Histogram Estimators of Bivariate Densities[1]

by

Joyce Ann Stevens Hüsemann

Technical Report 86-5, April 1986

[1]A Thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, Rice University.

ACKNOWLEDGEMENTS

I wish to thank my advisor, Dr. George Terrell, for his generous suggestions and kind encouragement; Dr. Robert Stevens for the important grid algorithm and for the density graphics; Dr. Hans Müller for the excellent histogram graphics; and my husband, Michael, for many hours of typing, for encouragement and understanding, and especially for *demanding* that I obtain my Ph.D.

This research was supported in part by the Army Research Office (Durham) and by the Office of Naval Research under Grant No. DAAG29-85-K-0212 and Grant No. N0001485-K-0100 respectively.

ABSTRACT

HISTOGRAM ESTIMATORS OF BIVARIATE DENSITIES

by

Joyce Ann Stevens Hüsemann

One-dimensional fixed-interval histogram estimators of univariate probability density functions are less efficient than the analogous variable-interval estimators which are constructed from intervals whose lengths are determined by the criterion of integrated mean squared error (IMSE) minimization. Similarly, two-dimensional fixed-cell-size histogram estimators of bivariate probability density functions are less efficient than variable cell size estimators whose cell sizes are determined from IMSE minimization. Only estimators whose cell sides are parallel to the coordinate axes are examined.

The estimators are classified according to the functional dependence of their cell dimensions upon x and y : each cell dimension of the Minimally Restricted Mesh depends upon both x and y ; one cell dimension of the Semi-fixed-dimension Mesh is fixed, and the other depends upon either x alone or y alone; one cell dimension of the Variable-dimension Mesh I depends upon x and the other upon y; one cell dimen-

sion of the Variable-dimension Mesh II depends upon x alone or y alone and the other depends upon both x and y. The Minimally Restricted Mesh results in the smallest IMSE of the four types, but is not implementable. The other meshes are implementable and are listed above in order of decreasing IMSE. Random vectors from Dirichlet, mixed bivariate and elliptical bivariate normal distributions were generated and used to construct optimal histograms. The Variable-dimension Mesh II produced histograms having IMSEs from 20 to 90 percent smaller than those from histograms based upon optimal fixed-dimension meshes. The most substantial improvements were observed for mixed bivariate normal densities having strongly unequal variances. Modest improvements (20%) were observed for skewed densities and slightly elliptical densities, but no improvements were observed in cases of highly elliptical densities whose axes were rotated 45% from the coordinate axes.

TABLE OF CONTENTS

DEDICATION

I dedicate this work to the memory of Joyce Alberta and Hermione.

$(x_{k-1}, x_k]$. A histogram estimate is constructed by placing a block of height $\hat{f}$ along each such interval.

The histogram estimate may also be derived in the following way (Rosenblatt, 1965): Since the probability density function is the derivative of the cumulative distribution function, we may estimate the density f by a central difference approximation to the derivative of F:

$$\hat{f}(x) = \frac{\hat{F}(x+h) - \hat{F}(x-h)}{2h}$$

$$= \frac{\left[\frac{\nu(x+h)}{n} - \frac{\nu(x-h)}{n}\right]}{2h}$$

$$= \frac{\nu(x+h) - \nu(x-h)}{2nh}$$

$$= \frac{1}{nh} \sum_{i=1}^{n} K_o\left(\frac{x - x_i}{h}\right) \qquad (1.1.1)$$

where K_o is a function of $\frac{x - x_i}{h}$. If $\frac{x - x_i}{h} > 1$, the sample point x_i lies outside the interval $(x-h, x+h]$ and K_o is set equal to zero, i.e., the point does not contribute to the estimate of f at x. If $\frac{x - x_i}{h} \leq 1$, x_i lies within the interval and K_o is set equal to $1/2$ so that x_i contributes $\frac{1}{2nh}$ to

1. Introduction

Non-parametric density estimation refers to the estimation of probability density functions whose general form as well as parameters are unknown. Of the several methods of non-parametric probability density estimation, we will discuss the four most frequently encountered and, for our purposes, the four most instructive: the kernel, the k-nearest-neighbor, the series, and the histogram.

1.1 Important non-parametric estimators of probability density

We begin with a random sample of size n from a population whose underlying density is the object of our study. Let $\nu(x)$ be the number of sample points having values less than or equal to x. Then a natural approximation to the cumulative distribution function F is

$$\hat{F} = \frac{\nu(x)}{n} .$$

Similarly, a natural approximation to the probability density function f is

$$\hat{f} = \frac{\nu(x_k) - \nu(x_{k-1})}{n(x_k - x_{k-1})}$$

where the x_k are points defined by a mesh on the real line and where $\nu_k = \nu(x_k) - \nu(x_{k-1})$ is the number of sample points falling in the *kth* interval

'rough' to represent f adequately since it will reflect random fluctuations in the data. It is apparent, therefore, that in both types of estimation, a particular choice of h may not be optimal throughout the entire domain of support of f. In addition, an optimal choice of h will depend upon sample size, since a smaller sample size requires a larger h so that a sequence of separate peaks is not obtained; on the other hand, a larger sample size will accommodate a smaller h.

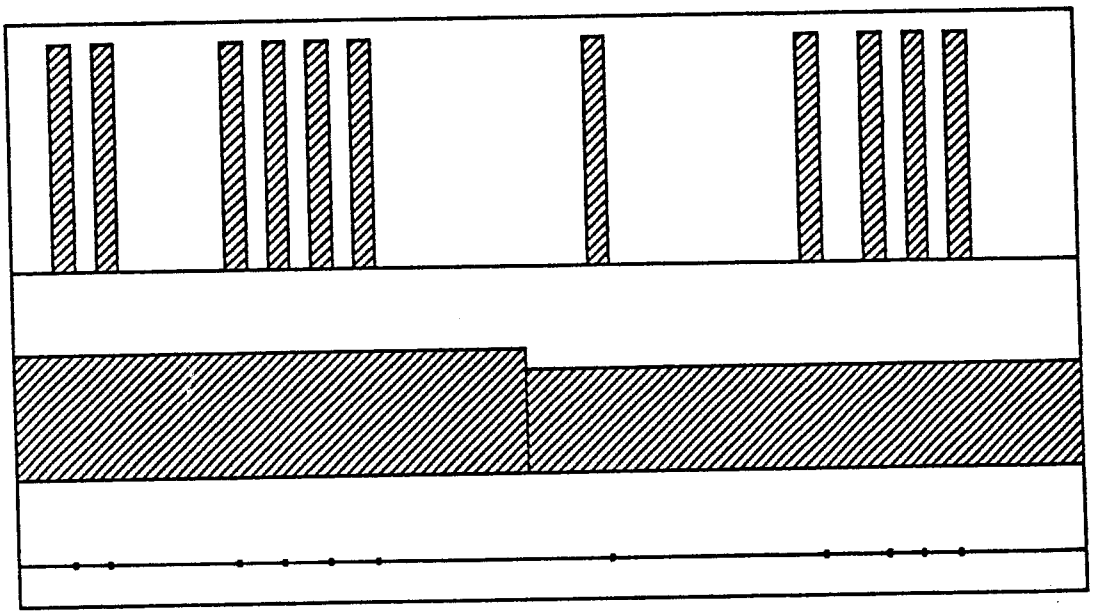

Figure 1.1. Above: example of histogram in which h is too small so that $\hat{f}$ is too rough (variance is large and bias is small). Below: example of histogram in which h is too large so that $\hat{f}$ is too smooth (variance is small and bias is large).

The k nearest-neighbor method is a fixed-frequency approach (Breiman, et al., 1977) in which the number of sample points $k = \nu(x+h) - \nu(x-h)$ falling in an interval is fixed and the sizes of the intervals $(x-h, x+h]$ vary

the estimate $\hat{f}(x)$. It should be apparent that this procedure will produce a histogram estimate of f where $\hat{f}$ at x is represented by the height of the histogram block which is centered at x and of width $2h$. The shape of the histogram will depend upon the choice of h, the parameter of interval width (also called window, cell, or bin size).

The kernel method is a straightforward generalization of the above, see Parzen (1962) and Rosenblatt (1965). If one inquires whether it is really desirable that every point falling into a given interval contribute equally to the estimate there when, perhaps, those points falling near the boundaries of the interval are of less importance to the estimate at x, an estimator which weights the contribution of points according to their distance from x recommends itself. Let $\frac{x - x_i}{h} = z$ and, as an example, let $K_o = \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{-z^2}{2}}$ in (1.1.1) so that all points of the sample contribute to the estimate at x but in inverse proportion to their distances from x. K_o is called the kernel function and may be any function which satisfies conditions which insure estimator consistency.

In the kernel approach, the shape of the estimate depends upon the choice of kernel and of 'smoothing parameter' h. In both histogram and kernel estimation, h may be considered to be a smoothing parameter since for large h, $\hat{f}$ will appear 'smoother' than when h is small, variance is large, bias is small and $\hat{f}$ follows f more closely. If h is too small, $\hat{f}$ will be too

$$\int_{-\infty}^{\infty} \left[f(x) - \sum_{i=1}^{s} a_i \phi_i(x) \right]^2 dx.$$

Use of this particular criterion leads to the following definition of the a_i:

$$\hat{a}_i = E\left[\phi_i(x)\right]$$

$$= \frac{1}{n} \sum_{k=1}^{n} \phi_i(x_k)$$

where x_k, $k = 1,\ldots,n$, is a sample from f. The density estimate becomes:

$$\hat{f}(x) = \sum_{i=1}^{m} \left[\frac{1}{n} \sum_{k=1}^{n} \phi_i(x_k) \right] \phi_i(x)$$

$$= \frac{1}{nh} \sum_{k=1}^{n} h \left[\sum_{i=1}^{m} \phi_i(x_k) \phi_i(x) \right]$$

$$= \frac{1}{nh} \sum_{k=1}^{n} K_o \left(\frac{x - x_k}{h} \right).$$

We observe that this estimator depends upon the choice of series and upon the number of terms m which are to be included in the approximation. Since the series is a global estimator with adaptability being limited to the selection of an optimal m, the approach does not readily lend itself to

in order to accommodate the constant number of points. The estimate $\hat{f}(x) = \frac{1}{nh_k} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h_k}\right)$, where h_k is the radius of the window with k neighbors, depends upon the choice of k and is subject to less random variation (i.e., has a smaller variance) when k is large but is also subject to smaller errors due to averaging (i.e., has a smaller bias) when the interval length , and therefore k, is smaller. Again, the optimal choice of parameter, k in this case, may not be constant throughout the domain of support of f.

Like the histogram and k-nearest-neighbor methods, the series method (Čencov, 1962) may also be interpreted as a special case of kernel estimation if $K_o\left(\frac{x - x_k}{h}\right)$ is set equal to $h\sum_{i=1}^{s} \phi_i(x_k)\phi_i(x)$ with ϕ_i defined as follows. The density function f can be considered to be a waveform which may be approximated by a series of orthonormal basis functions such as Fourier series, Legendre polynomials, Hermite polynomials and others. If $\{\phi_i\}$ is the selected set of orthonormal basis functions, then the density f is approximated by:

$$\hat{f}(x) = \sum_{i=1}^{\infty} a_i \phi_i(x)$$

where the series is truncated so that there are a finite number of terms m and where the a_i are approximated by $\hat{a}_i$ which are determined by minimizing an error criterion such as

will also develop some practical criteria for the construction of optimal histograms in applied problems in which the form of the underlying density is known or assumed. The problem of constructing pure data-based histograms is beyond the scope of this paper.

modifications designed to provide optimality within particular subregions of the domain of support.

1.2 Choice of methods

In any applied problem, the choice of approach depends upon practical as well as theoretical considerations. We have selected the histogram estimator for investigation, because it is conceptually the simplest, the easiest to apply and, therefore, the most commonly used of all probability density estimators.

1.3 Research objectives

In all of the density estimation approaches presented above, there is a recurring problem of simultaneously controlling both variance and bias through judicious choice of parameter as well as the difficulty created by the fact that the optimal parameter in one region of the domain of support may very well not be optimal in another region. Both of these questions have been addressed for probability densities which are functions of only one variable (Scott, 1979, 1982), (Scott and Terrell, 1983). New complications arise when densities of more than one variable are considered (Terrell, 1983, 1984, 1986), (Scott, 1985). It is the purpose of the present paper to address the questions of parameter choice and parameter variability in the case of histogram estimators of bivariate probability density functions. Results of this research should provide a basis for future research in higher dimensions. We

$$IMSE[\hat{f}(x)] = \int_{-\infty}^{\infty} Var[\hat{f}(x)]\,dx + \int_{-\infty}^{\infty} Bias^2[\hat{f}(x)]\,dx.$$

The optimal fixed interval length is derived as follows. Let $f'(x)$ be square Riemann integrable and defined over the entire real line. Let h be the length of each interval and let n be the sample size so that $h(n) \to 0$ as $n \to \infty$ and $nh(n) \to \infty$ as $n \to \infty$. Let $\nu_i(x)$ be the number of points falling in the ith interval $(y_o+ih,\ y_o+(i+1)h]$ where $y_o+ih=y_i$. If we have an independent random sample, then $\nu_i(x)$ has a binomial $(n,\ p_i)$, distribution where p_i is the probability that a sample point will fall in the ith interval, *i.e.*,

$$p_i = \int_{y_i}^{y_i+h} f(x)\,dx.$$

Then the histogram estimator of f at x becomes

$$\hat{f}(x) = \frac{\nu_i(x)}{nh} \qquad x \in (y_o+ih,\ y_o+(i+1)h]$$

which is the proportion of sample points falling in the interval divided by the size of the interval.

We will derive an expression for the integrated mean squared error beginning with the bias term (c.f., Scott, 1986).

2. Theory

2.1 One-dimensional estimation

In 1979, Scott addressed the question of optimal interval length (binwidth) for histograms which were constructed to approximate the probability density function of one random variable and whose interval lengths remained constant throughout the domain of support. The integrated mean squared error (discussed below) was suggested as a global measure of histogram error and, thereby, introduced a rigorous treatment of variance and bias into the context of histogram estimation. A global measure is preferred since it is the shape of the density that is of interest, so that an optimal histogram is one which best approximates the form of the true distribution.

Let $f(x)$ be the probability density function at a point x, and let $\hat{f}(x)$ be its estimator. Then the mean squared error (MSE) of $\hat{f}(x)$ is defined as:

$$MSE[\hat{f}(x)] = E[\,(\,\hat{f}(x) - f(x)\,)^2\,]$$

$$= E[\,(\,\hat{f}(x) - E[\hat{f}(x)]\,)^2\,] + [\,E[\hat{f}(x)] - f(x)\,]^2$$

$$= Var[\hat{f}(x)] + Bias^2[\hat{f}(x)]$$

so that the integrated mean squared error (IMSE) becomes

$$= \int_{y_i}^{y_i+h} [\, f(y) - f(y_i)\,]^2 \, dy \; - \; \frac{1}{h} \left[\int_{y_i}^{y_i+h} [\, f(y) - f(y_i)\,] \, dy \right]^2$$

$$= \int_{y_i}^{y_i+h} (y-y_i)^2 \left[\frac{f(y) - f(y_i)}{y - y_i} \right]^2 dy \; - \; \frac{1}{h} \left\{ \int_{y_i}^{y_i+h} (y-y_i) \left[\frac{f(y) - f(y_i)}{y-y_i} \right] dy \right\}^2$$

$$= \left[\frac{f(\xi_1) - f(y_i)}{\xi_1 - y_i} \right]^2 \int_{y_i}^{y_i+h} (y-y_i)^2 \, dy \; - \; \frac{1}{h} \left[\frac{f(\xi_2) - f(y_i)}{\xi_2 - y_i} \right]^2 \left[\int_{y_i}^{y_i+h} (y-y_i) \, dy \right]^2$$

$$= \left[f'(\xi_3) \right]^2 \frac{h^3}{3} \; - \; \left[f'(\xi_4) \right]^2 \frac{h^3}{4} \qquad (2.1.1)$$

where $\xi_1, \xi_2, \xi_3, \xi_4 \in (y_i, y_i+h]$.

so that over all the intervals of the histogram we have

$$\int_{-\infty}^{\infty} Bias^2[\hat{f}(x)] = \sum_{i=-\infty}^{\infty} \left[f'(\xi_{3i}) \right]^2 h \, \frac{h^2}{3} \; - \; \sum_{i=-\infty}^{\infty} \left[f'(\xi_{4i}) \right] h \, \frac{h^2}{4}$$

$$= \left[\frac{h^2}{3} - \frac{h^2}{4} \right] \int_{-\infty}^{\infty} \left[f'(y) \right]^2 dy \; + \; o(h^2)$$

$$= \frac{h^2}{12} \int_{-\infty}^{\infty} \left[f'(y) \right]^2 dy \; + \; o(h^2)$$

$$Bias[\hat{f}(y)] = E[\hat{f}(y) - f(y)]$$

$$= \frac{E(\nu_i)}{nh} - f(y)$$

$$= \frac{1}{h} \int_{y_i}^{y_i+h} f(x)\,dx - f(y)$$

$$= \frac{1}{h} \int_{y_i}^{y_i+h} [f(x) - f(y)]\,dx \qquad \text{for } y \in (y_i, y_i+h]$$

$$= \frac{1}{h} \int_{y_i}^{y_i+h} [f(x) - f(y_i)]\,dx - [f(y) - f(y_i)]$$

so that

$$\int_{y_i}^{y_i+h} Bias^2[\hat{f}(x)] = \frac{1}{h}\left[\int_{y_i}^{y_i+h} [f(x) - f(y_i)]\,dx\right]^2$$

$$- \frac{2}{h} \int_{y_i}^{y_i+h} [f(y) - f(y_i)]\,dy \int_{y_i}^{y_i+h} [f(x) - f(y_i)]\,dx$$

$$+ \int_{y_i}^{y_i+h} [f(y) - f(y_i)]^2\,dy$$

$$= \frac{1}{nh} + o\left(\frac{1}{n}\right).$$

by the Riemann integrability of $f^2(x)$ as $n \to \infty$ and $nh \to \infty$. Therefore, the integrated mean squared error becomes:

$$IMSE = \frac{1}{nh} + \frac{h^2}{12} \int_{-\infty}^{\infty} \left[f'(x)\right]^2 dx + o(h^2) + o\left(\frac{1}{n}\right).$$

Differentiating the above expression with respect to h and setting the result equal to zero, we obtain asymptotically

$$0 = \frac{1}{nh^2} + \frac{h}{6} \int_{-\infty}^{\infty} \left[f'(x)\right]^2 dx$$

so that

$$h^* = \left[\frac{6}{n \int_{-\infty}^{\infty} \left[f'(x)\right]^2 dx}\right]^{\frac{1}{3}}.$$

When the optimal constant interval length h^* is used, the following minimal *IMSE* will be obtained

$$IMSE^* = \frac{3}{2} 6^{-\frac{1}{3}} n^{-\frac{2}{3}} \left[\int_{-\infty}^{\infty} \left[f'(x)\right]^2\right]^{\frac{1}{3}}.$$

The problem of optimality in different regions of the domain of support

by the Riemann integrability of $\left[f'(y)\right]^2$ as $n \to \infty$ and $h \to 0$.

Now, considering the variance term in the expression for the integrated mean squared error, we have

$$Var[\hat{f}(y)] = \frac{Var(\nu_i)}{n^2h^2}$$

$$= \frac{np_i(1-p_i)}{n^2h^2}$$

$$= \frac{1}{nh^2}\int_{y_i}^{y_i+h} f(x)\,dx - \frac{1}{nh^2}\left[\int_{y_i}^{y_i+h} f(x)\,dx\right]^2$$

$$= \frac{1}{nh^2}\int_{y_i}^{y_i+h} f(x)\,dx - \frac{h^2}{nh^2}f^2(\xi_o) \qquad (2.2.2)$$

where $\xi_o \in (y_i\ ,\ y_i+h]$.

Then

$$\int_{-\infty}^{\infty} Var[\hat{f}(y)] = \frac{1}{nh^2}\sum_{-\infty}^{\infty}\int_{y_i}^{y_i+h} f(x)\,dx\cdot h - \frac{1}{n}\sum_{-\infty}^{\infty} f^2(\xi_o)\cdot h$$

$$= \frac{1}{nh}\int_{-\infty}^{\infty} f(x)\,dx - \frac{1}{n}\int_{-\infty}^{\infty} f^2(x)\,dx + o\left(\frac{1}{n}\right)$$

$$= \frac{f(\xi_5)}{nh(x)} - \frac{1}{n} f^2(\xi_o)$$

and

$$B = \frac{1}{h(x)} \left[\left[f'(\xi_3) \right]^2 \frac{h^3(x)}{3} - \left[f'(\xi_4) \right]^2 \frac{h^3(x)}{4} \right]$$

$$= \left[f'(\xi_3) \right]^2 \frac{h^2(x)}{3} - \left[f'(\xi_4) \right]^2 \frac{h^2(x)}{4} \quad .$$

Therefore

$$MSE_{(y_i, y_i + h]} = \frac{f(\xi_5)}{nh} - \frac{1}{n} f^2(\xi_o) + \left[f'(\xi_3) \right]^2 \frac{h^2}{3} - \left[f'(\xi_4) \right]^2 \frac{h^2}{4}$$

(2.1.3)

using (2.1.1) and (2.1.2) above.

The mean squared error averaged over an interval will no longer remain constant but will vary according to the size and location of the interval. Therefore, for a particular location x in the domain of support, we wish to obtain an optimal interval length $h(x)$ which will minimize the mean squared error averaged over that particular interval. The minimal integrated mean squared error will be obtained only when the entire sequence of optimal interval lengths is employed.

remains, however.

In 1982 and 1983, Terrell and Scott addressed this issue for probability densities of one random variable. If the interval lengths are allowed to vary, that is, if h becomes a function of x, then the mean squared error over one interval becomes

$$MSE_{(y_i, y_i+h(x))} = \frac{1}{h(x)} \int_{y_i}^{y_i+h(x)} Var[\hat{f}(x)]\, dx$$

$$+ \frac{1}{h(x)} \int_{y_i}^{y_i+h(x)} Bias^2[\hat{f}(x)]\, dx = A + B$$

where

$$A = \frac{1}{h(x)} \int_{y_i}^{y_i+h(x)} \left[\frac{1}{nh^2(x)} \int_{y_i}^{y_i+h(x)} f(x)\, dx - \frac{1}{n} f^2(\xi_o) \right]$$

$$+ \frac{1}{h(x)} \left[\left[f'(\xi_3) \right]^2 \frac{h^3(x)}{3} - \left[f'(\xi_4) \right]^2 \frac{h^3(x)}{4} \right]$$

$$= \frac{1}{h(x)} \int_{y_i}^{y_i+h(x)} \left[\frac{1}{nh^2(x)} h(x) f(\xi_5) - \frac{1}{n} f^2(\xi_o) \right]$$

$$= \frac{1}{h(x)} \left[\frac{f(\xi_5)}{nh^2(x)} h^2(x) - \frac{1}{n} f^2(\xi_o)\, h(x) \right]$$

That $h_i^*(n,x) \longrightarrow 0$ for each interval i as $n \longrightarrow \infty$ in the case of a normal density can be seen from the following argument. Let us truncate our estimate at x_o and x_k so that $\hat{f}(x) = 0$ for $x < x_o$, and $x \geq x_k$.

The integrated mean squared error

$$IMSE = E\left[\int_{-\infty}^{\infty} (f(x) - \hat{f}(x))^2 \, dx\right]$$

$$= E\left[\int_{x_o}^{x_k} (f(x) - \hat{f}(x))^2 \, dx\right] + E\left[\int_{-\infty}^{x_o} (f(x) - \hat{f}(x))^2 \, dx\right]$$

$$+ E\left[\int_{x_k}^{\infty} (f(x) - \hat{f}(x))^2 \, dx\right]$$

becomes

$$= E\left[\int_{x_o}^{x_k} (f(x) - \hat{f}(x))^2 \, dx\right] + \int_{-\infty}^{x_o} [f(x)]^2 \, dx + \int_{x_k}^{\infty} [f(x)]^2 \, dx$$

if $|x_k - x_o|$ is sufficiently large so that an arbitrarily small number of points fall outside the interval $(x_o, x_k]$.

