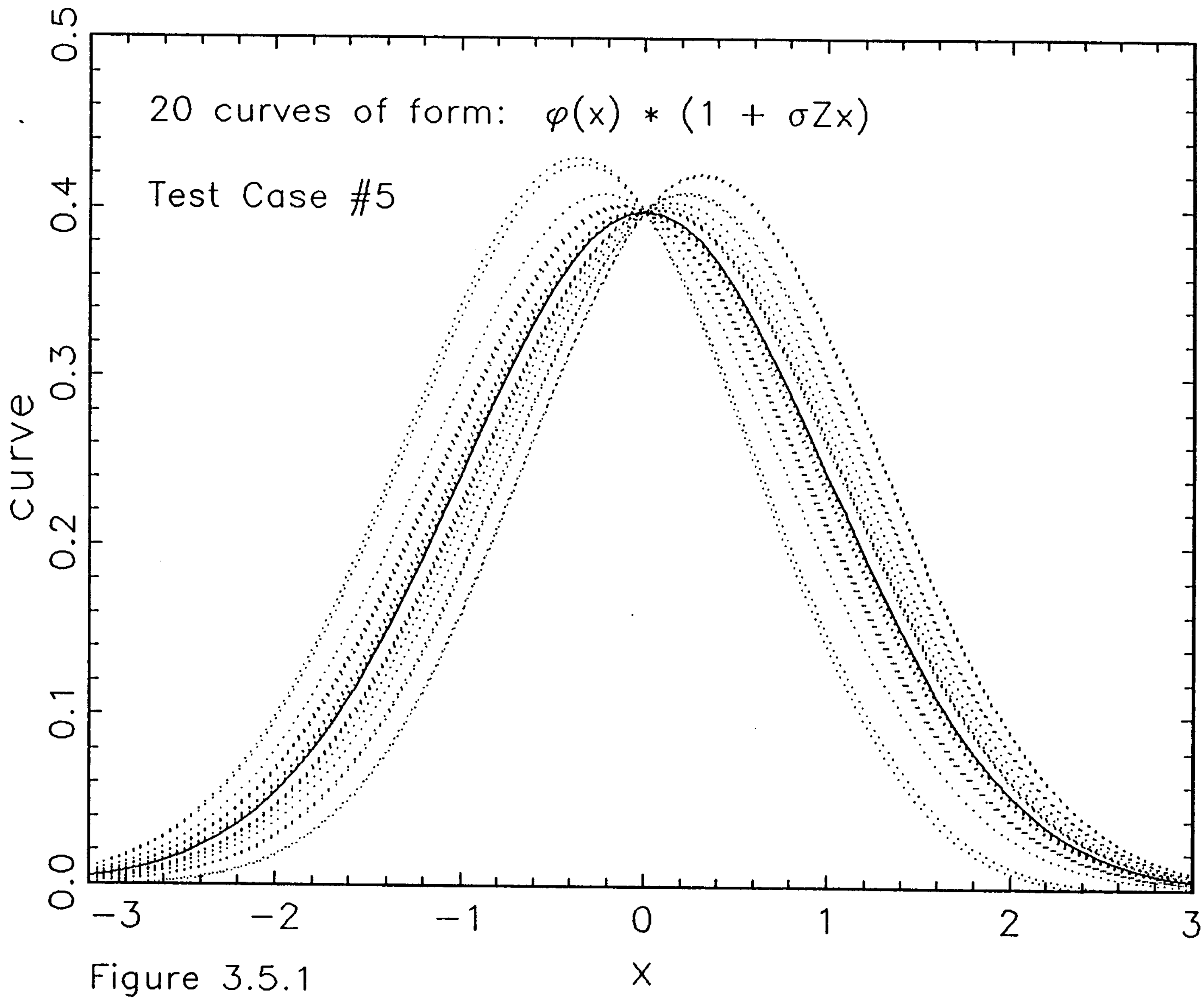
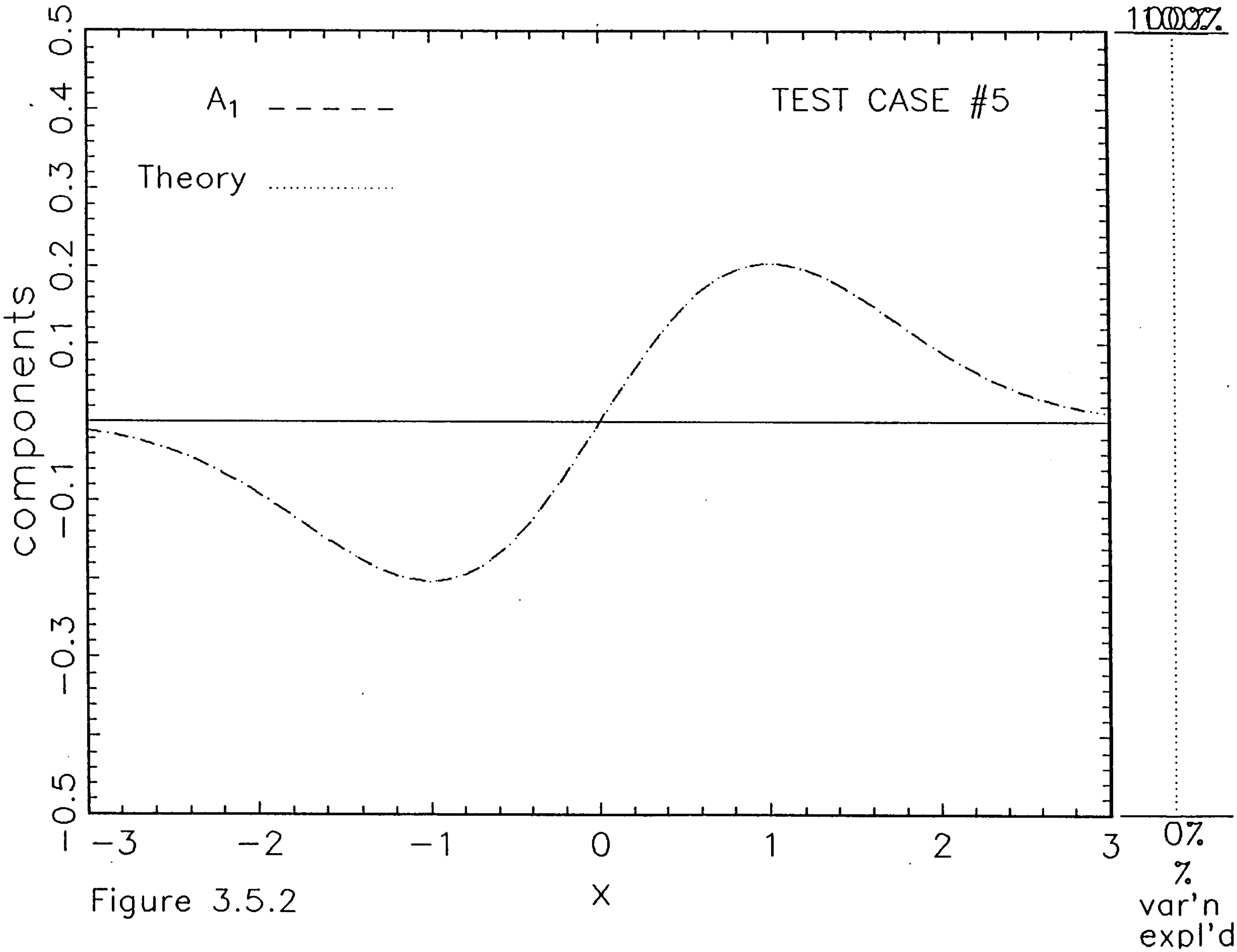

components
-0.5
-0.3
-0.1
0.1
0.2
0.3
0.4
0.5
A$_1$
Theory
TEST CASE #4
-3
-2
-1
0
1
2
3
x
0%
% var'n expl'd

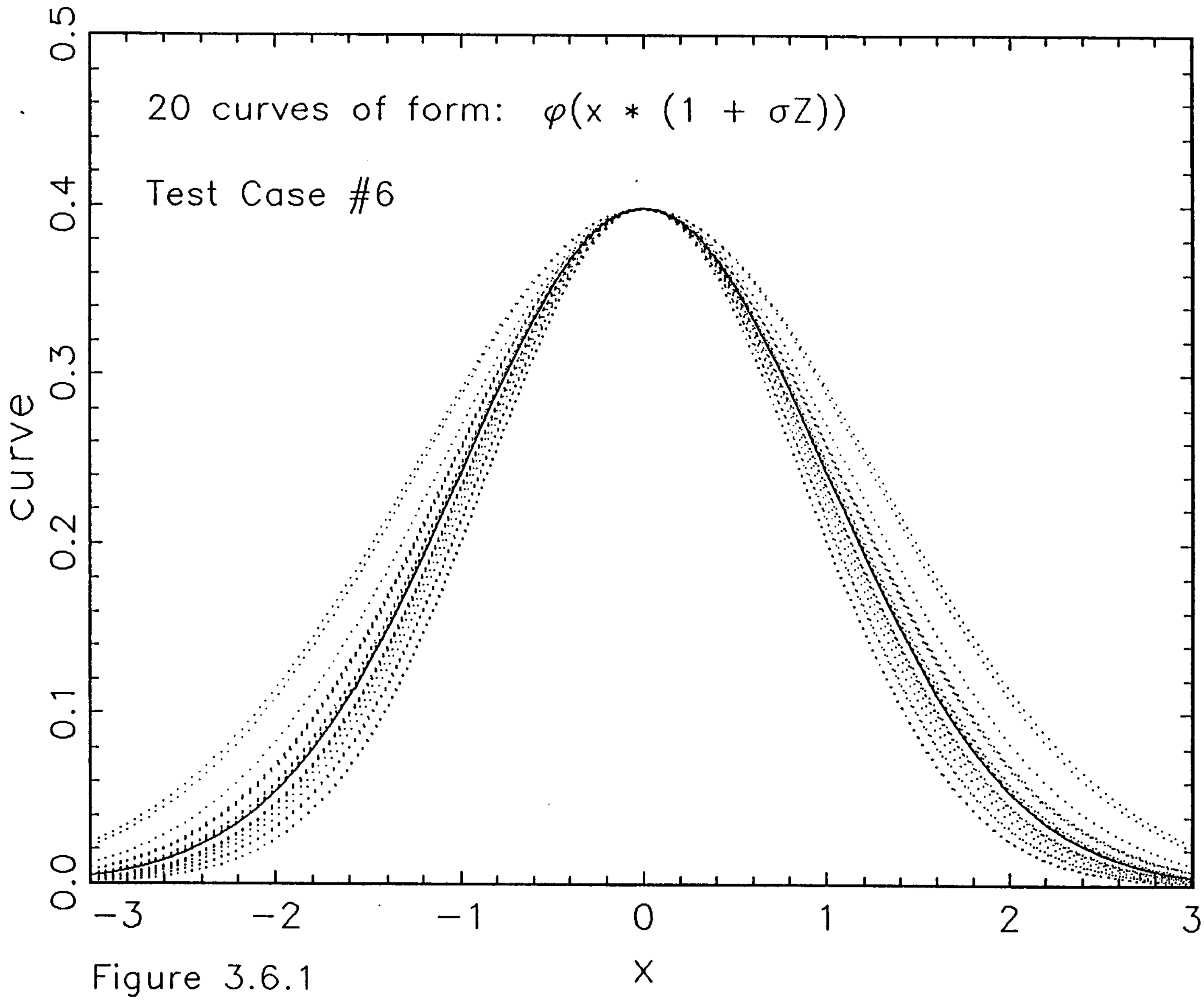
Figure 3.4.2


20 curves of form: φ(x) * (1 + σZx)
Test Case #5
curve
0.0
0.1
0.2
0.3
0.4
0.5
−3
−2
−1
0
1
2
3
x
Figure 3.5.1



A1
Theory
TEST CASE #5
components
0.5
0.4
0.3
0.2
0.1
−0.1
−0.3
−0.5
−3
−2
−1
0
1
2
3
x
100%
0%
%
var'n
expl'd

Figure 3.5.2


20 curves of form: φ(x * (1 + σZ))
Test Case #6
curve
0.0
0.1
0.2
0.3
0.4
0.5
−3
−2
−1
0
1
2
3
x
Figure 3.6.1

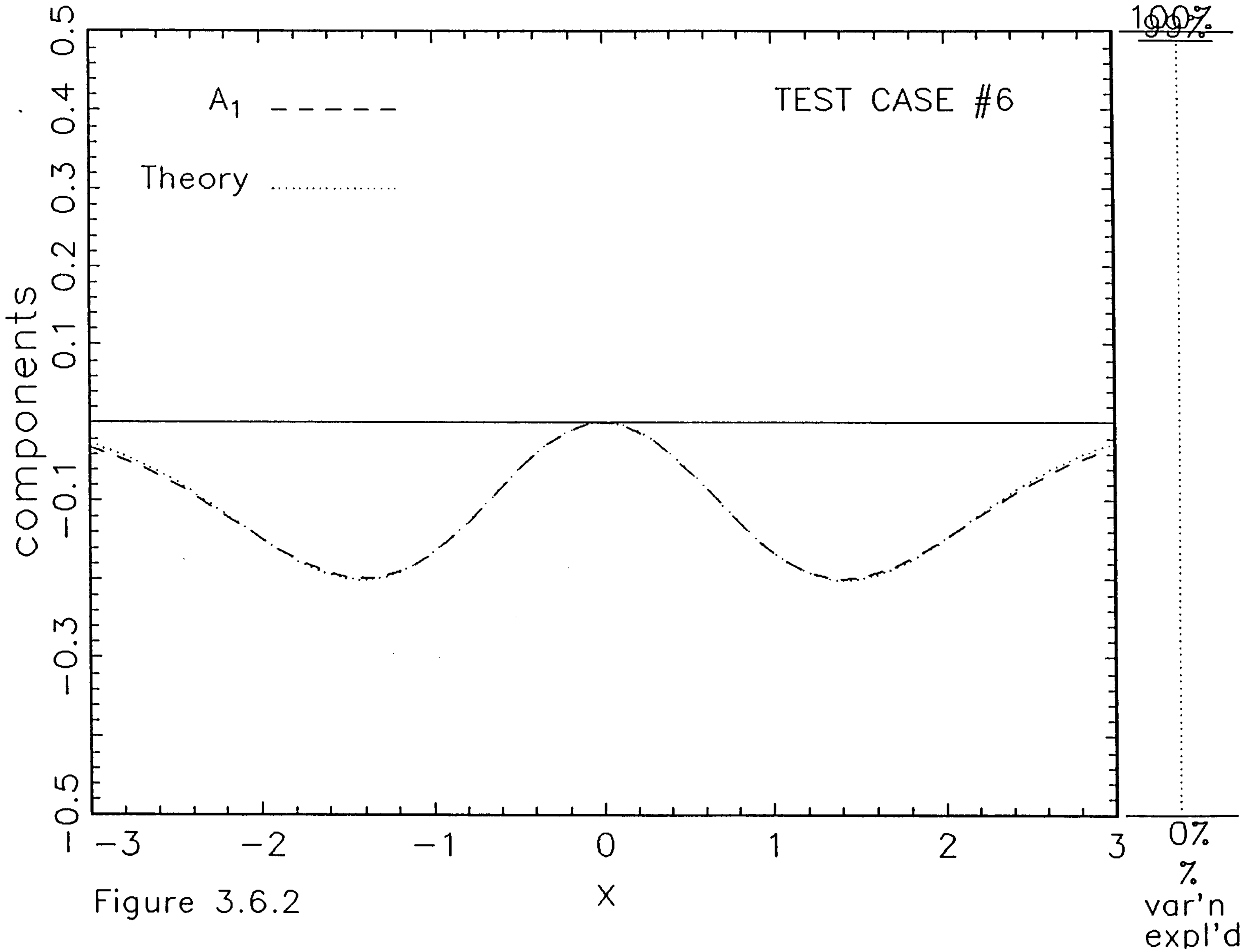

components
x
A_1
Theory
TEST CASE #6
0.5
0.4
0.3
0.2
0.1
−0.1
−0.3
−0.5
−3
−2
−1
0
1
2
3
100%
99%
0%
% var'n expl'd

Figure 3.6.2

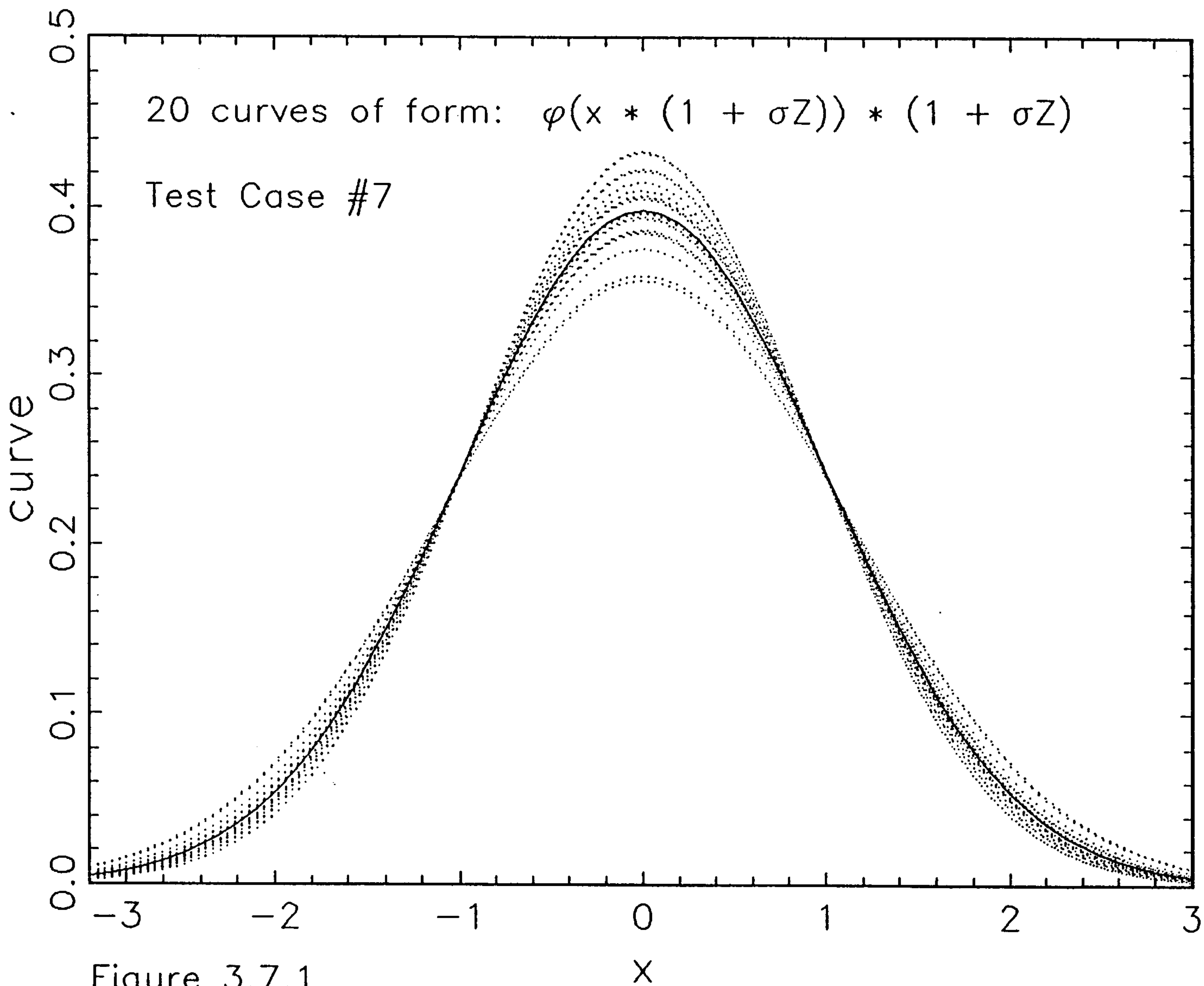
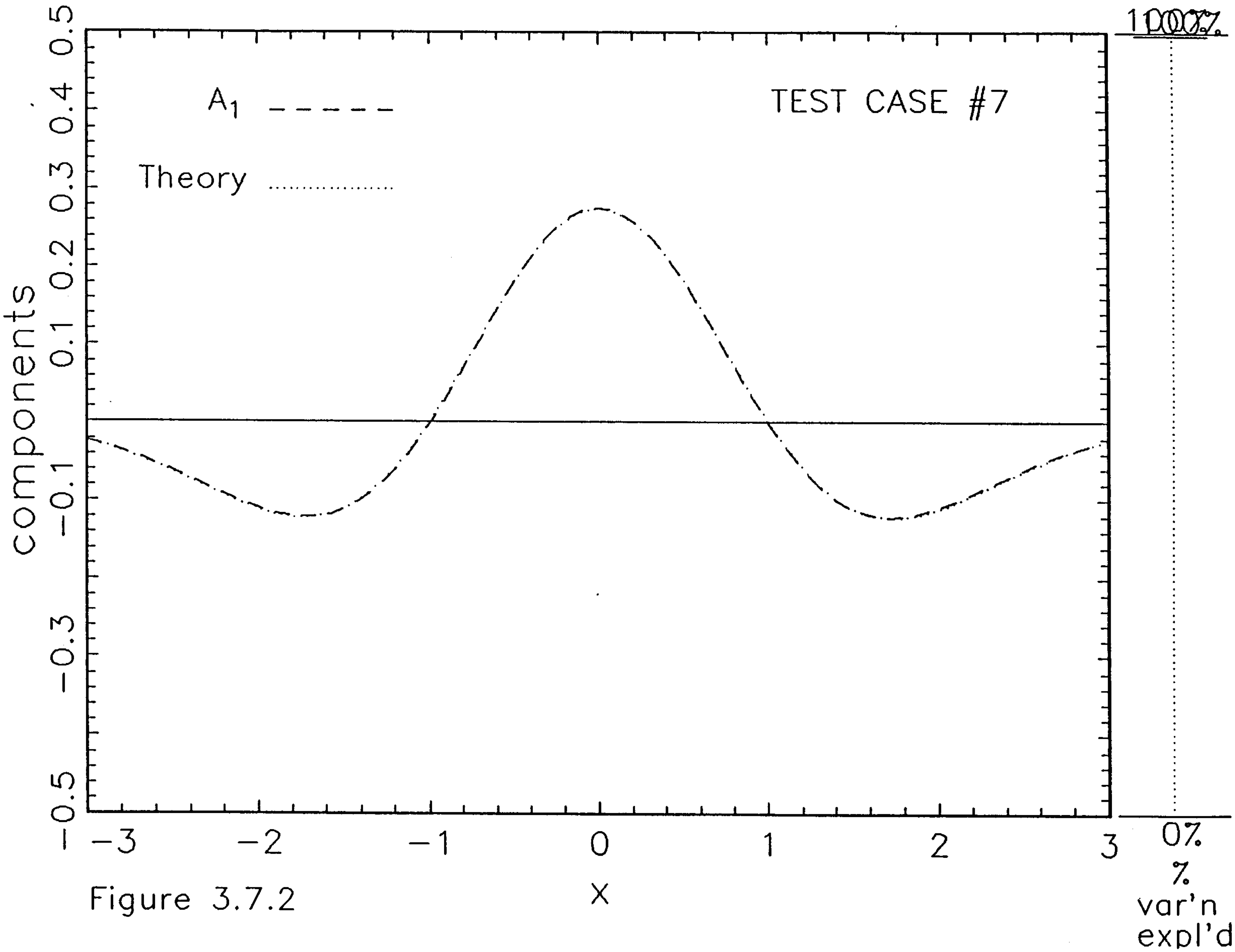

20 curves of form: φ(x * (1 + σZ)) * (1 + σZ)
Test Case #7
curve
0.0
0.1
0.2
0.3
0.4
0.5
−3
−2
−1
0
1
2
3
X

Figure 3.7.1



components
0.5
0.4
0.3
0.2
0.1
−0.1
−0.3
−0.5
A1
Theory
TEST CASE #7
−3
−2
−1
0
1
2
3
X
100%
0%
%
var'n
expl'd