For

Holding x constant and minimizing the mean squared error (2.1.3) over $(y_i, y_i+h]$ by differentiating with respect to h and setting the result equal to zero, we have

$$0 = -\frac{f(\xi_5)}{nh^2(x)} + \frac{2}{3}h(x)\left[f_x'(\xi_3)\right]^2 - \frac{1}{2}h(x)\left[f_x'(\xi_4)\right]^2$$

$$\frac{f(\xi_5)}{nh^2(x)} = h(x)\left[\frac{2}{3}\left[f_x'(\xi_3)\right]^2 - \frac{1}{2}\left[f_x'(\xi_4)\right]^2\right]$$

$$h^*(x) = \left[\frac{f(\xi_5)}{n\left[\frac{2}{3}\left[f_x'(\xi_3)\right]^2 - \frac{1}{2}\left[f_x'(\xi_4)\right]^2\right]}\right]^{\frac{1}{3}}$$

for a particular value of x. We assume that $h^*(x)$ is approximately constant over the interval $(y_i, y_i+h]$ and substitute $h^*(x)$ into (2.1.3) above to obtain the mean squared error over the interval:

$$MSE^*_{(y_i, y_i+h]} = \frac{3}{2}n^{-\frac{2}{3}}\left\{f^{\frac{2}{3}}(\xi_5)\left[\frac{2}{3}\left[f_x'(\xi_3)\right]^2 - \frac{1}{2}\left[f_x'(\xi_4)\right]^2\right]^{\frac{1}{3}}\right\} - \frac{1}{n}f^2(\xi_o)$$

and the integrated means squared error:

$$IMSE^* = \frac{3}{2}n^{-\frac{2}{3}}\sum_{i=-\infty}^{\infty}\left\{f^{\frac{2}{3}}(\xi_{5i})\left[\frac{2}{3}\left[f'(\xi_{3i})\right]^2 - \frac{1}{2}\left[f'(\xi_{4i})\right]^2\right]^{\frac{1}{3}}\right\}\cdot h_i^* + O\left(\frac{1}{n}\right).$$

variable interval case. If $f(x)$ is the normal distribution with mean equal to zero and variance equal to one,

$$h \approx 3.491\ n^{-\frac{1}{3}} \quad \text{for the fixed interval case}$$

and

$$h \approx 2.469\ |x|^{-\frac{2}{3}} \cdot e^{\frac{x^2}{6}} \cdot n^{-\frac{1}{3}} \quad \text{for the variable interval case.}$$

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}, \qquad 1 - F(x) \approx \frac{1}{x\sqrt{2\pi}} e^{\frac{-x^2}{2}} .$$

Choose x_n so that $1 - F(x_n) \approx \frac{1}{n}$. Then

$$n = \sqrt{2\pi}\, x_n e^{\frac{x^2}{2}}$$

and the optimal asymptotic interval width

$$h^*(x_n) = \left(\frac{6f(x_n)}{n[f'(x_n)]^2} \right)^{\frac{1}{3}} \approx \frac{1}{x_n^3} \to 0$$

as $n \to \infty$.

Since $\int_{x_n}^{\infty} f^2(x)\, dx \to 0$ as $n \to \infty$

and

$\int_{-x_n}^{x_n} [\hat{f}(x) - f(x)]^2\, dx \to \int_{-\infty}^{\infty} [\hat{f}(x) - f(x)]^2\, dx$, we have

$$IMSE^* = \frac{3}{2}\, 6^{-\frac{1}{3}}\, n^{-\frac{2}{3}} \int_{-\infty}^{\infty} \left[f(x)\, f'(x) \right]^{\frac{2}{3}} dx + o\left(\max h_i^2 \right) + O\left(\frac{1}{n} \right)$$

by the Riemann integrability of $f(y)$ and $\left[f'(y) \right]^2$.

Comparison of the minimal obtainable integrated mean squared error for the fixed and variable interval cases shows that the minimal *IMSE* for the fixed interval case is always greater than or equal to the minimal *IMSE* for the

Let $X=(x,y)$ be a bivariate random variable with joint probability density function $f(x,y)$. If the domain of support is subdivided into rectangles of the form $(x_i, x_i+g]\times(y_i, y_i+h]$ where $g>0$ and $h>0$ are the lengths of the sides, a histogram estimator of $f(x,y)$ at the point (x,y) may be defined in analogy with the univariate case as:

$$\hat{f}(x,y) = \frac{\nu(x,y)}{ngh} \qquad \text{for } x \in (x_i, x_i+g],\ y \in (y_i, y_i+h]$$

where $\nu(x,y)$ is the number of sample points falling in the rectangle and where g and h may be constant, functions of x_i or y_i alone, or functions of both x_i and y_i. The integrated mean squared error is defined as before and in the bivariate case becomes:

$$IMSE[\hat{f}(x,y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E[\,(\hat{f}(x,y) - f(x,y))^2\,]\,dxdy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left\{E[\,\hat{f}(x,y) - E[\hat{f}(x,y)]^2] + [\,E[\hat{f}(x,y)] - f(x,y)\,]^2\right\} dxdy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[\,Var[\hat{f}(x,y)] + Bias^2[\hat{f}(x,y)\,]\,]\,dxdy.$$

Once again $\nu(x,y)$ has a binomial (n,p) distribution with p the probability that a sample point (x,y) lies in the above rectangle centered at

2.2 Two-dimensional estimation

In our discussion of the one-dimensional case of histogram estimation it was apparent that there were only two possible types of mesh from which a histogram might be constructed: h could remain constant throughout the domain of support or could vary according to some criterion of optimality. However, when the concept of histogram estimation is extended to two dimensions, the number of possible grid types becomes infinite: the plane may be partitioned into sets of any shape as long as the sets are mutually exclusive and the subdivision is exhaustive. Since most histograms based upon arbitrary partitions would not be implementable, Terrell (1983) confined his investigation to rectangular meshes with cell sides parallel to the coordinate axes. Scott (1985) studied a number of other mesh types, some having triangular and others hexagonal shapes, but found that hexagons resulted in only slightly improved estimates at a cost of some difficulty in implementation and that regular triangles, which were also difficult to implement, resulted in worse estimates than did Terrell's variable-dimension rectangles. Terrell also demonstrated that a partition of the plane by a rhomboidal mesh produced histogram estimators equivalent to those produced by a rectangular mesh. In view of the above, the mesh types in the present paper have been confined to rectangular grids having cell sides of variable length and width parallel to the coordinate axes. The problem of grid orientation is a separate topic of research and will not be treated here.

$$+ [\, f(s,t) - f(x_i,y_i)\,]^2,$$

so that the bias over one rectangle is

$$\int_{y_i}^{y_i+h}\int_{x_i}^{x_i+g} Bias^2[\,\hat{f}(x,y)] = \frac{1}{gh}\left[\int_{y_i}^{y_i+h}\int_{x_i}^{x_i+g} [\, f(x,y) - f(x_i,y_i)\,]\, dxdy\right]^2$$

$$- \frac{2}{gh}\left[\int_{y_i}^{y_i+h}\int_{x_i}^{x_i+g} [\, f(x,y) - f(x_i,y_i)\,]\, dxdy\right]$$

$$\cdot \left[\int_{y_i}^{y_i+h}\int_{x_i}^{x_i+g} [\, f(s,t) - f(x_i,y_i)\,]\, dsdt\right]$$

$$+ \int_{y_i}^{y_i+h}\int_{x_i}^{x_i+g} [\, f(s,t) - f(x_i,y_i)\,]^2\, dsdt$$

$$= \int_{y_i}^{y_i+h}\int_{x_i}^{x_i+g} [\, f(x,y) - f(x_i,y_i)\,]^2\, dxdy$$

$$- \frac{1}{gh}\left[\int_{y_i}^{y_i+h}\int_{x_i}^{x_i+g} [\, f(x,y) - f(x_i,y_i)\,]\, dxdy\right]^2$$

$\left(x_i+\frac{g}{2},\ y_i+\frac{h}{2}\right)$, so that

$$p = \int_{y_i}^{y_i+h} \int_{x_i}^{x_i+g} f(x,y)\, dxdy \qquad where \quad x \in (x_i,\ x_i+g] \text{ and } y \in (y_i,\ y_i+h]$$

and

$$Bias[\hat{f}(x,y)] = E[\ \hat{f}(x,y) - f(x,y)\ ]$$

$$= E\left[\frac{\nu(x,y)}{ngh}\right] - f(x,y)$$

$$= \frac{n}{ngh} \int_{y_i}^{y_i+h} \int_{x_i}^{x_i+g} f(x,y)\, dxdy - f(x,y)$$

$$= \frac{1}{gh} \int_{y_i}^{y_i+h} \int_{x_i}^{x_i+g} [\ f(x,y) - f(x_i,y_i)\ ]\, dxdy \ - \ [\ f(x,y) - f(x_i,y_i)\ ]$$

and

$$Bias^2[\hat{f}(x,y)] = \frac{1}{g^2h^2}\left[\int_{y_i}^{y_i+h} \int_{x_i}^{x_i+g} [\ f(x,y) - f(x_i,y_i)\ ]\, dxdy\right]^2$$

$$- \frac{2}{gh}[\ f(s,t) - f(x_i,y_i)\ ]\left[\int_{y_i}^{y_i+h} \int_{x_i}^{x_i+g} [\ f(x,y) - f(x_i,y_i)\ ]\, dxdy\right]$$

$$+ 2\frac{\partial f}{\partial x}(\xi_{i2},\xi_{j2})\frac{\partial f}{\partial y}(\xi_{i3},\xi_{j3})\int_{y_i}^{y_i+h}\int_{x_i}^{y_i+g}(x-x_i)(y-y_i)\,dxdy$$

$$+ \left[\frac{\partial f}{\partial y}(\xi_{i4},\xi_{j4})\right]^2\int_{y_i}^{y_i+h}\int_{x_i}^{y_i+g}(y-y_i)^2\,dxdy$$

$$- \frac{1}{gh}\left[\frac{\partial f}{\partial x}(\xi_{i5},\xi_{j5})\int_{y_i}^{y_i+h}\int_{x_i}^{y_i+g}(x-x_i)\,dxdy + \frac{\partial f}{\partial y}(\xi_{i6},\xi_{j6})\int_{y_i}^{y_i+h}\int_{x_i}^{y_i+g}(y-y_i)\,dxdy\right]^2$$

by the integral form of the mean value theorem where each ξ_{ik} , k=1,..,6, is a particular value of ξ_x for some x and where each ξ_{jk} , k=1,..,6, is a particular value of ξ_y for some y, and where $x - x_i > 0$, $y - y_i > 0$ since $x \in (x_i, x_i+g]$, $y \in (y_i, y_i+h]$. The above then becomes

$$= \left[\frac{\partial f}{\partial x}(\xi_{i1},\xi_{j1})\right]^2\frac{g^3h}{3} + 2\frac{\partial f}{\partial x}(\xi_{i2},\xi_{j2})\frac{\partial f}{\partial y}(\xi_{i3},\xi_{j3})\frac{g^2h^2}{4} + \left[\frac{\partial f}{\partial y}(\xi_{i4},\xi_{j4})\right]^2\frac{gh^3}{3}$$

$$- \frac{1}{gh}\left[\frac{\partial f}{\partial x}(\xi_{i5},\xi_{j5})\frac{g^2h}{2} + \frac{\partial f}{\partial y}(\xi_{i6},\xi_{j6})\frac{gh^2}{2}\right]^2$$

$$= \frac{g^3h}{3}\left[\frac{\partial f}{\partial x}(\xi_{i1},\xi_{j1})\right]^2 + \frac{g^2h^2}{2}\frac{\partial f}{\partial x}(\xi_{i2},\xi_{j2})\frac{\partial f}{\partial y}(\xi_{i3},\xi_{j3}) + \frac{gh^3}{3}\left[\frac{\partial f}{\partial y}(\xi_{i4},\xi_{j4})\right]^2$$

$$- \frac{1}{gh}\,\frac{g^2h^2}{4}\left[g\frac{\partial f}{\partial x}(\xi_{i5},\xi_{j5}) + h\frac{\partial f}{\partial y}(\xi_{i6},\xi_{j6})\right]^2 .$$

$$= \int_{y_i}^{y_i+h} \int_{x_i}^{x_i+g} \left[(x-x_i)\frac{\partial f(a)}{\partial x} + (y-y_i)\frac{\partial f(b)}{\partial y} \right]^2 dxdy$$

$$- \frac{1}{gh} \left[\int_{y_i}^{y_i+h} \int_{x_i}^{x_i+g} \left[(x-x_i)\frac{\partial f(a)}{\partial x} + (y-y_i)\frac{\partial f(b)}{\partial y} \right] dxdy \right]^2$$

where $a = (x_i+\xi(x-x_i)\ ,\ y_i+\xi(y-y_i))$ and $b = (x_i+\xi(x-x_i)\ ,\ y_i+\xi(y-y_i))$ and where $0 < \xi < 1$, by the mean value theorem.

Let $\xi_x = x_i+\xi(x-x_i)$ and $\xi_y = y_i+\xi(y-y_i)$.

Then the above becomes

$$= \int_{y_i}^{y_i+h} \int_{x_i}^{y_i+g} (x-x_i)^2 \left[\frac{\partial f}{\partial x}(\xi_x,\xi_y) \right]^2 + 2(x-x_i)(y-y_i) \left[\frac{\partial f}{\partial x}(\xi_x,\xi_y) \right] \left[\frac{\partial f}{\partial y}(\xi_x,\xi_y) \right]$$

$$+ (y-y_i)^2 \left[\frac{\partial f}{\partial y}(\xi_x,\xi_y) \right]^2 dxdy$$

$$- \frac{1}{gh} \left[\int_{y_i}^{y_i+h} \int_{x_i}^{y_i+g} \left[(x-x_i)\frac{\partial f}{\partial x}(\xi_x,\xi_y) + (y-y_i)\frac{\partial f}{\partial y}(\xi_x,\xi_y) \right] dxdy \right]^2$$

$$= \left[\frac{\partial f}{\partial x}(\xi_{i1},\xi_{j1}) \right]^2 \int_{y_i}^{y_i+h} \int_{x_i}^{y_i+g} (x-x_i)^2\ dxdy$$

$$+\sum_{j=-\infty}^{\infty}\sum_{i=-\infty}^{\infty} gh\left[\frac{gh}{2}\left(\frac{\partial f}{\partial x}(\xi_{i2},\xi_{j2})\right)\left(\frac{\partial f}{\partial y}(\xi_{i3},\xi_{j3})\right)-\frac{gh}{2}\left(\frac{\partial f}{\partial x}(\xi_{i5},\xi_{j5})\right)\left(\frac{\partial f}{\partial y}(\xi_{i6},\xi_{j6})\right)\right]$$

$$+\sum_{j=-\infty}^{\infty}\sum_{i=-\infty}^{\infty} gh\left[\frac{h^2}{3}\left(\frac{\partial f}{\partial y}(\xi_{i4},\xi_{j4})\right)^2-\frac{h^2}{4}\left(\frac{\partial f}{\partial y}(\xi_{i6},\xi_{j6})\right)^2\right].$$

As $n \longrightarrow \infty$ and $g,h \longrightarrow 0$, we have by the Riemann integrability of partial derivatives :

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} Bias^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{g^2}{12}\left(\frac{\partial f}{\partial x}\right)^2 + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{h^2}{12}\left(\frac{\partial f}{\partial y}\right)^2$$

The variance of $\hat{f}(x,y)$ is :

$$Var[\hat{f}(x,y)] = Var\left[\frac{\nu(x,y)}{ngh}\right]$$

$$= \frac{1}{n^2g^2h^2} Var[\nu(x,y)]$$

$$= \frac{n}{n^2g^2h^2} p(1-p)$$

$$= \frac{1}{ng^2h^2}\int_{y_i}^{y_i+h}\int_{x_i}^{y_i+g} f(x,y)\,dxdy \;-\; \frac{1}{ng^2h^2}\left[\int_{y_i}^{y_i+h}\int_{x_i}^{y_i+g} f(x,y)\,dxdy\right]^2$$

$$= \frac{g^3h}{3}\left[\frac{\partial f}{\partial x}(\xi_{i1},\xi_{j1})\right]^2 + \frac{g^2h^2}{2}\frac{\partial f}{\partial x}(\xi_{i2},\xi_{j2})\frac{\partial f}{\partial y}(\xi_{i3},\xi_{j3}) + \frac{gh^3}{3}\left[\frac{\partial f}{\partial y}(\xi_{i4},\xi_{j4})\right]^2$$

$$- \frac{gh}{4}\left[g^2\left[\frac{\partial f}{\partial x}(\xi_{i5},\xi_{j5})\right]^2 + 2gh\frac{\partial f}{\partial x}(\xi_{i5},\xi_{j5})\frac{\partial f}{\partial y}(\xi_{i6},\xi_{j6}) + h^2\left[\frac{\partial f}{\partial y}(\xi_{i6},\xi_{j6})\right]^2\right]$$

$$= \frac{g^3h}{3}\left[\frac{\partial f}{\partial x}(\xi_{i1},\xi_{j1})\right]^2 + \frac{g^2h^2}{2}\frac{\partial f}{\partial x}(\xi_{i2},\xi_{j2})\frac{\partial f}{\partial y}(\xi_{i3},\xi_{j3}) + \frac{gh^3}{3}\left[\frac{\partial f}{\partial y}(\xi_{i4},\xi_{j4})\right]^2$$

$$- \frac{g^3h}{4}\left[\frac{\partial f}{\partial x}(\xi_{i5},\xi_{j5})\right]^2 - \frac{g^2h^2}{2}\frac{\partial f}{\partial x}(\xi_{i5},\xi_{j5})\frac{\partial f}{\partial y}(\xi_{i6},\xi_{j6}) - \frac{gh^3}{4}\left[\frac{\partial f}{\partial y}(\xi_{i6},\xi_{j6})\right]^2$$

$$= gh\left[\frac{g^2}{3}\left[\frac{\partial f}{\partial x}(\xi_{i1},\xi_{j1})\right]^2 - \frac{g^2}{4}\left[\frac{\partial f}{\partial x}(\xi_{i5},\xi_{j5})\right]^2\right]$$

$$+ gh\left[\frac{gh}{2}\left[\frac{\partial f}{\partial x}(\xi_{i2},\xi_{j2})\right]\left[\frac{\partial f}{\partial y}(\xi_{i3},\xi_{j3})\right] - \frac{gh}{2}\left[\frac{\partial f}{\partial x}(\xi_{i5},\xi_{j5})\right]\left[\frac{\partial f}{\partial y}(\xi_{i6},\xi_{j6})\right]\right]$$

$$+ gh\left[\frac{h^2}{3}\left[\frac{\partial f}{\partial y}(\xi_{i4},\xi_{j4})\right]^2 - \frac{h^2}{4}\left[\frac{\partial f}{\partial y}(\xi_{i6},\xi_{j6})\right]^2\right],$$

so that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} Bias^2[\hat{f}(x,y)] = \sum_{j=-\infty}^{\infty}\sum_{i=-\infty}^{\infty} gh\left[\frac{g^2}{3}\left[\frac{\partial f}{\partial x}(\xi_{i1},\xi_{j1})\right]^2 - \frac{g^2}{4}\left[\frac{\partial f}{\partial x}(\xi_{i5},\xi_{j5})\right]^2\right]$$

ting the expression above equal to zero, we obtain the following conditions :

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \eta_x \left[\frac{-f}{ng^2h} + \frac{g}{6}\left(\frac{\partial f}{\partial x}\right)^2 \right] dxdy = 0$$

and (2.2.2)

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \eta_y \left[\frac{-f}{ngh^2} + \frac{h}{6}\left(\frac{\partial f}{\partial y}\right)^2 \right] dxdy = 0 \ .$$

If we solve the above equations simultaneously for g and h and choose the rectangular dimensions accordingly, we will obtain the minimum integrated mean squared error. Different solutions to the above equations are obtained depending upon whether g and h are constant, functions of only one variable or functions of both variables, each case reflecting a different type of mesh.

2.3 Minimally restricted (Free) mesh

If g and h are functions of both x and y, we obtain a subdivision of the domain of support by rectangles of arbitrary dimensions.

$$= \frac{1}{ng^2h^2} \int_{y_i}^{y_i+h} \int_{x_i}^{y_i+g} f(x,y)\, dxdy \; - \; \frac{gh}{ng^2h^2} f^2(\xi_i,\xi_j)$$

where $\xi_i \in (y_i,\ x_i+g]$, $\xi_j \in (y_i,\ y_i+h]$, so that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} Var[\hat{f}(x,y)]\, dxdy = \frac{1}{ng^2h^2} \sum_{j=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} \int_{y_i}^{y_i+h} \int_{x_i}^{y_i+g} f(x,y)\, dxdy \cdot gh$$

$$- \frac{1}{ngh} \sum_{j=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} f^2(\xi_i,\xi_j) \cdot gh$$

$$= \frac{1}{ngh} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\, dxdy \; - \; \frac{1}{n} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f^2(x,y)\, dxdy \; + \; o\left(\frac{1}{n}\right)$$

as $n \to \infty$.

Thus the integrated mean squared error becomes :

$$IMSE = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\frac{f(x,y)}{ngh} + \frac{g^2}{12} \left(\frac{\partial f}{\partial x} \right)^2 + \frac{h^2}{12} \left(\frac{\partial f}{\partial y} \right)^2 \right] dxdy \qquad (2.2.1)$$

$$+ \; o(g^2) \; + \; o(h^2) \; + \; O\left(\frac{1}{n}\right)$$

Now, treating g and h as if they were continuously varying functions in x and y, we take the derivative with respect to g in the arbitrary direction η_x and the derivative with respect to h in the arbitrary direction η_y and set-

$$g(x,y) = \left[\frac{6f(x,y)\left|\frac{\partial f}{\partial y}\right|}{n\left|\frac{\partial f}{\partial y}\right|^3} \right]^{\frac{1}{4}}$$

and

$$h(x,y) = \left[\frac{6f(x,y)\left|\frac{\partial f}{\partial x}\right|}{n\left|\frac{\partial f}{\partial y}\right|^3} \right]^{\frac{1}{4}}$$

which yield an optimal integrated mean squared error

$$IMSE^* = 2 \cdot 6^{-\frac{1}{2}} n^{-\frac{1}{2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [f(x,y)]^{\frac{1}{2}} \left[\left|\frac{\partial f}{\partial x}\right| \cdot \left|\frac{\partial f}{\partial y}\right| \right]^{\frac{1}{2}} dxdy \ .$$

2.4 Fixed-dimension (Regular) mesh

If g and h remain constant throughout the domain of support, we obtain a subdivision which is mutually exclusive, exhaustive, and easily implementable.

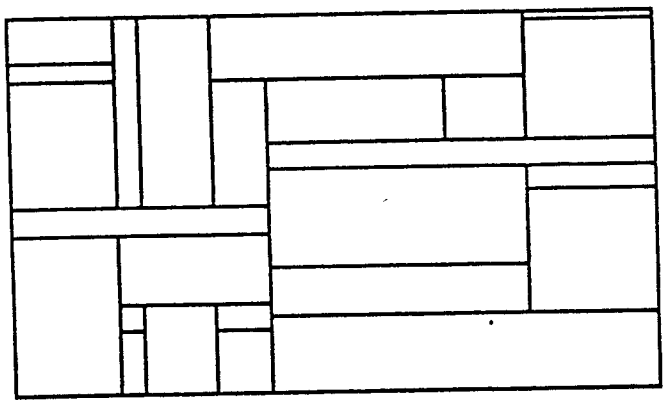

Figure 2.3. Minimally restricted (Free) mesh

Unfortunately, these will not necessarily be either mutually exclusive or exhaustive if we try to specify the optimal width and height in each region of the plane. Such a scheme is clearly not implementable but is rather of theoretical interest because it provides a lower bound for the integrated mean squared error for all possible rectangular meshes whoses cell sides are parallel to the coordinate axes. Since η_x and η_y are arbitrary functions of x and y, the equations (2.2.2) hold if and only if

$$\frac{f(x,y)}{ng^2(x,y)h(x,y)} = \frac{1}{6}g(x,y)\left[\frac{\partial f}{\partial x}\right]^2$$

and

$$\frac{f(x,y)}{ng(x,y)h^2(x,y)} = \frac{1}{6}h(x,y)\left[\frac{\partial f}{\partial y}\right]^2$$

The solution to these equations is easily obtained as

$$g = \left[\frac{36 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{\partial f}{\partial y} \right)^2 dxdy}{n^2 \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{\partial f}{\partial x} \right)^2 dxdy \right]^3} \right]^{\frac{1}{8}}$$

and

$$h = \left[\frac{36 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{\partial f}{\partial x} \right)^2 dxdy}{n^2 \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{\partial f}{\partial y} \right)^2 dxdy \right]^3} \right]^{\frac{1}{8}} .$$

Substitution of these values into (2.2.1) yields the minimal integrated mean squared error for this mesh type :

$$IMSE^* = 2{\cdot}6^{-\frac{1}{2}} n^{-\frac{1}{2}} \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{\partial f}{\partial x} \right)^2 dxdy \right]^{\frac{1}{4}} \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{\partial f}{\partial y} \right)^2 dxdy \right]^{\frac{1}{4}} .$$

A similar argument is found in Nezames (1980). As will be demonstrated in later sections, more efficient yet easily implementable mesh types may be designed.

2.5 Semi-fixed-dimension (Semiregular) mesh

If g is a function of x, and h remains constant throughout the domain of support of f, we obtain a partition in which, for example, the cell widths

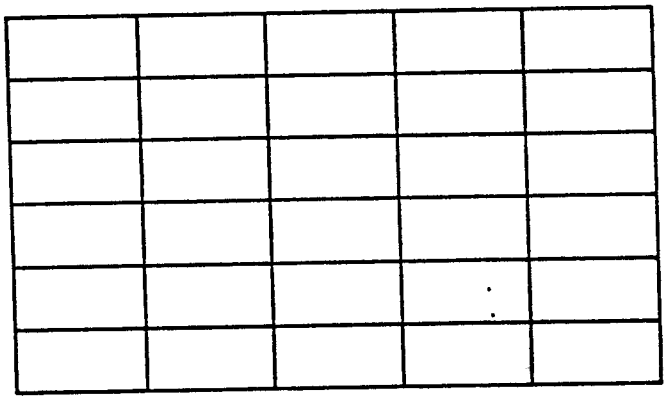

Figure 2.4. Fixed-dimension (Regular) mesh

The histogram estimator produced by this scheme, however, is less efficient relative to that produced by the minimally restricted mesh than others which will be proposed later. When g and h are both constant, equations (2.2.2) become

$$\frac{1}{ng^2h} = \frac{1}{6}g \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left|\frac{\partial f}{\partial x}\right|^2 dxdy$$

and

$$\frac{1}{ngh^2} = \frac{1}{6}h \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left|\frac{\partial f}{\partial y}\right|^2 dxdy$$

with solution

$$g(x) = \frac{6^{\frac{1}{4}}}{n^{\frac{1}{4}}} \frac{\left[\int_{-\infty}^{\infty} f(x,y)\,dy\right]^{\frac{1}{3}} \left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\frac{\partial f}{\partial y}\right)^2 dx dy\right]^{\frac{1}{8}}}{\left[\int_{-\infty}^{\infty} \left(\frac{\partial f}{\partial x}\right)^2 dy\right]^{\frac{1}{3}} \left[\int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(x,y)\,dy\right]^{\frac{2}{3}} \left[\int_{-\infty}^{\infty} \left(\frac{\partial f}{\partial x}\right)^2 dy\right]^{\frac{1}{3}} dx\right]^{\frac{1}{8}}}$$

and

$$h = \frac{6^{\frac{1}{4}}}{n^{\frac{1}{4}}} \left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\frac{\partial f}{\partial y}\right)^2 dx dy\right]^{-\frac{3}{8}} \left[\int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(x,y)\,dy\right]^{\frac{2}{3}} \left[\int_{-\infty}^{\infty} \left(\frac{\partial f}{\partial x}\right)^2 dy\right]^{\frac{1}{3}} dx\right]^{\frac{3}{8}}$$

which yields the following

$$IMSE^* = 2 \cdot 6^{-\frac{1}{2}} n^{-\frac{1}{2}} \left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\frac{\partial f}{\partial y}\right)^2 dx dy\right]^{\frac{1}{4}}$$

$$\cdot \left[\int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(x,y)\,dy\right]^{\frac{2}{3}} \left[\int_{-\infty}^{\infty} \left(\frac{\partial f}{\partial x}\right)^2 dy\right]^{\frac{1}{3}} dx\right]^{\frac{3}{4}}.$$

A similar semi-fixed-dimension mesh may be obtained if g remains constant and h becomes a function of y. The expression for g, $h(y)$, and the *IMSE* are identical to those for the above h, $g(x)$, and *IMSE*, respectively, except that $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$, dx and dy are exchanged throughout.

2.6 Variable-dimension (Grid) mesh I: $g(x)$, $h(y)$

remain constant while the lengths vary in order to accommodate changes in the form of the probability density function in different regions of its domain of support. This scheme produces a histogram estimator which is more efficient than the fixed-dimension mesh.

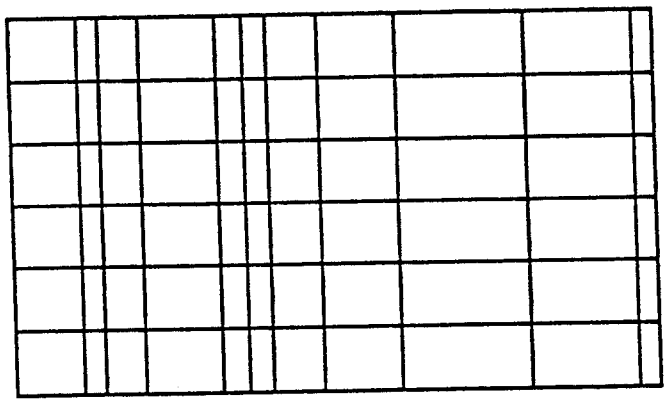

Figure 2.5. Semi-fixed-dimension (Semiregular) mesh

With the above assumptions, equations (2.2.2) become

$$\frac{1}{nh(y)} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{f(x,y)}{g^2(x)}\, dxdy = \frac{1}{6} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x) \left[\frac{\partial f}{\partial x}\right]^2 dxdy$$

and

$$\frac{1}{nh^2(y)} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{f(x,y)}{g(x)}\, dxdy = \frac{1}{6} h(y) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\frac{\partial f}{\partial y}\right]^2 dxdy$$

with solution

$$\frac{1}{nh^2(y)} \int_{-\infty}^{\infty} \frac{f(x,y)}{g(x)}\, dx = \frac{h(y)}{6} \int_{-\infty}^{\infty} \left[\frac{\partial f}{\partial y}\right]^2 dx \ .$$

If $f(x,y)$ can be written as the product of two functions, each of which is a function of only one variable, i.e. , if

$$f(x,y) = r(x)\, s(y)$$

then analytic solutions of the following form may be obtained :

$$g(x) = \frac{6^{\frac{1}{4}} \cdot \left\{ \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(x,y)\, dx \right]^{\frac{2}{3}} \left[\int_{-\infty}^{\infty} \left(\frac{\partial f}{\partial y} \right)^2 dx \right]^{\frac{1}{3}} dy \right\}^{\frac{3}{8}} \left[\int_{-\infty}^{\infty} f(x,y)\, dy \right]^{\frac{1}{3}}}{n^{\frac{1}{4}} \cdot \left\{ \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(x,y)\, dy \right]^{\frac{2}{3}} dy \left[\int_{-\infty}^{\infty} \left(\frac{\partial f}{\partial x} \right)^2 dy \right]^{\frac{1}{3}} dx \right\}^{\frac{1}{8}} \left[\int_{-\infty}^{\infty} \left(\frac{\partial f}{\partial x} \right)^2 dy \right]^{\frac{1}{3}}}$$

$$h(y) = \frac{6^{\frac{1}{4}} \cdot \left\{ \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(x,y)\, dy \right]^{\frac{2}{3}} \left[\int_{-\infty}^{\infty} \left(\frac{\partial f}{\partial x} \right)^2 dy \right]^{\frac{1}{3}} dx \right\}^{\frac{3}{8}} \left[\int_{-\infty}^{\infty} f(x,y)\, dx \right]^{\frac{1}{3}}}{n^{\frac{1}{4}} \cdot \left\{ \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(x,y)\, dx \right]^{\frac{2}{3}} \left[\int_{-\infty}^{\infty} \left(\frac{\partial f}{\partial y} \right)^2 dx \right]^{\frac{1}{3}} dy \right\}^{\frac{1}{8}} \left[\int_{-\infty}^{\infty} \left(\frac{\partial f}{\partial y} \right)^2 dx \right]^{\frac{1}{3}}}$$

yielding an optimal integrated mean squared error

If g is a function of x and h is a function of y , we obtain a mesh which is adaptable in both dimensions to the form of the density function and which is almost as easily implemented as the fixed-dimension mesh.

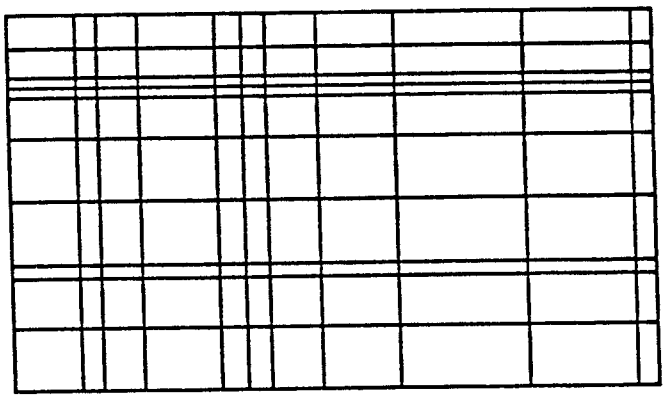

Figure 2.6. Variable-dimension (Grid) mesh I: $g(x),h(y)$

In several respects this type of mesh is optimal since it combines both adaptability and ease of implementation. Marginal histograms as well as histograms along any strip in either direction may easily be obtained. Although each such histogram is not itself optimal, the set of all such histograms so constructed is optimal on the average. Difficulties arise, however, when an attempt is made to solve the equations deriving from (2.2.2) under the above restrictions on g and h :

$$\frac{1}{ng^2(x)} \int_{-\infty}^{\infty} \frac{f(x,y)}{h(y)}\, dy \;=\; \frac{g(x)}{6} \int_{-\infty}^{\infty} \left[\frac{\partial f}{\partial x}\right]^2 dy \qquad (2.6.1)$$

and

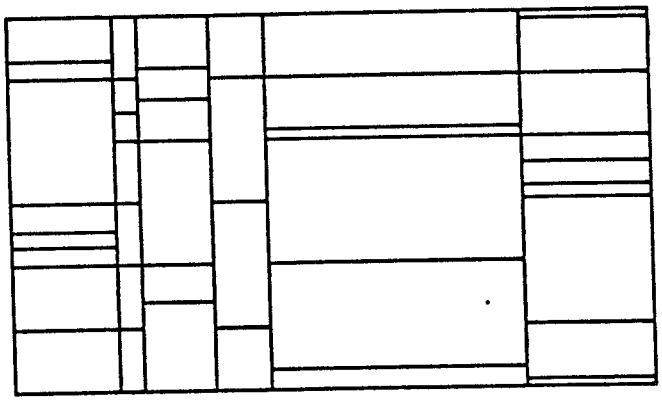

Figure 2.7. Variable-dimension (Semigrid) mesh II: $g(x),h(x,y)$

An analytic solution to the equations deriving from (2.2.2) under the present restrictions is, however, readily obtainable as :

$$g(x) = 6^{\frac{1}{4}} n^{-\frac{1}{4}} \left[\frac{\int_{-\infty}^{\infty} \left[f(x,y) \frac{\partial f}{\partial y} \right]^{\frac{2}{3}} dy}{\int_{-\infty}^{\infty} \left(\frac{\partial f}{\partial x} \right)^2 dy} \right]^{\frac{3}{8}}$$

and

$$h(x,y) = 6^{\frac{1}{4}} n^{-\frac{1}{4}} \left[f(x,y) \right]^{\frac{1}{3}} \cdot \left[\frac{\partial f}{\partial y} \right]^{-\frac{2}{3}} \cdot \left[\frac{\int_{-\infty}^{\infty} \left(\frac{\partial f}{\partial x} \right)^2 dy}{\int_{-\infty}^{\infty} \left[f(x,y) \frac{\partial f}{\partial y} \right]^{\frac{2}{3}} dy} \right]^{\frac{1}{8}}$$

with

$$IMSE^* = 2{\cdot}6^{-\frac{1}{2}}\, n^{-\frac{1}{2}} \left\{ \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(x,y)\, dx \right]^{\frac{2}{3}} \left[\int_{-\infty}^{\infty} \left(\frac{\partial f}{\partial y} \right)^2 dx \right]^{\frac{1}{3}} dy \right\}^{\frac{3}{4}}$$

$$\cdot \left\{ \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(x,y)\, dy \right]^{\frac{2}{3}} \left[\int_{-\infty}^{\infty} \left(\frac{\partial f}{\partial x} \right)^2 dy \right]^{\frac{1}{3}} dx \right\}^{\frac{3}{4}}.$$

In the general case when $f(x,y)$ is not separable, a numerical rather than an analytic solution has been obtained. The algorithm will be treated in section 3 of this paper.

2.7 Variable-dimension (Semigrid) mesh II: $g(x)$, $h(x,y)$

If g is a function of x and h is a function of both x and y, we obtain a mesh which performs better in terms of the integrated mean squared error than the variable-dimensioned mesh described in section 2.6 but at a cost of considerable difficulties in implementation.

3. Implementation

3.1 Introduction

In the previous section, several types of two-dimensional meshes were introduced together with expressions for their optimal cell dimensions and for the expected minimal integrated mean squared error. Theoretically optimal histograms were constructed on the basis of these expressions using a variety of probability density functions. The integrated mean squared error was calculated in each case and the *IMSEs* compared in terms of their efficiencies relative to the least restrictive mesh previously introduced (the mesh in which each cell dimension is a function of both x and y). No optimal histograms were constructed for the latter case since, as noted previously, a completely freely varying mesh is not implementable. Since the theoretical mean squared error also depends upon sample size, a variety of sample sizes (from 50 to 5000) were used and their effect upon the *IMSE*, relative efficiency, and form of the optimal histogram were observed. Finally, simulated data were used to test several of the theoretical constructs.

3.2 Cell dimensions g and h

In cases where the bin dimensions g and h remain constant throughout the domain of support, where g is a function of x and h remains constant, where h is a function of y and g remains constant, where g is a function of

$$IMSE^{*} = 2\cdot 6^{-\frac{1}{2}}\, n^{-\frac{1}{2}} \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \left(\frac{\partial f}{\partial x} \right)^{2} dy \right]^{\frac{1}{4}} \left[\int_{-\infty}^{\infty} \left[f(x,y) \frac{\partial f}{\partial y} \right]^{\frac{2}{3}} dy \right]^{\frac{3}{4}} dx \ .$$

Since for each of the mesh types above we are minimizing the integrated mean squared error under increasingly tight constraints, it may be shown that the minimum *IMSE* becomes larger as more constraints are placed upon the mesh, in particular :

IMSE (free) $\leq$ *IMSE* (semigrid) $\leq$ *IMSE* (grid) $\leq$ *IMSE* (semiregular) $\leq$ *IMSE* (regular).

given point the average number of histogram intervals per unit distance along a particular axis. Solving the first equation for $G(x)$ and the second equation for $H(x)$, the following numerical approach may be constructed :

Let H_o be any positive real number between 10^{-28} and 10^{28}. Then for $n=1,...,N$, where N is a positive integer, let

$$G_n(x) = \left[\frac{\frac{1}{6} \int_{-\infty}^{\infty} \left(\frac{\partial f}{\partial x} \right)^2 dy}{\int_{-\infty}^{\infty} H_n(y)\, f(x,y)\, dy} \right]^{\frac{1}{3}}$$

and

(3.2.1)

$$H_{n+1}(y) = \left[\frac{\frac{1}{6} \int_{-\infty}^{\infty} \left(\frac{\partial f}{\partial y} \right)^2 dx}{\int_{-\infty}^{\infty} G_n(x)\, f(x,y)\, dx} \right]^{\frac{1}{3}}$$

and replace $H_n(y)$ in the first equation by $H_{n+1}(y)$ upon each iteration. When the analytic forms of $g(x)$ and $f(y)$ are known, as in the case of separable $f(x,y)$, the algorithm was observed to produce convergence to the correct value with four place accuracy within five iterations. That convergence should occur may be seen from the following argument.

Let equations (3.2.1) define the sequences $\{G_n\}$ and $\{H_n\}$ and let

x alone and h is a function of both x and y, or where g is a function of both x and y and h is a function of y alone, we have seen that analytic solutions to the equations for optimal binwidth are obtainable. The implementation of optimal histograms on the basis of these is straightforward. If g is a function of x alone and h is a function of y alone, the analytic solution given in section 2.6 has been obtained for the case in which $f(x,y)$ may be written as a product of functions $r(x)$ and $s(y)$. Again, implementation is straightforward. However, in the general case in which $f(x,y)$ is not separable, no analytic solution to the simultaneous integral equations (2.6.1) has been obtained. Rather, a numerical procedure involving a functional iteration was used to determine the optimal $g(x)$ and $h(y)$. Since the solution depends upon both the functional form of the density and the sample size, the iteration must be performed whenever these are modified. Let the equation (2.6.1) be written in the following form:

$$\int_{-\infty}^{\infty} H(y) f(x,y)\, dy = \frac{n}{6} G^3(x) \int_{-\infty}^{\infty} \left(\frac{\partial f}{\partial x} \right)^2 dy$$

$$\int_{-\infty}^{\infty} G(x) f(x,y)\, dx = \frac{n}{6} H^3(y) \int_{-\infty}^{\infty} \left(\frac{\partial f}{\partial y} \right)^2 dx$$

where $G(x)=\frac{1}{g(x)}$ and $H(y)=\frac{1}{h(y)}$ so that G and H may each be interpreted as a " bin density function " , i.e. , a function which determines at a

$$H_{n+1}(y) = \frac{\left[\int_{-\infty}^{\infty} H_n(t)\, s(t)\, dt\right]^{\frac{1}{9}}}{\left[\int_{-\infty}^{\infty} [A(x)]^{\frac{1}{3}}\, [r(x)]^{\frac{2}{3}}\, dx\right]^{\frac{1}{3}}} \left[\frac{B(y)}{s(y)}\right]^{\frac{1}{3}} \tag{3.2.3}$$

Multiplying equation (3.2.2) by $r(x)$ and integrating yields

$$\int_{-\infty}^{\infty} G_{n+1}(x)\, r(x)\, dx = \frac{\int_{-\infty}^{\infty} [A(x)]^{\frac{1}{3}}\, [r(x)]^{\frac{2}{3}}\, dx}{\left[\int_{-\infty}^{\infty} [B(y)]^{\frac{1}{3}}\, [s(y)]^{\frac{2}{3}}\, dy\right]^{\frac{1}{3}}} \cdot \left[\int_{-\infty}^{\infty} G_n(u)\, r(u)\, du\right]^{\frac{1}{9}} . \tag{3.2.4}$$

Similarly, from (3.2.3) :

$$\int_{-\infty}^{\infty} H_{n+1}(y)\, s(y)\, dy = \frac{\int_{-\infty}^{\infty} [B(y)]^{\frac{1}{3}}\, [s(y)]^{\frac{2}{3}}\, dy}{\left[\int_{-\infty}^{\infty} [A(x)]^{\frac{1}{3}}\, [r(x)]^{\frac{2}{3}}\, dx\right]^{\frac{1}{3}}} \cdot \left[\int_{-\infty}^{\infty} H_n(t)\, s(t)\, dt\right]^{\frac{1}{9}} . \tag{3.2.5}$$

If we let

$$A(x) = \frac{1}{6}\int_{-\infty}^{\infty}\left(\frac{\partial f}{\partial x}\right)^2 dy$$

$$B(y) = \frac{1}{6}\int_{-\infty}^{\infty}\left(\frac{\partial f}{\partial y}\right)^2 dx.$$

If these sequences converge, their limits will satisfy equations (2.6.1). If $f(x,y) = r(x)\, s(y)$, we have

$$G_{n+1} = \left[\int_{-\infty}^{\infty} H_{n+1}(y)\, s(y)\, dy\right]^{-\frac{1}{3}}\left[\frac{A(x)}{r(x)}\right]^{\frac{1}{3}}$$

$$= \left[\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} G_n(x)\, r(x)\, dx\right]^{-\frac{1}{3}}\left[\frac{B(y)}{s(y)}\right]^{\frac{1}{3}} s(y)\, dy\right]^{-\frac{1}{3}}\left[\frac{A(x)}{r(x)}\right]^{\frac{1}{3}}$$

$$= \frac{\left[\int_{-\infty}^{\infty} G_n(u)\, r(u)\, du\right]^{\frac{1}{9}}}{\left[\int_{-\infty}^{\infty}[B(y)]^{\frac{1}{3}}\,[s(y)]^{\frac{2}{3}}\, dy\right]^{\frac{1}{3}}}\left[\frac{A(x)}{r(x)}\right]^{\frac{1}{3}}. \qquad (3.2.2)$$

Similarly,

so that

$$|G_{n+1}(x) - G_n(x)| = \left|\left[(w_{n+1})^{-\frac{1}{3}} - (w_n)^{-\frac{1}{3}}\right]\left[\frac{A(x)}{r(x)}\right]^{\frac{1}{3}}\right|$$

and

$$|H_{n+1}(y) - H_n(y)| = \left|\left[(v_{n+1})^{-\frac{1}{3}} - (v_n)^{-\frac{1}{3}}\right]\left[\frac{B(y)}{s(y)}\right]^{\frac{1}{3}}\right| .$$

Thus on any interval in which $r(x)$ and $s(y)$ are bounded away from zero, $\left[\frac{A(x)}{r(x)}\right]^{\frac{1}{3}}$ and $\left[\frac{B(y)}{s(y)}\right]^{\frac{1}{3}}$ are bounded and, if the sequences $\{v_n\}$, $\{w_n\}$ converge,

$$\lim_{n\to\infty} |G_{n+1}(x) - G_n(x)| = 0$$

and

$$\lim_{n\to\infty} |H_{n+1}(y) - H_n(y)| = 0.$$

We show that the sequence v_n converges. By (3.2.4) and (3.2.5),

$$v_{n+1} = K_1 \cdot v_n^{\frac{1}{9}} \tag{3.2.6}$$

and

$$K_1 = \frac{\int_{-\infty}^{\infty} [A(x)]^{\frac{1}{3}} [r(x)]^{\frac{2}{3}} \, dx}{\left[\int_{-\infty}^{\infty} [B(y)]^{\frac{1}{3}} [s(y)]^{\frac{2}{3}} \, dy\right]^{\frac{1}{3}}} \quad ,$$

$$K_2 = \frac{\int_{-\infty}^{\infty} [B(y)]^{\frac{1}{3}} [s(y)]^{\frac{2}{3}} \, dy}{\left[\int_{-\infty}^{\infty} [A(x)]^{\frac{1}{3}} [r(x)]^{\frac{2}{3}} \, dx\right]^{\frac{1}{3}}} \quad ,$$

$$v_n = \int_{-\infty}^{\infty} G_n(x) \, r(x) \, dx \quad ,$$

and

$$w_n = \int_{-\infty}^{\infty} H_n(y) \, s(y) \, dy \quad ,$$

we can write

$$G_{n+1}(x) = (w_{n+1})^{-\frac{1}{3}} \cdot \left[\frac{A(x)}{r(x)}\right]^{\frac{1}{3}}$$

and

$$H_{n+1}(y) = (v_{n+1})^{-\frac{1}{3}} \cdot \left[\frac{B(y)}{S(y)}\right]^{\frac{1}{3}}$$

$$v = \lim_{n\to\infty} v_{n+1} = \lim_{n\to\infty} K_1 \cdot v_n^{\frac{1}{9}} = K_1 \cdot v^{\frac{1}{9}}$$

so that $v^{\frac{8}{9}} = K_1$ and $v = K_1^{\frac{9}{8}}$. Similarly, the sequence (3.2.7) converges to $K_2^{\frac{9}{8}}$.

In cases where $f(x,y)$ is not separable, the proof of convergence depends upon the solution of difficult cubic equations and is not yet available.