Figure 3.7.2

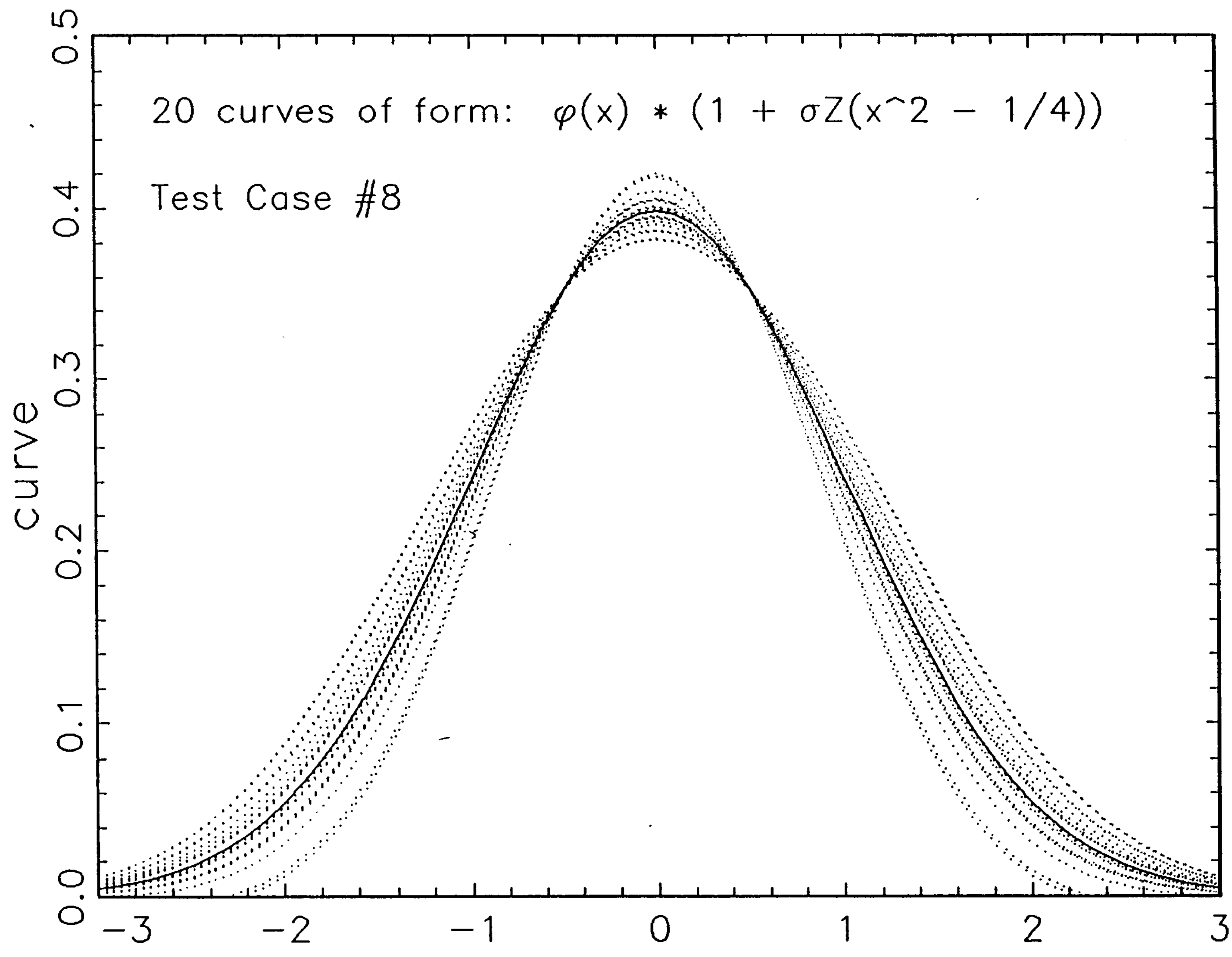
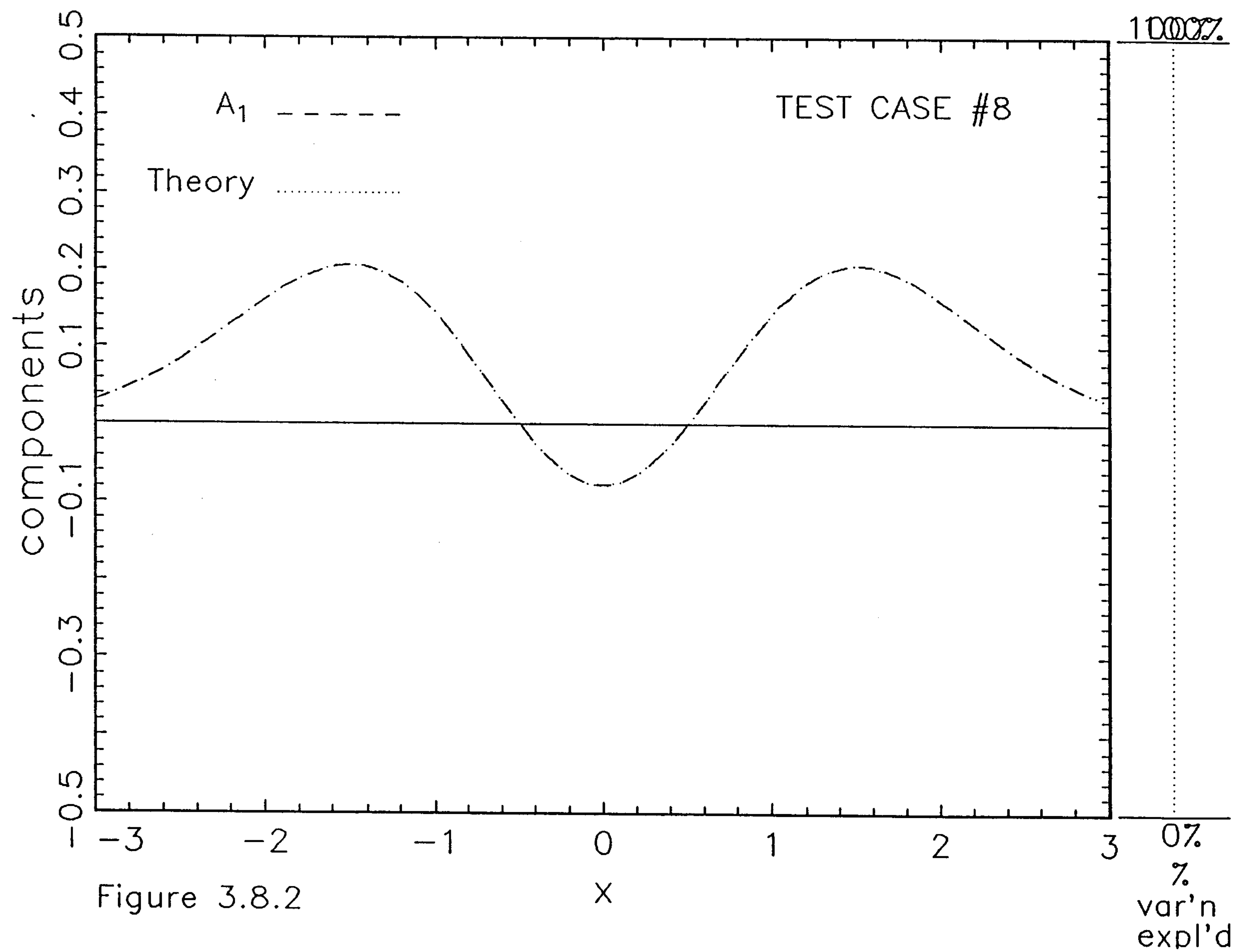
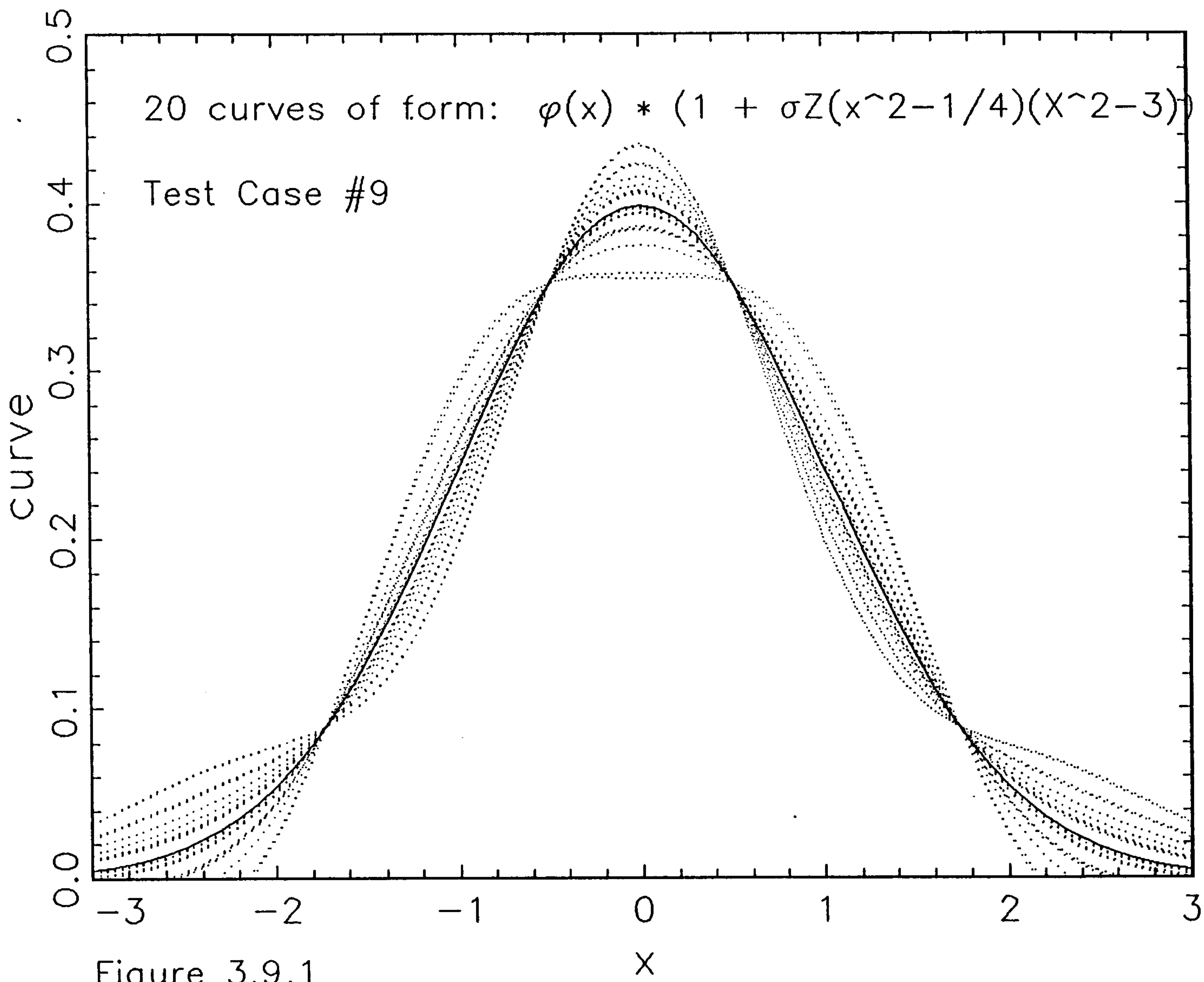

20 curves of form: φ(x) * (1 + σZ(x^2 - 1/4))
Test Case #8
curve
0.0
0.1
0.2
0.3
0.4
0.5
-3
-2
-1
0
1
2
3
x

Figure 3.8.1



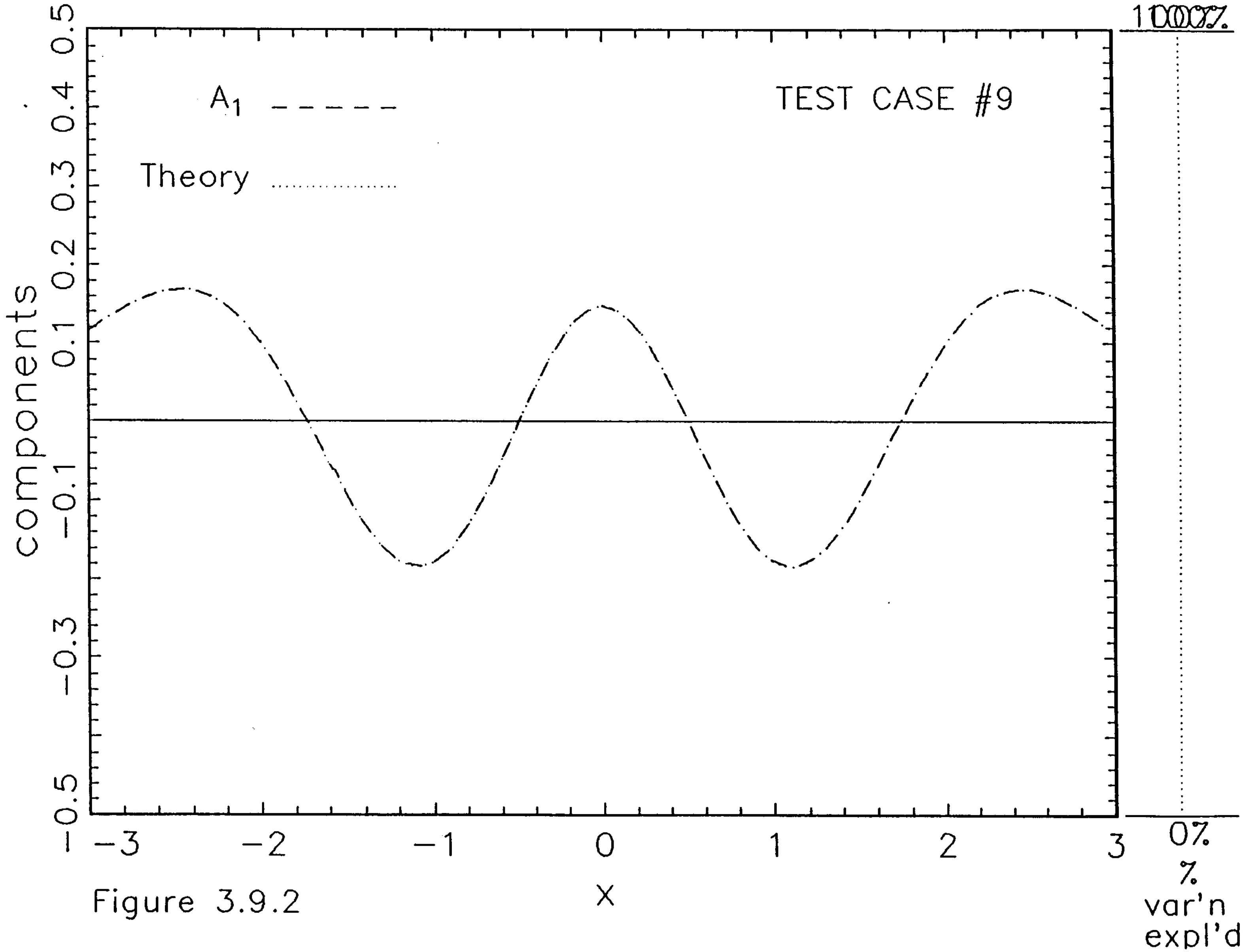
A1 ----
Theory ....
TEST CASE #8
components
0.5
0.4
0.3
0.2
0.1
-0.1
-0.3
-0.5
-3
-2
-1
0
1
2
3
X
100%
0%
% var'n expl'd

Figure 3.8.2



20 curves of form: φ(x) * (1 + σZ(x^2−1/4)(X^2−3))
Test Case #9
curve
0.0
0.1
0.2
0.3
0.4
0.5
−3
−2
−1
0
1
2
3
X

Figure 3.9.1



components
0.5
0.4
0.3
0.2
0.1
0.1
−0.1
−0.3
−0.5
A1
Theory
TEST CASE #9
−3
−2
−1
0
1
2
3
X
0%
%
var'n
expl'd

Figure 3.9.2

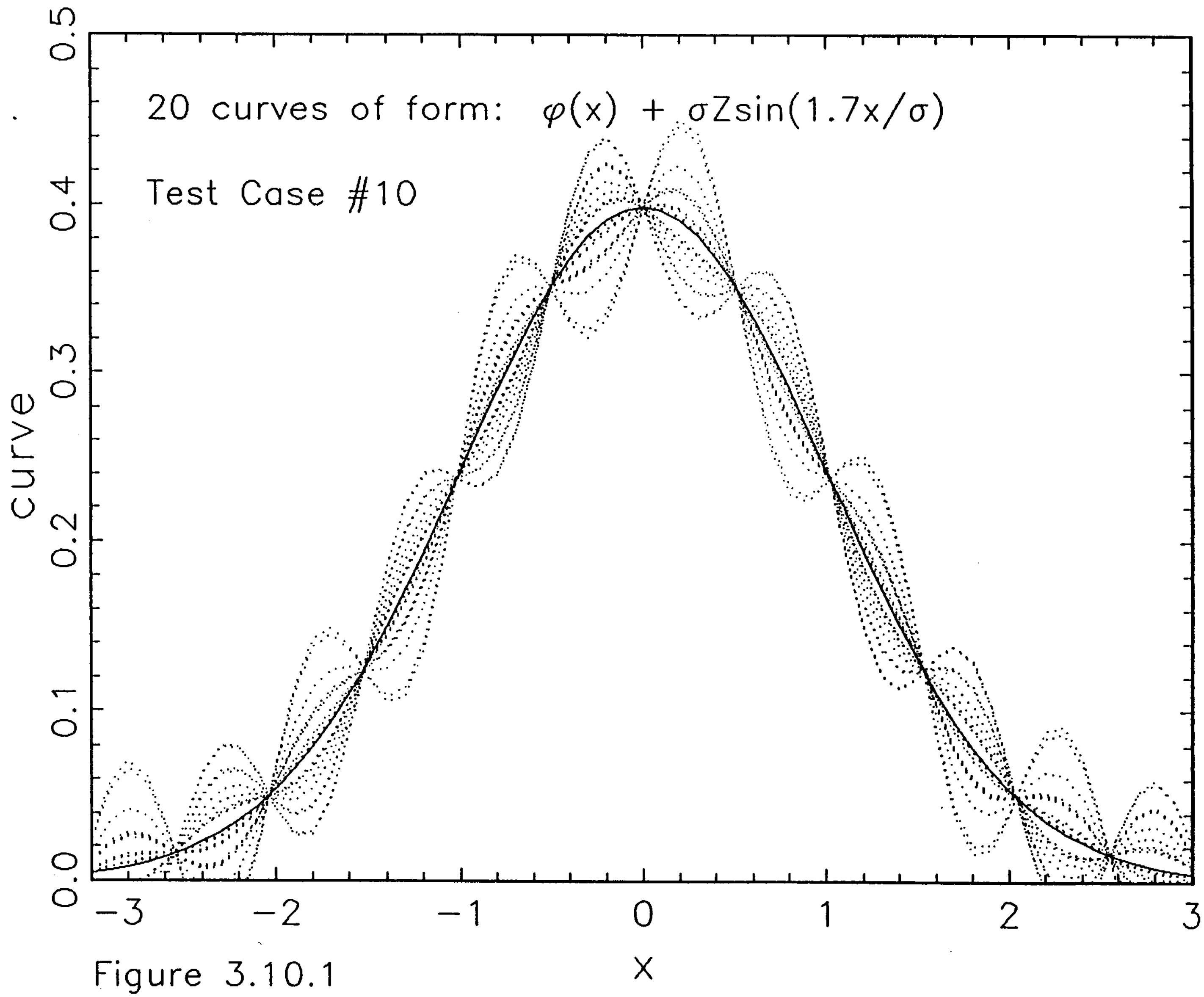
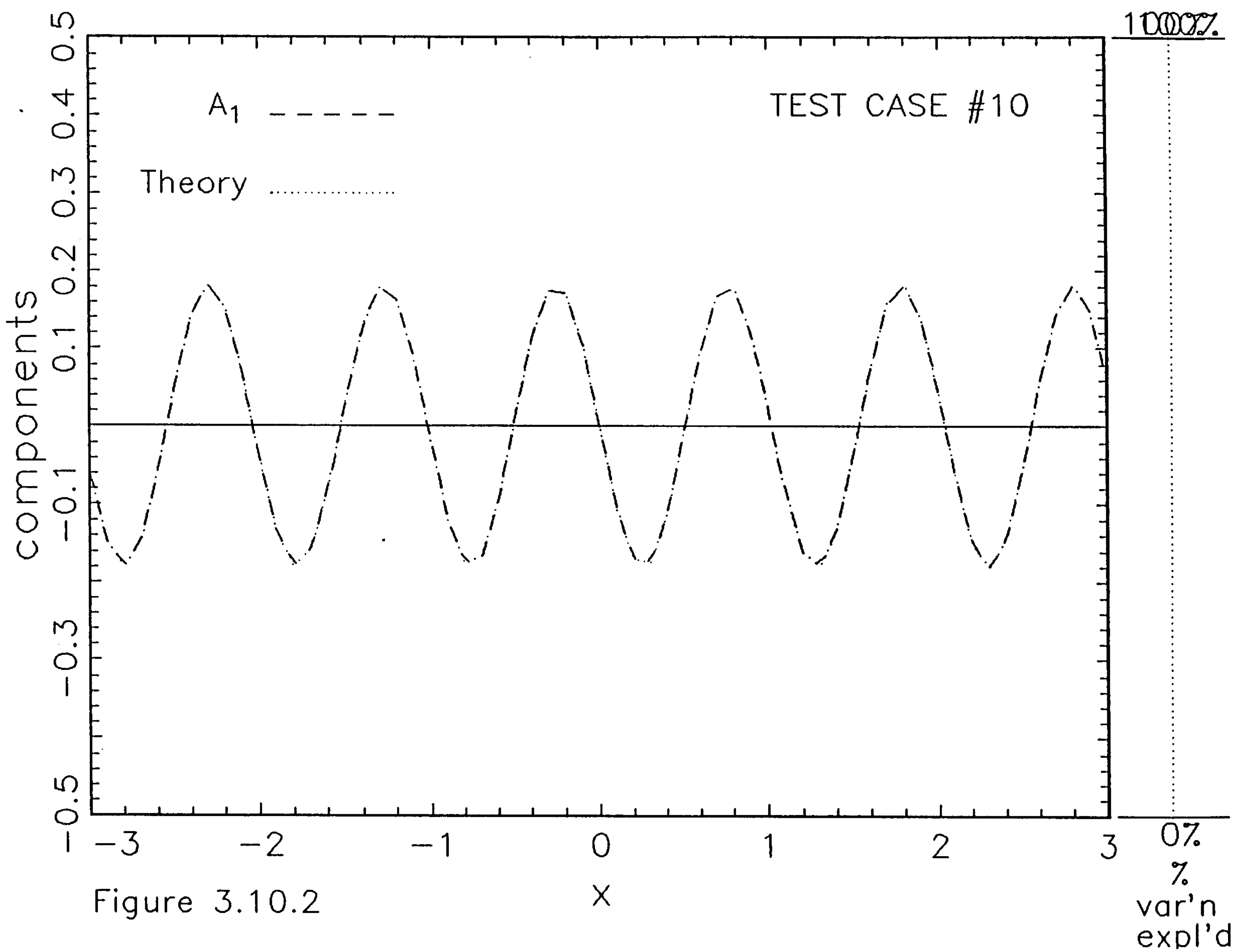
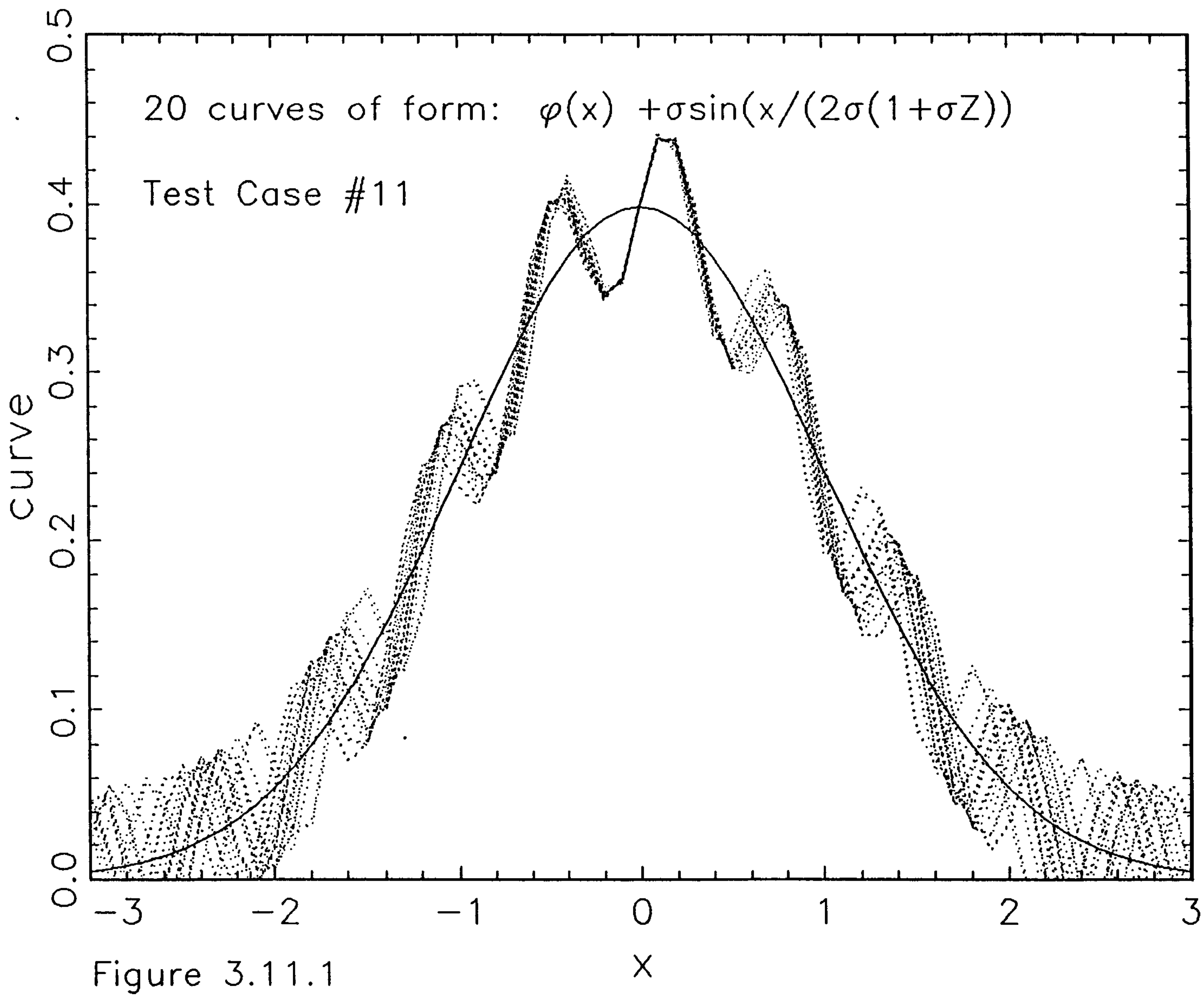

20 curves of form: φ(x) + σZsin(1.7x/σ)
Test Case #10
curve
0.0
0.1
0.2
0.3
0.4
0.5
−3
−2
−1
0
1
2
3
x

Figure 3.10.1



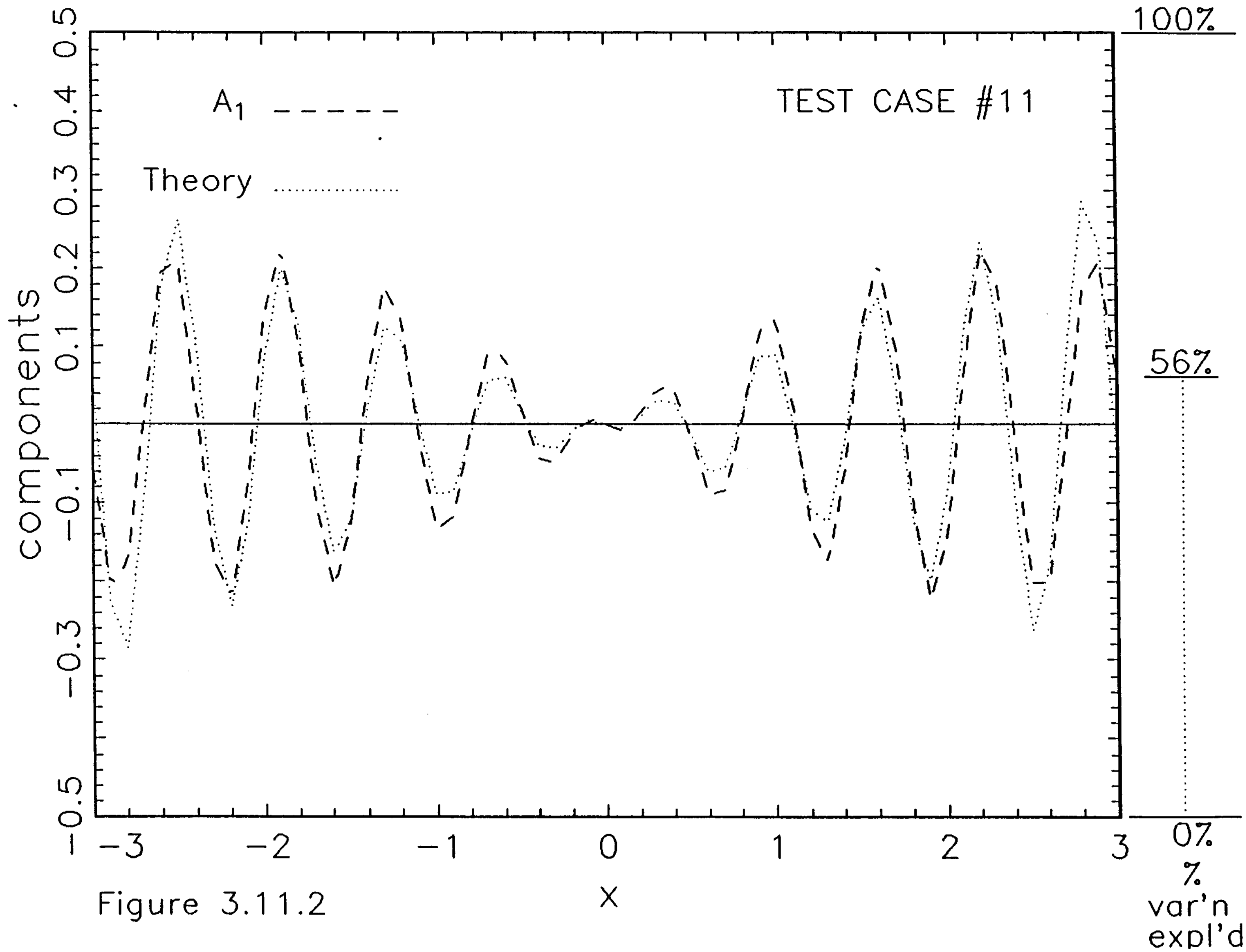
A1
Theory
TEST CASE #10
components
0.5
0.4
0.3
0.2
0.1
−0.1
−0.3
−0.5
−3
−2
−1
0
1
2
3
X
100%
0%
%
var'n
expl'd

Figure 3.10.2



20 curves of form: φ(x) +σsin(x/(2σ(1+σZ))
Test Case #11
curve
0.0
0.1
0.2
0.3
0.4
0.5
−3
−2
−1
0
1
2
3
x

Figure 3.11.1



components
0.5
0.4
0.3
0.2
0.1
−0.1
−0.3
−0.5
A1
Theory
TEST CASE #11
−3
−2
−1
0
1
2
3
x
100%
56%
0%
%
var'n
expl'd

Figure 3.11.2

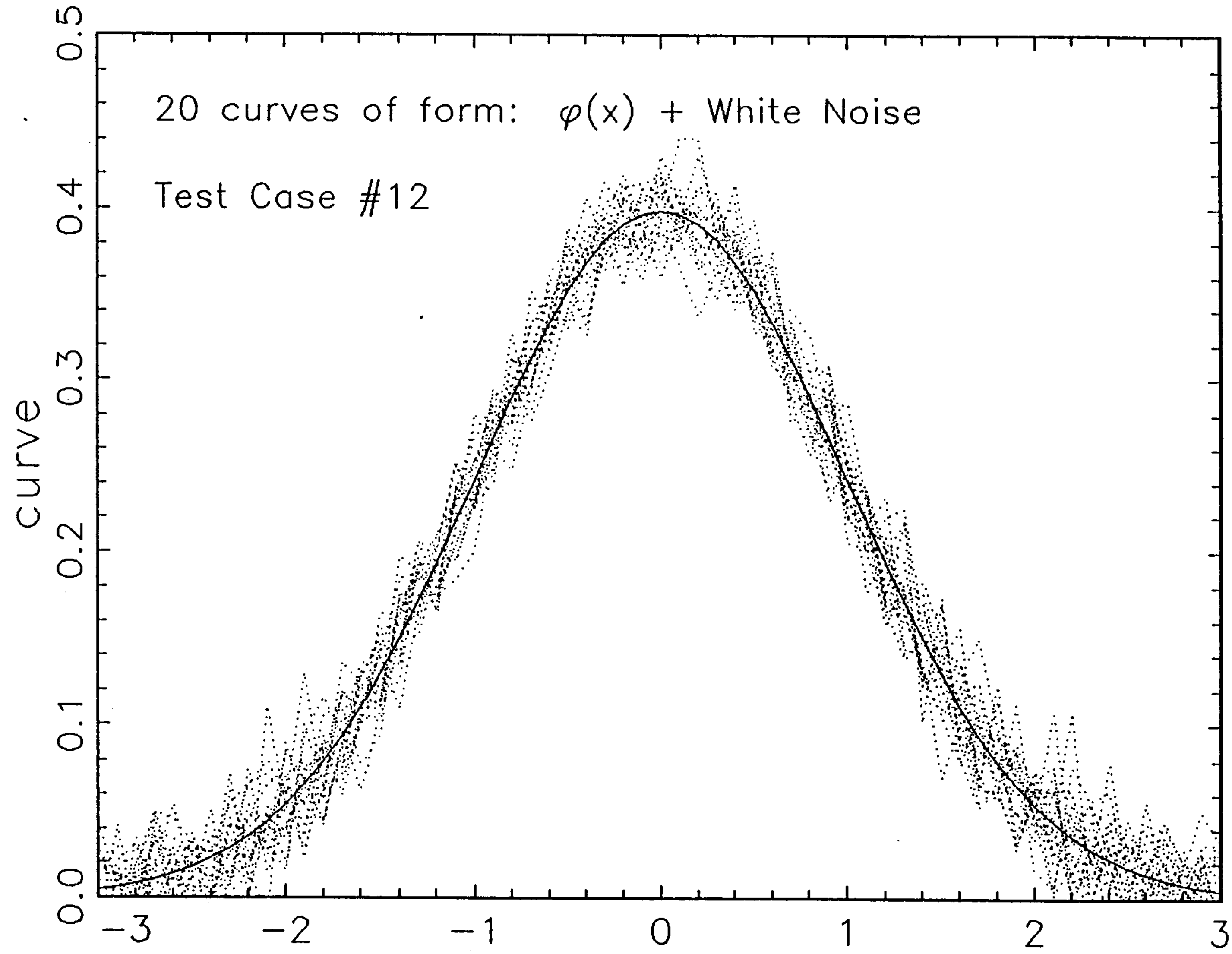

20 curves of form: φ(x) + White Noise
Test Case #12
curve
0.0
0.1
0.2
0.3
0.4
0.5
−3
−2
−1
0
1
2
3
x

Figure 3.12.1

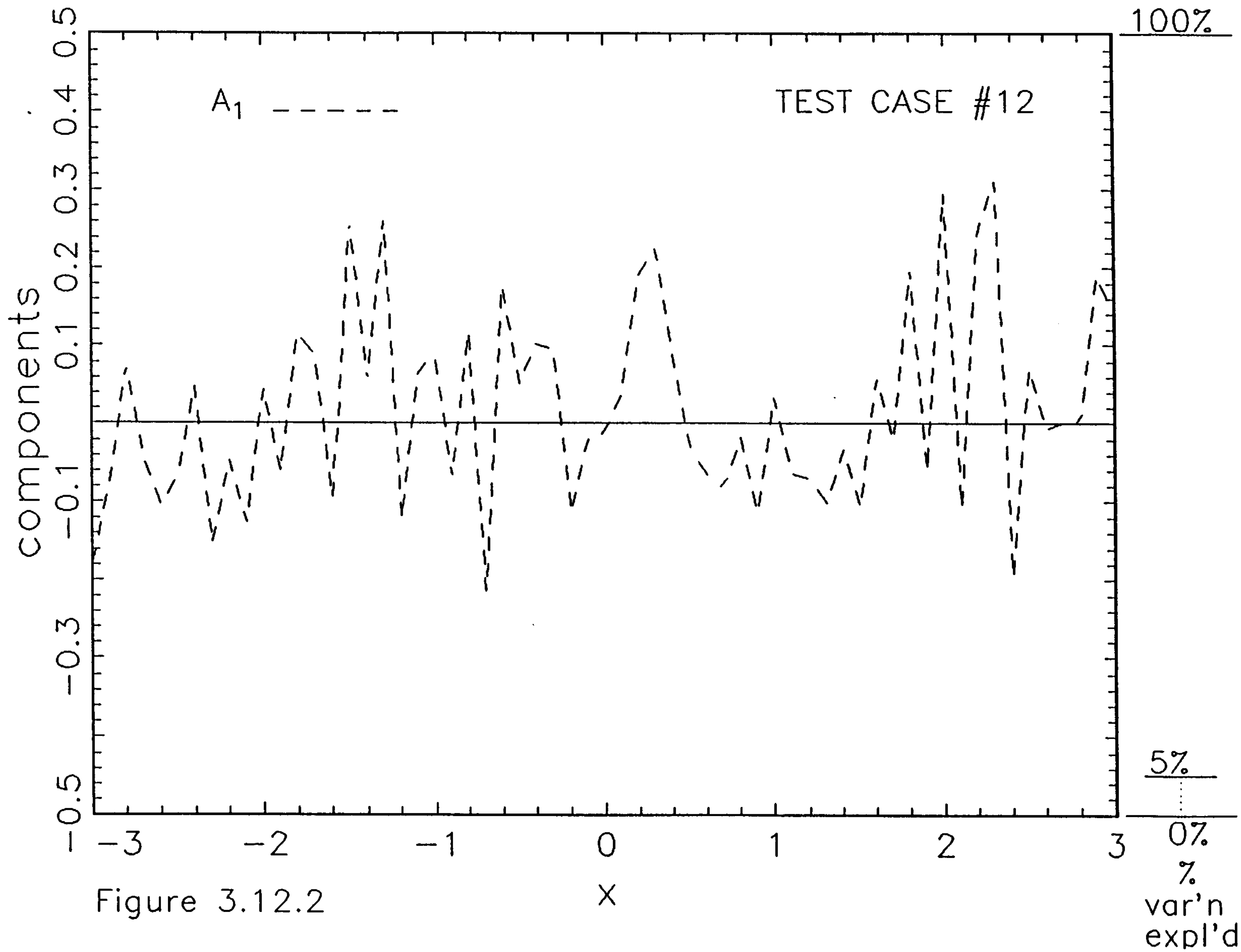
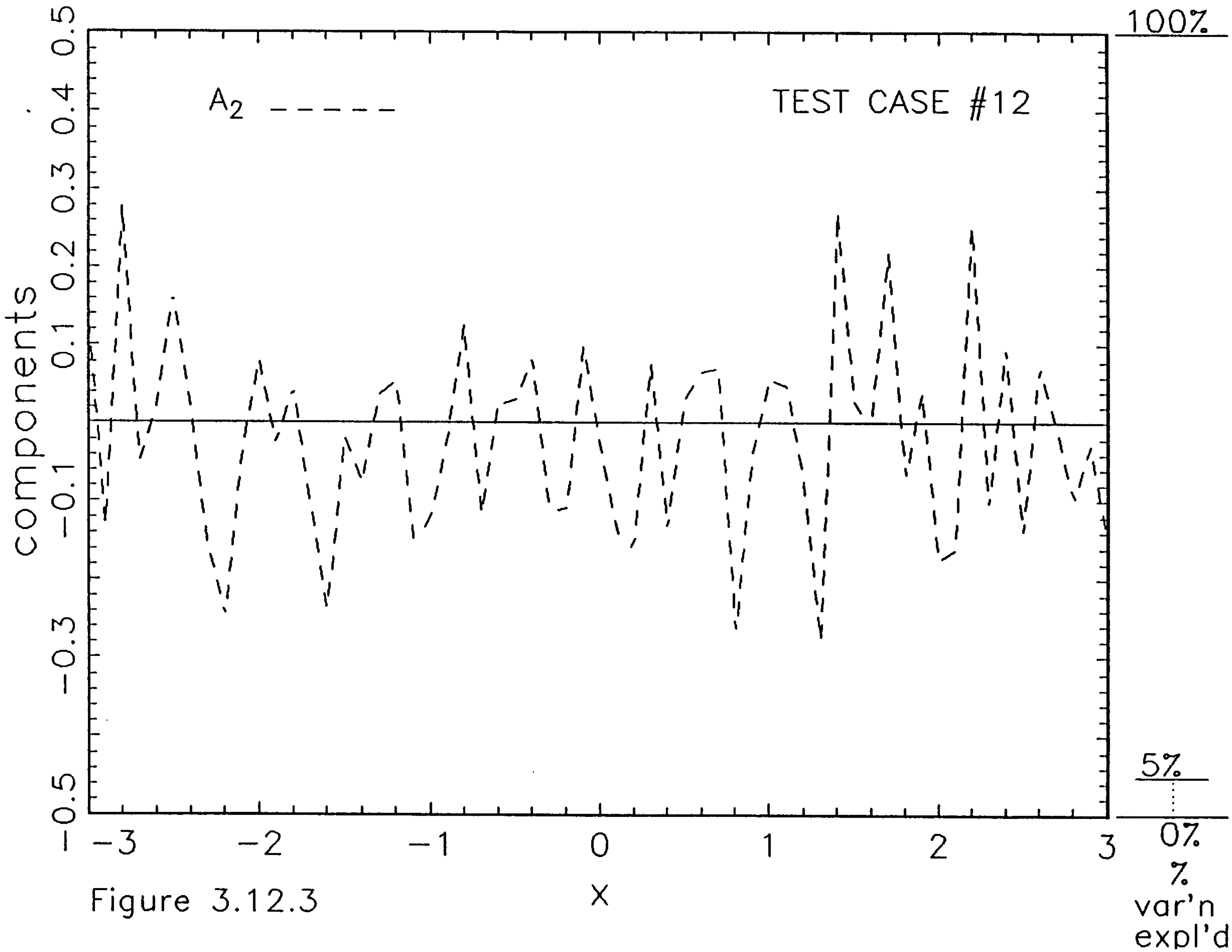
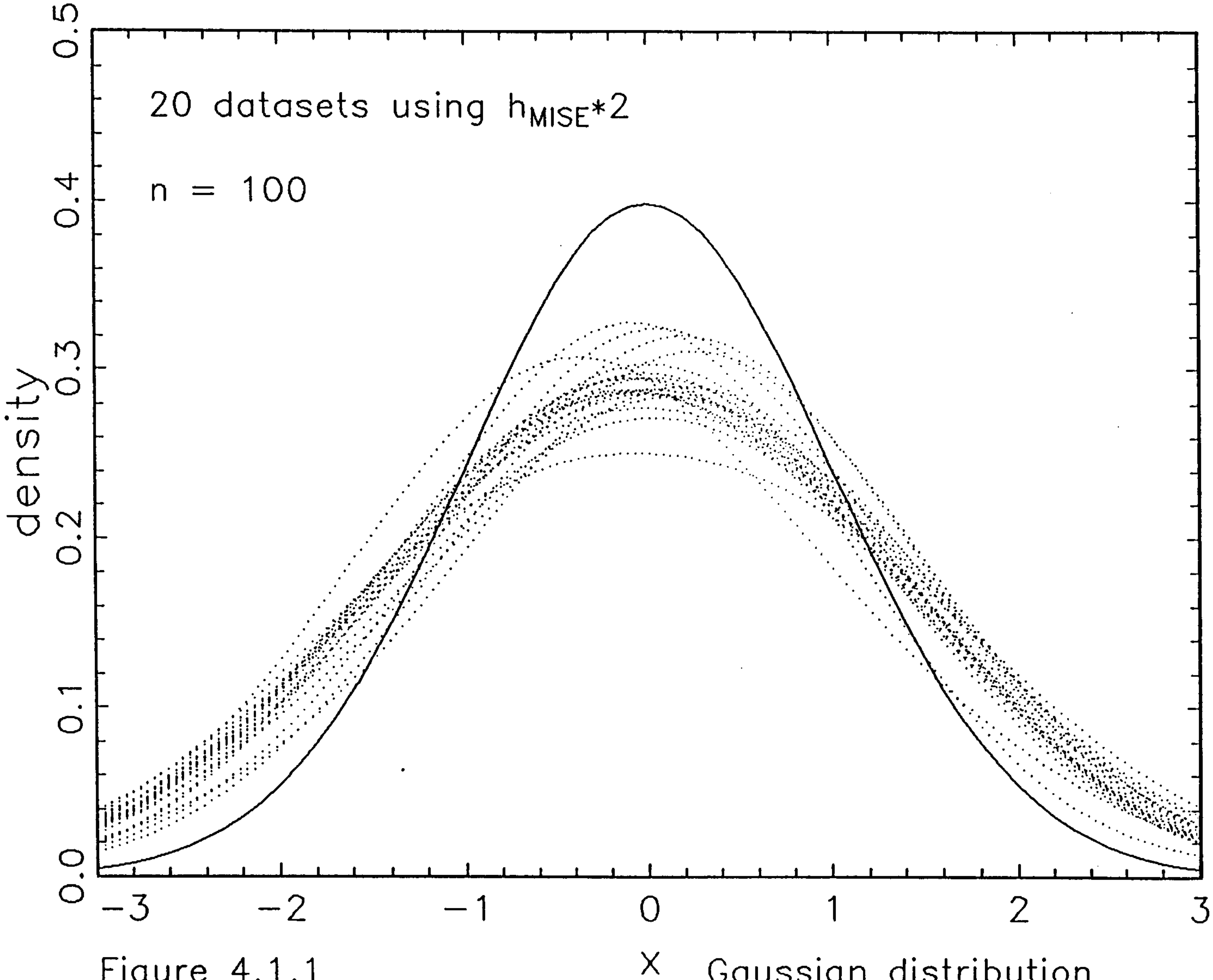

components
0.5
0.4
0.3
0.2
0.1
−0.1
−0.3
−0.5
A_1
TEST CASE #12
−3
−2
−1
0
1
2
3
X
100%
5%
0%
%
var'n
expl'd

Figure 3.12.2



A2 ------
TEST CASE #12
components
0.5
0.4
0.3
0.2
0.1
−0.1
−0.3
−0.5
−3
−2
−1
0
1
2
3
X
100%
5%
0%
%
var'n
expl'd

Figure 3.12.3



20 datasets using h_{MISE}*2
n = 100
density
0.0
0.1
0.2
0.3
0.4
0.5
−3
−2
−1
0
1
2
3
X
Gaussian distribution