3.3 Determination of cell boundaries

Whether determined analytically or numerically, once the optimal cell dimensions v and w have been obtained, a mesh must be constructed in the plane. The direct use of optimal cell dimensions for the construction of this mesh may lead to serious complications: after placing the first cell boundary in one of the dimensions, the center of the next adjacent cell must be determined by solving a difficult equation which may not even have a unique solution. An alternative method suggested in 1983 in the Terrell-Scott investigation of optimal interval width for one-dimensional histograms is to integrate the bin density function $G(x)$ or $H(y)$ with respect to x or y respectively, to place a boundary at the point along the particular axis at which the corresponding bin density function attains the value one, and to repeat the process beginning each integration at the most recently determined boundary. In view of our asymptotic arguments it is apparent that

$$w_{n+1} = K_2 \cdot w_n^{\frac{1}{9}} \quad . \tag{3.2.7}$$

If $v \geq K_1^{\frac{9}{8}}$, then

$$v^{\frac{8}{9}} > K_1$$

$$v \cdot v^{-\frac{1}{9}} > K_1$$

$$v > K_1 \cdot v^{\frac{1}{9}} > K_1^{\frac{9}{8}} .$$

If $0 < v \leq K_1^{\frac{9}{8}}$, then

$$v^{\frac{8}{9}} \leq K_1$$

$$v \cdot v^{-\frac{1}{9}} \leq K_1$$

$$v \leq K_1 \cdot v^{\frac{1}{9}} \leq K_1^{\frac{9}{8}}$$

$$v \leq K_1 \cdot v^{\frac{1}{9}} \leq K_1^{\frac{9}{8}} .$$

Thus the sequence defined by (3.2.6) is monotonic and bounded and, therefore, convergent. If the limit is denoted by v, we have

where the parameters μ_1, μ_2, μ_3, μ_4, σ_1, σ_2, σ_3, σ_4, ρ_1, ρ_2, and p were varied in order to test our methods on wide variety of distributions.

Based upon a discussion by H.A. David (1981), a special routine was designed for the generation of random vectors from the Dirichlet distribution:

$$f(x,y) = \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\,\Gamma(\beta)\,\Gamma(\gamma)}\, x^{\alpha-1}\, y^{\beta-1}\,(1-x-y)^{\gamma-1} \qquad (3.4.2)$$

where $0 \leq x \leq 1$, $0 \leq y \leq 1$, $\alpha, \beta, \gamma > 0$, and $x + y = 1$.

If n uniform (0,1) random variables are ordered so that $0 \leq u_{(1)} \leq u_{(2)} \leq \cdots \leq u_{(n)} \leq 1$, the joint density of $u_{(i)}$ and $u_{(j)}$, $i \leq j$, is given by

$$\frac{n!}{(i-1)!\,(j-i-1)!\,(n-j)!}\, u_{(i)}^{\,i-1} \left[u_{(j)} - u_{(i)}\right]^{j-i-1} \left[1 - u_{(j)}\right]^{n-j}.$$

If $x = u_{(i)}$, $y = u_{(j)} - u_{(i)}$, the joint density of x and y is given by

$$\frac{n!}{(i-1)!\,(j-i-1)!\,(n-j)!}\, x^{i-1}\, y^{j-i-1}\,(1-x-y)^{n-j}$$

where $x + y = 1$ and $x, y \geq 0$ which is a Dirichlet distribution with parameters $\alpha = i$, $\beta = j - i$, and $\gamma = n - j + 1$. Therefore, Dirichlet(α, β, γ) random vectors may be obtained by generating a random sample of $n = \alpha + \beta + \gamma - 1$ uniform (0,1) random numbers, ordering them as described above, and setting $x = u_{(\alpha)}$, $y = u_{(\alpha+\beta)} - u_{(\alpha)}$.

the initial point for this procedure is not important as sample size increases.

3.4 Heights above cells

Having constructed an optimal mesh, optimal histograms may be produced by generating n random vectors (x,y) from the same bivariate distribution which was used to construct the mesh by counting the number of points ν which fall in each cell, and by forming the estimate $\hat{f}$, the height of the histogram block above the cell, where $\hat{f}$ is set equal to ν divided by the product of n with the area of the cell.

A library routine was used together with other code for the generation of random vectors from a bivariate normal density:

$$f(x,y) = \frac{p}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_1^2}}$$

$$\cdot \exp\left(-\frac{1}{2(1-\rho_1^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho_1\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right)$$

$$+ \frac{(1-p)}{2\pi\sigma_3\sigma_4\sqrt{1-\rho_2^2}}$$

$$\cdot \exp\left(-\frac{1}{2(1-\rho_2^2)}\left[\left(\frac{x-\mu_3}{\sigma_3}\right)^2 - 2\rho_2\left(\frac{x-\mu_3}{\sigma_3}\right)\left(\frac{y-\mu_4}{\sigma_4}\right) + \left(\frac{y-\mu_4}{\sigma_4}\right)^2\right]\right) \tag{3.4.1}$$

These procedures were implemented, and the resulting *IMSEs*, efficiencies, and graphs are presented in Section 4.

3.5 Simulations

Simulations were performed in which an optimal mesh was obtained for a given probability density function f and a given sample of size n, followed by the generation of a large number of samples from f. The empirical *IMSE*:

$$IMSE_a = \sum_{j=1}^{m} \sum_{i=1}^{n} \int_{y_j}^{y_j+h_j} \int_{x_i}^{x_i+g_i} \left(\hat{f}(x_i,y_i) - f(x,y) \right)^2 \ ,$$

where n is the number of interval boundaries in the x-direction and m is the number of interval boundaries in the y-direction, was calculated for each sample and the average empirical *IMSE* obtained. A comparison of these with the theoretical *IMSEs* is presented in the next chapter.

4. Results

4.1 Hardware and Software

The algorithms described in the previous chapter were coded in FORTRAN 77 and implemented on a VAX 780 and subsequently on a VAX 750 and a UNIX pyramid. IMSL subroutines were used for sorting and random vector generation. Histogram graphics were produced on a MacIntosh Apple personal computer and an Imagen laser printer using Pascal programs written by Hausi Müller. Probability density graphics were produced by the Precision Visuals DI-3000 software package and Surface Display Library on a VAX 785 and a Versatec plotter.

4.2 Integrated mean squared errors

Circular, elliptical and bimodal bivariate normal distributions and Dirichlet distributions were used to test the concepts presented in the previous chapters. Arrays containing the bin dimensions g and h were calculated for each case, the resulting bin density function arrays were used to set bin boundaries in the plane, random vectors sampled from the appropriate distributions were distributed among the bins, and optimal histograms were constructed. The theoretical integrated mean squared error was calculated for each grid type for each distribution and the *IMSEs* compared on the basis of

their efficiencies relative to the free grid for each particular distribution. Simulations to test the effects upon real data were also run.

Figures 4.2.1, 4.2.2, 4.2.3, and 4.2.4 show the effect of increasing sample size upon the form of the bivariate histogram. The same bivariate density was used in each case. The effects of increasing sample size upon the integrated mean squared error may be found in Table 4.2.1.

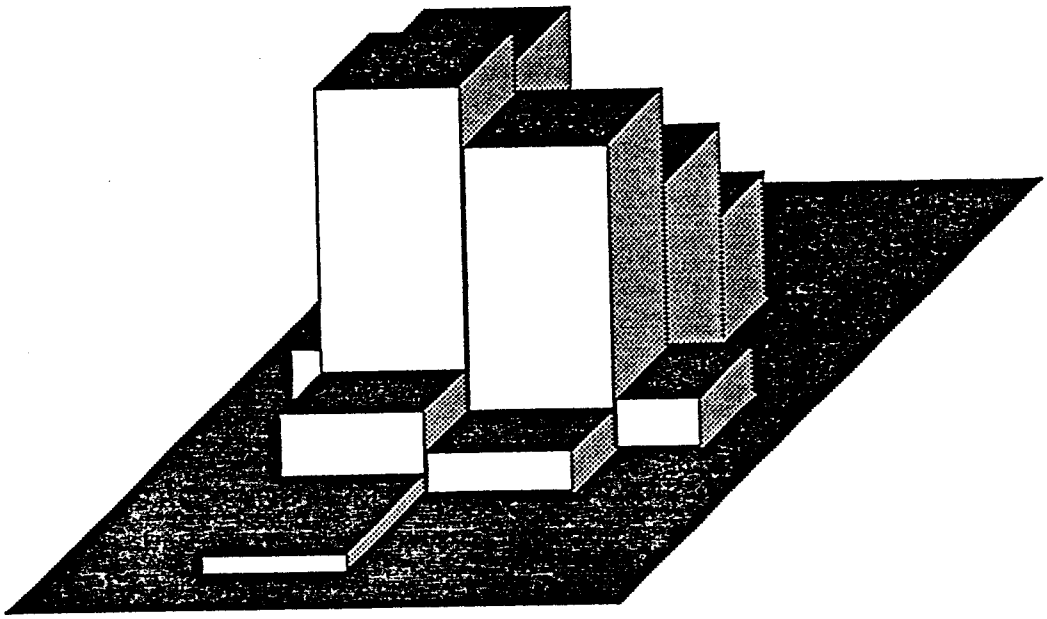

Figure 4.2.1. Grid histogram based upon a sample size of 50 from the bivariate normal density 3.4.1 with p = .5, $\rho_1 = \rho_2 = 0$, $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = 1$, $\mu_1 = -1$, $\mu_2 = 0$, $\mu_3 = 1$, $\mu_4 = 0$.

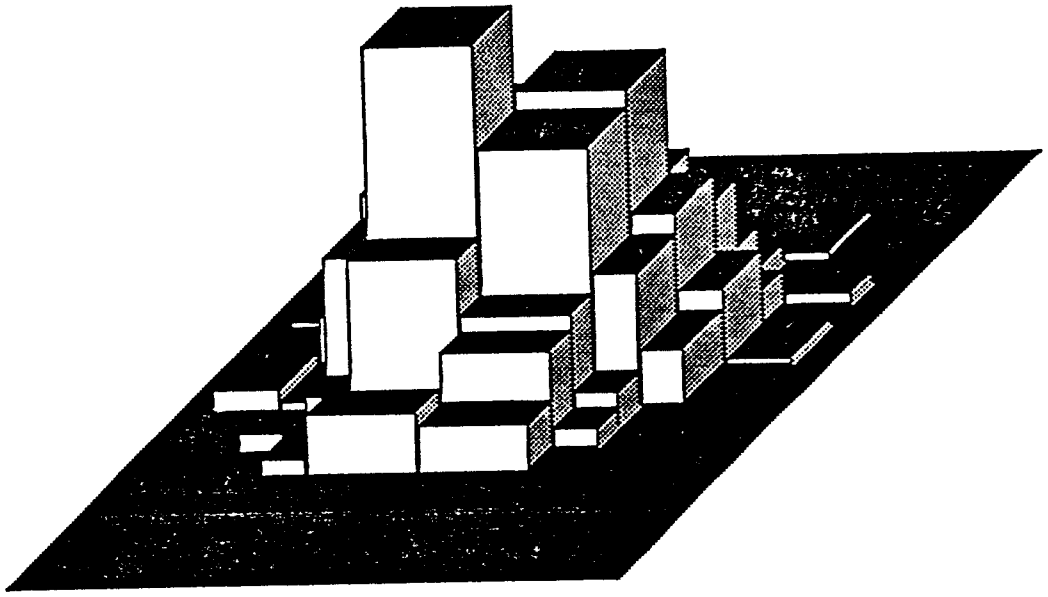

Figure 4.2.2 Grid histogram based upon a sample size of 500 (parameters same as in Figure 4.2.1).

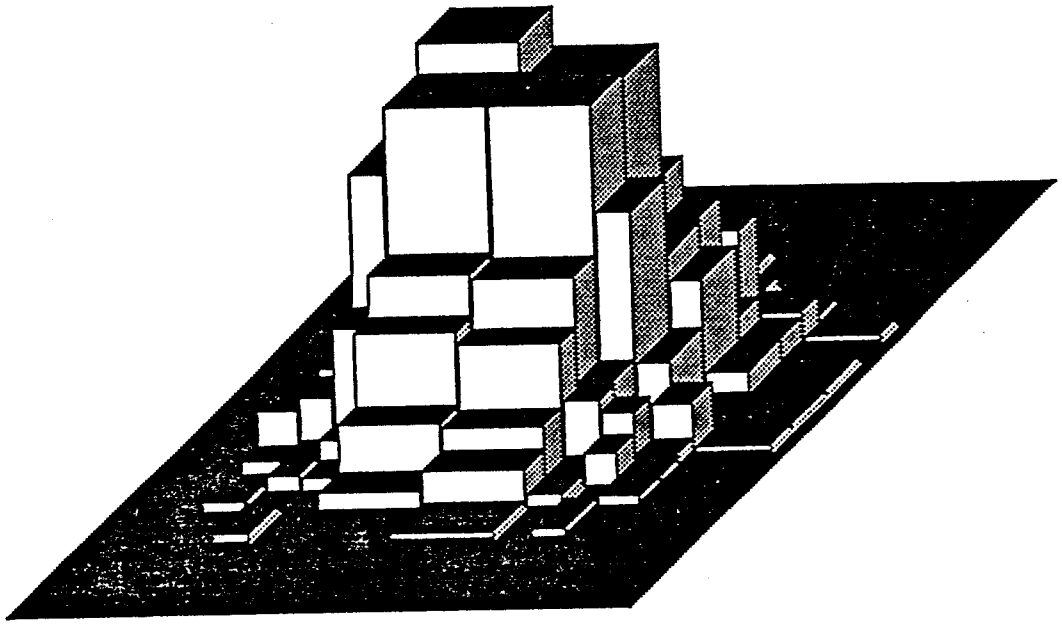

Figure 4.2.3. Grid histogram based upon a sample size of 1000 (parameters same as in Figure 4.2.1).

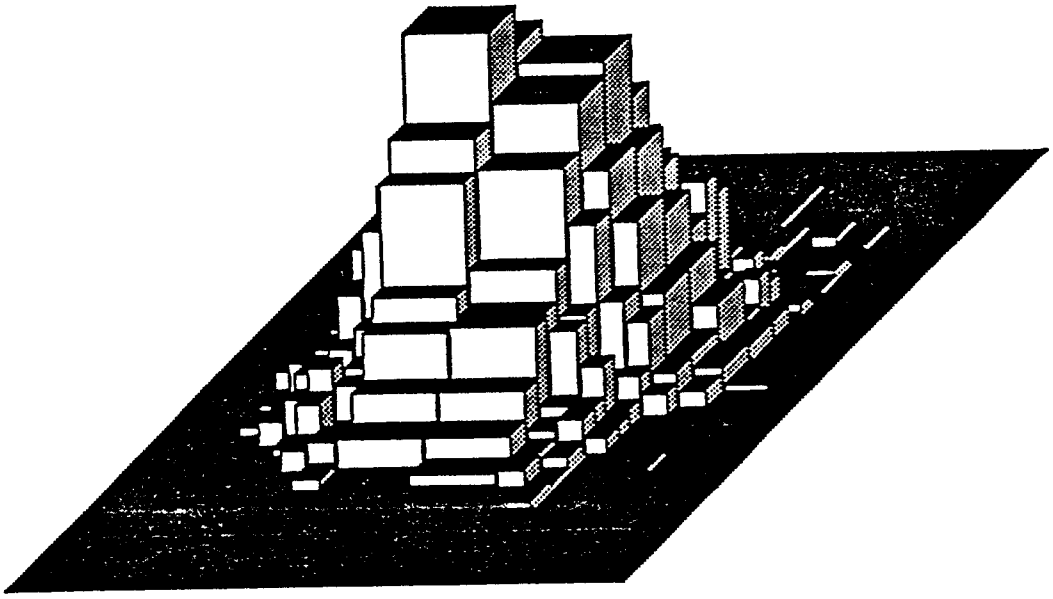

Figure 4.2.4. Grid histogram based upon a sample size of 5000 (parameters same as in Figure 4.2.1).

Normal $\sigma_1=\sigma_2=\sigma_3=\sigma_4=1 \quad \rho=.5 \quad \mu_1=-1 \quad \mu_2=0 \quad \mu_3=1 \quad \mu_4=0$			
Sample Size	Theoretical Grid IMSE	Theoretical Free IMSE	Efficiency of Grid Mesh
50	11.75	10.12	.8613
100	8.310	7.156	.8611
225	5.354	4.771	.8911
500	3.592	3.200	.8909
1000	2.540	2.263	.8909
1500	2.074	1.848	.8910
2000	1.796	1.600	.8908
5000	1.136	1.012	.8908

Table 4.2.1.

In every case the free grid performed best but, as noted previously, is not implementable and is of theoretical interest only because it provides a lower bound on the integrated mean squared error achievable by these methods. Also in every case, the errors for the regular mesh were largest, those for the semiregular mesh smaller, those for the grid still smaller, and those for the semigrid smallest for the implementable grid types. The following figures illustrate the four types of histogram for the case of the bivariate circular normal density.

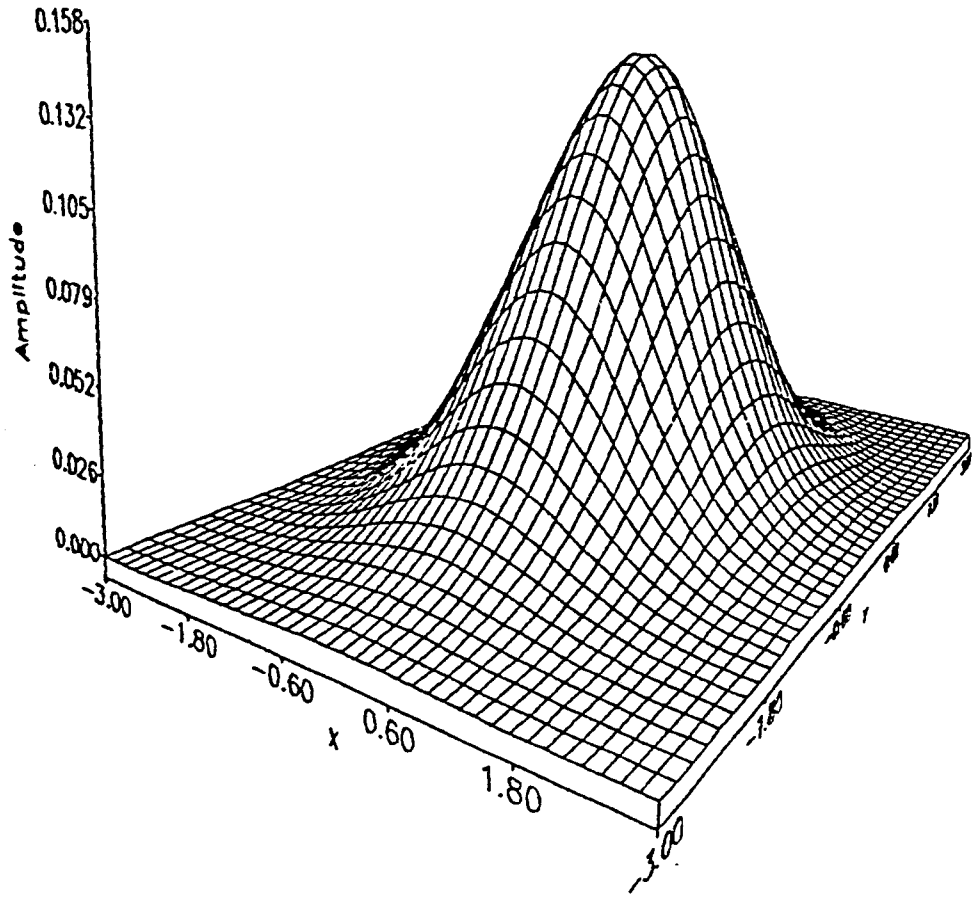

Figure 4.2.5. Bivariate normal density 3.4.1 with p = 1, $\rho_1 = \rho_2 = 0$, $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = 1$, $\mu_1 = \mu_2 = \mu_3 = \mu_4 = 0$.

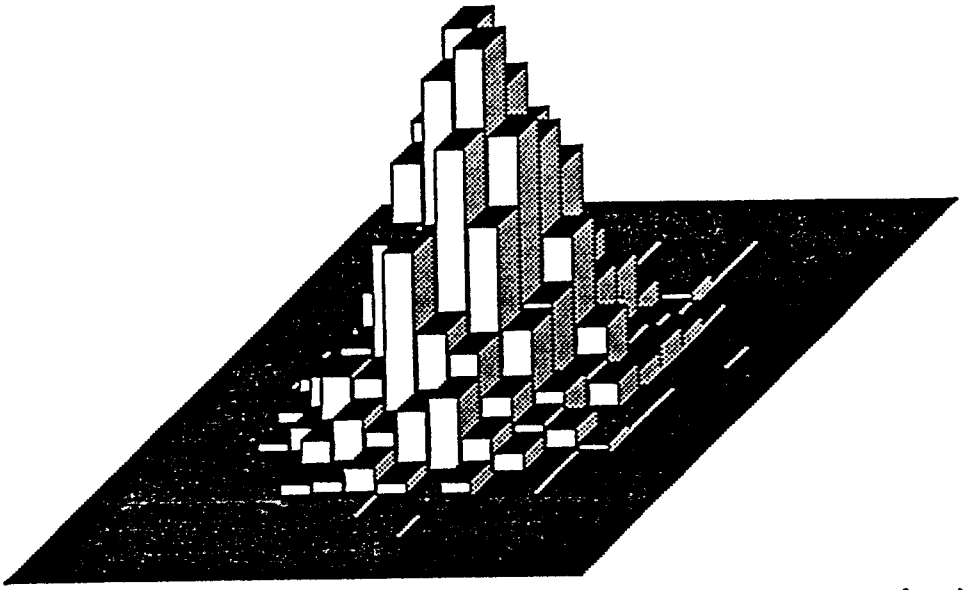

Figure 4.2.6. Regular histogram (density as in Figure 4.2.5, sample size = 2000).

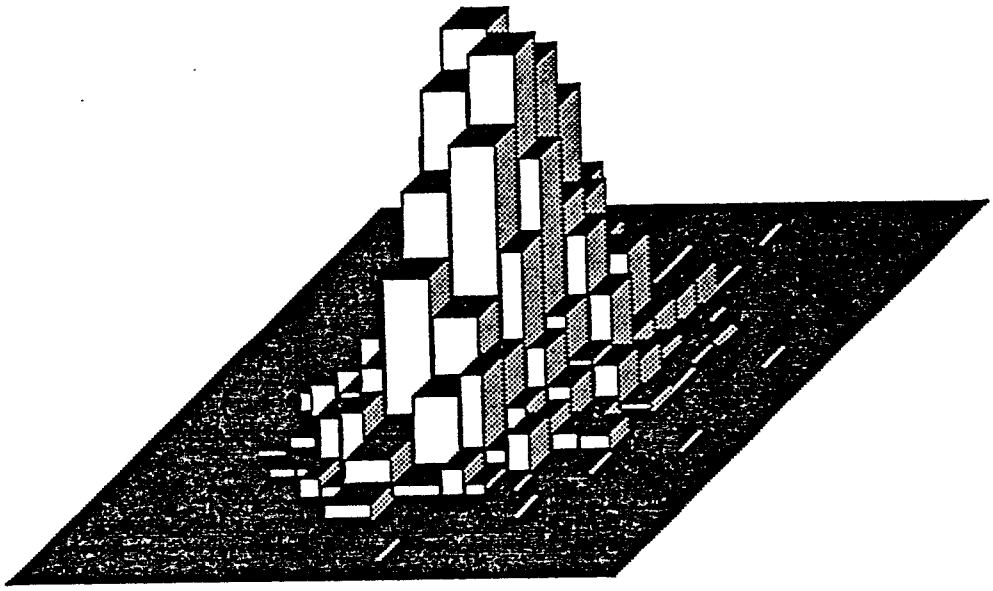

Figure 4.2.7. Semiregular histogram (density as in Figure 4.2.5, sample size = 2000).

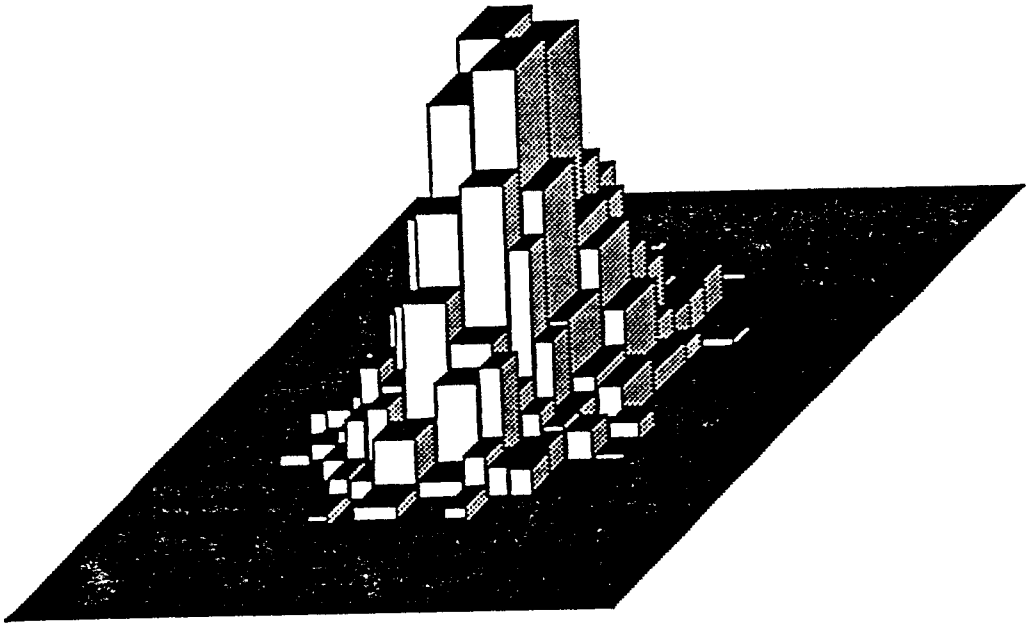

Figure 4.2.8. Grid histogram (density as in Figure 4.2.5, sample size = 2000).

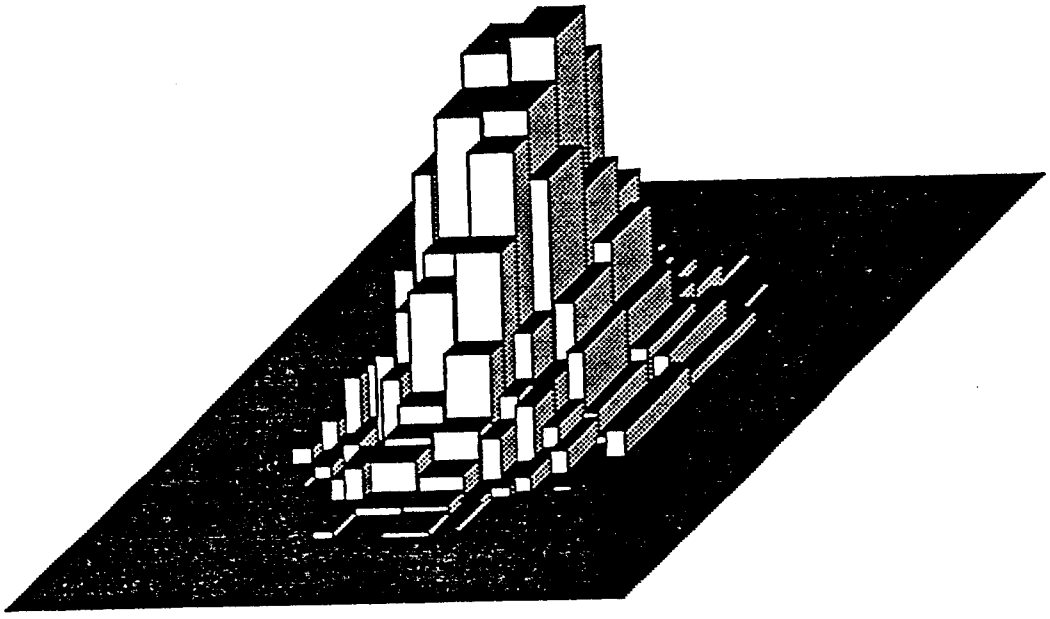

Figure 4.2.9. Semigrid histogram (density as in Figure 4.2.5, sample size = 2000).

Four sets of bimodal normal distributions were studied, each set having a different distance between modes. Within each set, the modes were first aligned along the x-axis, then along the y-axis, and finally along the line $x = y$. This was done in order to examine the difference in performance between the various grid types when presented with the same data under several rotations. Figures 4.2.10 through 4.2.24 illustrate the case of a mixed bivariate unimodal normal density, and figures 4.2.25 through 4.2.39 illustrate the case of a mixed bivariate bimodal normal density.

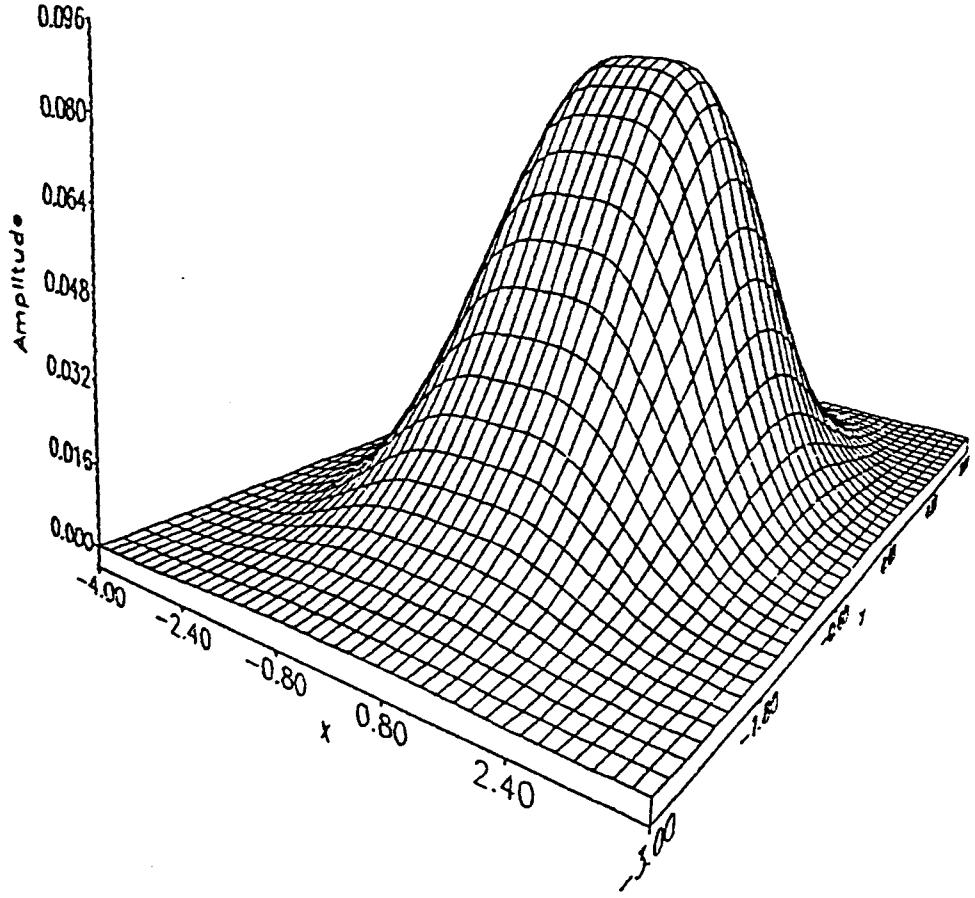

Figure 4.2.10. Bivariate normal density 3.4.1 with p = .5, $\rho_1 = \rho_2 = 0$, $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = 1$, $\mu_1 = -1$, $\mu_2 = 0$, $\mu_3 = 1$, $\mu_4 = 0$.

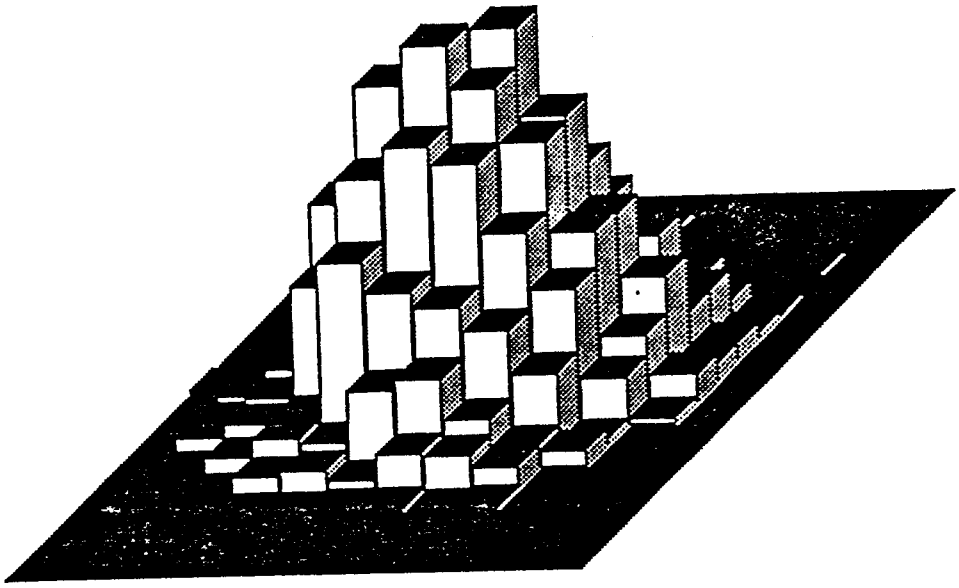

Figure 4.2.11. Regular histogram (density as in Figure 4.2.10, sample size = 2000).

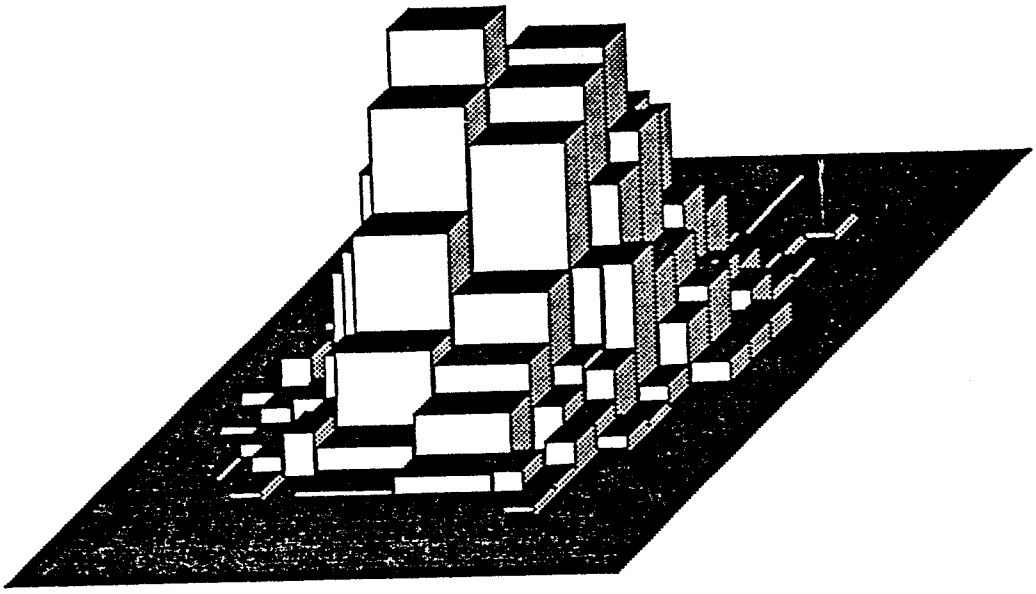

Figure 4.2.12. Semiregular histogram (density as in Figure 4.2.10, sample size = 2000).

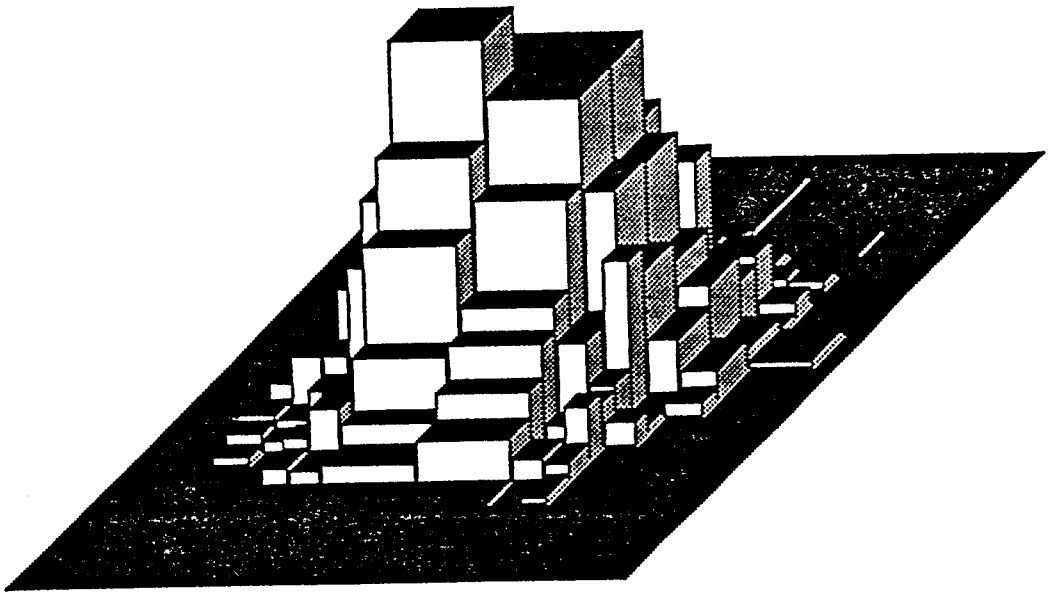

Figure 4.2.13. Grid histogram (density as in Figure 4.2.10, sample size = 2000).

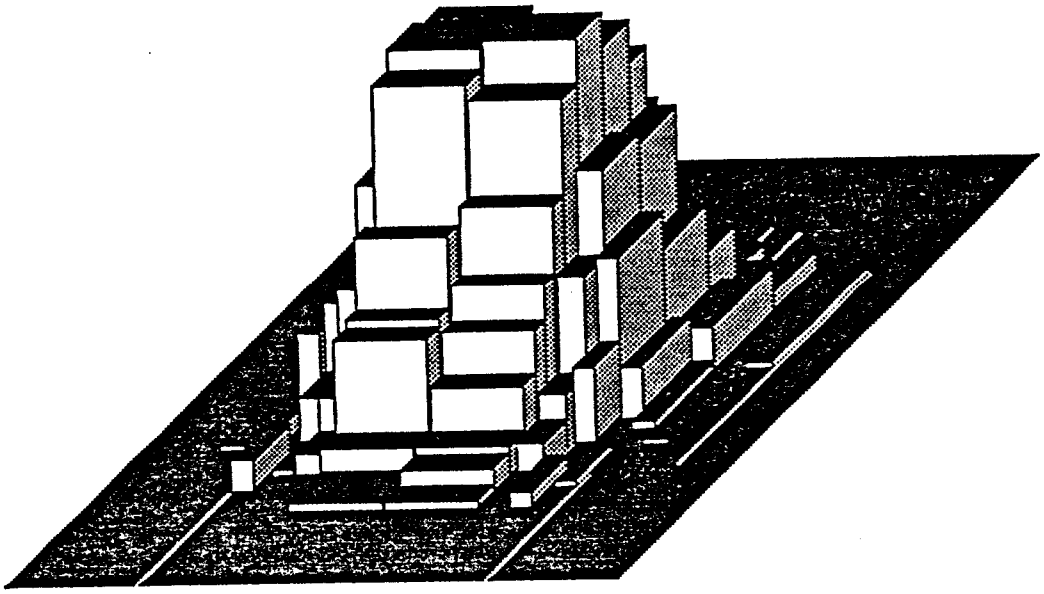

Figure 4.2.14. Semigrid histogram (density as in Figure 4.2.10, sample size = 2000).

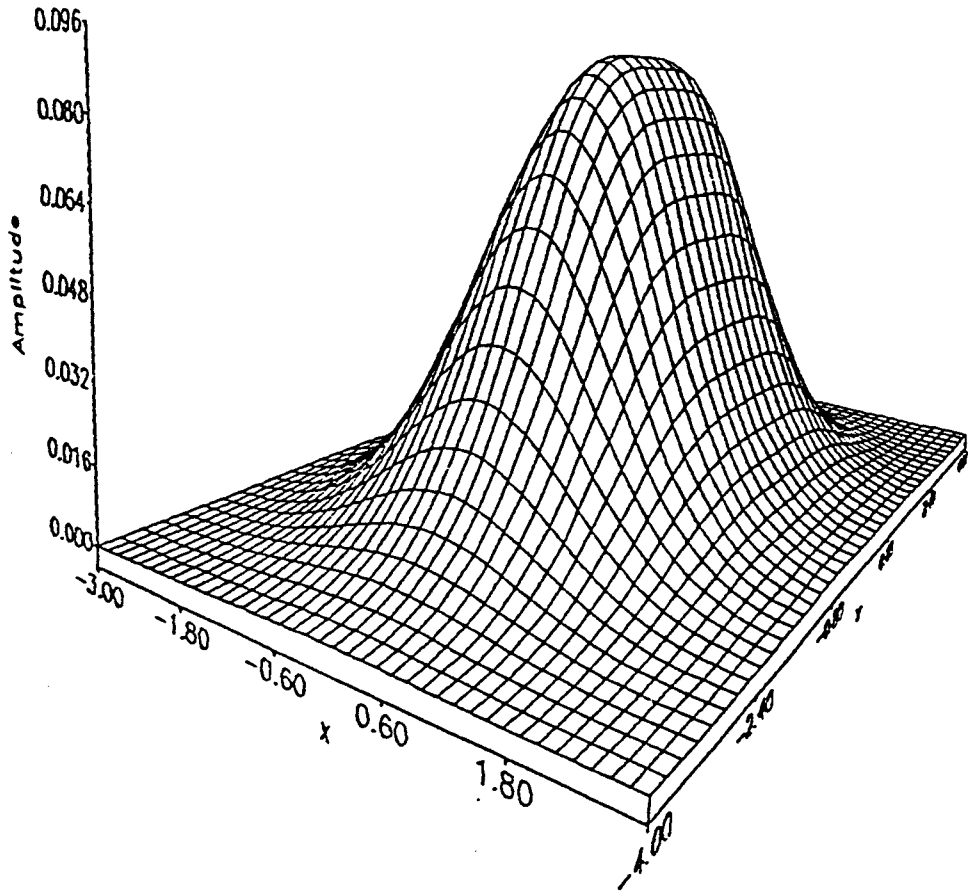

Figure 4.2.15. Bivariate normal density 3.4.1 with p = .5, $\rho_1 = \rho_2 = 0$, $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = 1$, $\mu_1 = 0$, $\mu_2 = -1$, $\mu_3 = 0$, $\mu_4 = 1$.

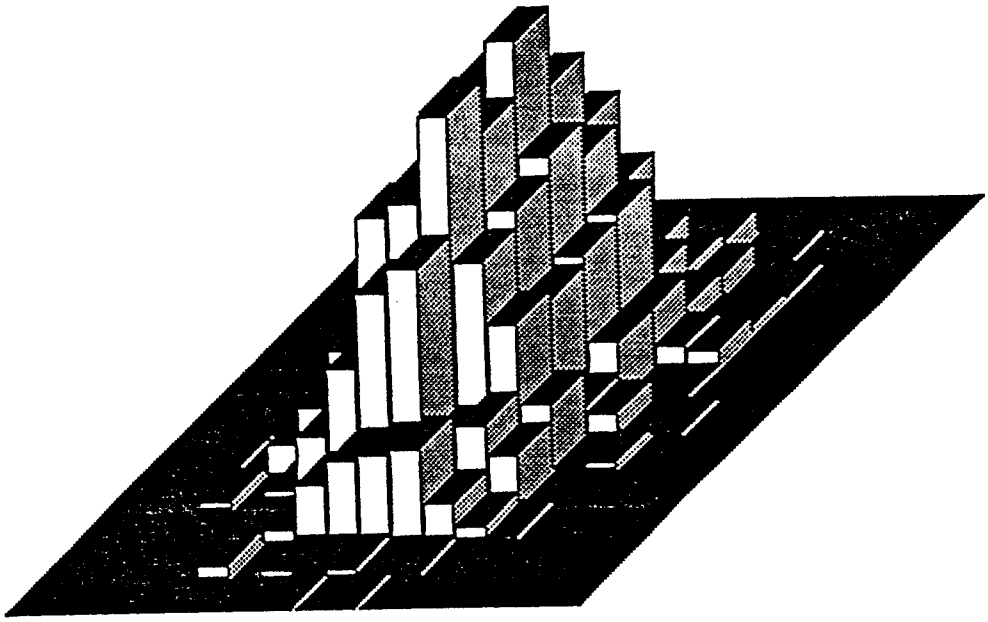

Figure 4.2.16. Regular histogram (density as in Figure 4.2.15, sample size = 2000).

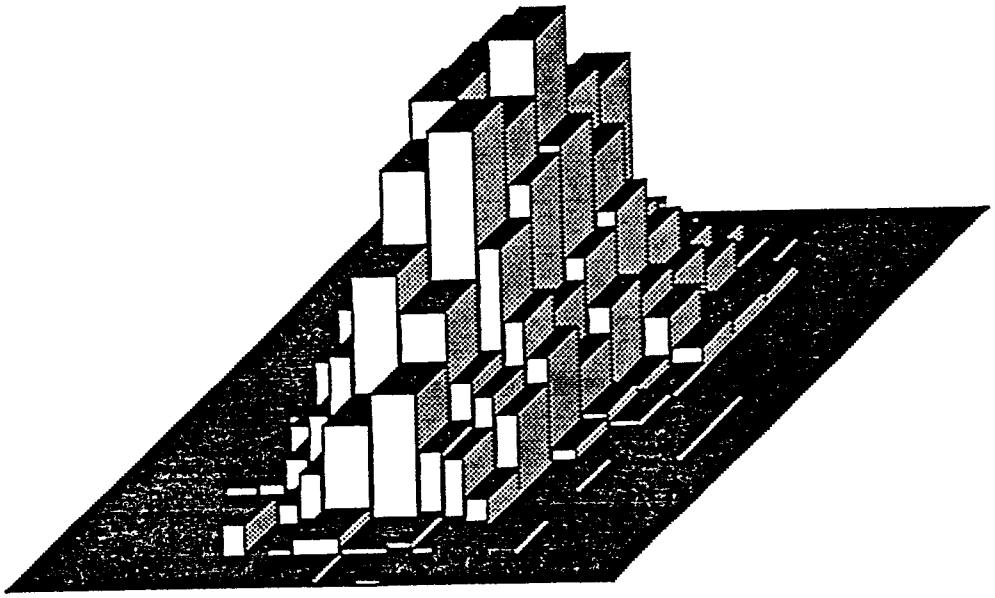

Figure 4.2.17. Semiregular histogram (density as in Figure 4.2.15, sample size = 2000).

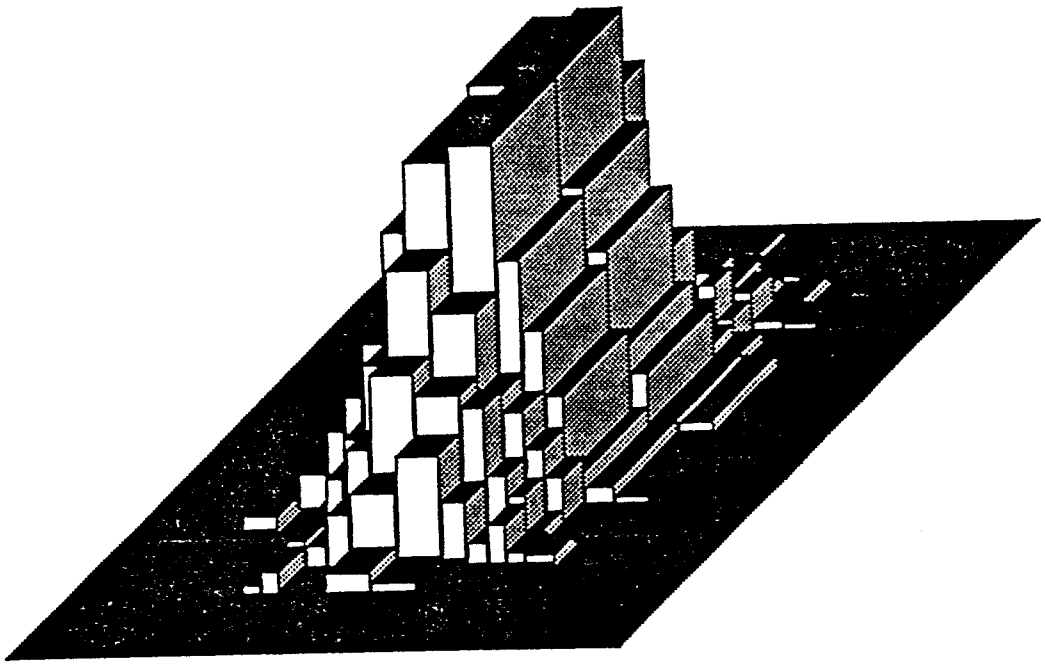

Figure 4.2.18. Grid histogram (density as in Figure 4.2.15, sample size = 2000).