Figure 4.1.1

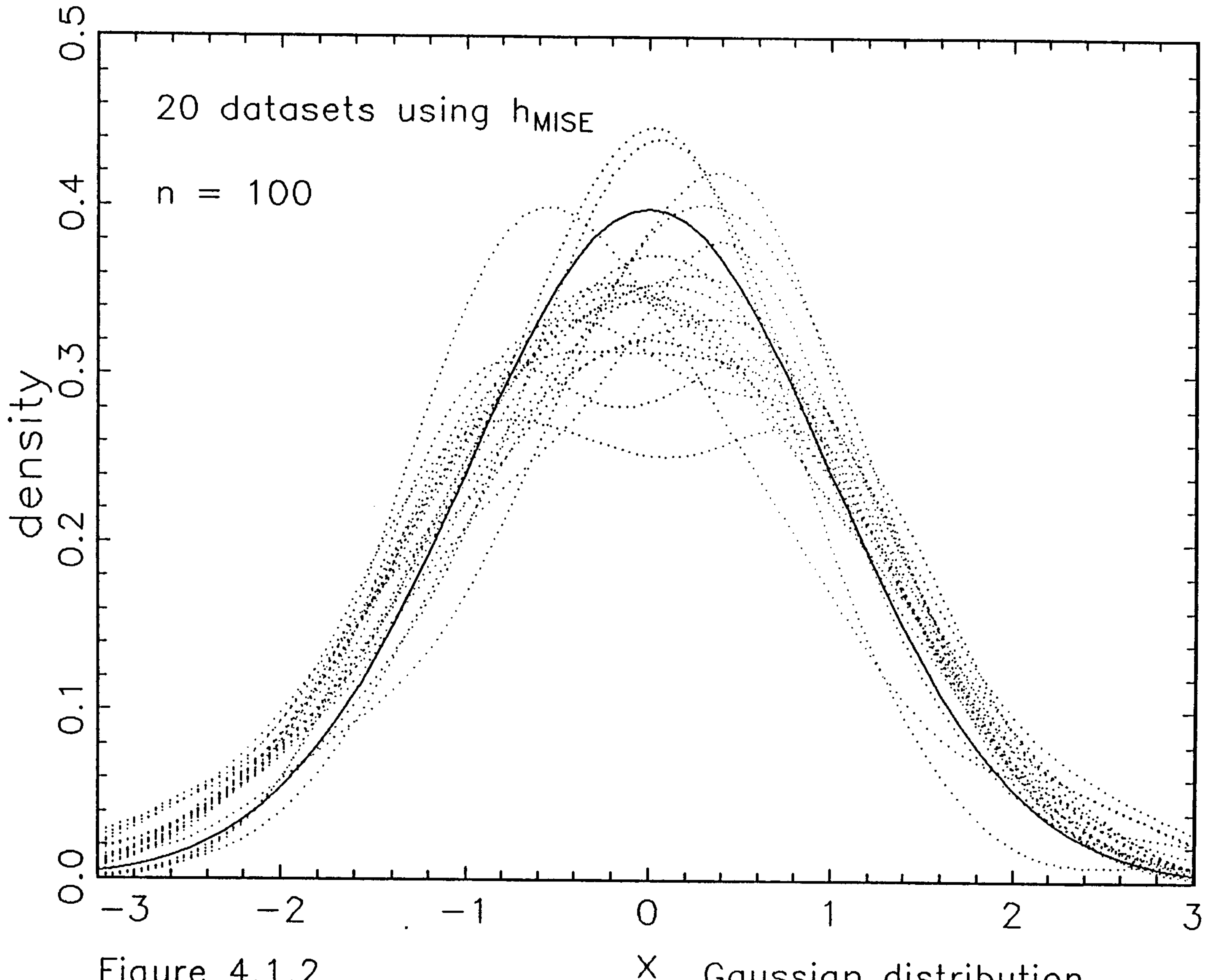


Figure 4.1.2 Gaussian distribution

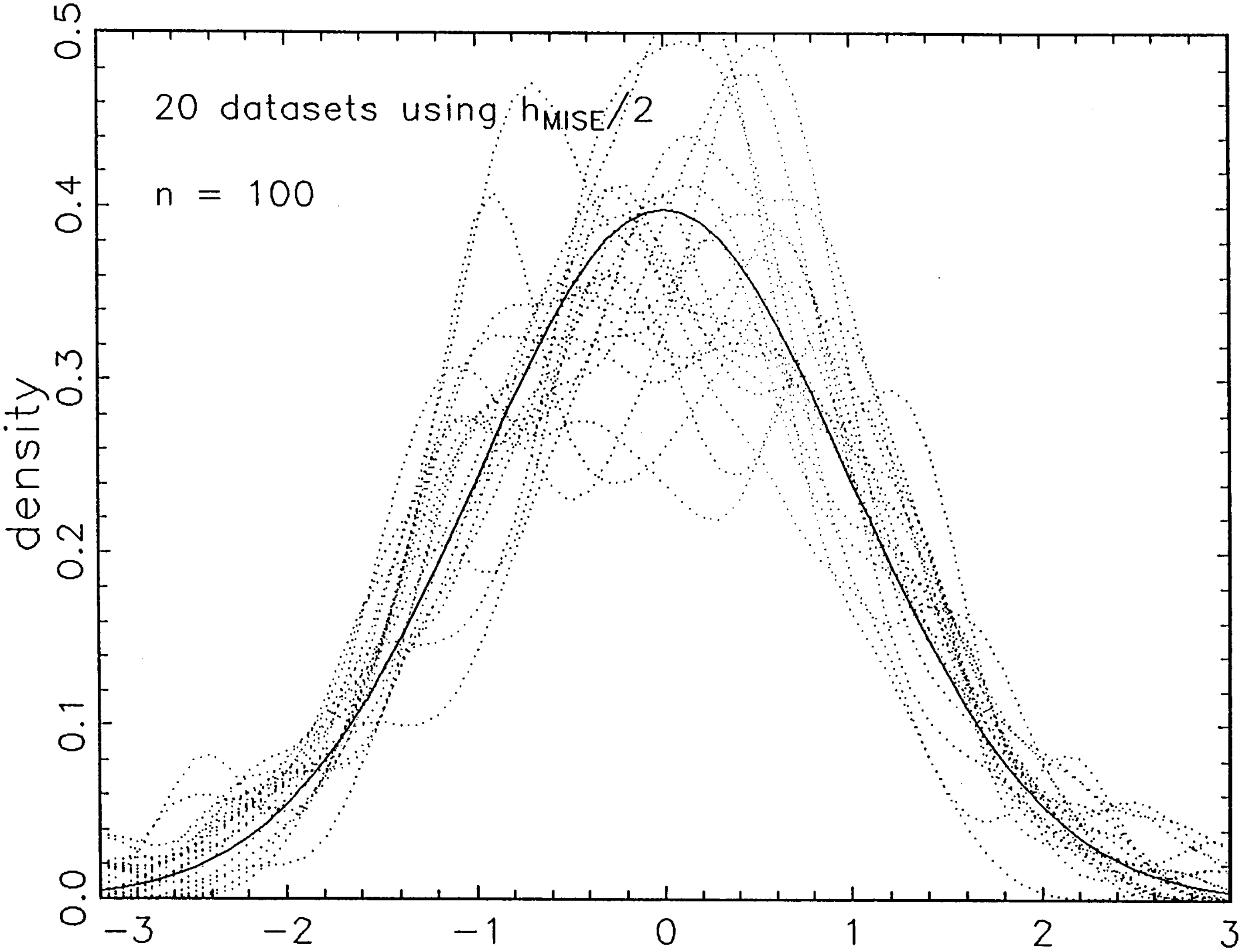


Figure 4.1.3 Gaussian distribution

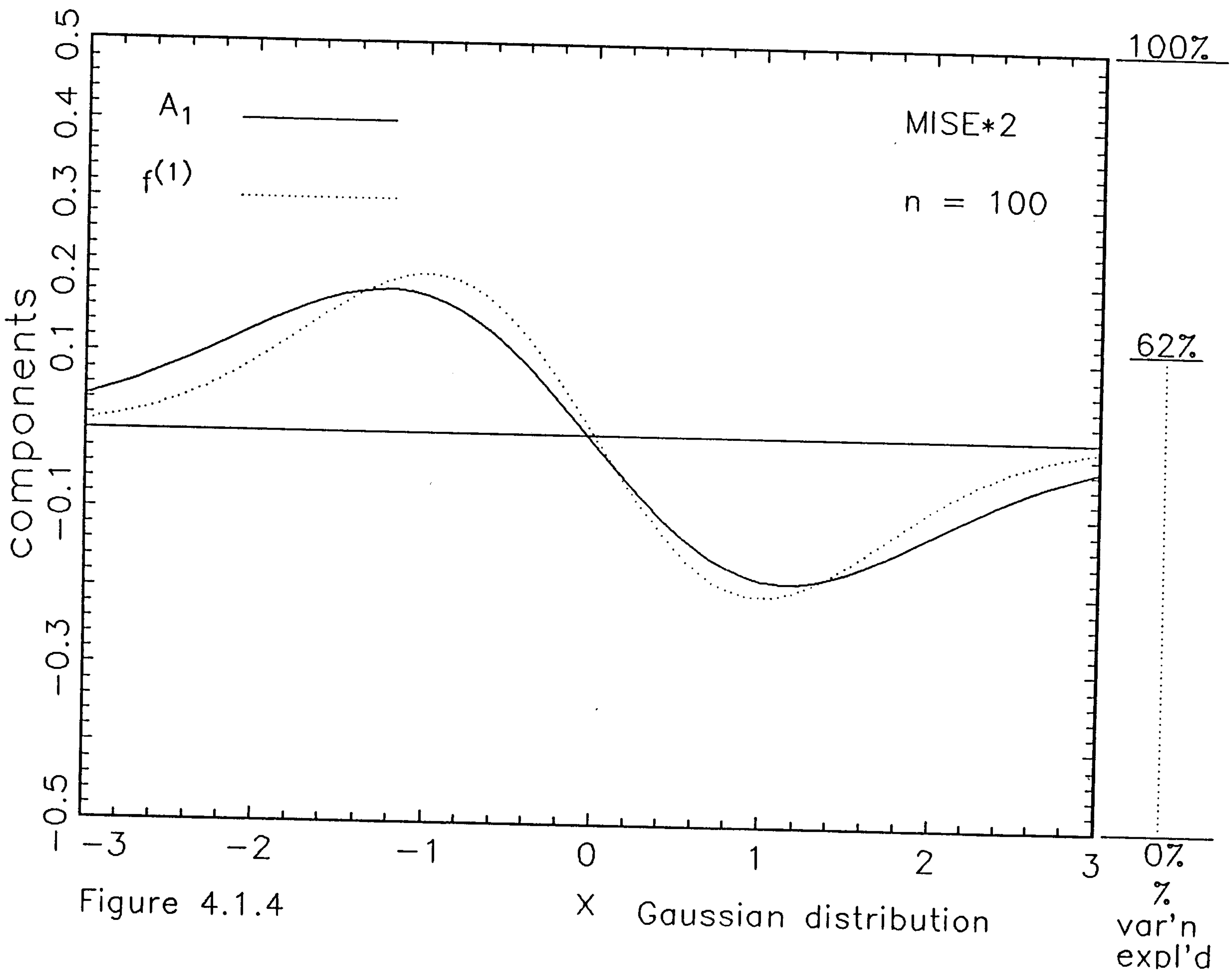


Figure 4.1.4

Gaussian distribution

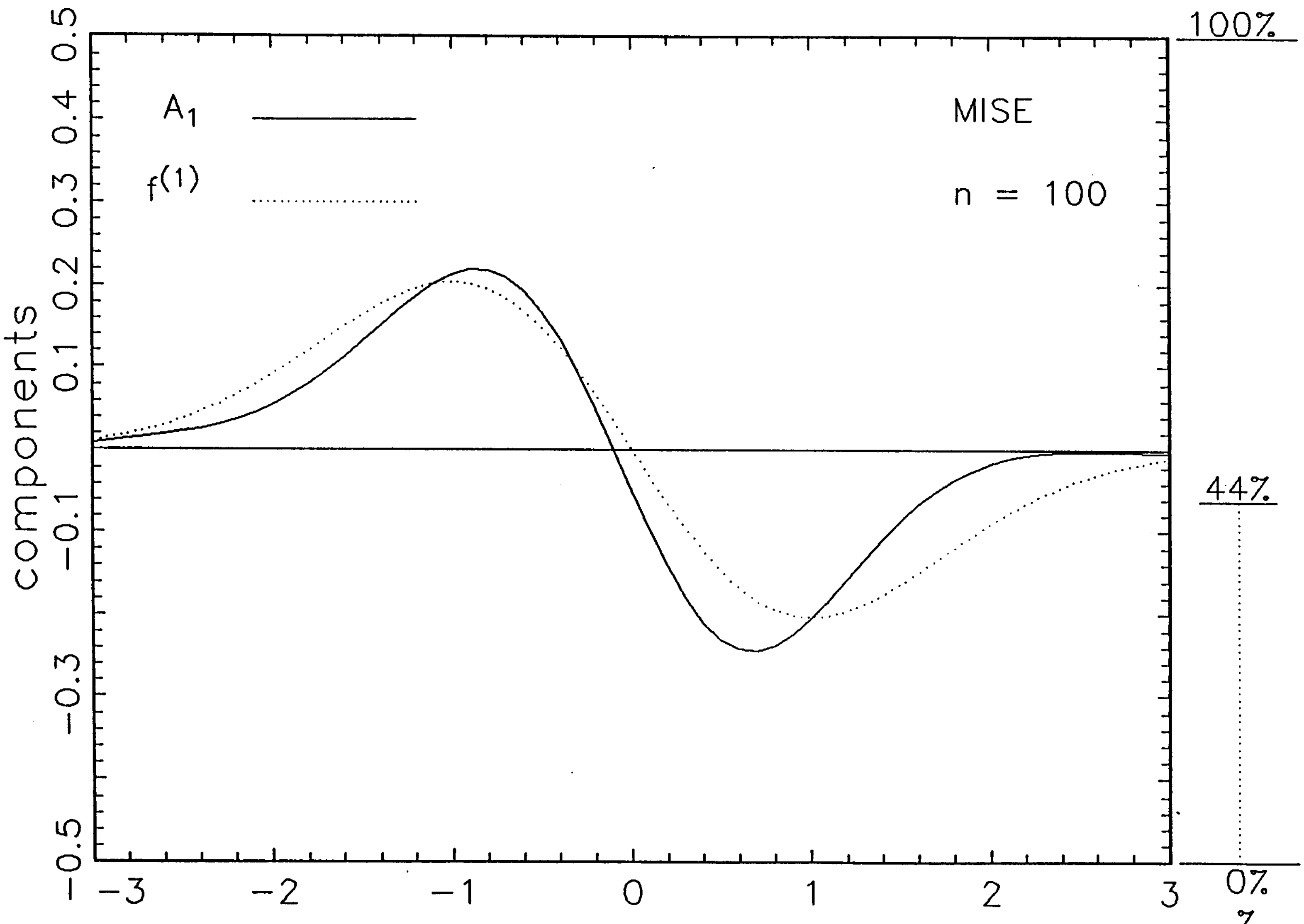
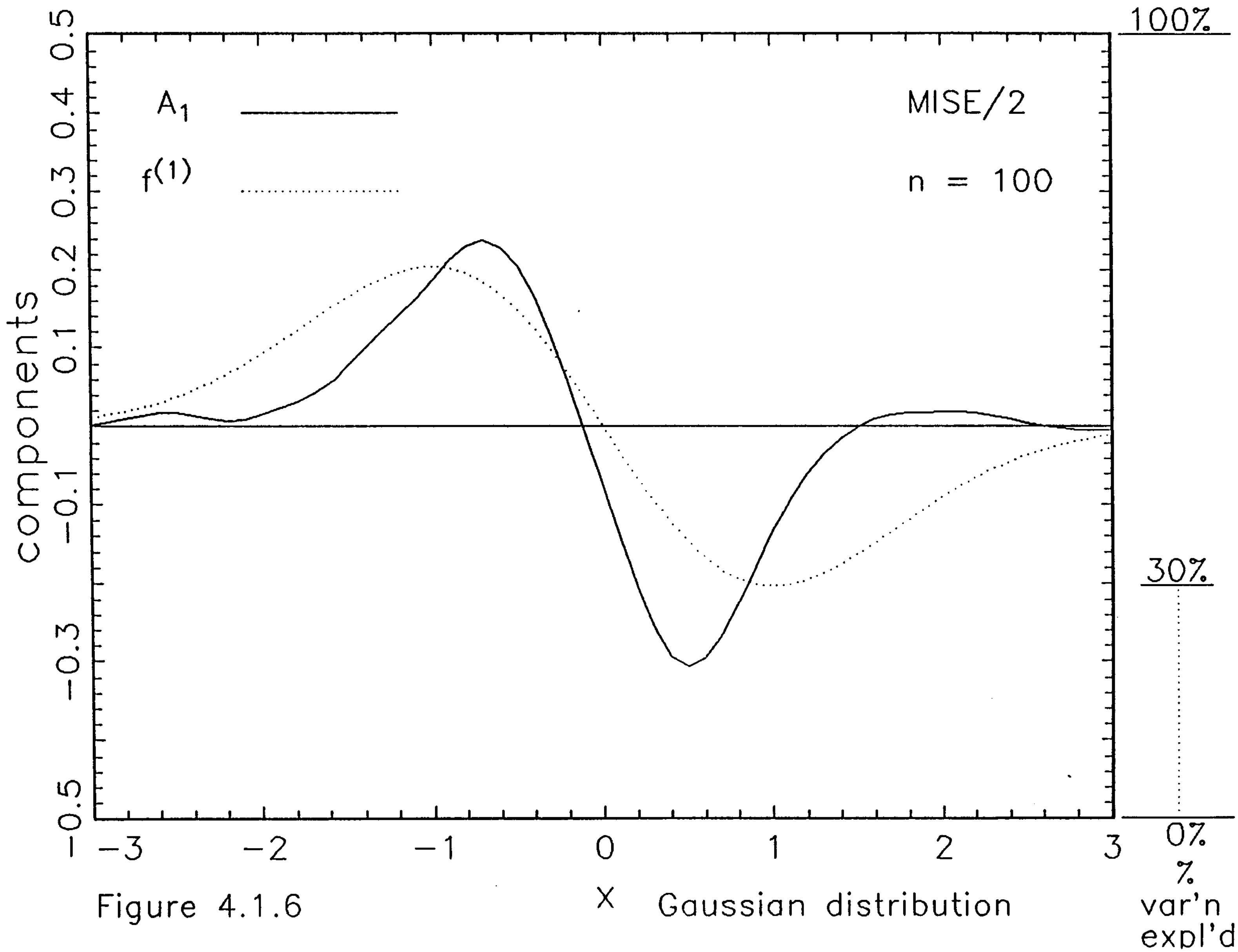


Figure 4.1.5

Gaussian distribution



components
0.5
0.4
0.3
0.2
0.1
−0.1
−0.3
−0.5
A1
f(1)
MISE/2
n = 100
100%
30%
0%
%
var'n
expl'd
−3
−2
−1
0
1
2
3
X Gaussian distribution

Figure 4.1.6

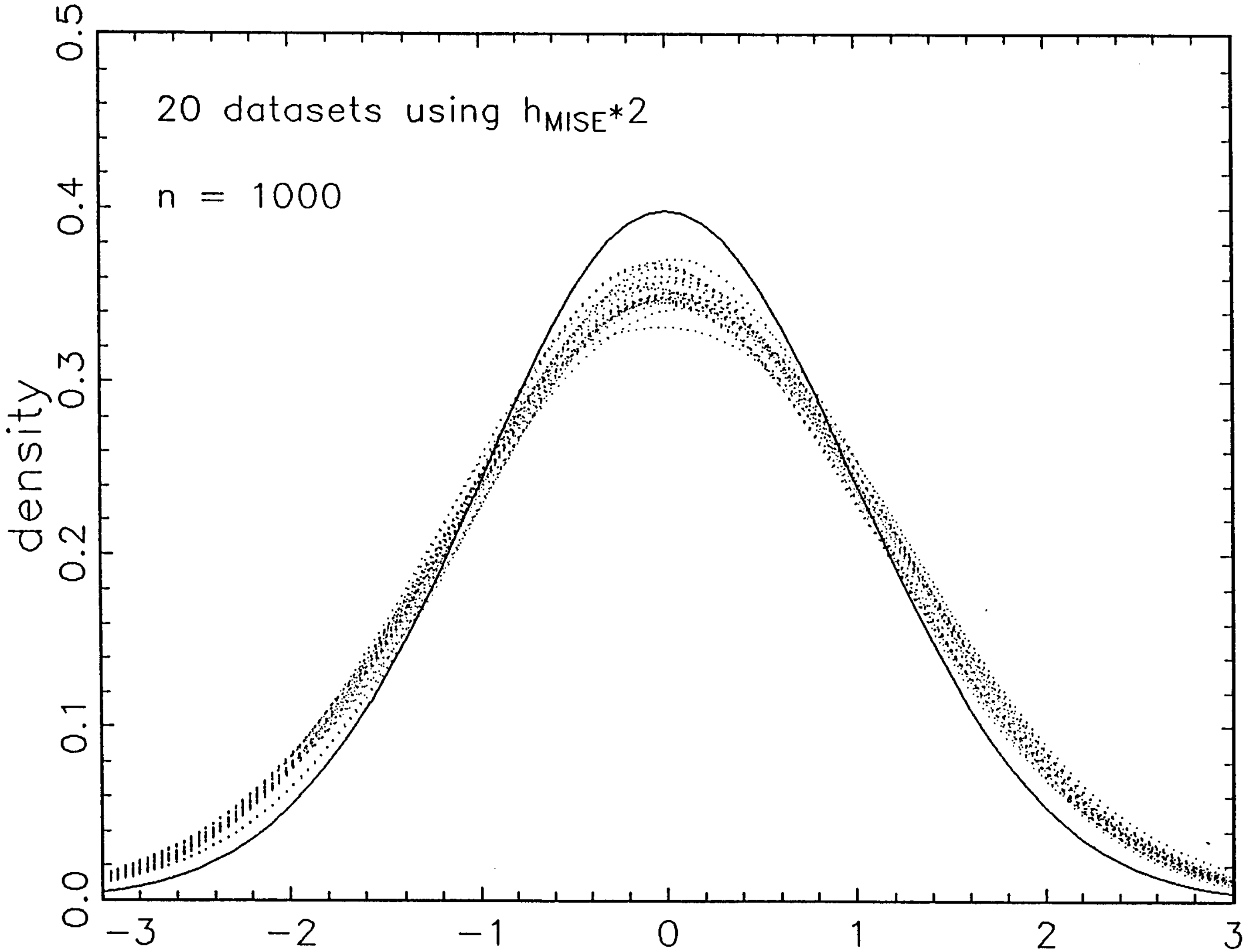


Figure 4.1.7 X Gaussian distribution

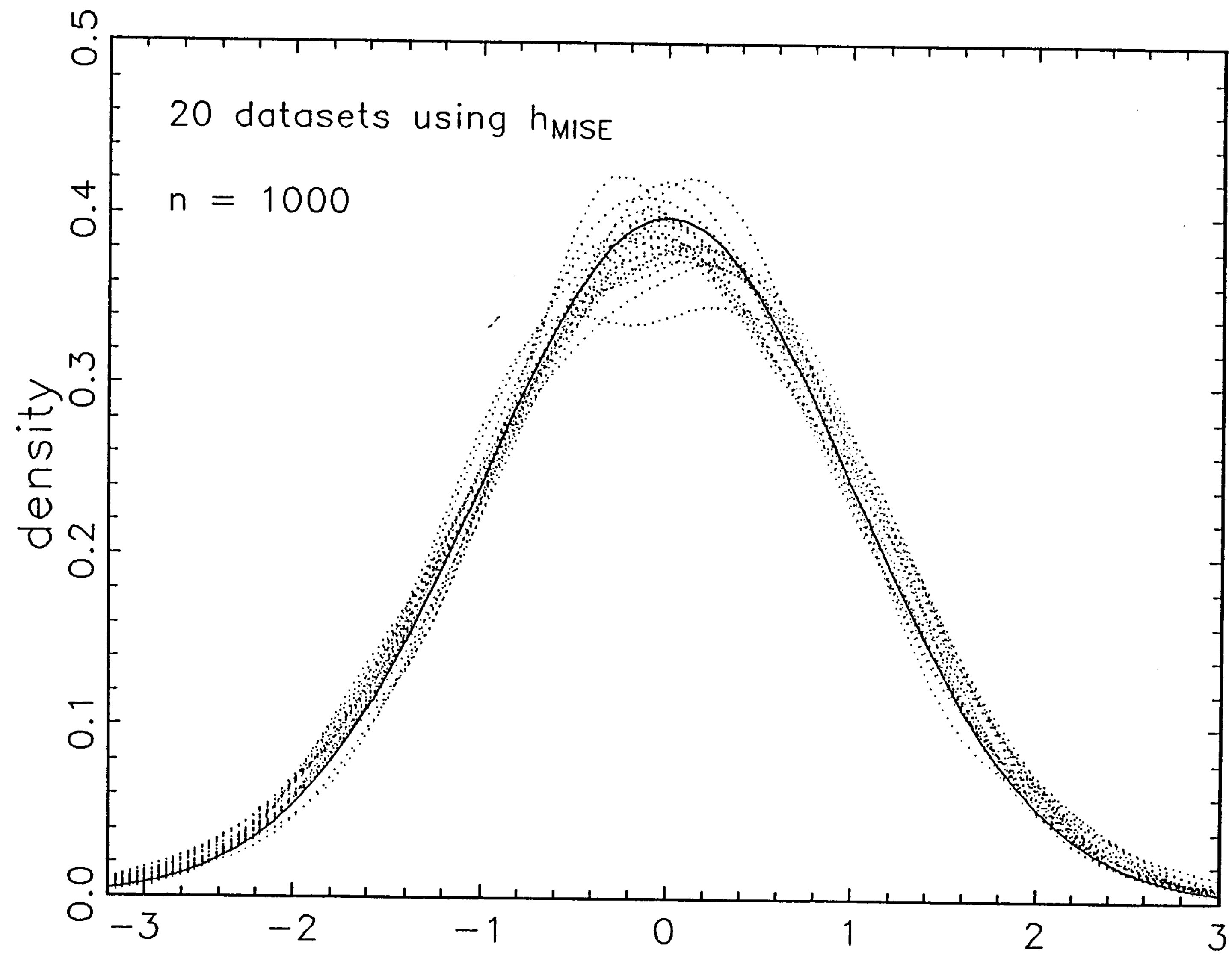


Figure 4.1.8 Gaussian distribution

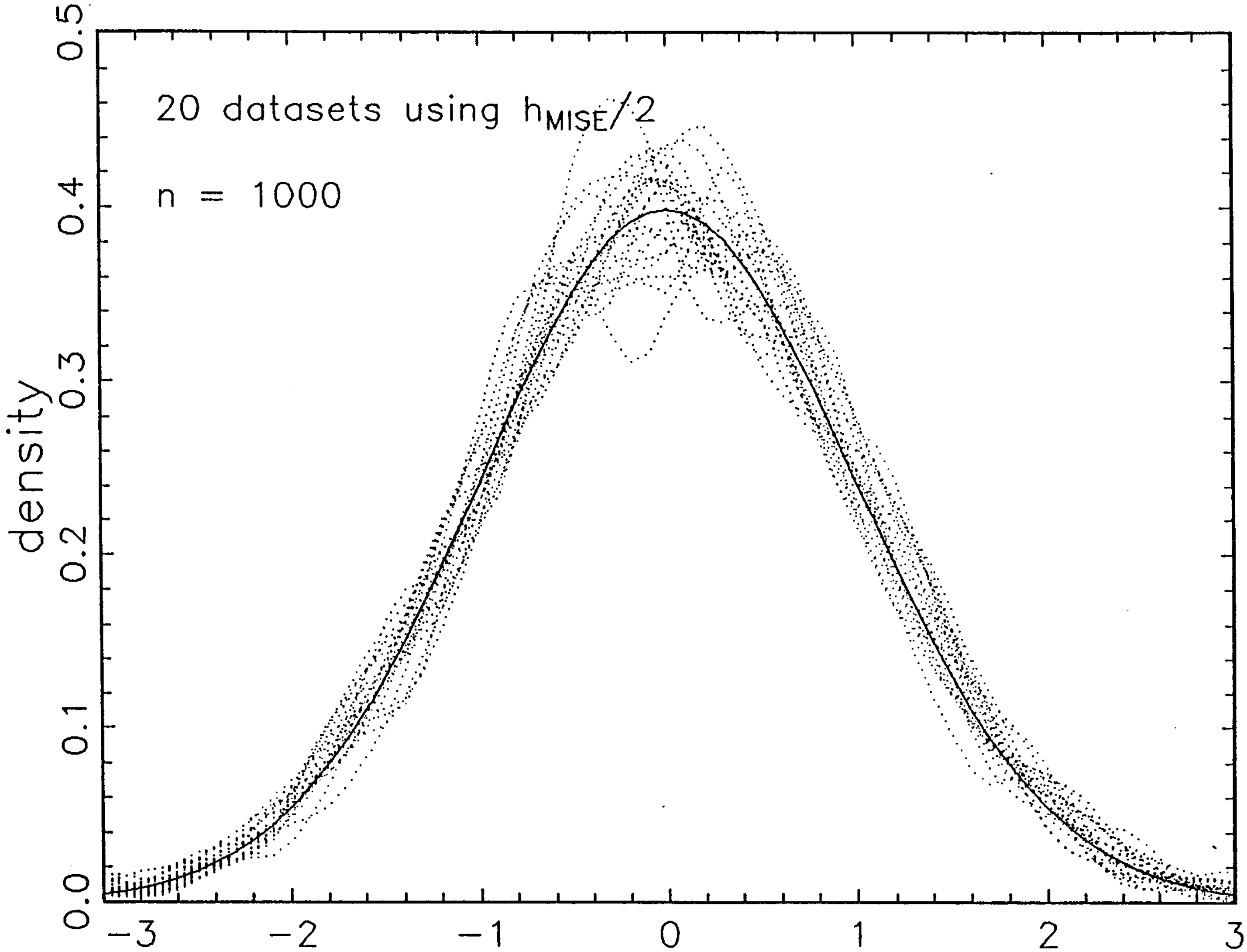


Figure 4.1.9 Gaussian distribution

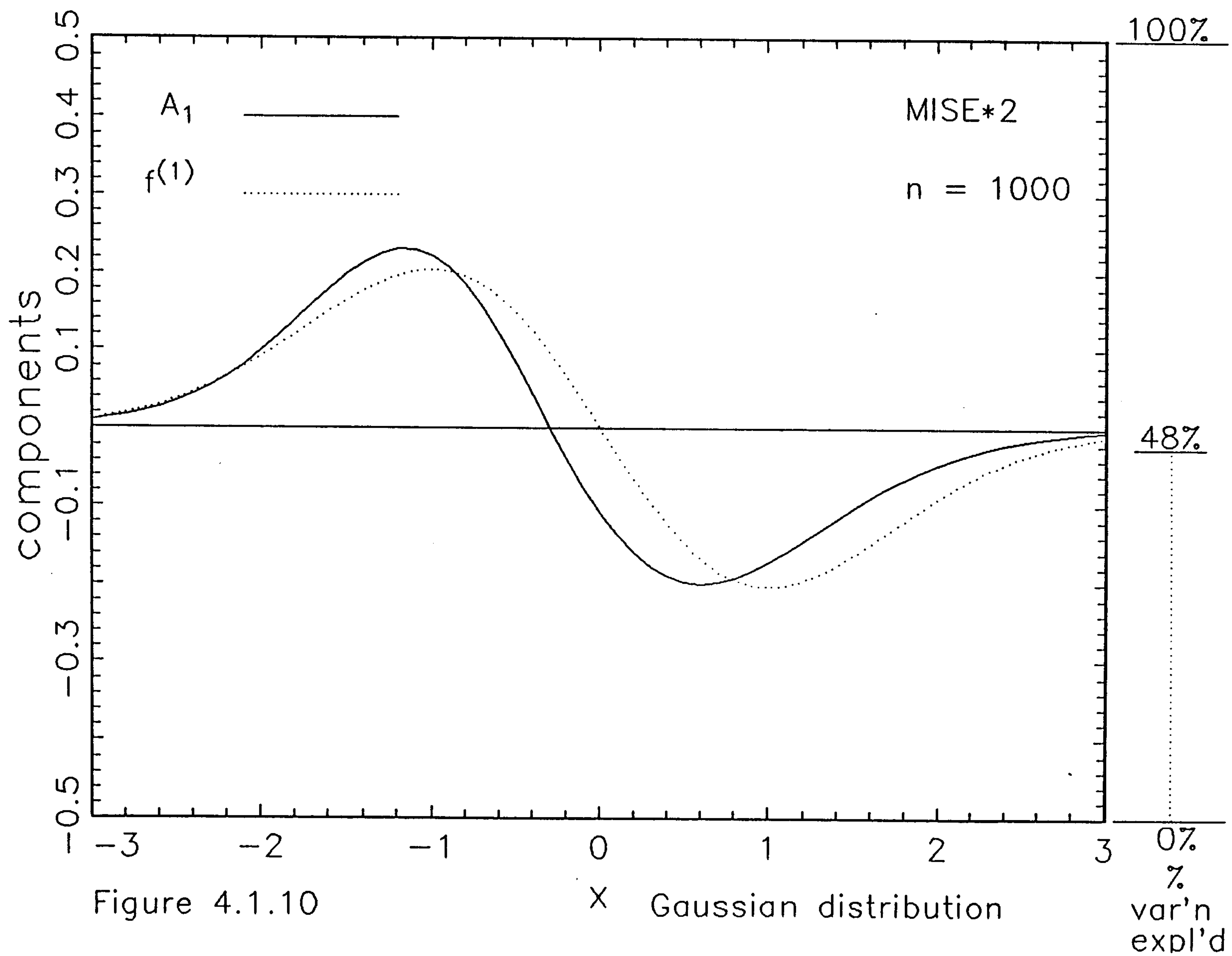


Figure 4.1.10 Gaussian distribution

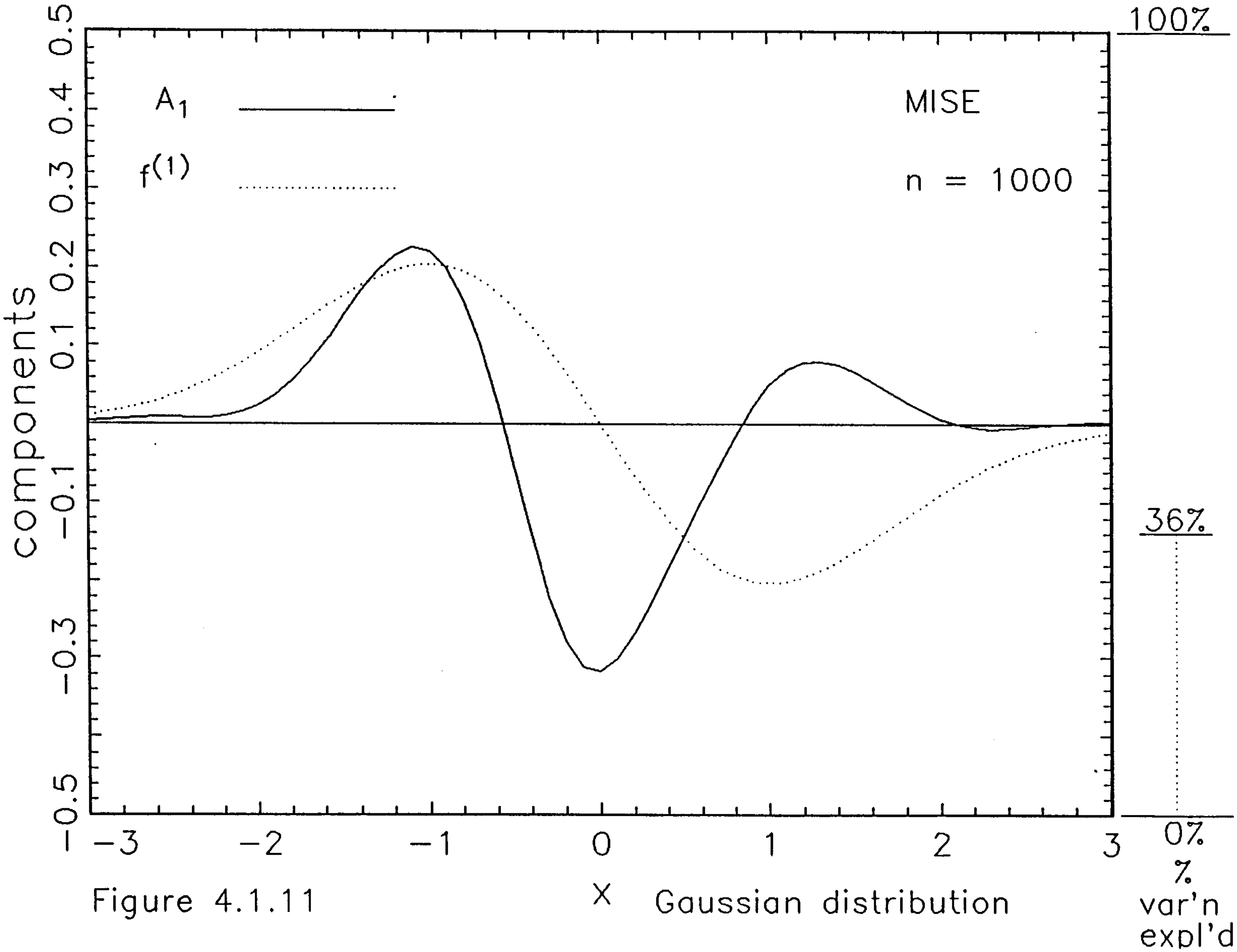


Figure 4.1.11

Gaussian distribution

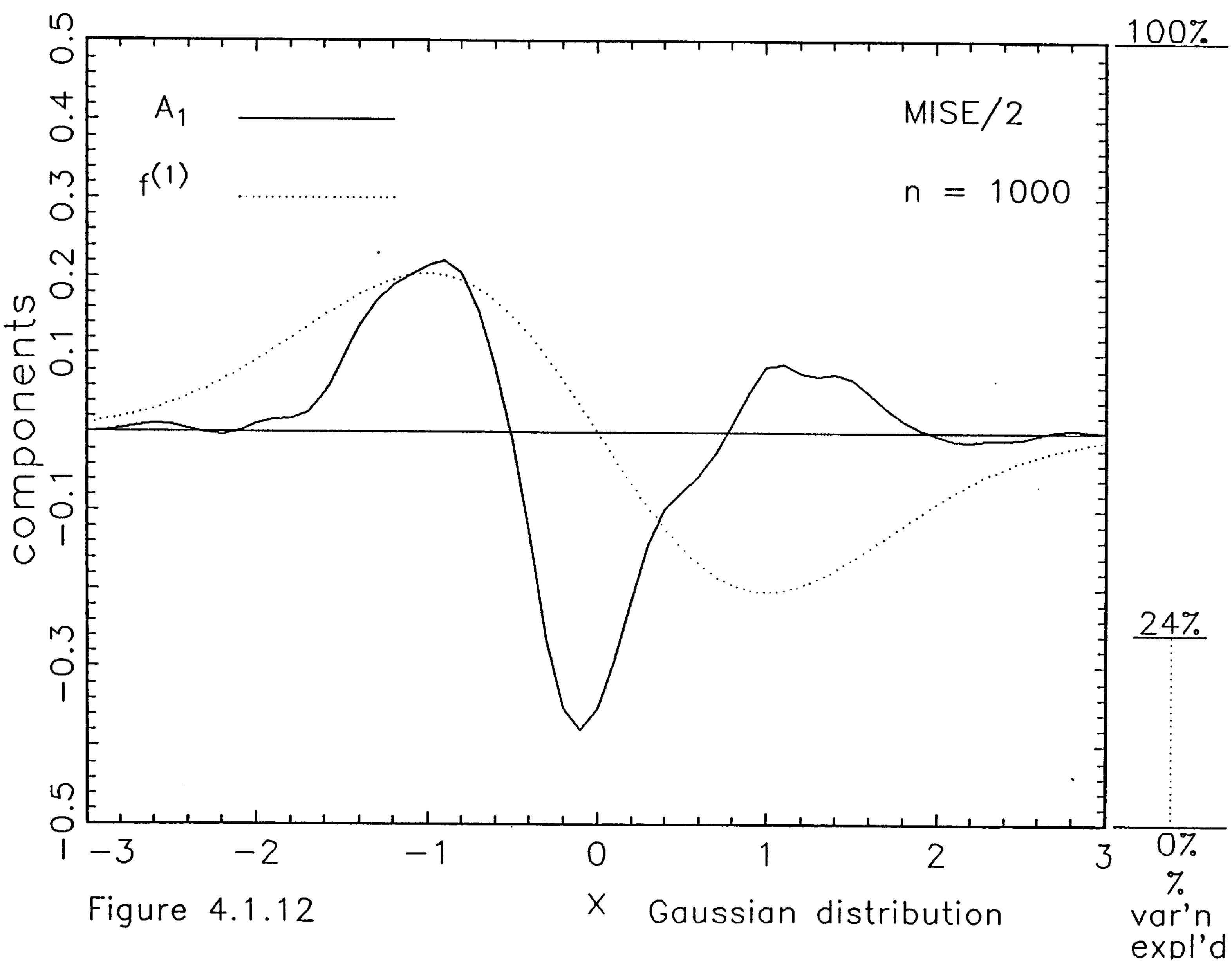


Figure 4.1.12 Gaussian distribution

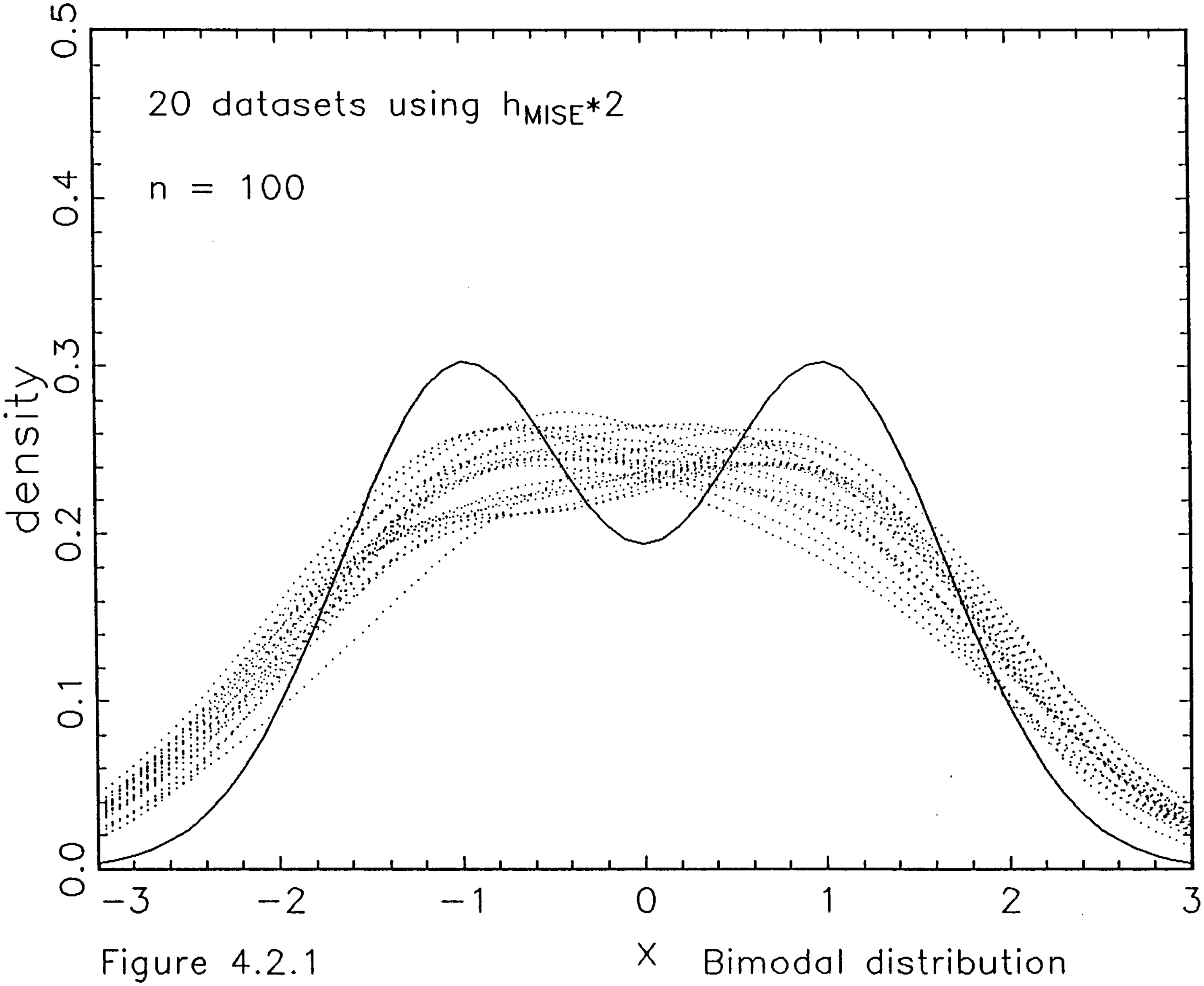


Figure 4.2.1 Bimodal distribution

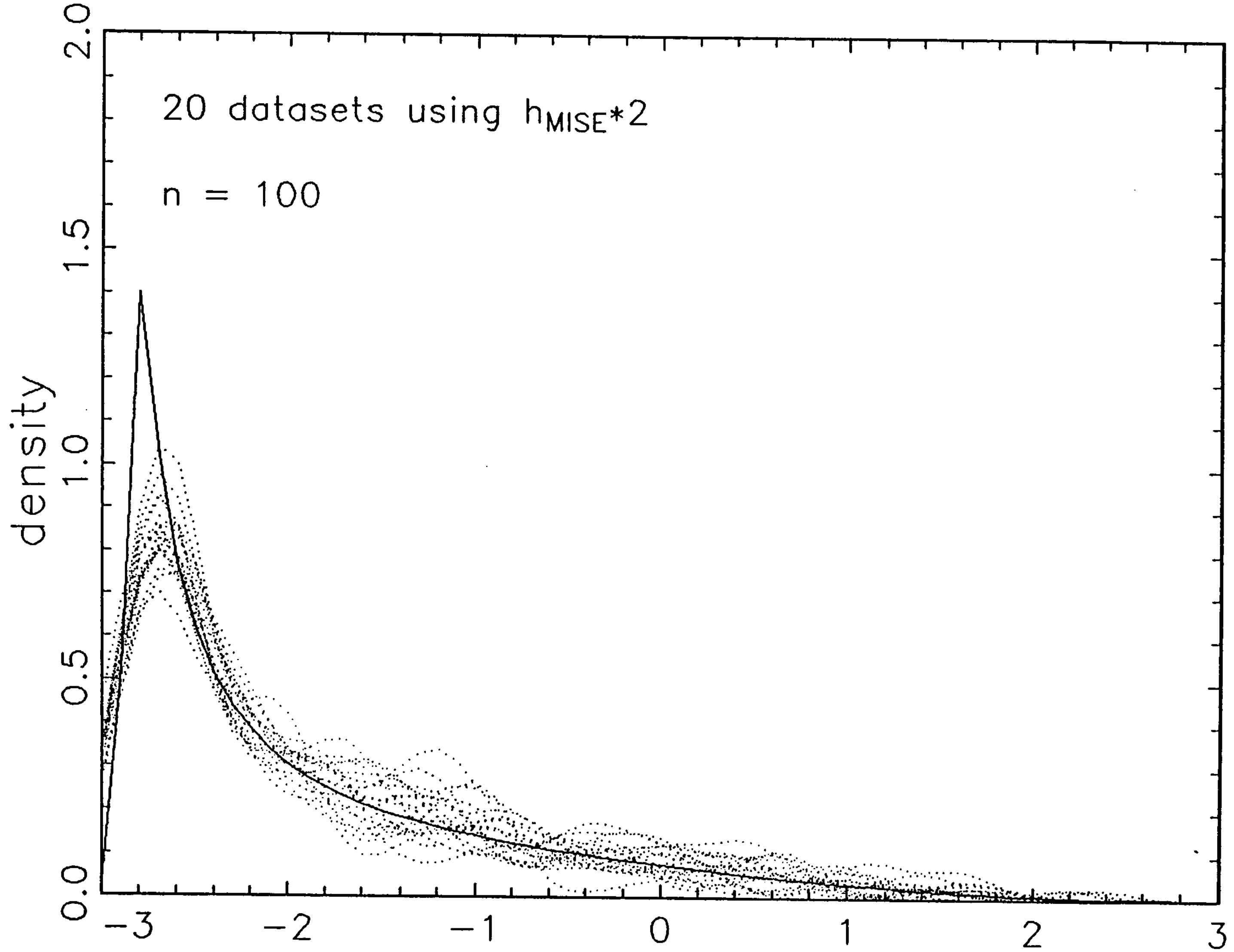


Figure 4.3.1 Strongly skewed distribution

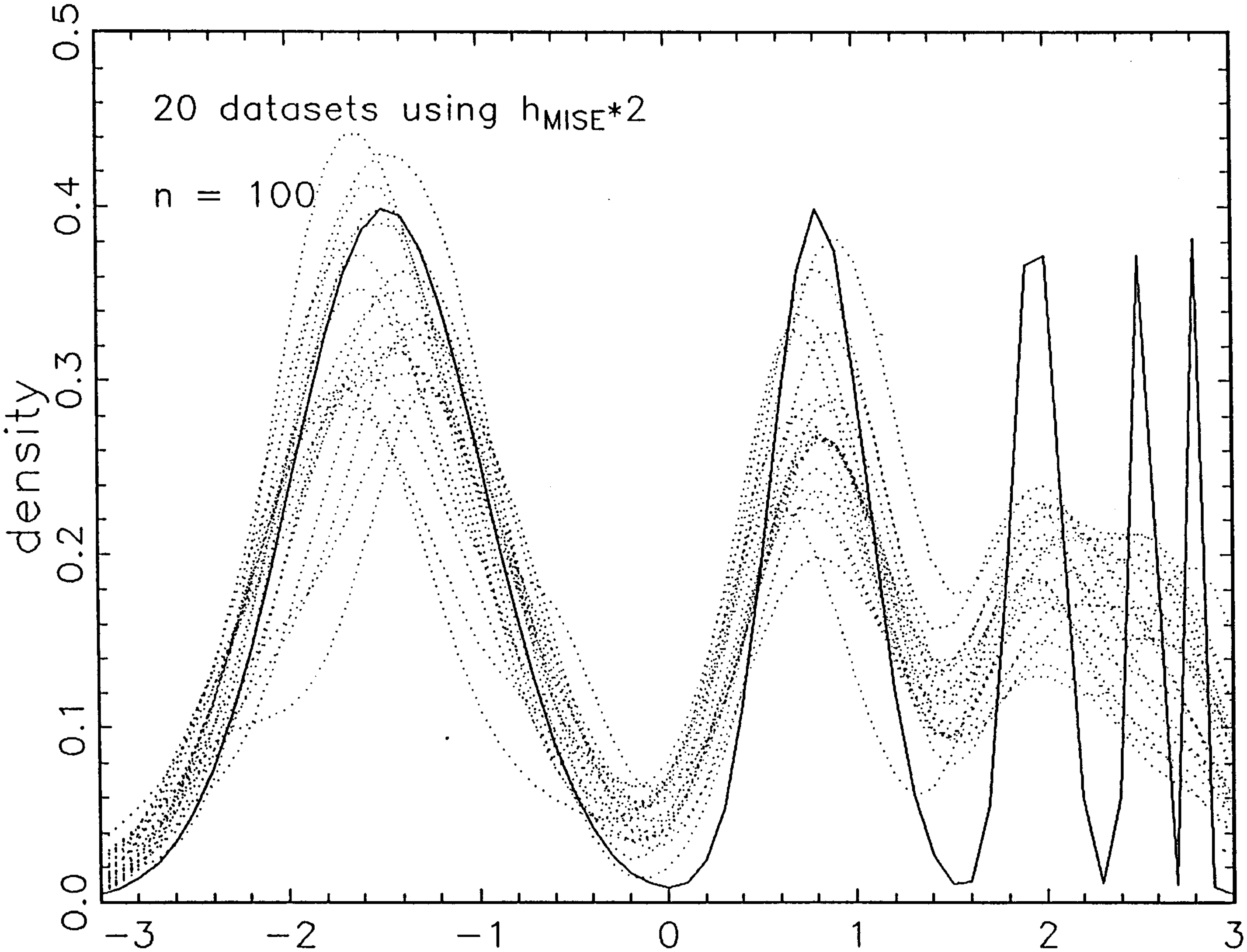


Figure 4.4.1 Smooth Comb distribution

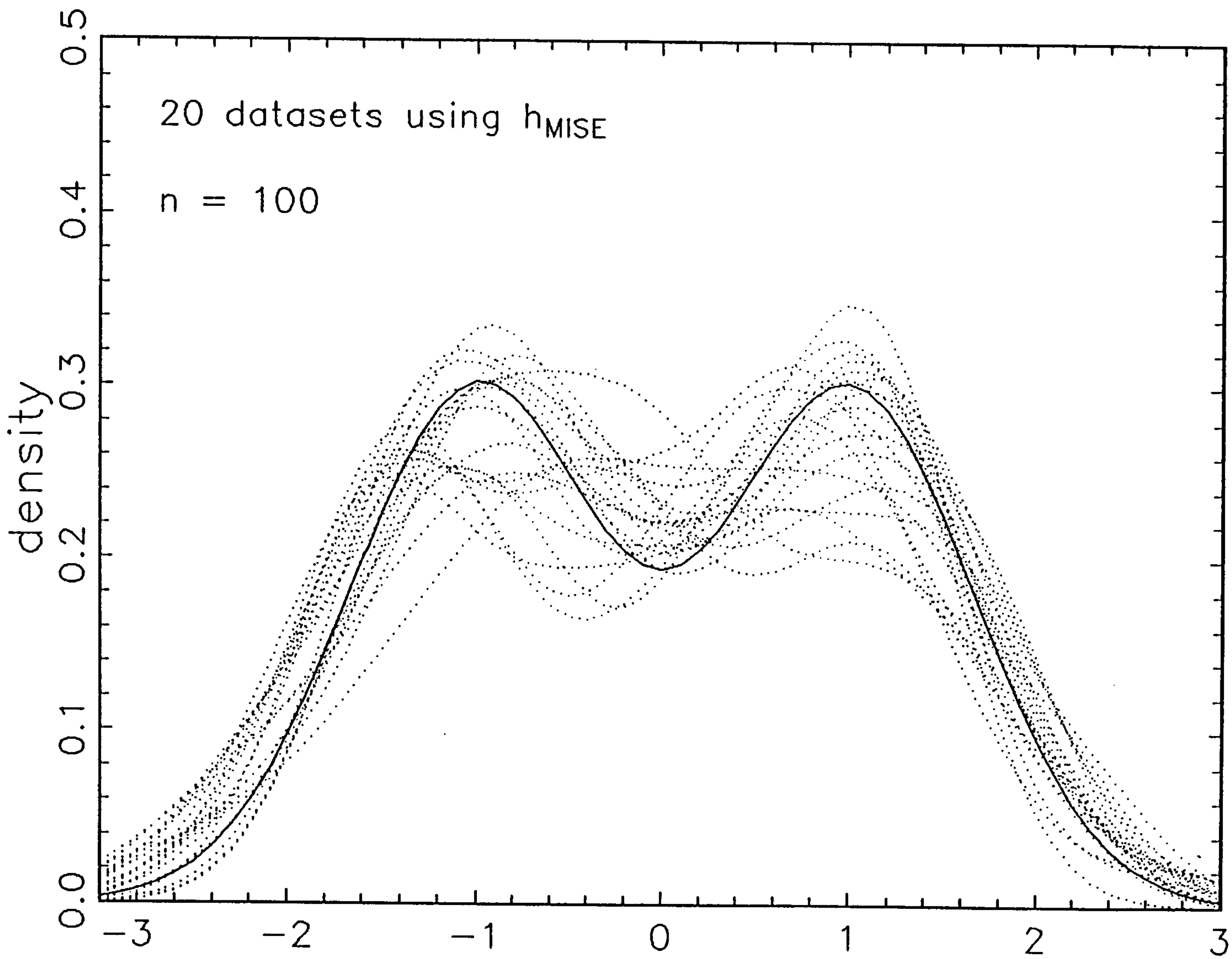


Figure 5.1 Bimodal distribution

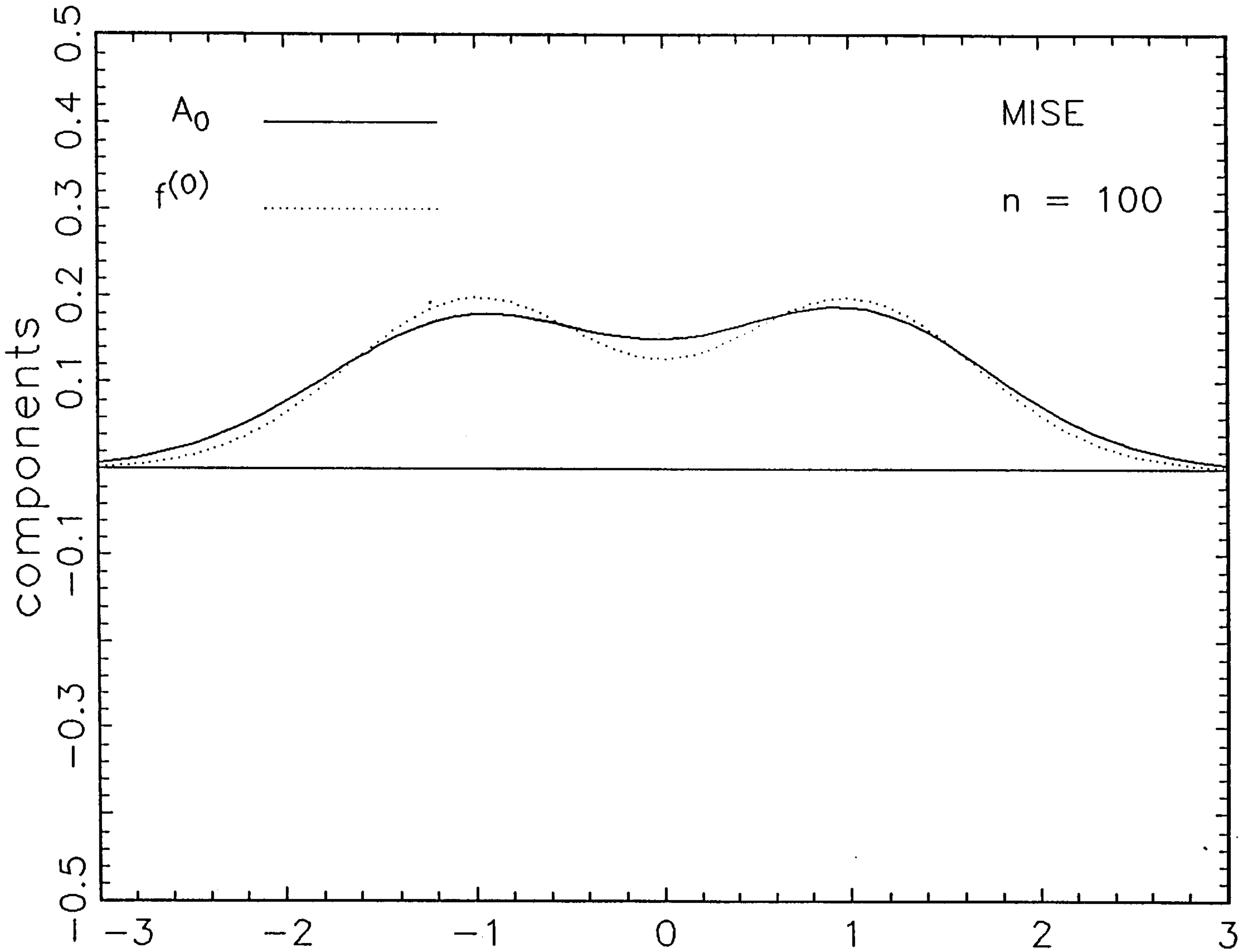