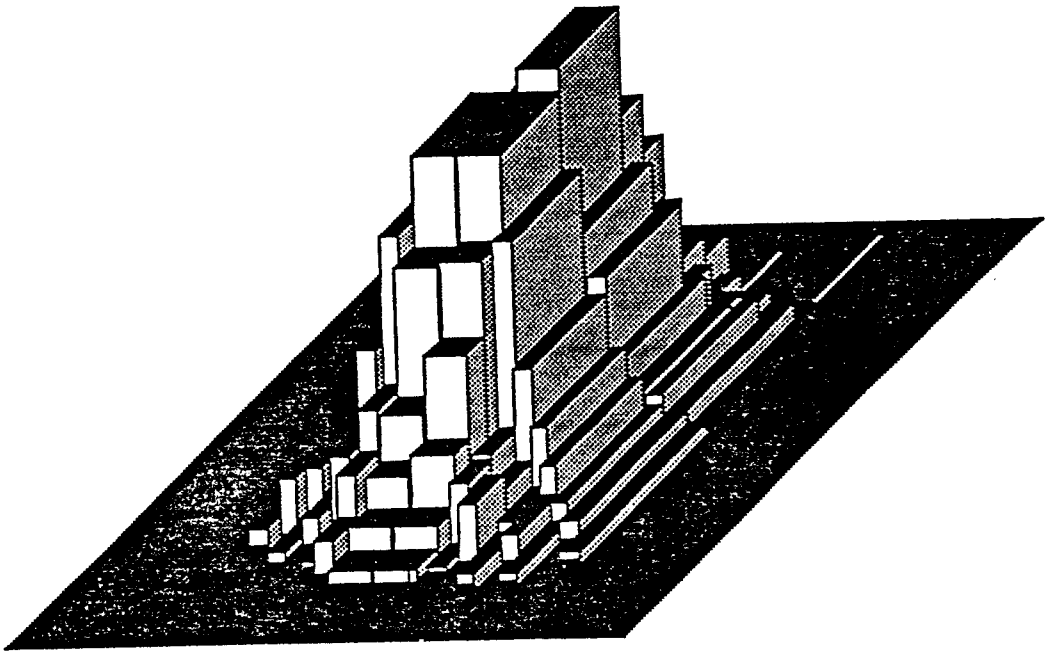

Figure 4.2.19. Semigrid histogram (density as in Figure 4.2.15, sample size = 2000).

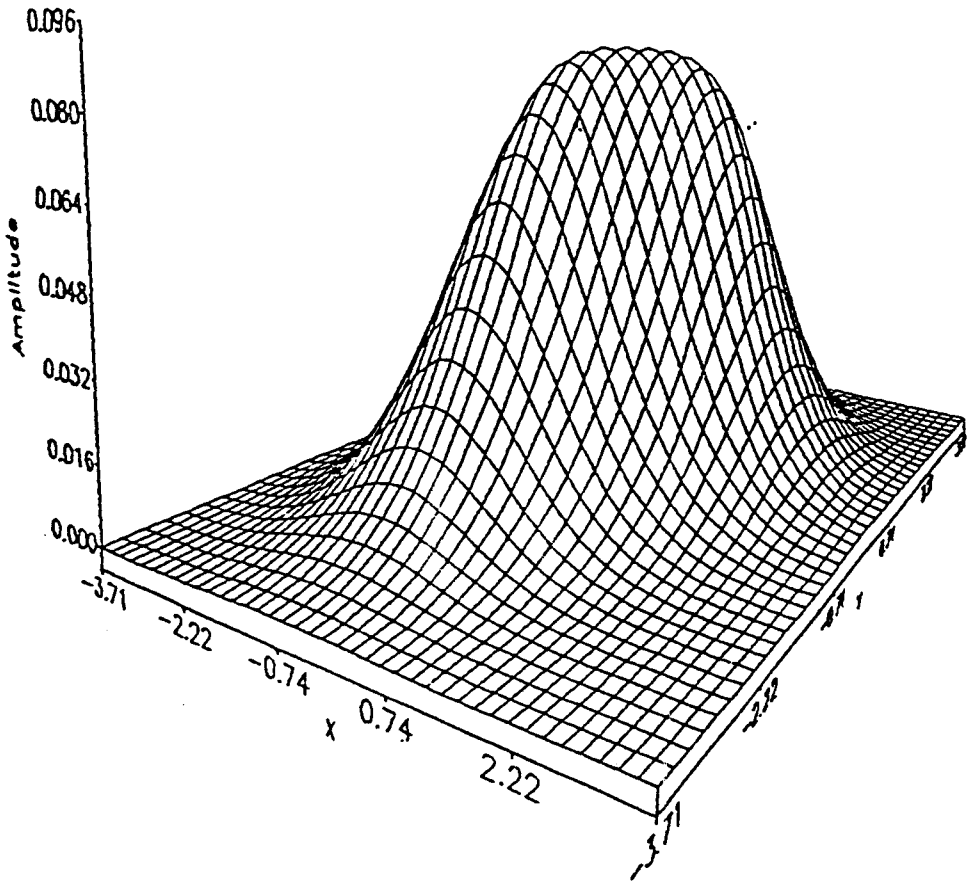

Figure 4.2.20. Bivariate normal density 3.4.1 with p = .5, $\rho_1 = \rho_2 = 0$, $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = 1$, $\mu_1 = \mu_2 = \mu_3 = \mu_4 = .7071$

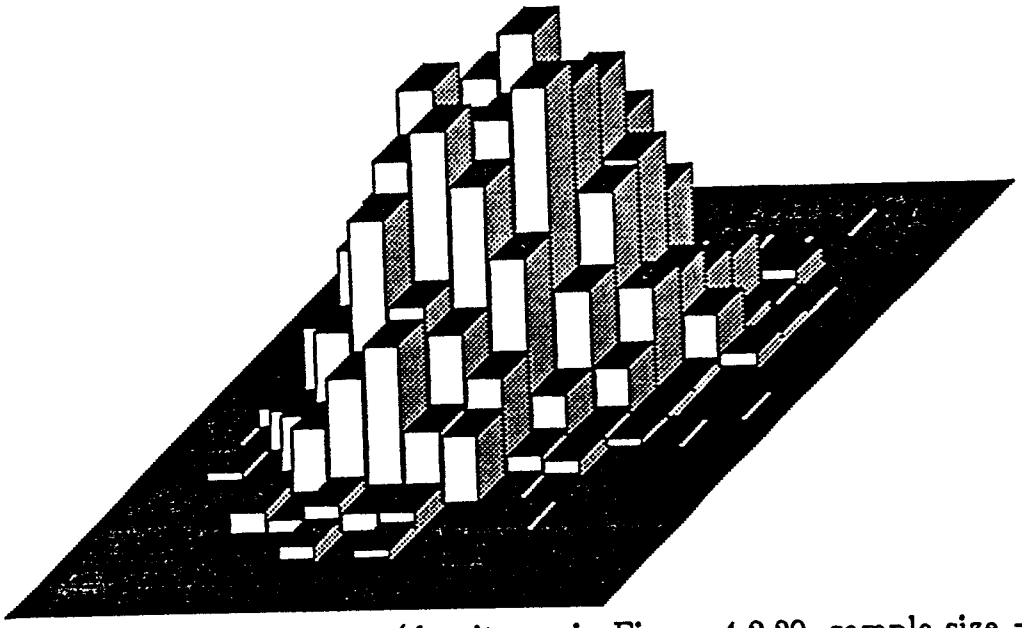

Figure 4.2.21. Regular histogram (density as in Figure 4.2.20, sample size = 2000).

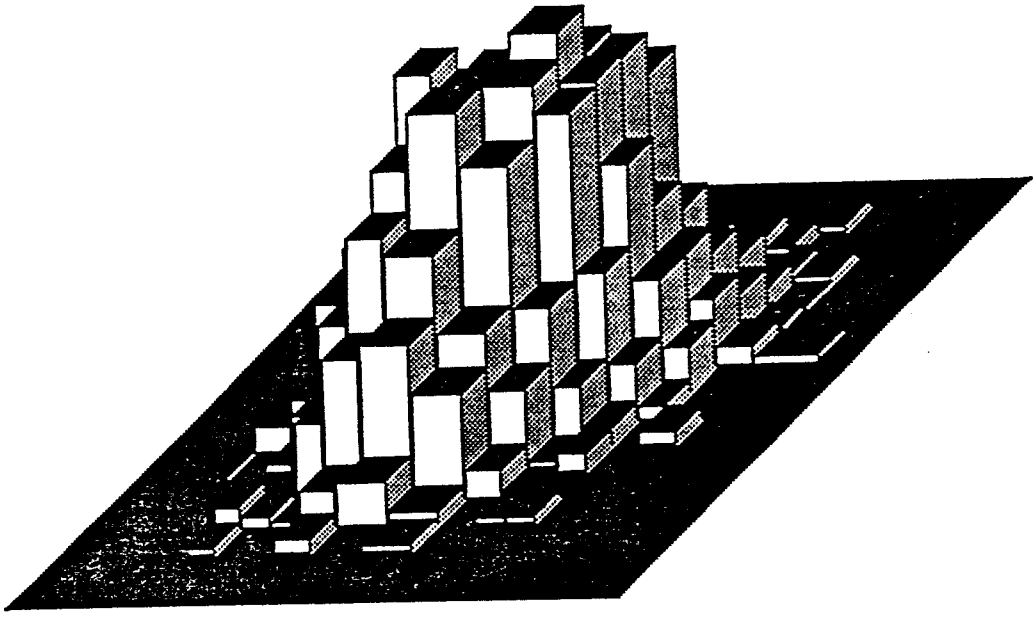

Figure 4.2.22. Semiregular histogram (density as in Figure 4.2.20, sample size = 2000).

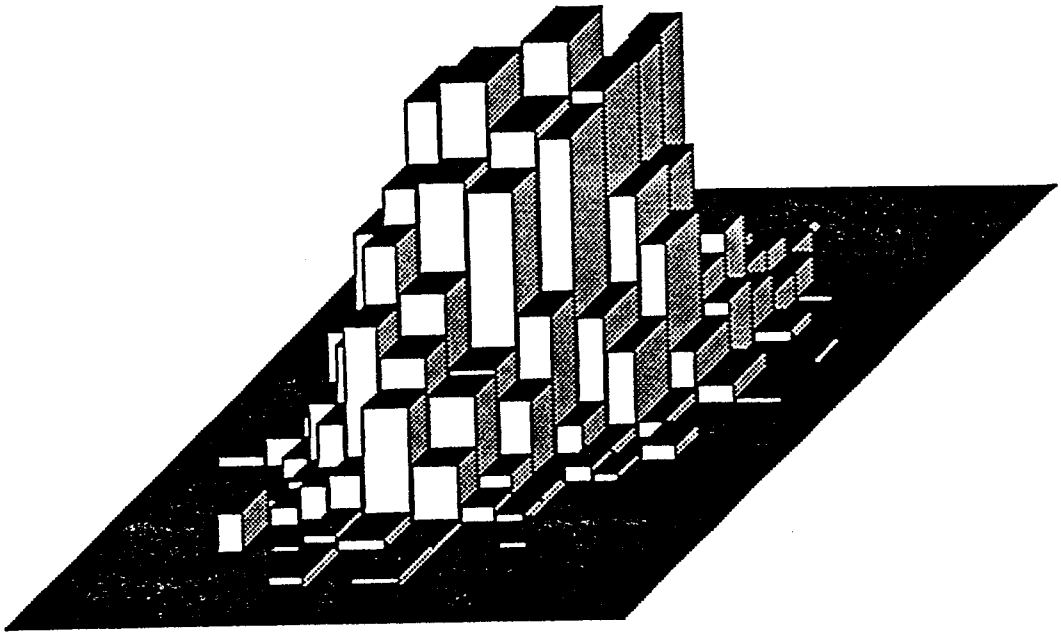

Figure 4.2.23. Grid histogram (density as in Figure 4.2.20, sample size = 2000).

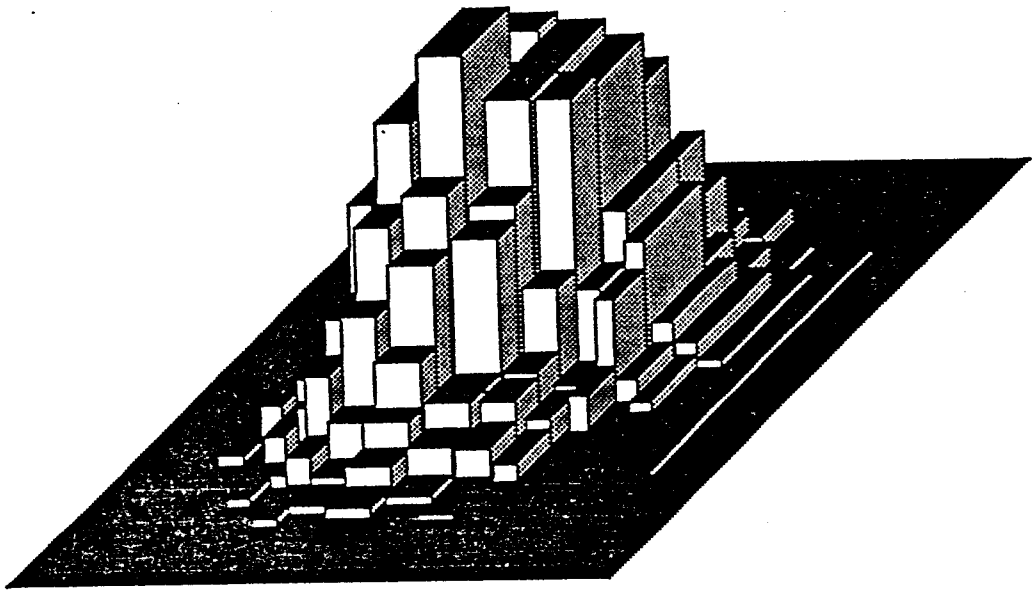

Figure 4.2.24. Semigrid histogram (density as in Figure 4.2.20, sample size = 2000).

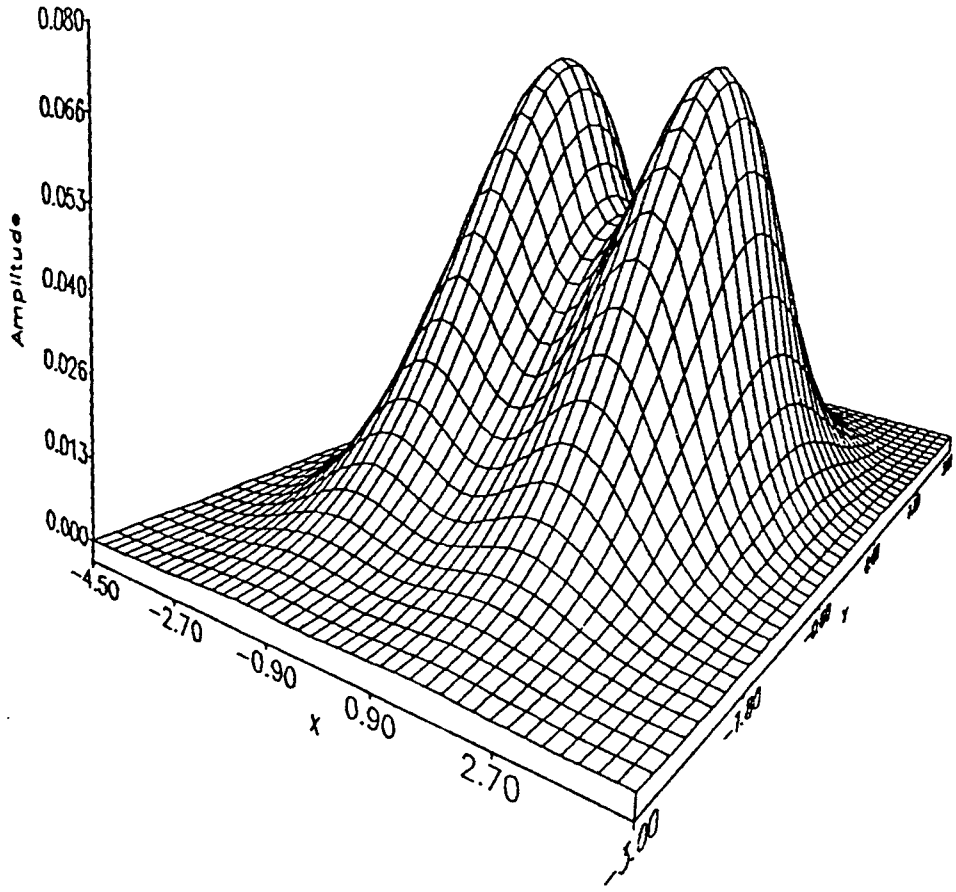

Figure 4.2.25. Bivariate normal density 3.4.1 with p = .5, $\rho_1 = \rho_2 = 0$, $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = 1$, $\mu_1 = -1.5$, $\mu_2 = 0$, $\mu_3 = 1.5$, $\mu_4 = 0$.

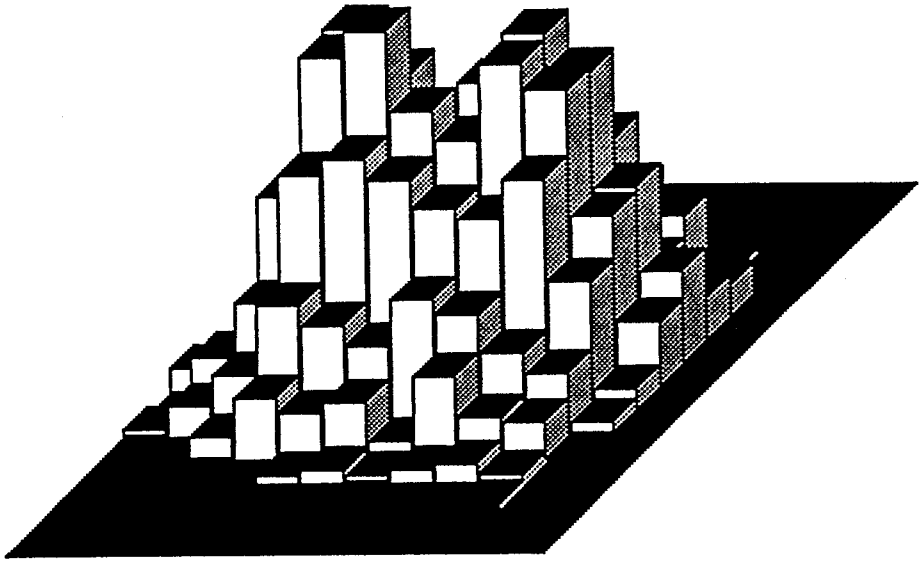

Figure 4.2.26. Regular histogram (density as in Figure 4.2.25, sample size = 2000).

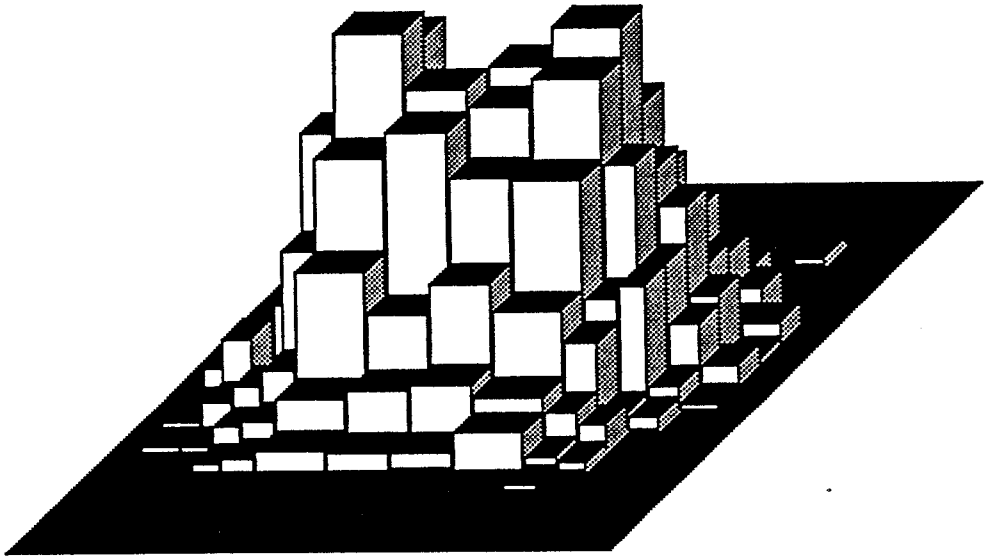

Figure 4.2.27. Semiregular histogram (density as in Figure 4.2.25, sample size = 2000).

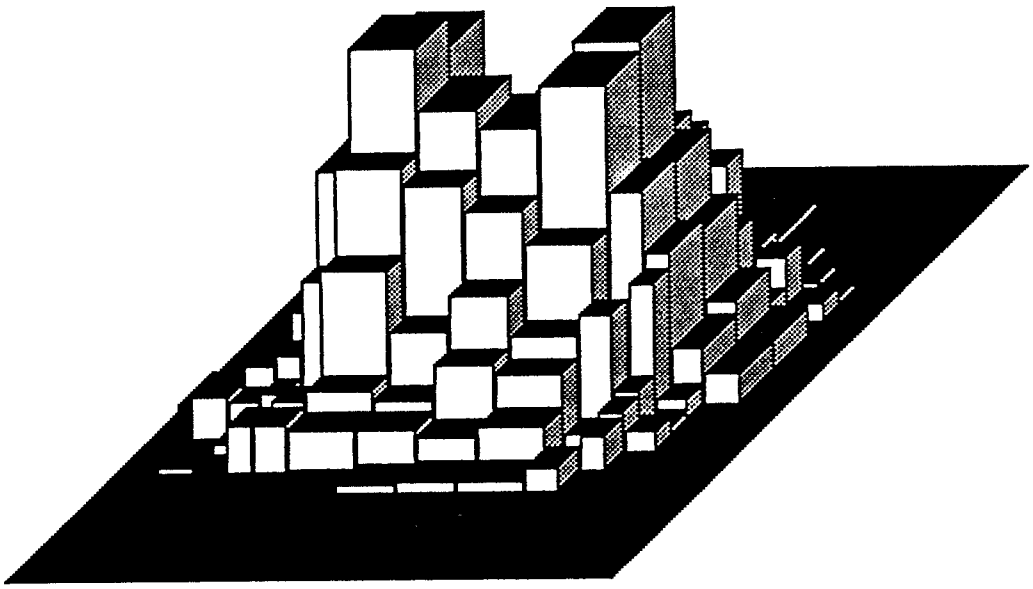

Figure 4.2.28. Grid histogram (density as in Figure 4.2.25, sample size = 2000).

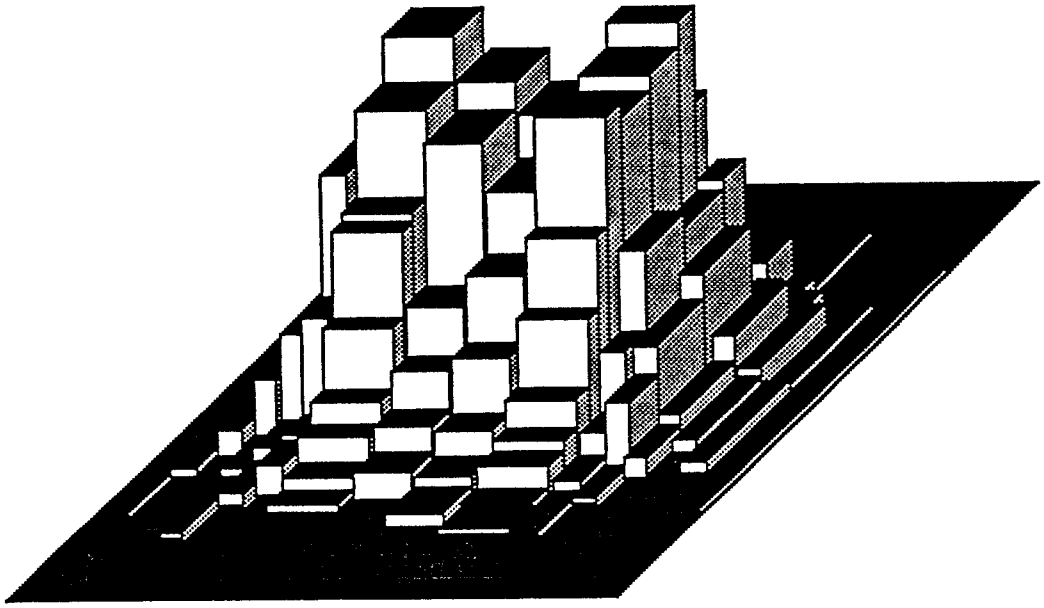

Figure 4.2.29. Semigrid histogram (density as in Figure 4.2.25, sample size = 2000).