Figure 5.2 Biomodal distribution

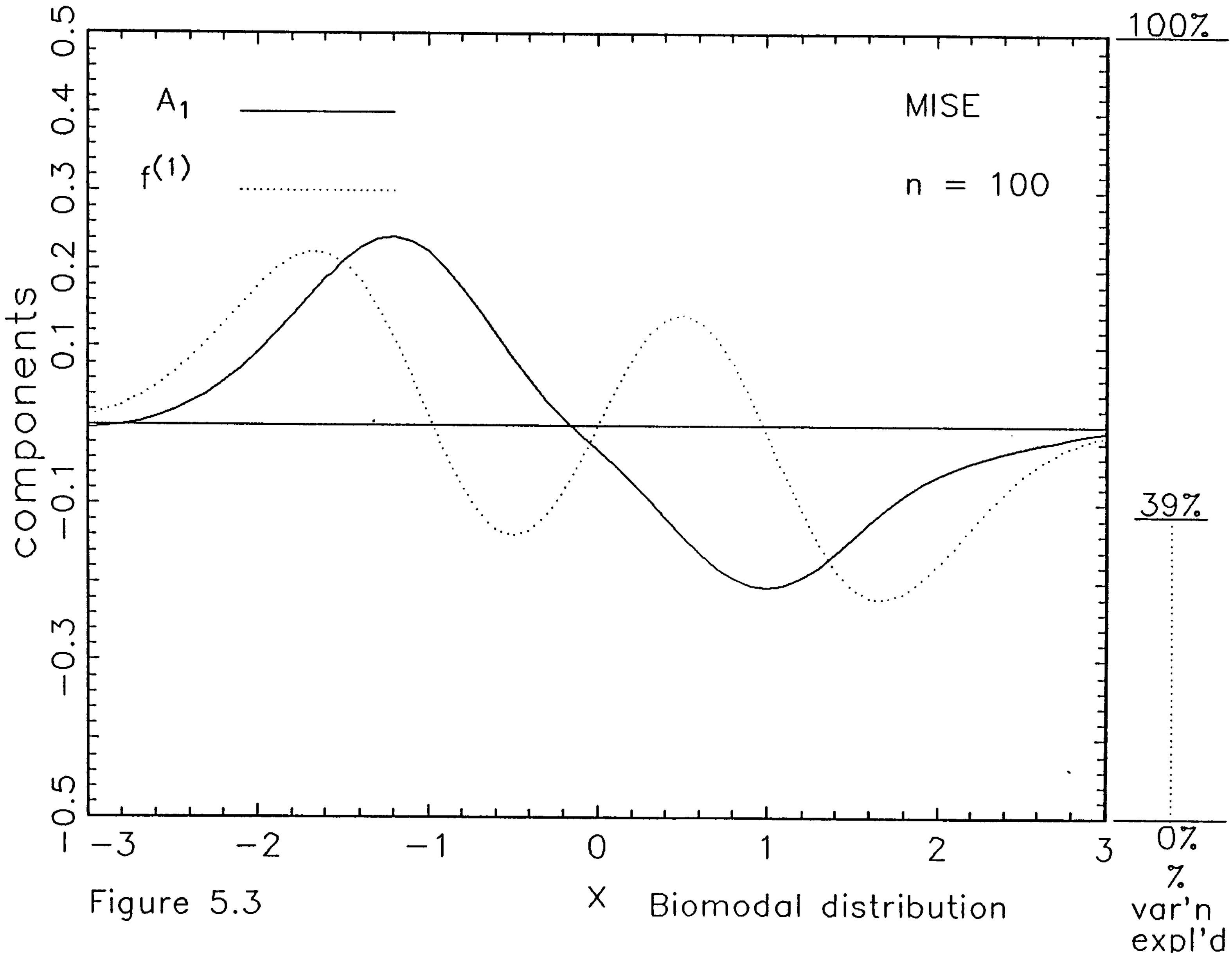

components
0.5
0.4
0.3
0.2
0.1
−0.1
−0.3
−0.5
A1
f(1)
MISE
n = 100
−3
−2
−1
0
1
2
3
X Biomodal distribution
100%
39%
0%
% var'n expl'd

Figure 5.3

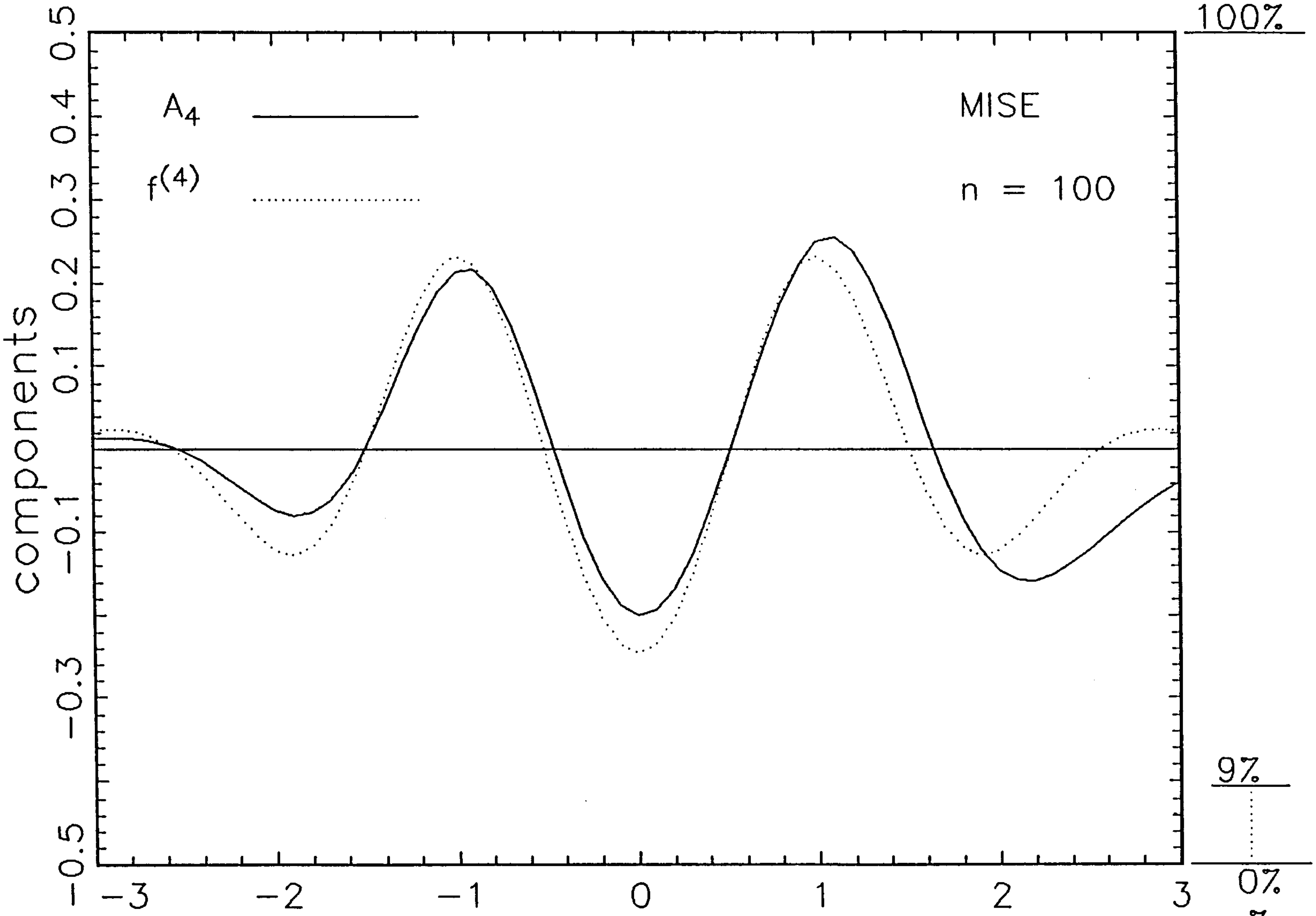
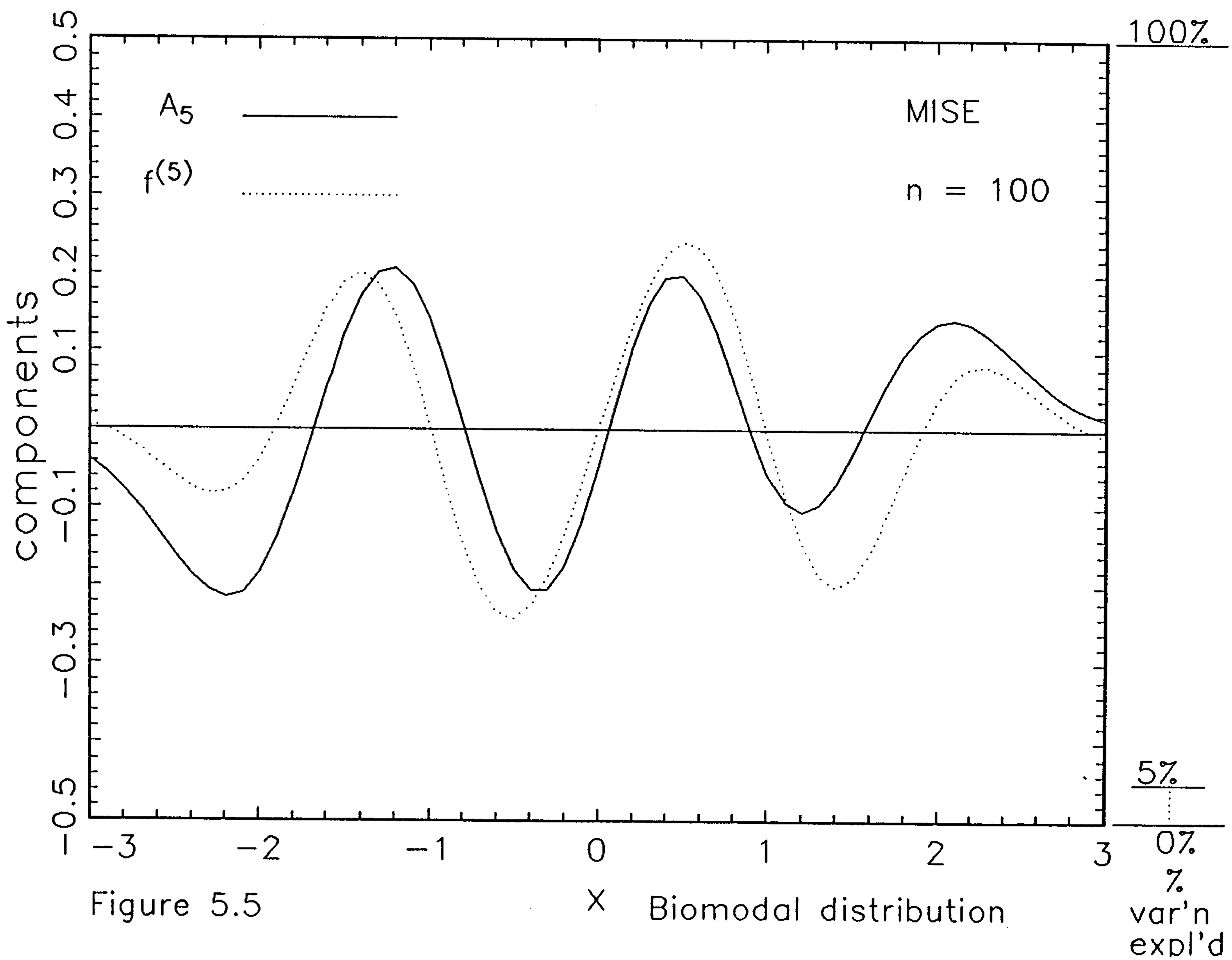

components
0.5
0.4
0.3
0.2
0.1
-0.1
-0.3
-0.5
A_4
$f^{(4)}$
MISE
n = 100
100%
9%
0%
% var'n expl'd
-3
-2
-1
0
1
2
3
X Biomodal distribution

Figure 5.4



components
0.5
0.4
0.3
0.2
0.1
−0.1
−0.3
−0.5
A_5
$f^{(5)}$
MISE
n = 100
100%
5%
0%
% var'n expl'd
−3
−2
−1
0
1
2
3
X
Biomodal distribution

Figure 5.5

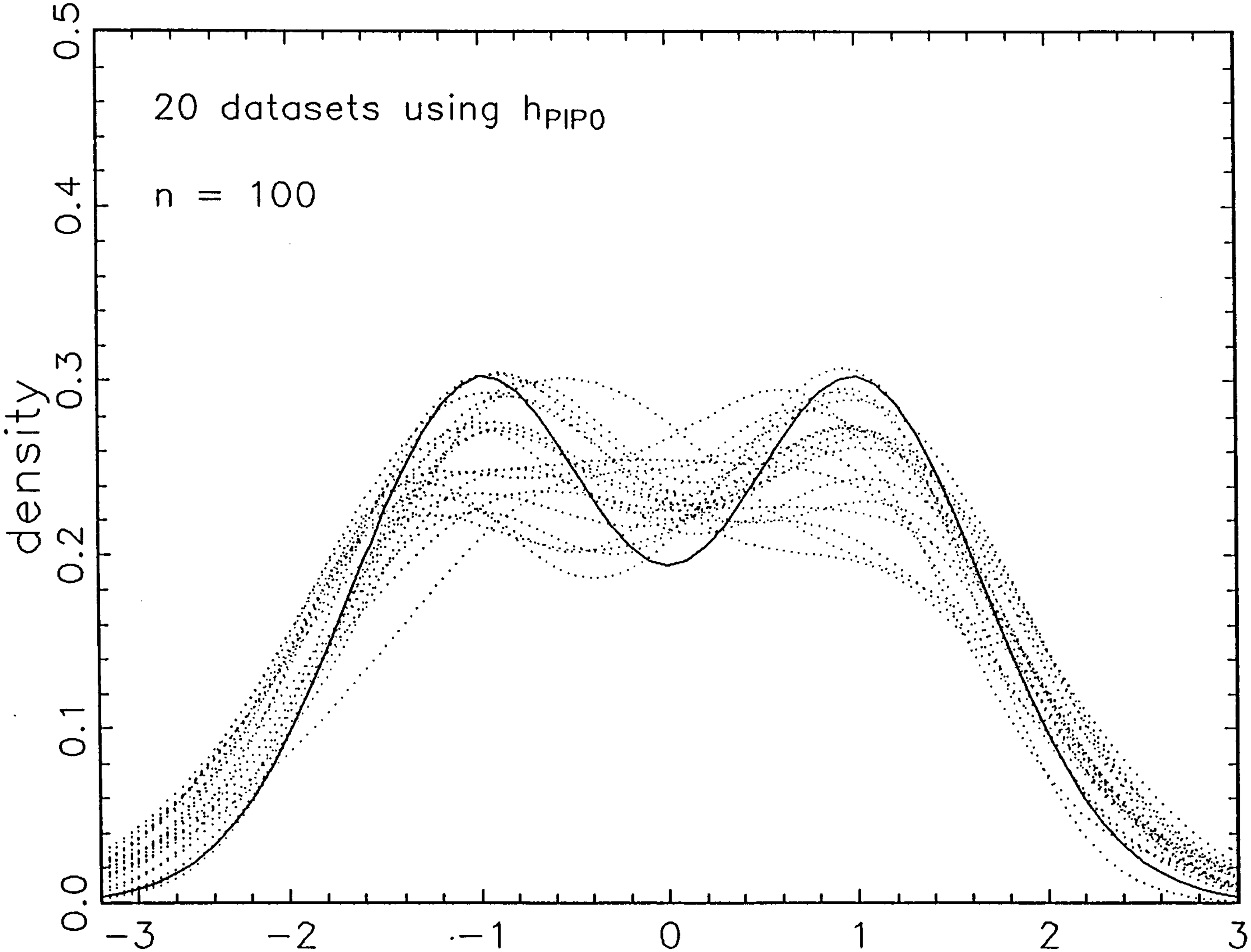


Figure 6.1 Bimodal distribution

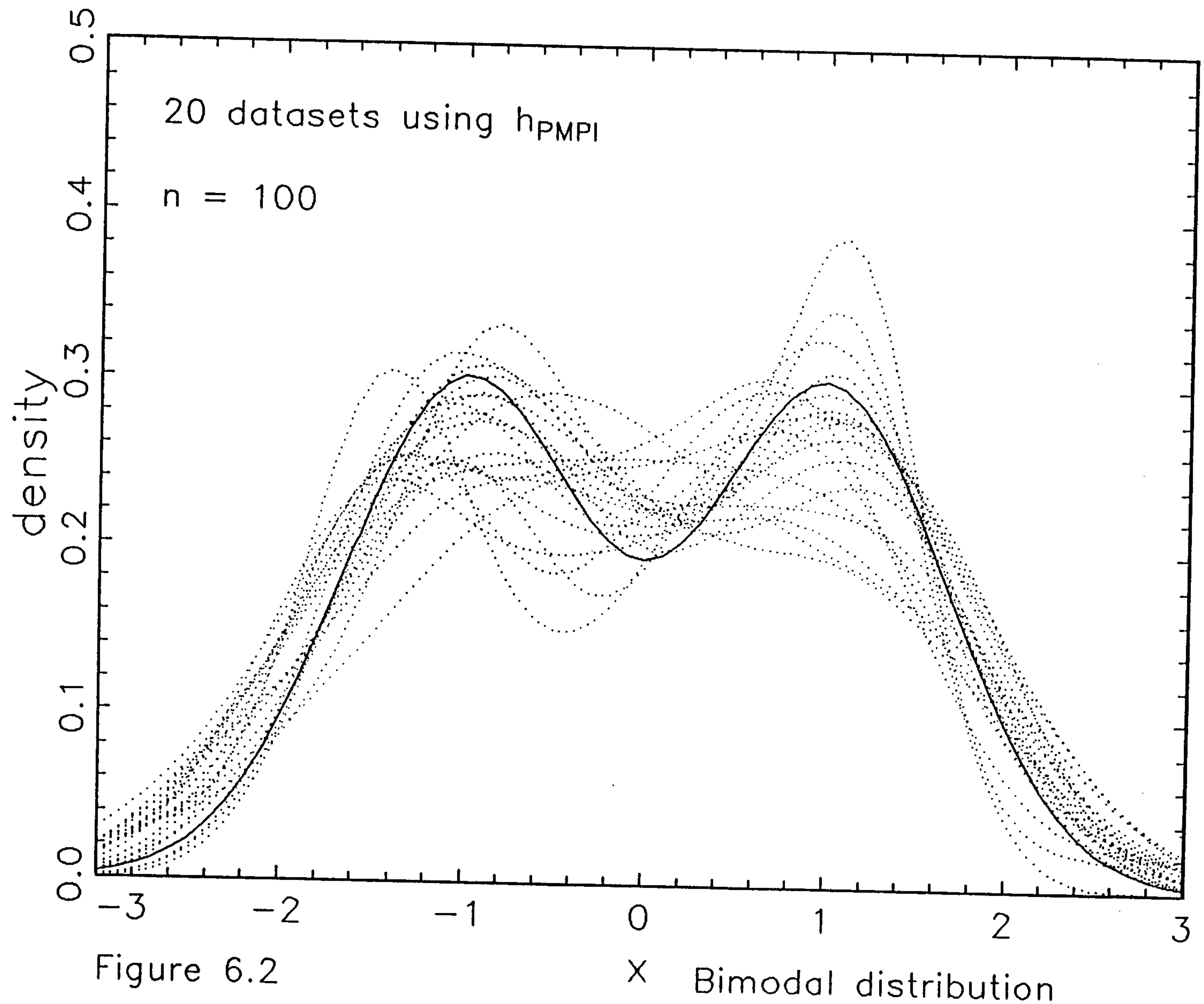

20 datasets using h_{PMPI}
n = 100
density
0.0
0.1
0.2
0.3
0.4
0.5
−3
−2
−1
0
1
2
3
X
Bimodal distribution