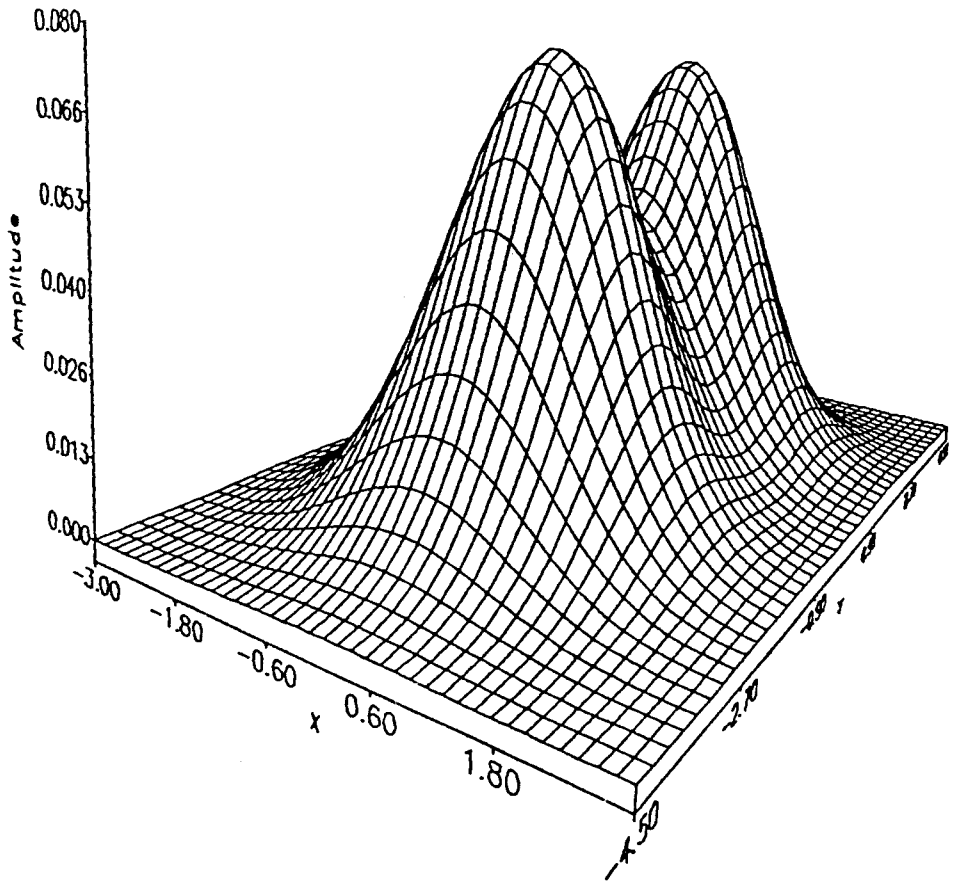

Figure 4.2.30. Bivariate normal density 3.4.1 with p = .5, $\rho_1 = \rho_2 = 0$, $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = 1$, $\mu_1 = 0$, $\mu_2 = -1.5$, $\mu_3 = 0$, $\mu_4 = 1.5$.

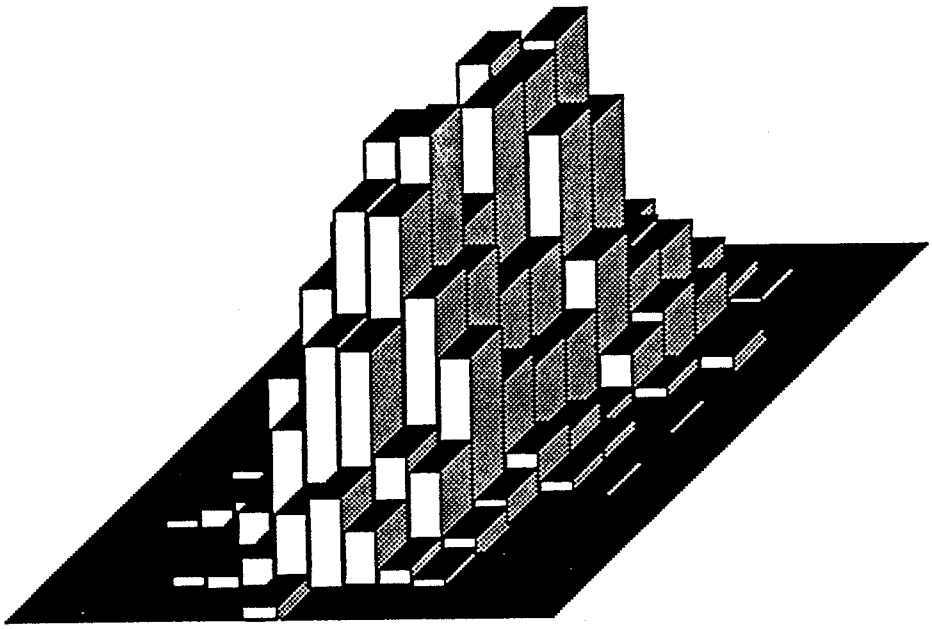

Figure 4.2.31. Regular histogram (density as in Figure 4.2.30, sample size = 2000).

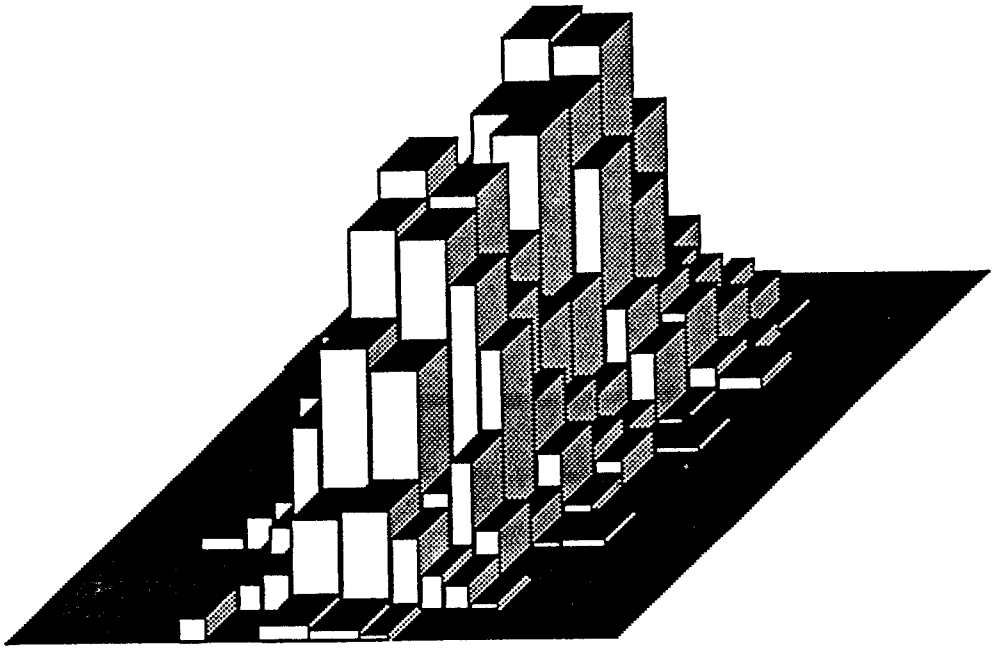

Figure 4.2.32. Semiregular histogram (density as in Figure 4.2.30, sample size = 2000).

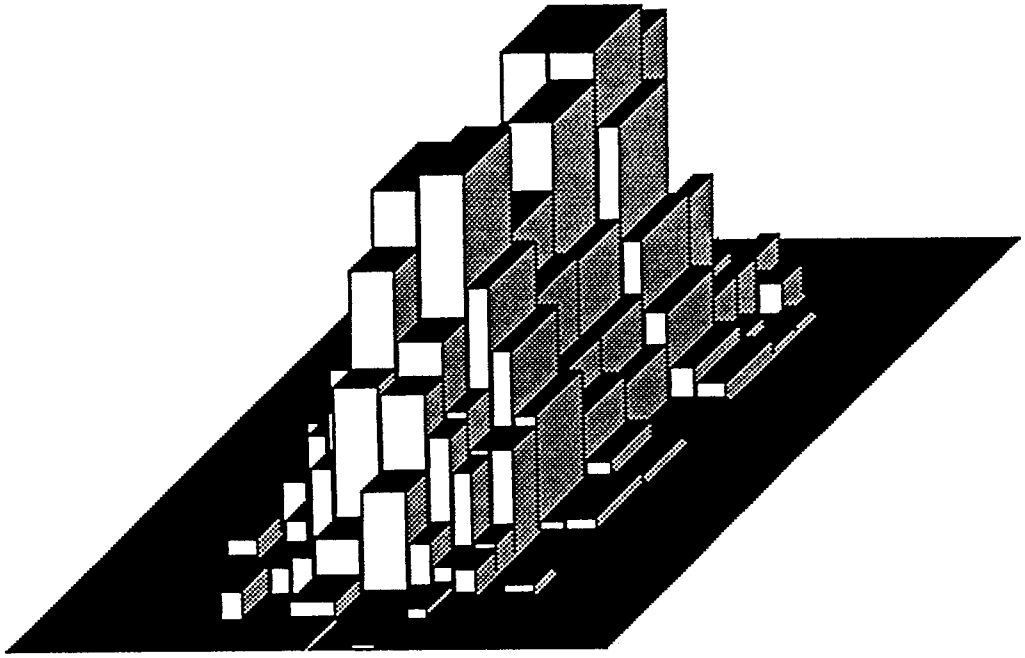

Figure 4.2.33. Grid histogram (density as in Figure 4.2.30, sample size = 2000).

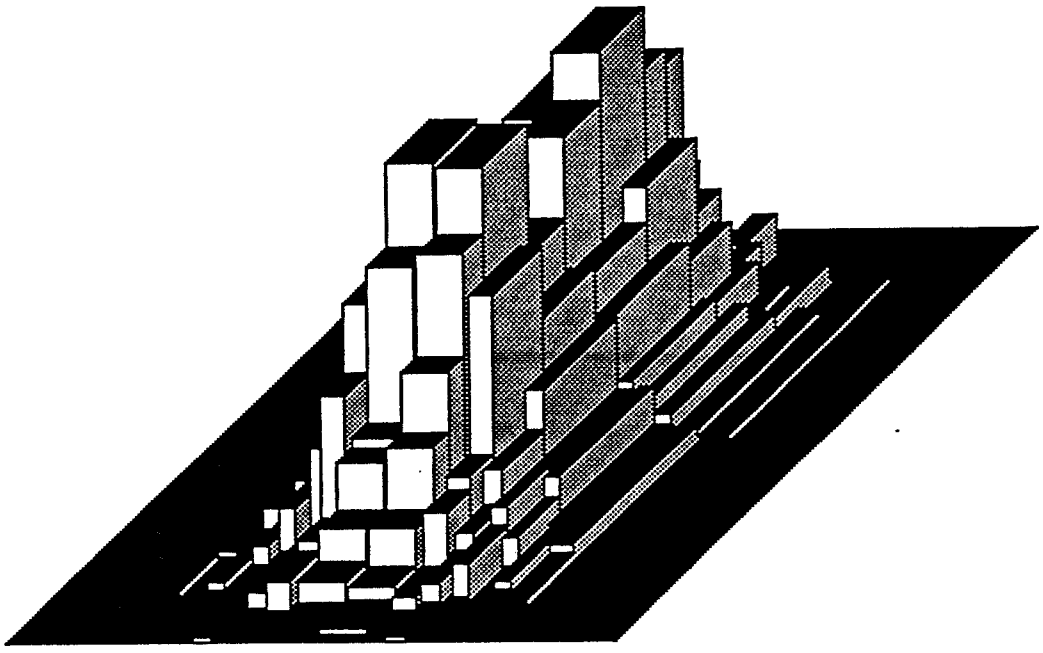

Figure 4.2.34. Semigrid histogram (density as in Figure 4.2.30, sample size = 2000).

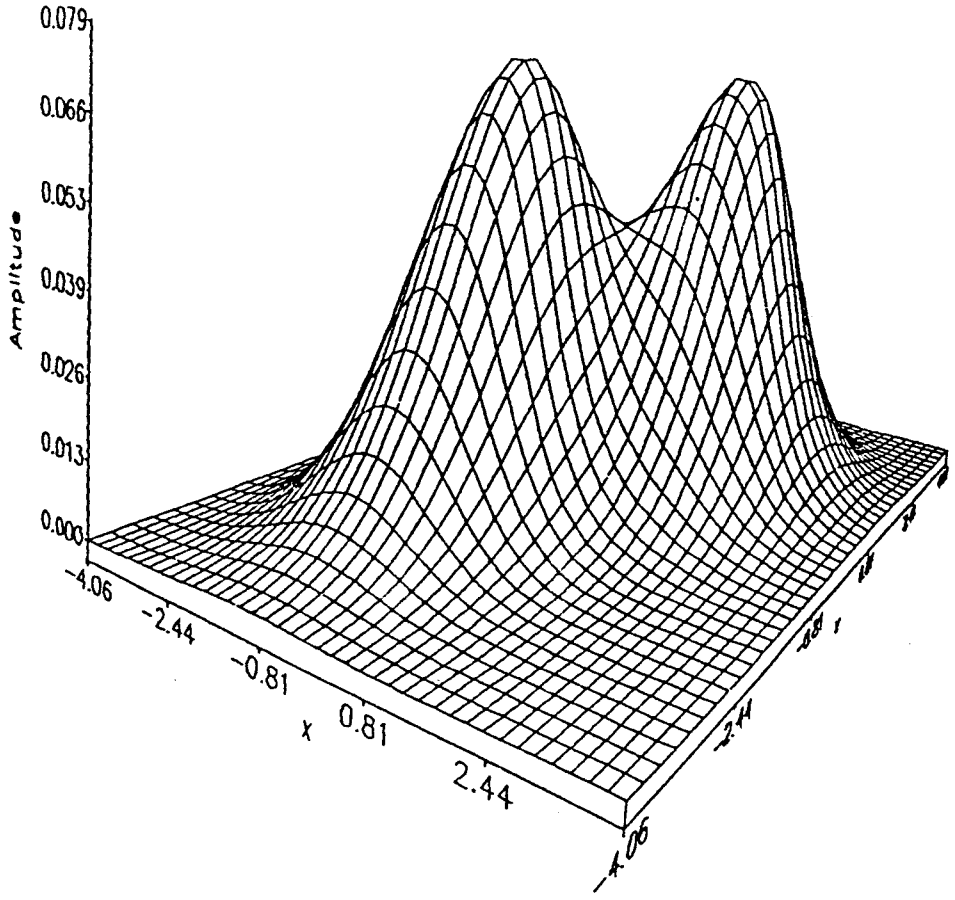

Figure 4.2.35. Bivariate normal density 3.4.1 with p = .5, $\rho_1 = \rho_2 = 0$, $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = 1$, $\mu_1 = \mu_2 = -1.060$, $\mu_3 = \mu_4 = 1.060$.

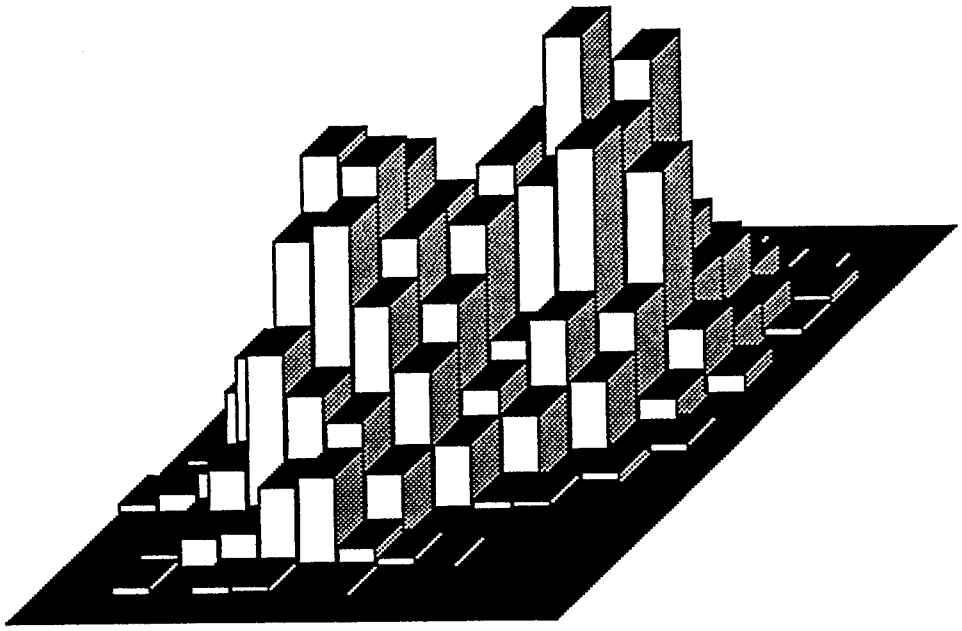

Figure 4.2.36. Regular histogram (density as in Figure 4.2.35, sample size = 2000).

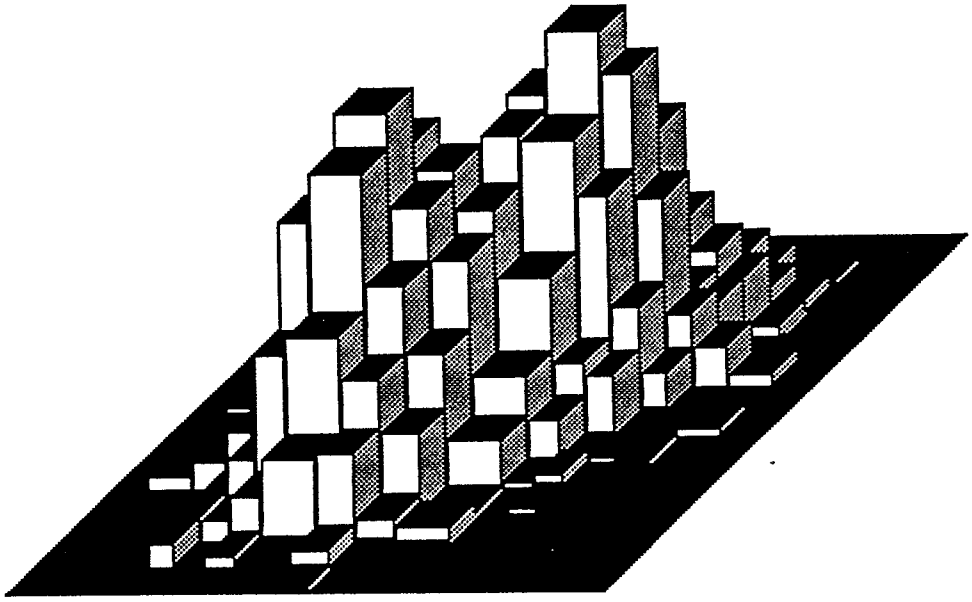

Figure 4.2.37. Semiregular histogram (density as in Figure 4.2.35, sample size = 2000).

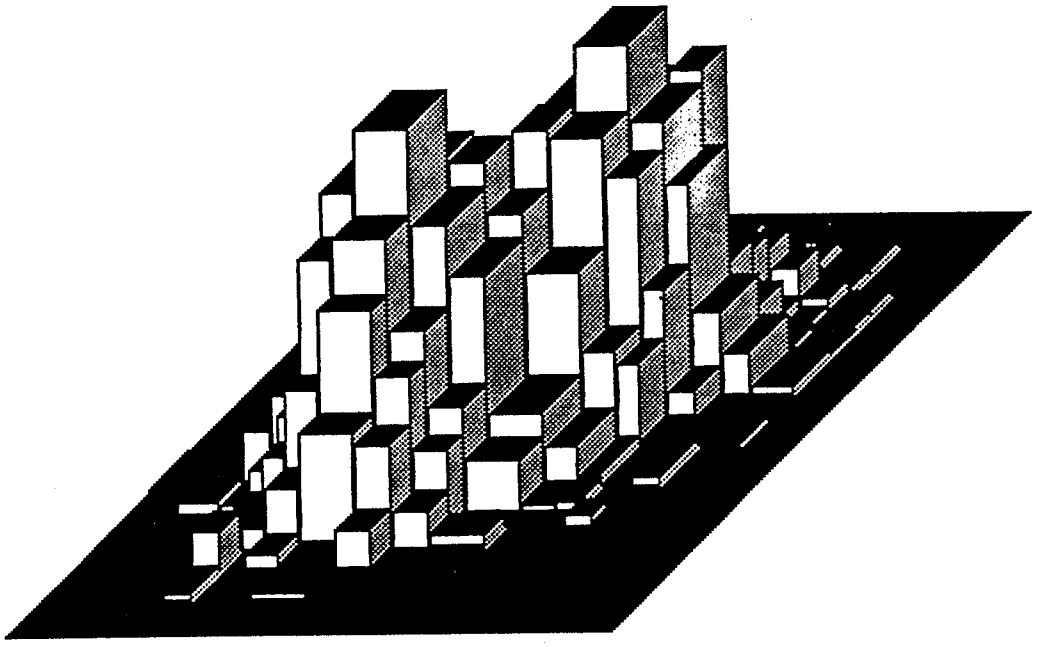

Figure 4.2.38. Grid histogram (density as in Figure 4.2.35, sample size = 2000).

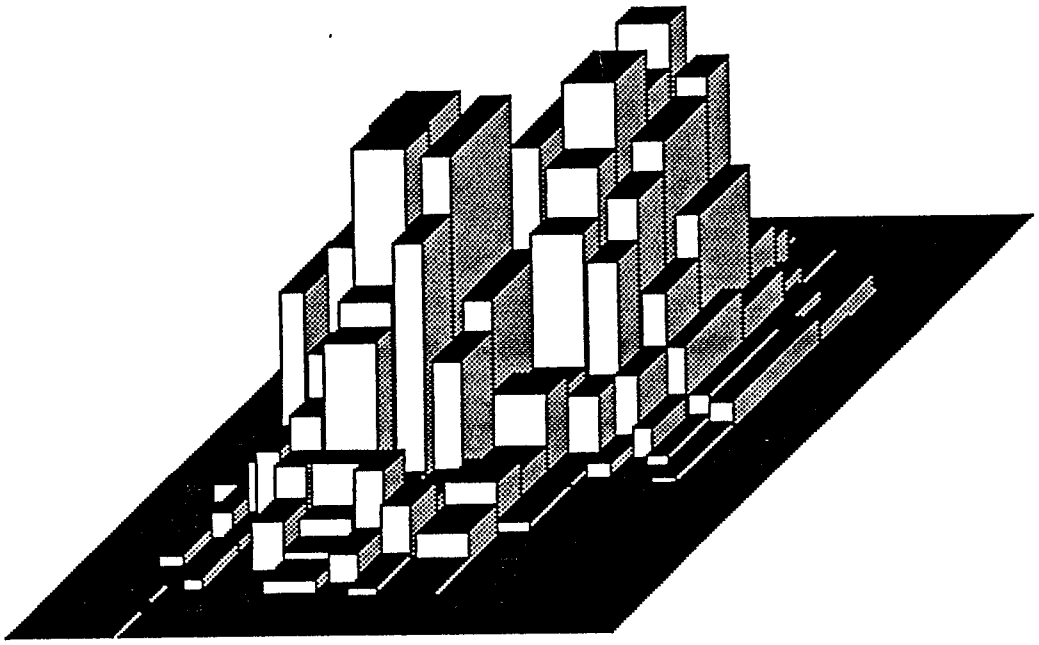

Figure 4.2.39. Semigrid histogram (density as in Figure 4.2.35, sample size = 2000).