Figure 6.2

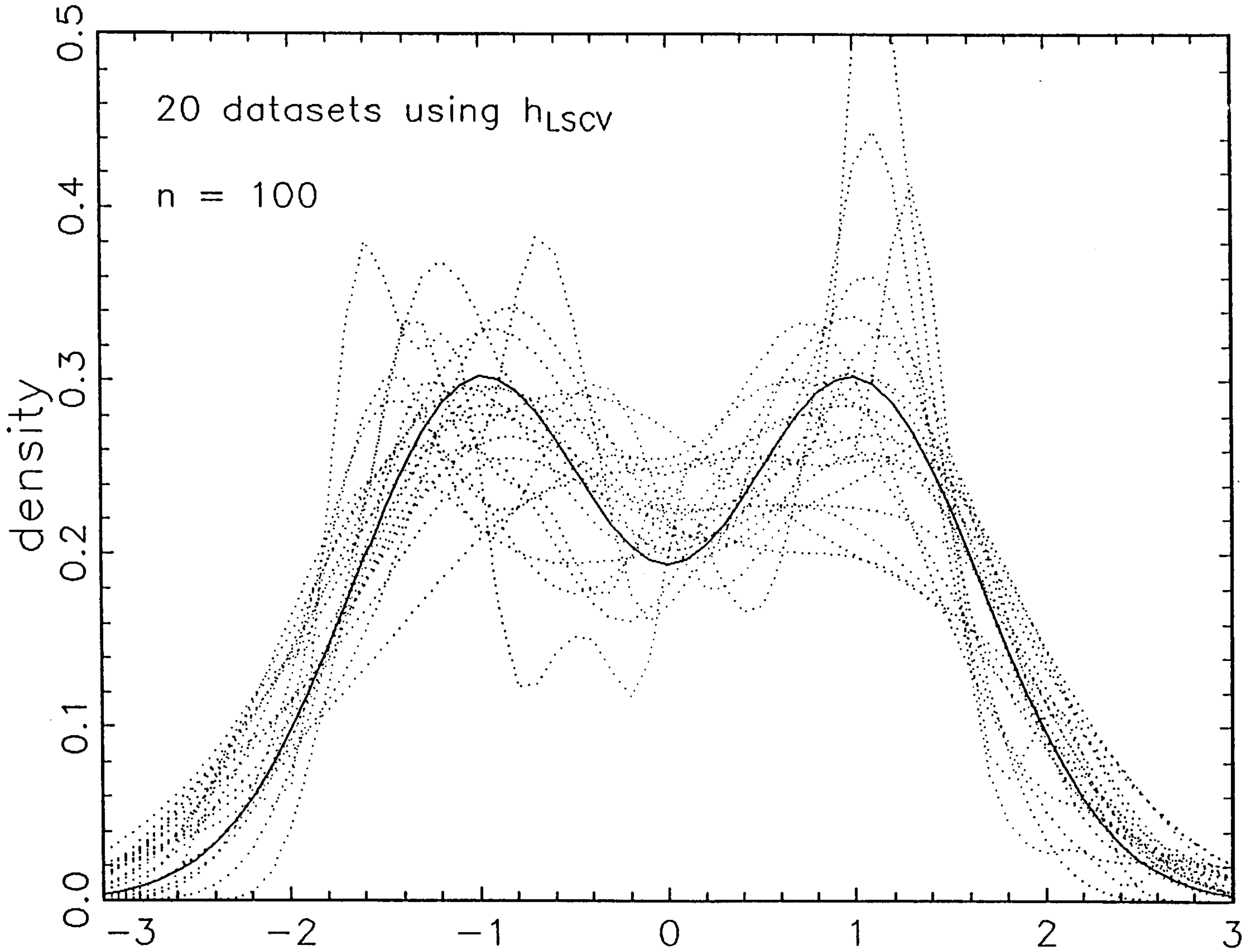


Figure 6.3 Bimodal distribution

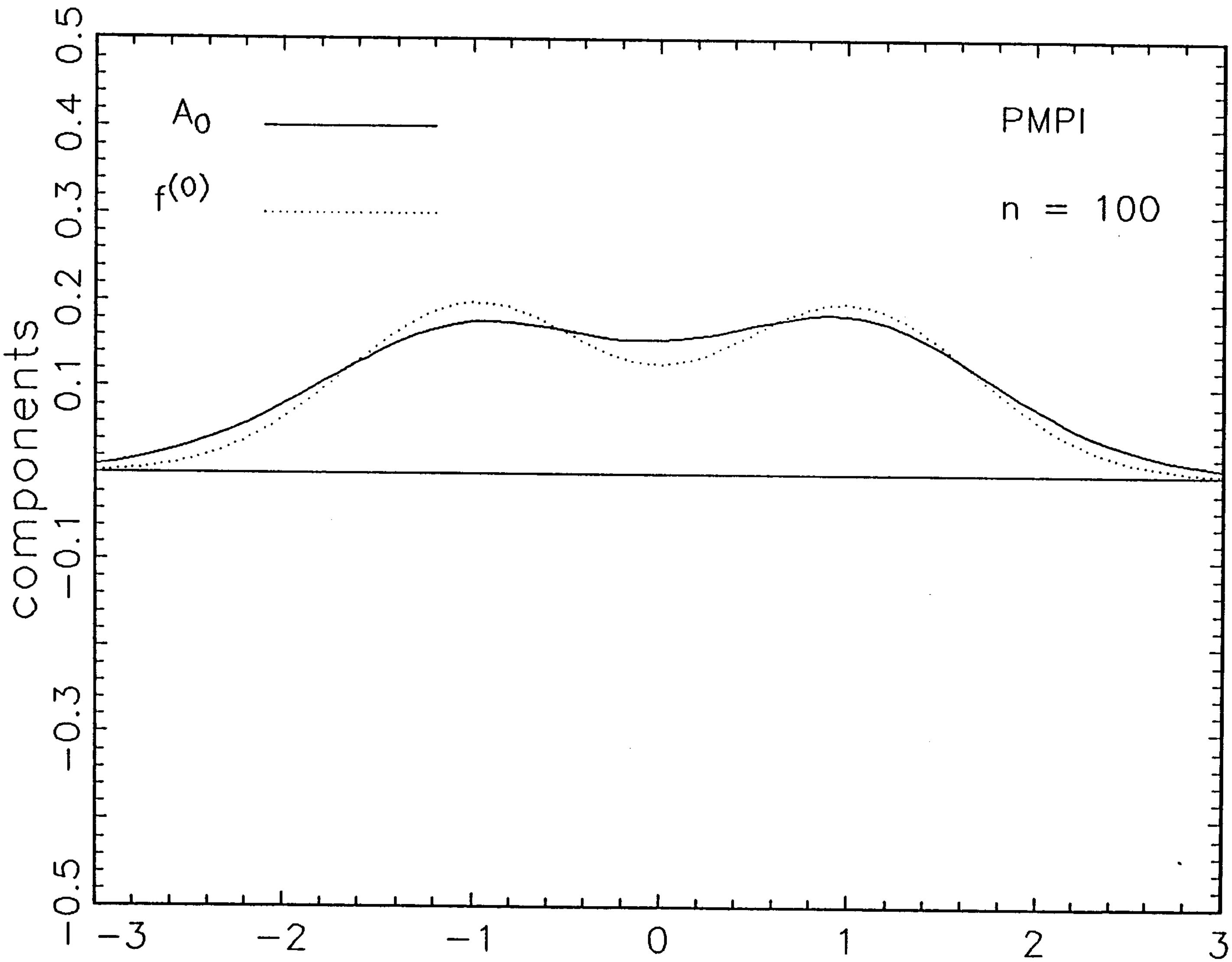


Figure 6.4 X Biomodal distribution

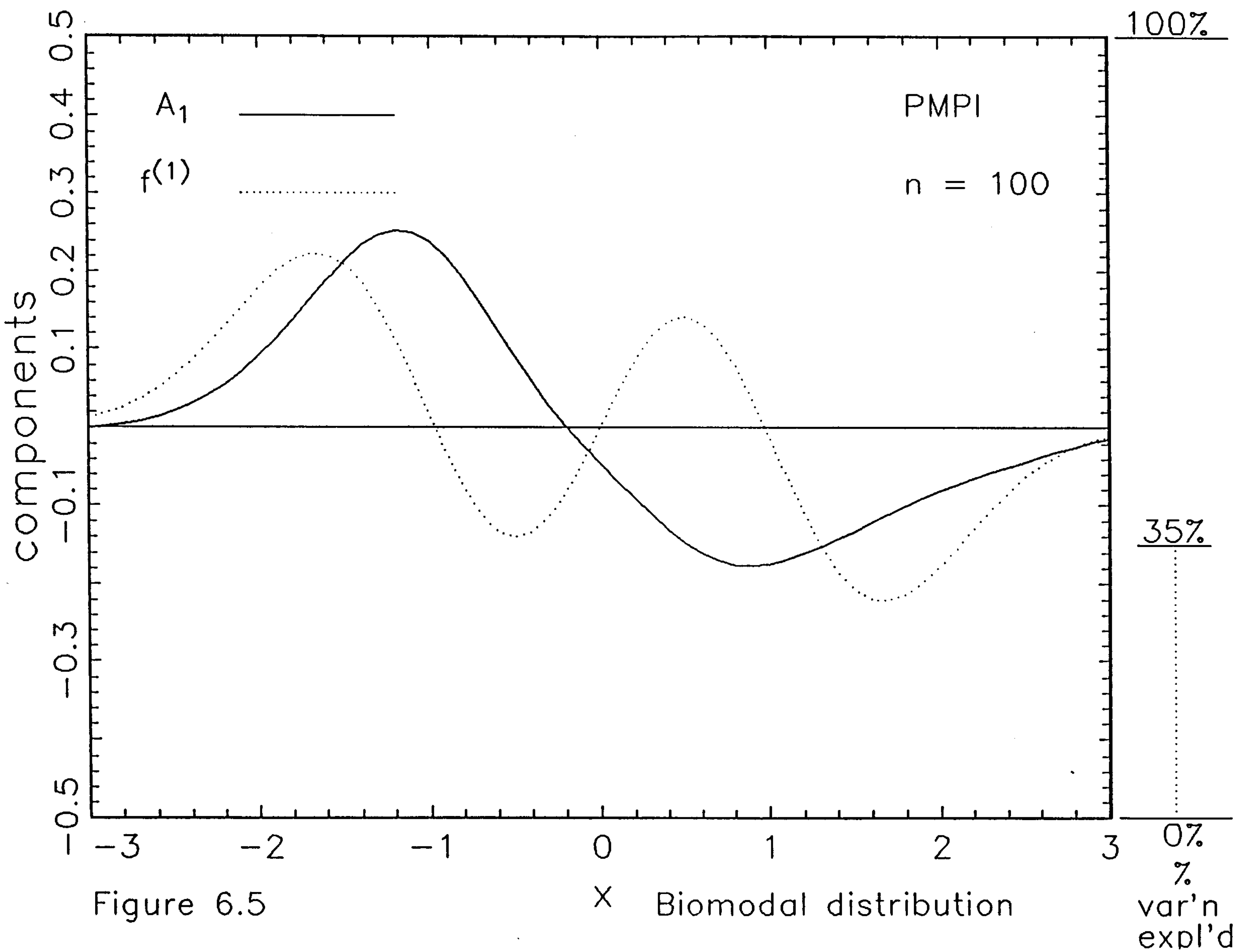


Figure 6.5 Biomodal distribution

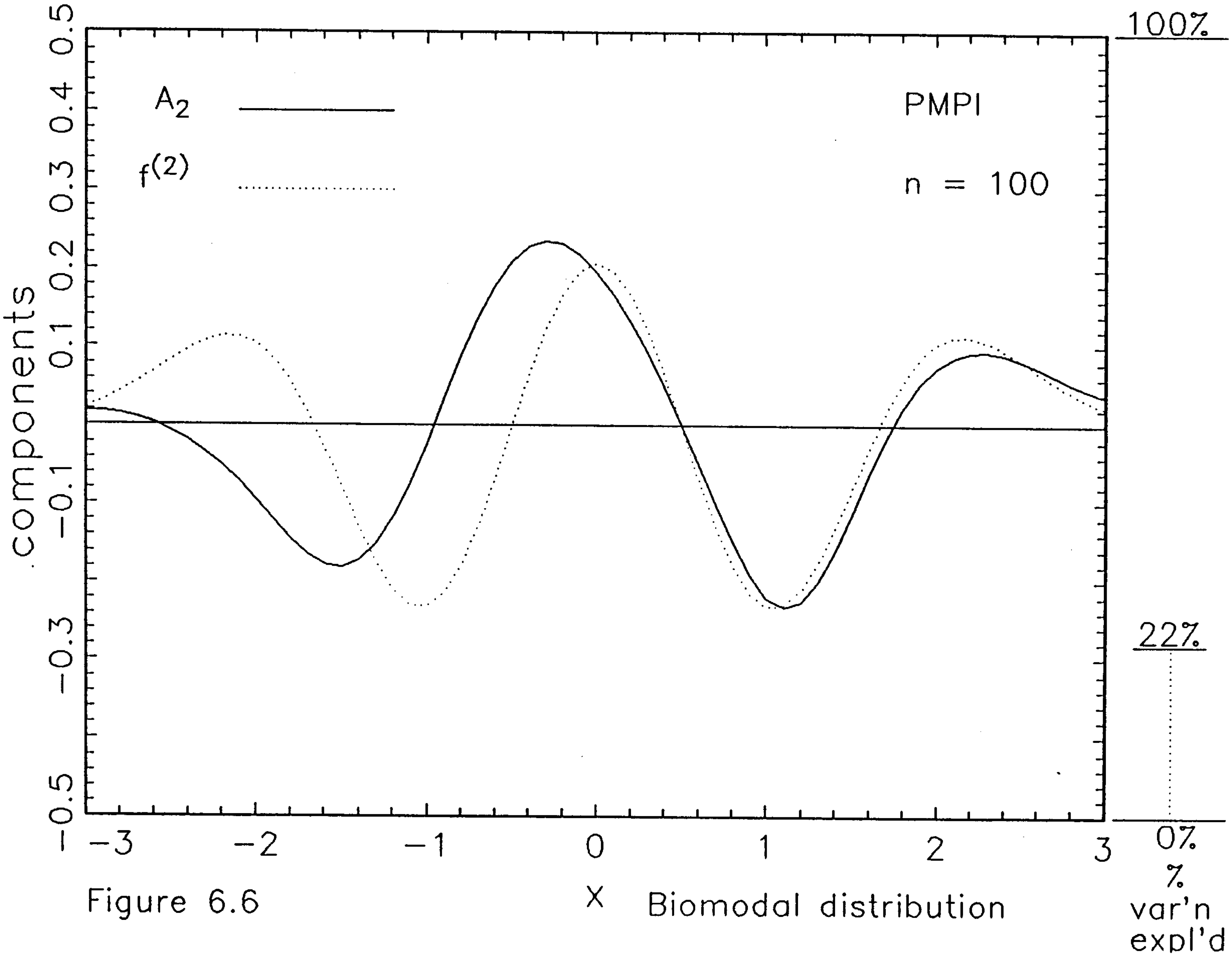
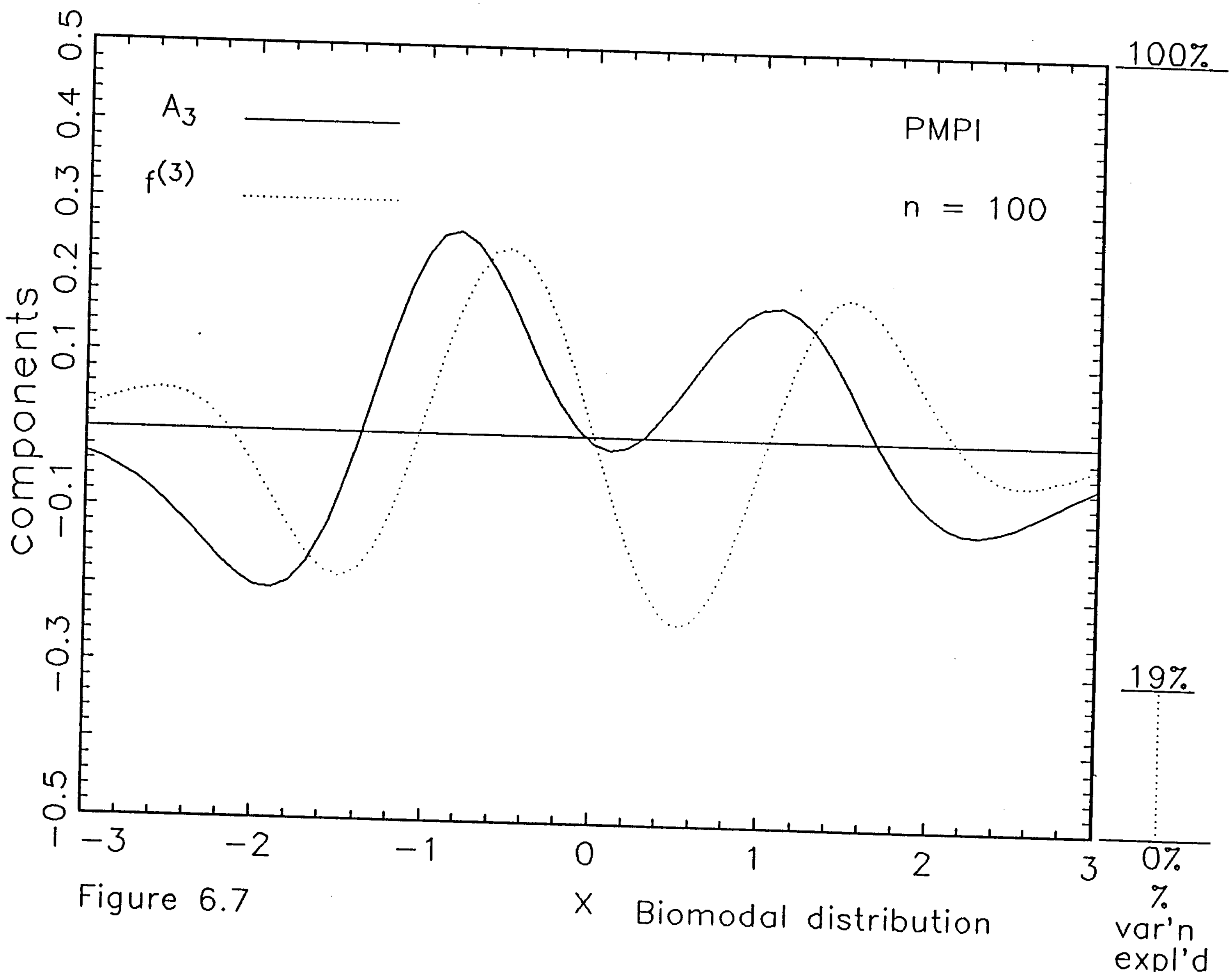

components
-0.5
-0.3
-0.1
0.1
0.2
0.3
0.4
0.5
-3
-2
-1
0
1
2
3
A_2
$f^{(2)}$
PMPI
n = 100
X Biomodal distribution
100%
22%
0%
%
var'n
expl'd

Figure 6.6



components
0.5
0.4
0.3
0.2
0.1
−0.1
−0.3
−0.5
A_3
$f^{(3)}$
PMPI
n = 100
−3
−2
−1
0
1
2
3
X Biomodal distribution
100%
19%
0%
%
var'n
expl'd

Figure 6.7

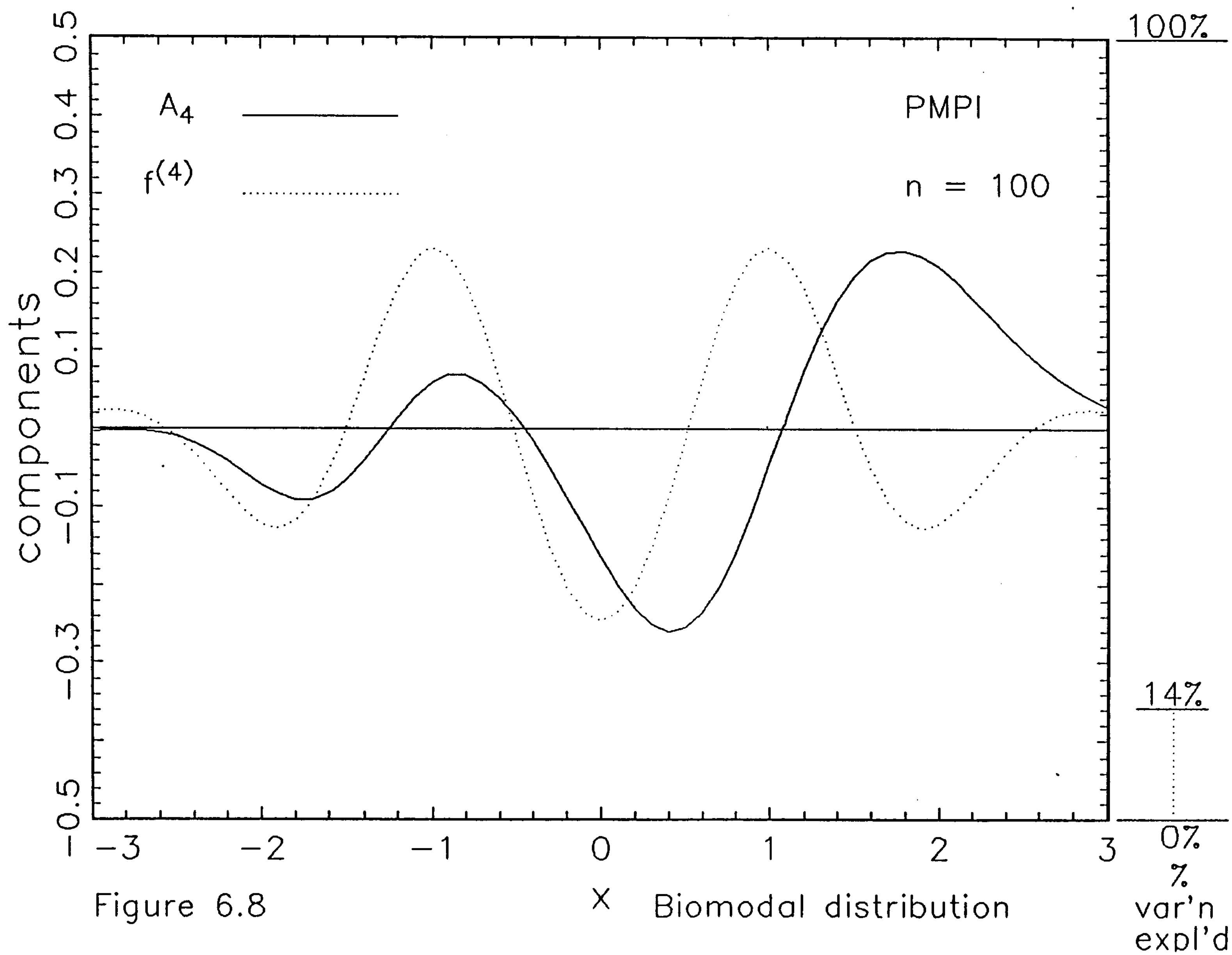
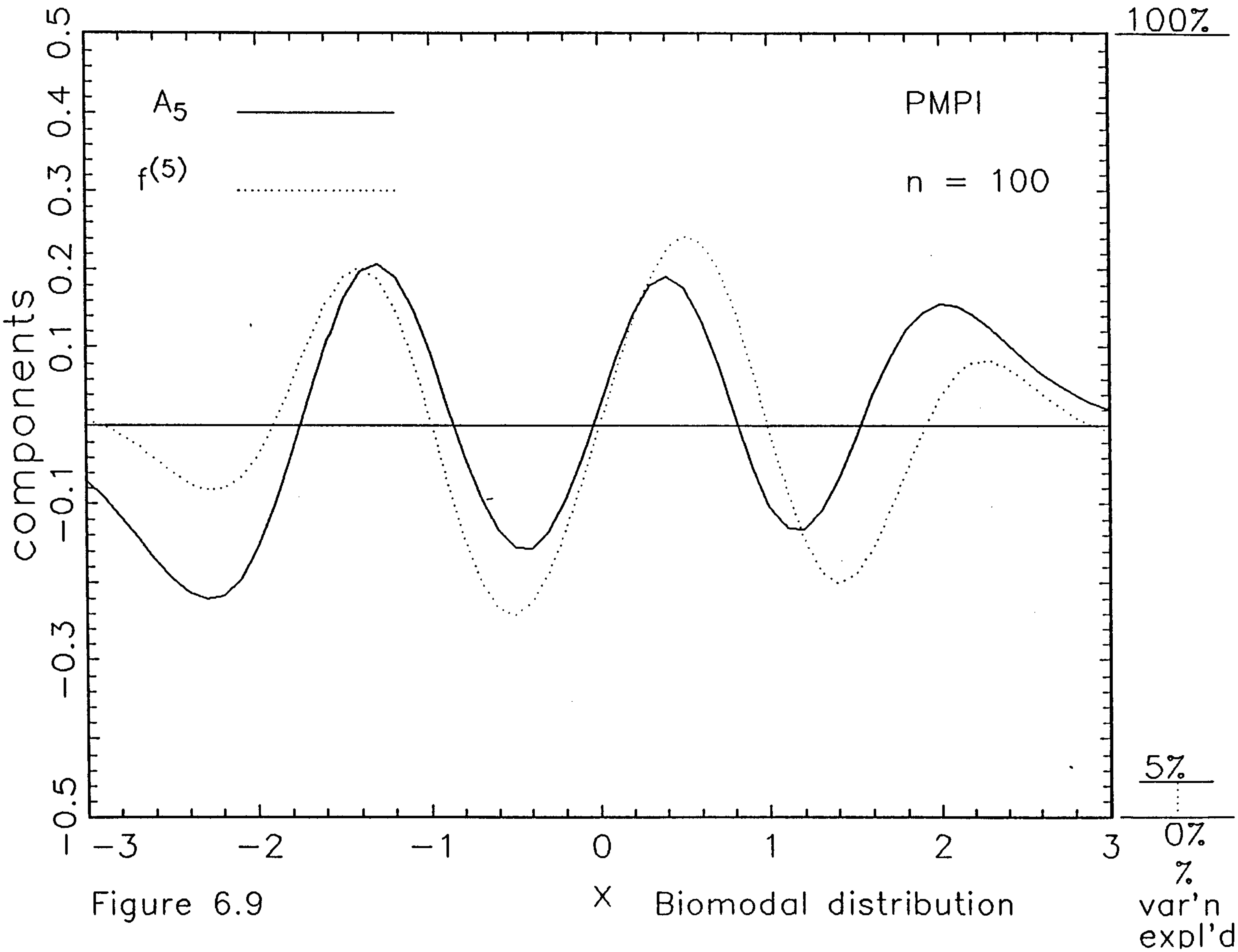


Figure 6.8

Biomodal distribution



components
0.5
0.4
0.3
0.2
0.1
−0.1
−0.3
−0.5
A_5
$f^{(5)}$
PMPI
n = 100
−3
−2
−1
0
1
2
3
X Biomodal distribution
100%
5%
0%
% var'n expl'd

Figure 6.9

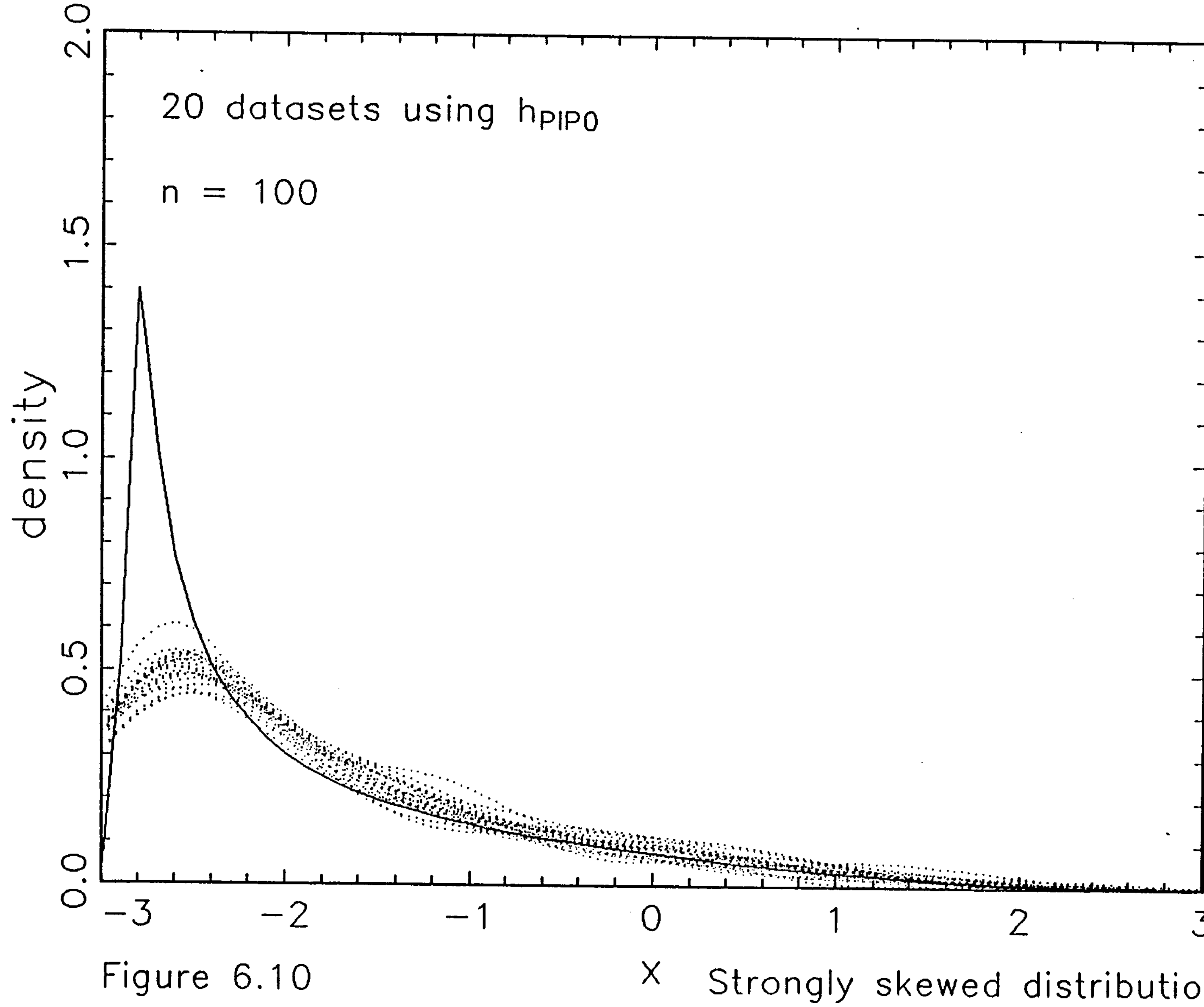


Figure 6.10 Strongly skewed distribution

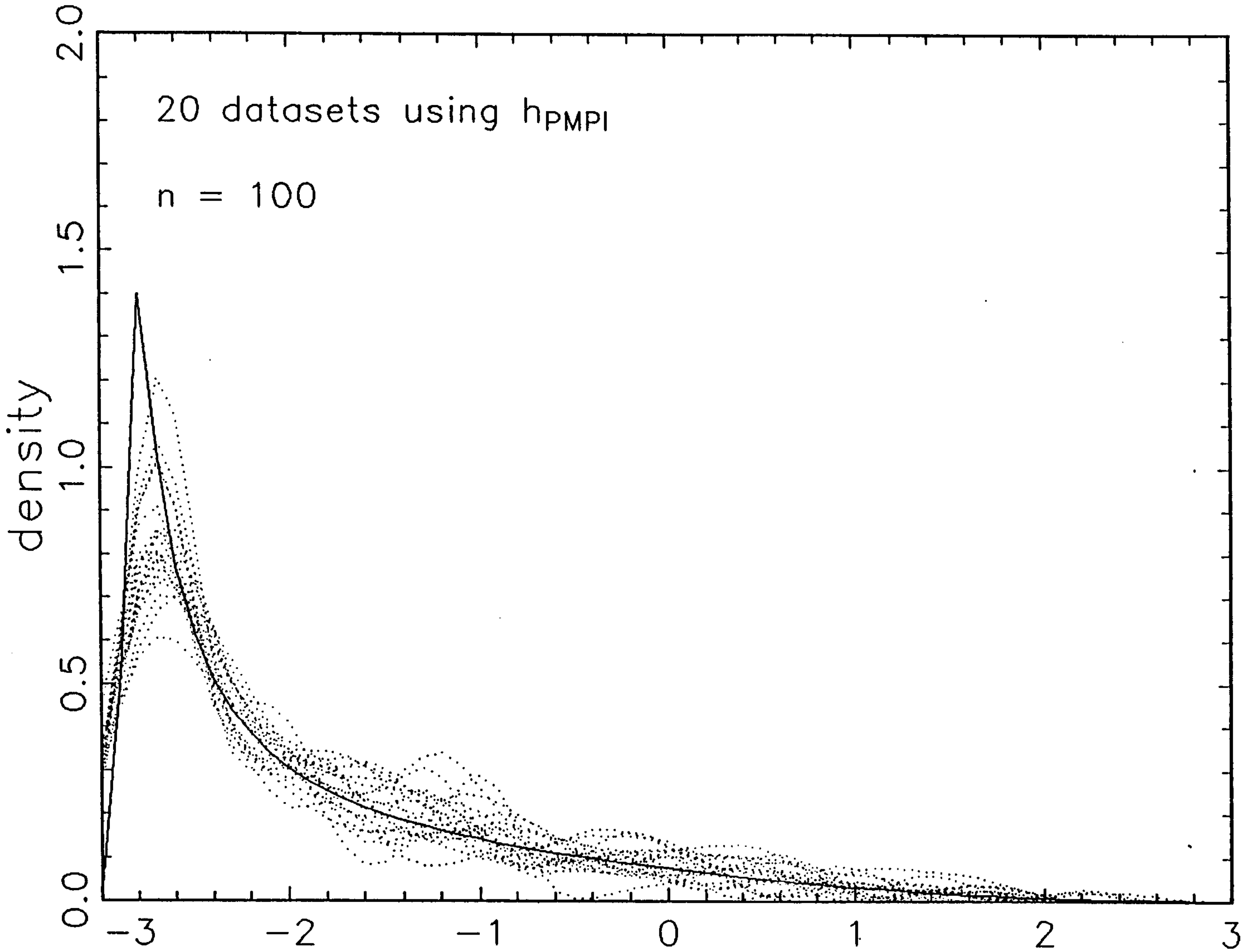


Figure 6.11 Strongly skewed distribution

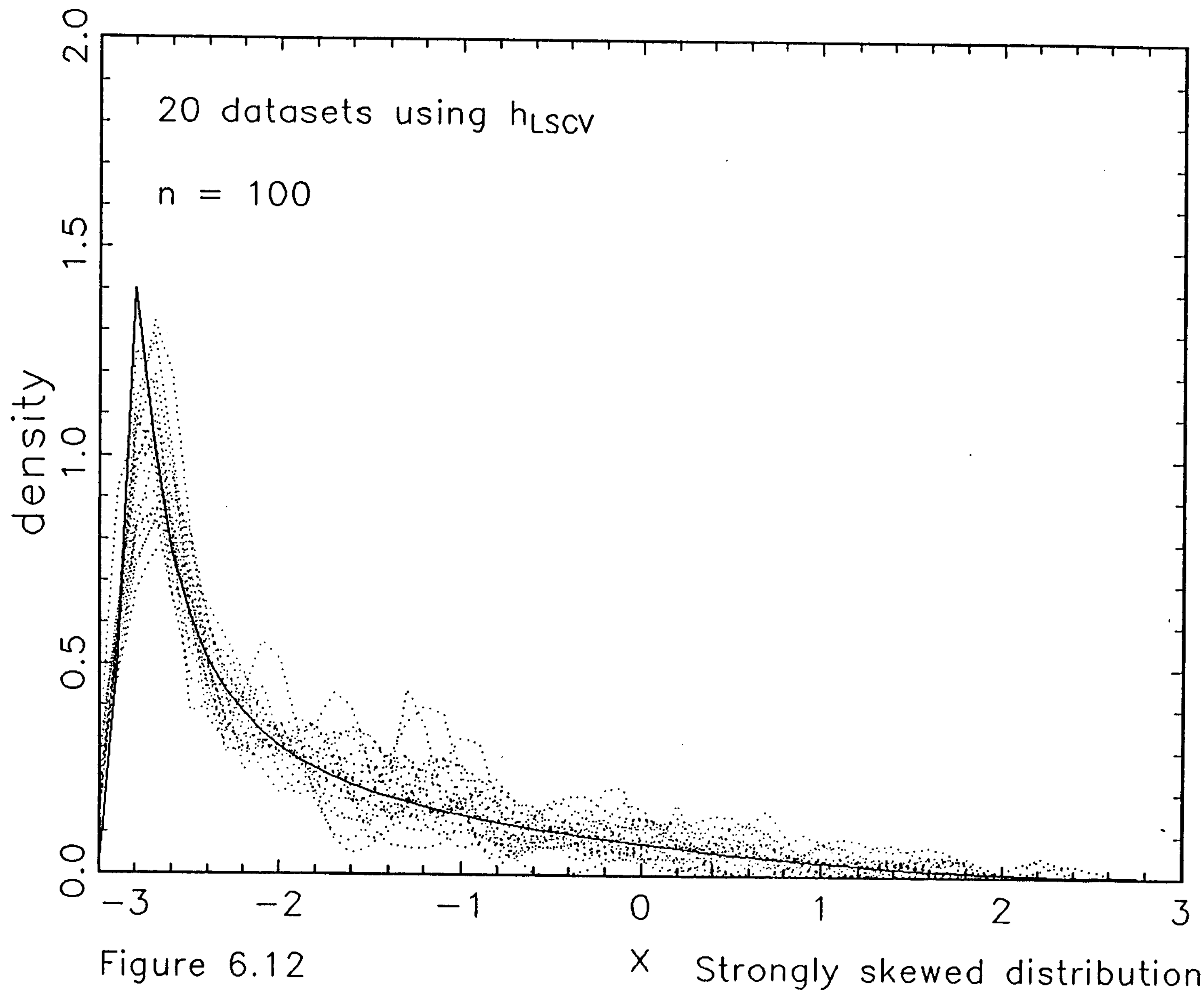


Figure 6.12 Strongly skewed distribution

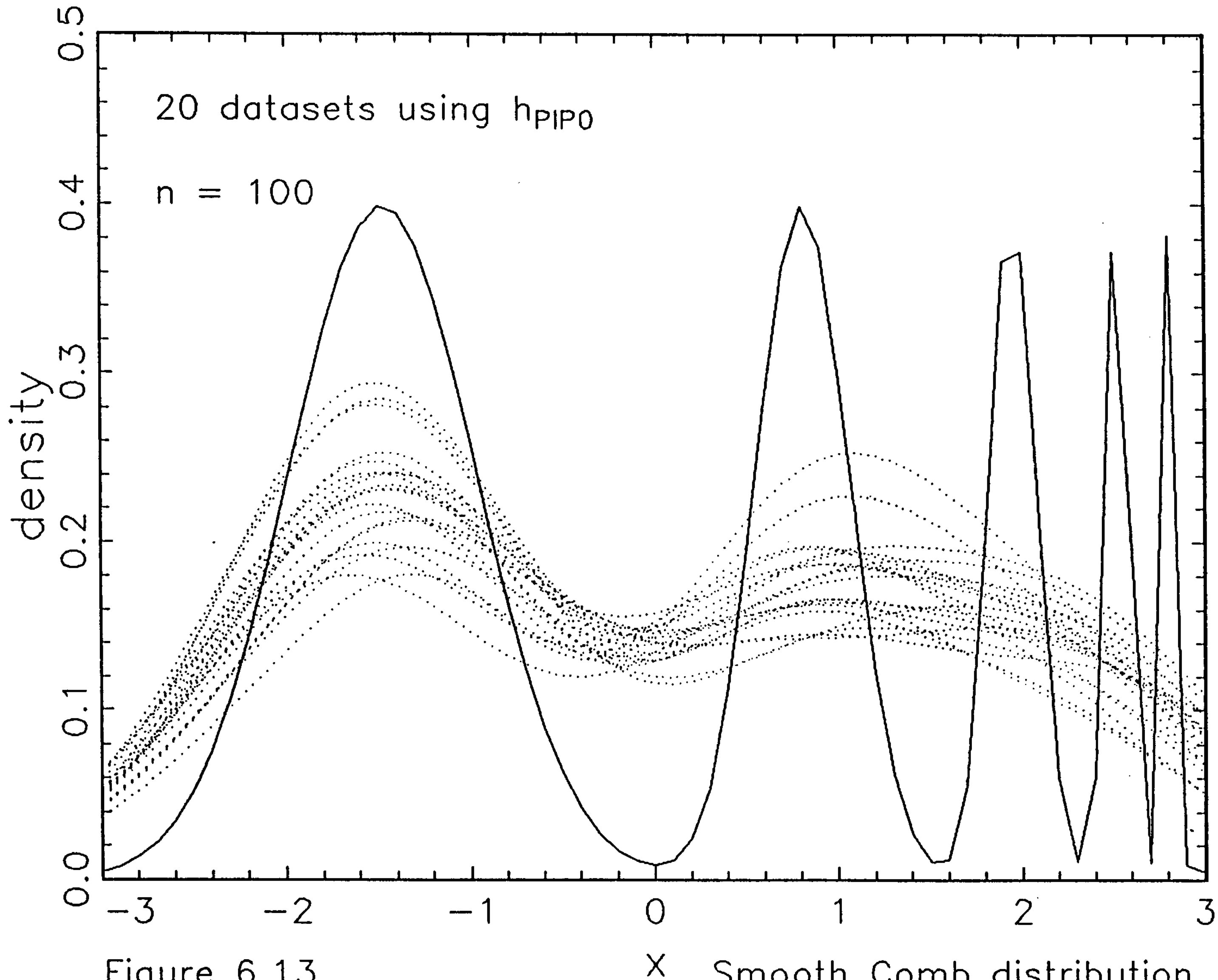


Figure 6.13 Smooth Comb distribution

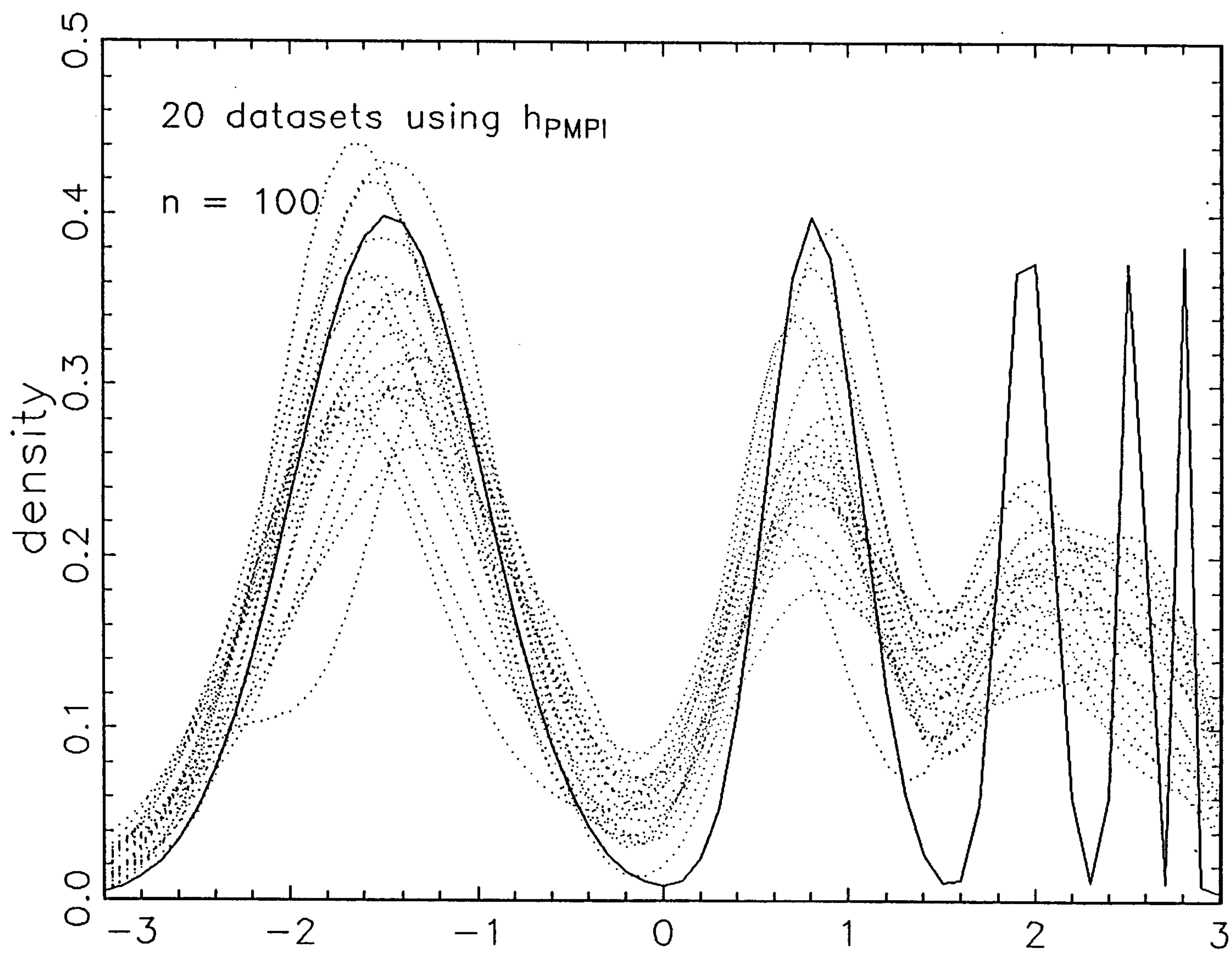


Figure 6.14 Smooth Comb distribution

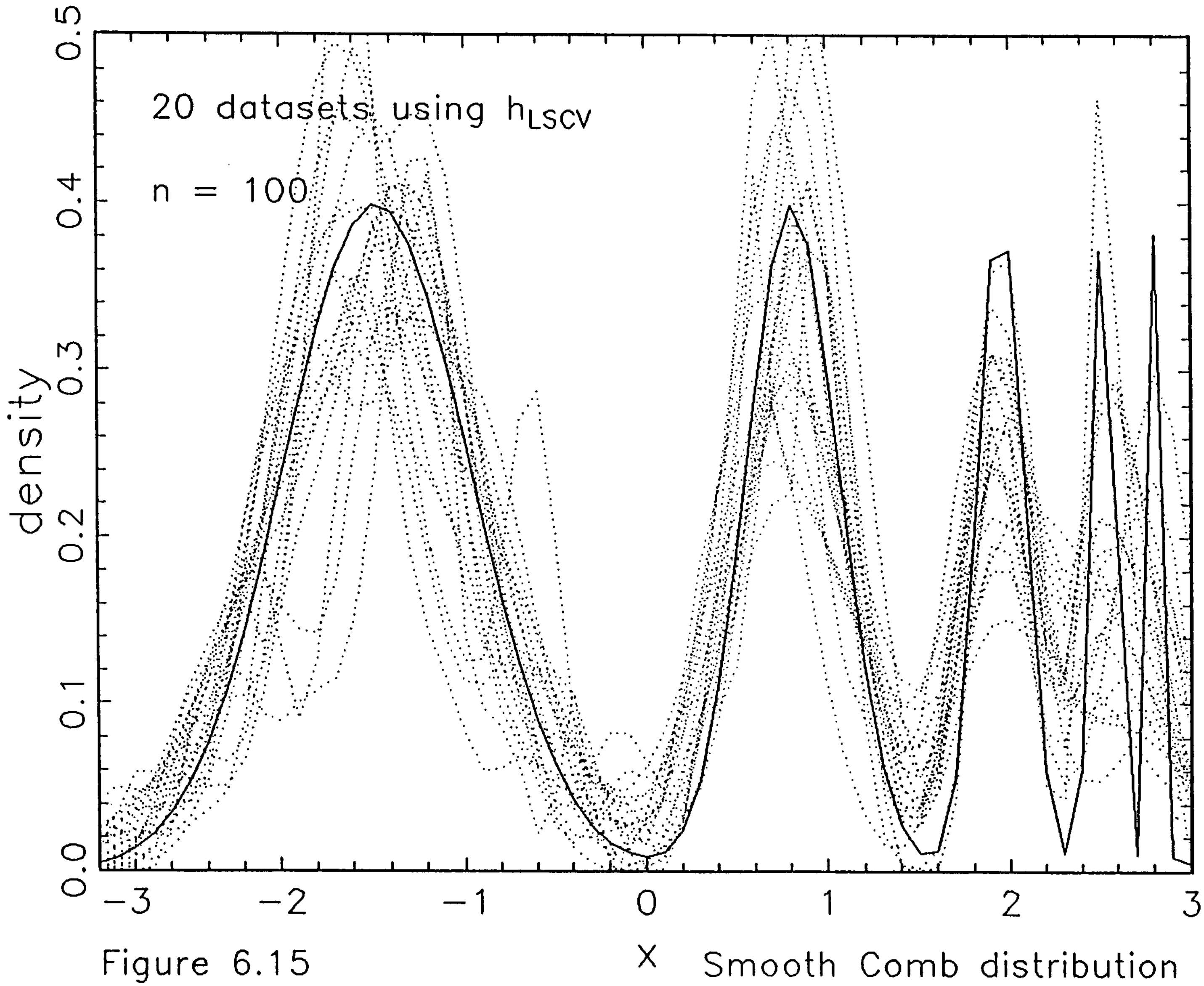


Figure 6.15

Smooth Comb distribution