There was, of course, no difference in performance by the regular, grid, and free meshes when alignment was made along either the x- or y-axes. There were differences, however, in the semiregular and semigrid cases. When bimodality is expressed in a particular direction, the adaptability of the model in that direction determines the relative size of the integrated mean squared error. Because of the structure of the semiregular grid used (in which the interval widths in the x-direction were variable and the interval widths in the y-direction were constant), the integrated mean squared errors were less when the modes were aligned along the x-axis than when they were aligned along the y-axis. Because of the structure of the semigrid mesh used (in which the interval widths in the x-direction were variable and the interval widths in the y-direction were variable within each strip), the integrated mean squared errors were greater when the modes were aligned along the x-axis than when they were aligned along the y-axis. This is due to the fact that this mesh provides greater adaptability in the y-direction by allowing bin boundaries to be set independently within each strip. In all cases, the integrated mean squared errors were greater when the modes were aligned along the line $x = y$. This occurs because of the basic structure of all of the meshes: cell sides are parallel to the x- and y-axes in all cases. The IMSE results and efficiencies relative to the free mesh for specific cases are found in Tables 4.2.3 and 4.2.4.

Bivariate Normal Distribution $\sigma_1=\sigma_2=\sigma_3=\sigma_4=1$ sample size = 2000							
Parameters of Normal Distribution							
No	p	ρ_1	ρ_2	μ_1	μ_2	μ_3	μ_4
1	1	0	0	0	0	0	0
2	.5	0	0	-.5	0	.5	0
3	.5	0	0	0	-.5	0	.5
4	.5	0	0	-.3536	-.3536	.3536	.3536
5	.5	0	0	-1	0	1	0
6	.5	0	0	0	-1	0	1
7	.5	0	0	-.7071	-.7071	.7071	.7071
8	.5	0	0	-1.5	0	1.5	0
9	.5	0	0	0	-1.5	0	1.5
10	.5	0	0	-1.061	-1.061	1.061	1.061
11	.5	0	0	-2	0	2	0
12	.5	0	0	0	-2	0	2
13	.5	0	0	-1.414	-1.414	1.414	1.414
14	1	.2	.2	0	0	0	0
15	1	.5	.5	0	0	0	0
16	1	.7	.7	0	0	0	0
17	1	.8	.8	0	0	0	0
18	.5	0	0	-.25	0	.25	0
19	.5	0	0	0	-.25	0	.25
20	.5	0	0	-.1768	-.1768	.1768	.1768
21	.5	0	0	-1.2	0	1.2	0
22	.5	0	0	0	-1.2	0	1.2
23	.5	0	0	-.8485	-.8485	.8485	.8485
$\sigma_1=.5$ $\sigma_2=.5$ $\sigma_3=2$ $\sigma_4=2$ sample size = 2000							
24	.5	0	0	-1.5	0	1.5	0
$\sigma_1=.2$ $\sigma_2=.2$ $\sigma_3=3$ $\sigma_4=3$ sample size = 2000							
25	.5	0	0	-1.5	0	1.5	0

Table 4.2.2. Identifying numbers and density parameters for Tables 4.2.3 and 4.2.4.

Bivariate Normal Distribution

$\sigma_1=\sigma_2=\sigma_3=\sigma_4=1$ sample size = 2000

Theoretical IMSE 10^{-3}					
No.	Regular	Semireg.	Grid	Semigrid	Free
1	3.643	3.336	3.054	2.840	2.643
2	3.229	2.947	2.699	2.509	2.334
3	3.229	2.958	2.699	2.509	2.334
4	3.242	3.014	2.802	2.612	2.401
5	2.484	1.961	1.796	1.785	1.600
6	2.484	2.274	1.796	1.776	1.600
7	2.576	2.491	2.410	2.226	2.027
8	2.355	2.090	1.915	1.802	1.681
9	2.355	2.156	1.915	1.772	1.681
10	2.399	2.290	2.182	2.007	1.816
11	2.498	2.281	2.089	1.939	1.830
12	2.498	2.287	2.089	1.904	1.830
13	2.504	2.314	2.139	1.966	1.812
14	3.757	3.531	3.321	3.068	2.983
15	4.519	4.421	4.329	3.909	3.645
16	6.035	5.999	5.966	5.369	5.076
17	7.837	7.811	7.790	7.023	6.467
18	3.531	3.233	2.963	2.755	2.413
19	3.531	3.233	2.963	2.755	2.422
20	3.533	3.247	2.985	2.788	2.428
21	2.343	1.880	1.722	1.552	1.036
22	2.343	2.147	1.722	1.598	1.171
23	2.437	2.366	2.296	2.116	1.732
24	7.200	5.539	4.928	4.416	3.763
25	43.040	30.120	25.420	22.540	19.900

Table 4.2.3.

Bivariate Normal Distribution

$\sigma_1=\sigma_2=\sigma_3=\sigma_4=1$ sample size = 2000

Theoretical Efficiencies relative to Free Grid					
No.	Regular	Semireg.	Grid	Semigrid	Free
1	.7255	.7923	.8654	.9306	1.000
2	.7228	.7920	.8647	.9303	1.000
3	.7228	.7890	.8647	.9303	1.000
4	.7406	.7966	.8569	.9192	1.000
5	.6441	.8159	.8909	.8963	1.000
6	.6441	.7036	.8909	.8502	1.000
7	.7869	.8137	.8411	.9106	1.000
8	.7138	.8043	.8778	.9329	1.000
9	.7138	.7797	.8778	.9486	1.000
10	.7570	.7930	.8323	.9048	1.000
11	.7326	.8023	.8760	.9438	1.000
12	.7326	.8002	.8760	.9611	1.000
13	.7236	.7831	.8471	.9217	1.000
14	.7940	.8448	.8982	.9723	1.000
15	.8066	.8245	.8420	.9325	1.000
16	.8411	.8461	.8508	.9454	1.000
17	.8252	.8279	.8302	.9208	1.000
18	.6834	.7464	.8144	.8759	1.000
19	.6859	.7491	.8174	.8791	1.000
20	.6872	.7478	.8134	.8709	1.000
21	.4422	.5510	.6016	.6675	1.000
22	.4998	.5454	.6800	.7328	1.000
23	.7107	.7320	.7544	.8185	1.000
24	.5226	.6794	.7636	.8521	1.000
25	.4624	.6607	.7828	.8829	1.000

Table 4.2.4

The integrated mean squared error of the elliptical bivariate normal distribution is quite similar to that of the circular normal when $\rho_1 = \rho_2 = 0.2$ but as ellipticity increases, the error increases rapidly and is tripled at $\rho_1 = \rho_2 = 0.8$. The differences in adaptability between the different mesh types nearly disappears at greater ellipticities, because all of the meshes consist of line segments which are parallel to the axes. The ability to adapt to highly elliptical distributions is limited primarily to decreasing cell size, so that not only is the error similar for each grid type, but also the forms of the optimal histogram are quite similar. Figures 4.2.41 and 4.2.42 demonstrate the similarity of form.

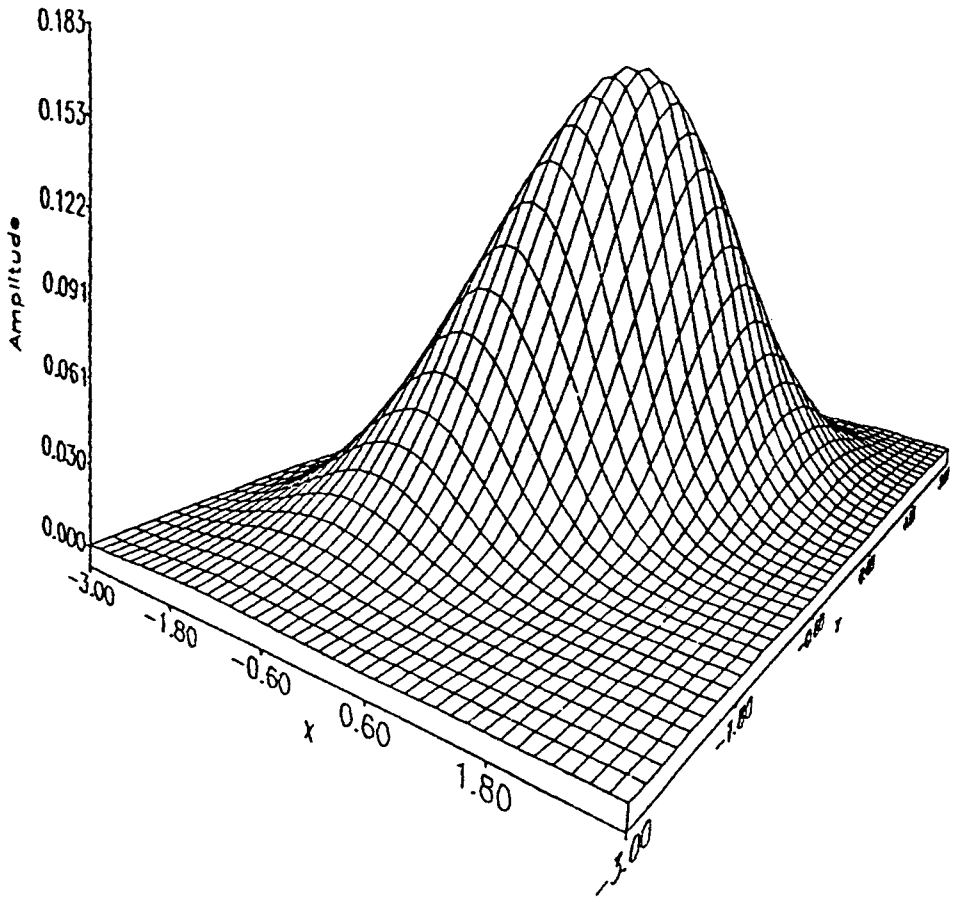

Figure 4.2.40. Bivariate normal density 3.4.1 with p = 1, $\rho_1 = .5$, $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = 1$, $\mu_1 = \mu_2 = \mu_3 = \mu_4 = 0$.

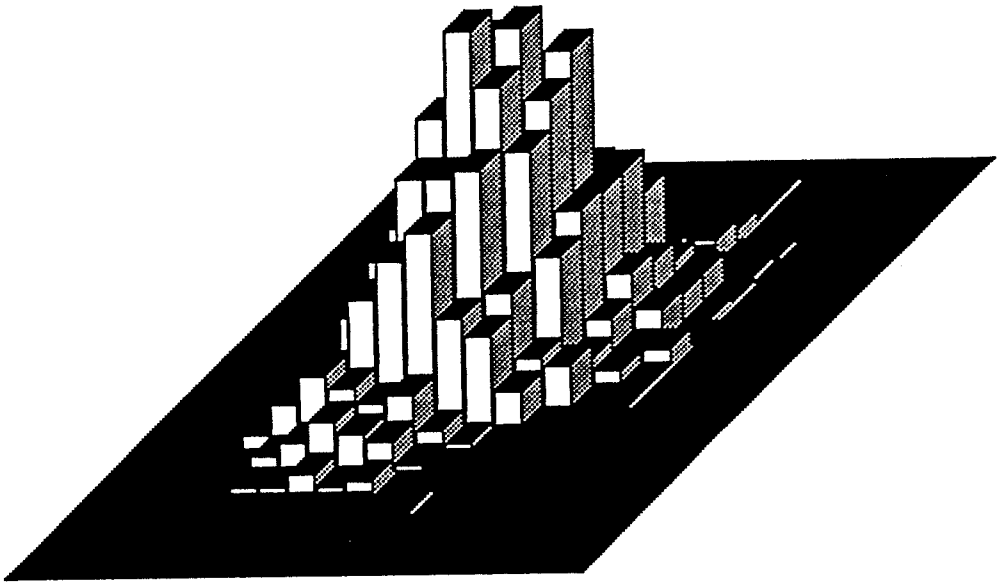

Figure 4.2.41. Regular histogram (density as in Figure 4.2.40, sample size = 2000).

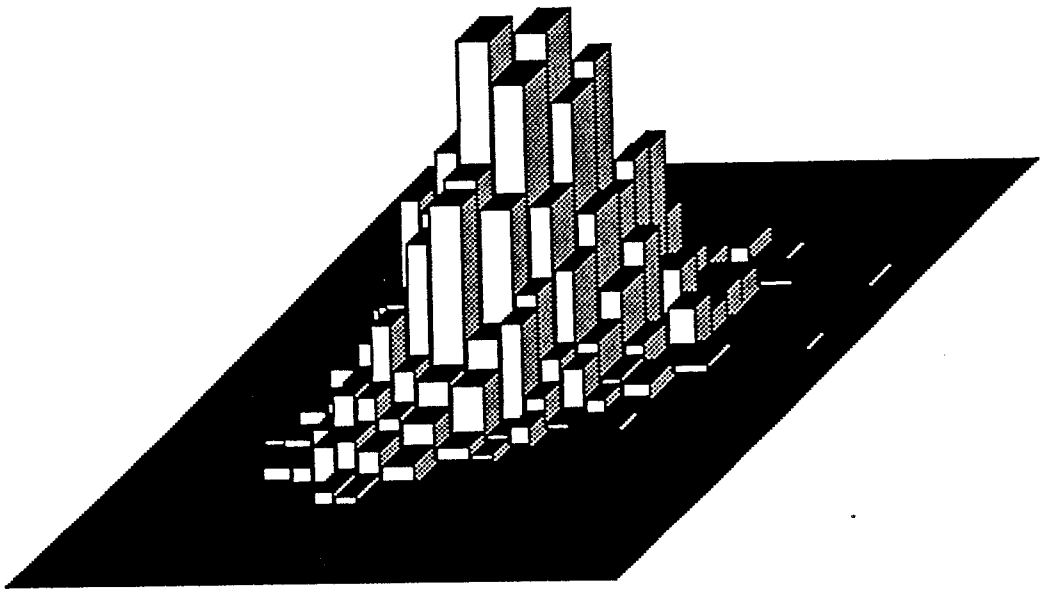

Figure 4.2.42. Grid histogram (density as in Figure 4.2.40, sample size = 2000).

Further improvement can be made only by the rotation of the entire mesh so that the cell sides become parallel to the major axis of an elliptical distribution or to the line joining the modes of a bimodal distribution. Terrell (1984) provided a solution to this problem of orientation in the case of a regular mesh. He defined an "information matrix"

$$I_f = \left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{\partial f}{\partial x} \right) \cdot \left(\frac{\partial f}{\partial y} \right) \right)_{ij}$$

and showed that the integrated mean squared error of a histogram based upon a regular mesh is minimal when the product of the diagonal elements of I_f is minimal. This occurs when the product of the diagonal elements is equal to the determinant of I_f , that is when I_f is diagonal. I_f is diagonalized by pre- and post- multiplying by an orthonormal matrix whose columns are the eigenvectors of I_f . Then the axes of the histogram are oriented in the direction of the eigenvectors of I_f and the constant bin dimensions $g = C_1$, $h = C_2$ are chosen according to the expressions given in section 2.2. The problem of orientation of general meshes is beyond the scope of this paper.

Simulations were performed in which the average empirical integrated mean squared error was obtained. These averages were slightly better than the theoretical *IMSEs* but always higher than the lower bound provided by the free mesh. The averaging of 100 simulations using the circular normal

distribution produced an empirical *IMSE* of 3.179 for the regular mesh, 2.992 for the semiregular mesh, and 2.807 for the grid mesh. The averaging of 100 simulations using density 3.4.1 with p = .5, $\rho_1 = \rho_2 = 0$, $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = 1$, $\mu_1 = -1.5$, $\mu_2 = 0$, $\mu_3 = 1.5$, $\mu_4 = 0$, produced an empirical *IMSE* of 1.928 for the regular mesh, 1.792 for the semiregular mesh, and 1.748 for the grid mesh. Terrell and Scott (1983) obtained empirical results for the one dimensional case which were also slightly better than those predicted by theory.

A set of Dirichlet distributions with a variety of parameters was studied in order to observe the performance of the different meshes with skewed distributions. One of these is illustrated in the next figure; the corresponding histograms follow. Tables 4.2.5 and 4.2.6 contain the theoretical IMSEs and relative efficiencies. Results were similar in general to the results from the normal distributions, the semigrid mesh showing a 20% improvement over the regular mesh.

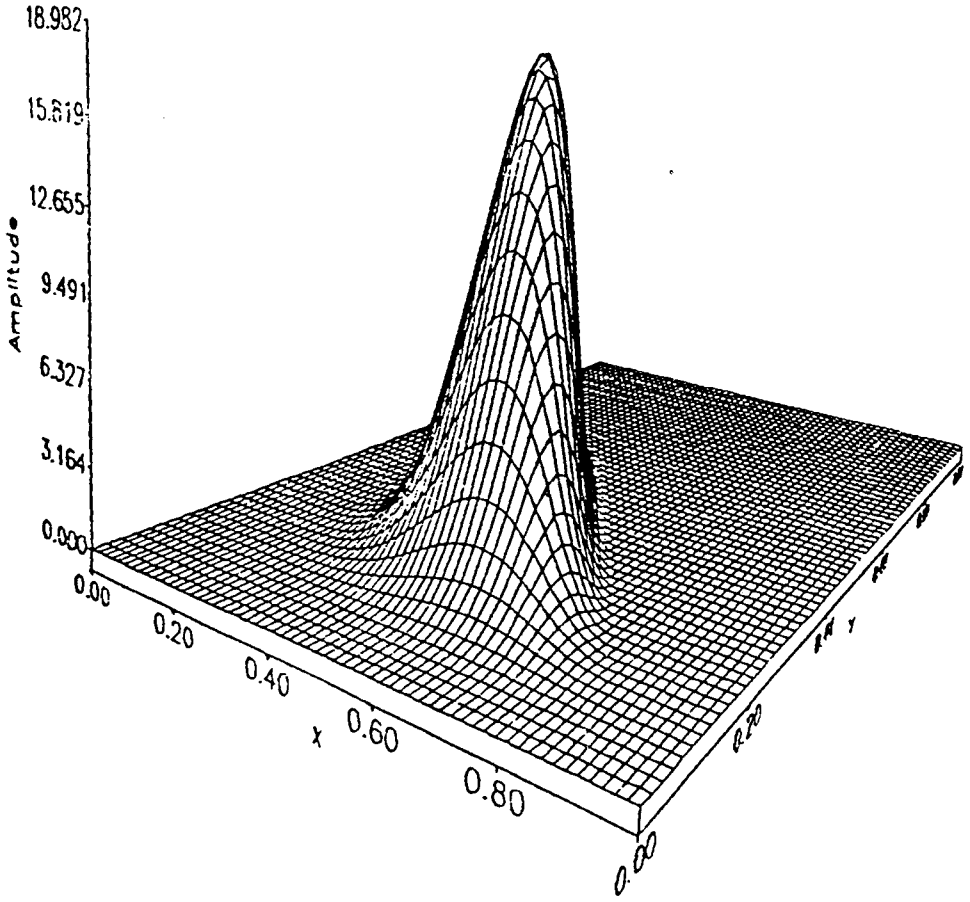

Figure 4.2.43. Dirichlet density 3.4.2 with $\alpha = 10$, $\beta = 7$, $\gamma = 5$.

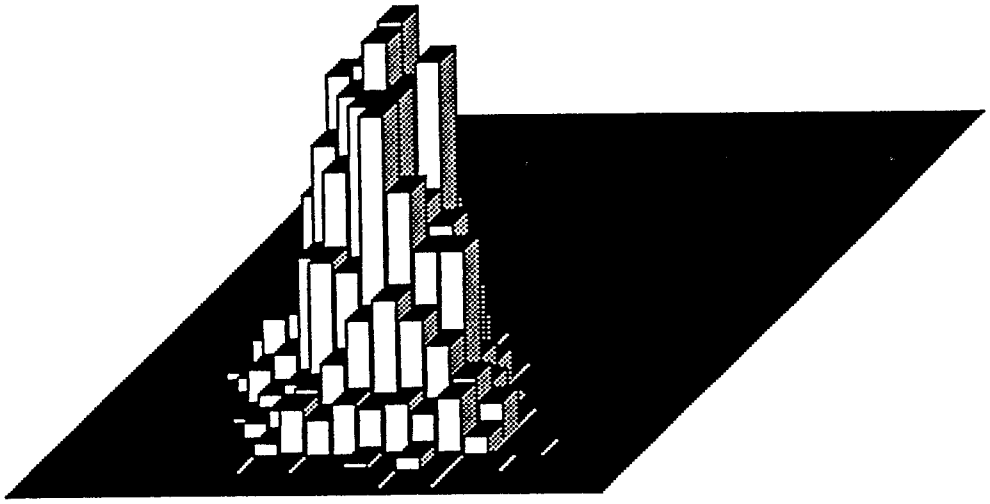

Figure 4.2.44. Regular histogram (density as in Figure 4.2.43, sample size = 2000).

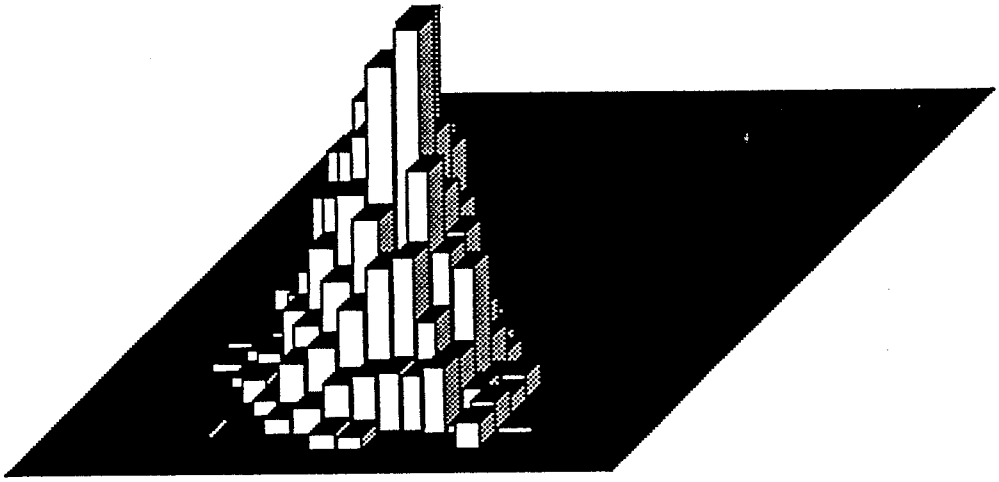

Figure 4.2.45. Semiregular histogram (density as in Figure 4.2.43, sample size = 2000).

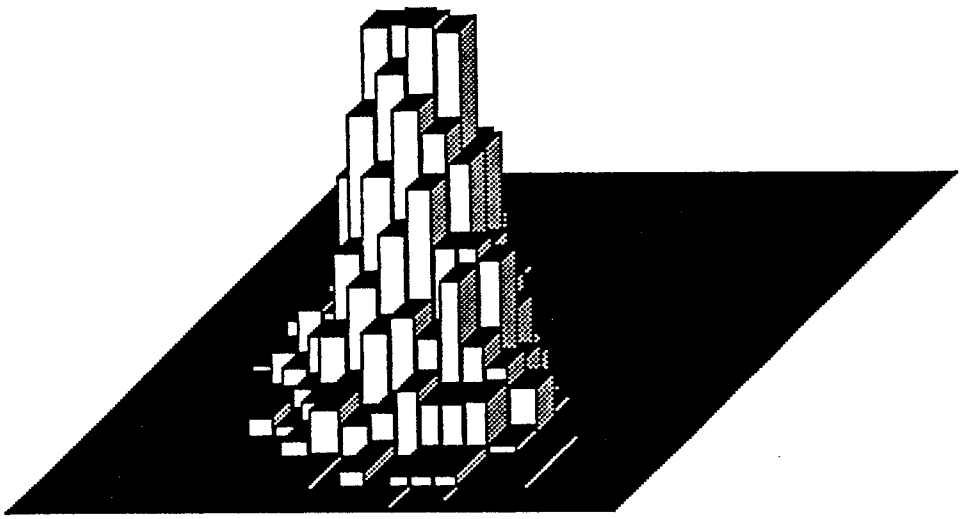

Figure 4.2.46. Grid histogram (density as in Figure 4.2.43, sample size = 2000).

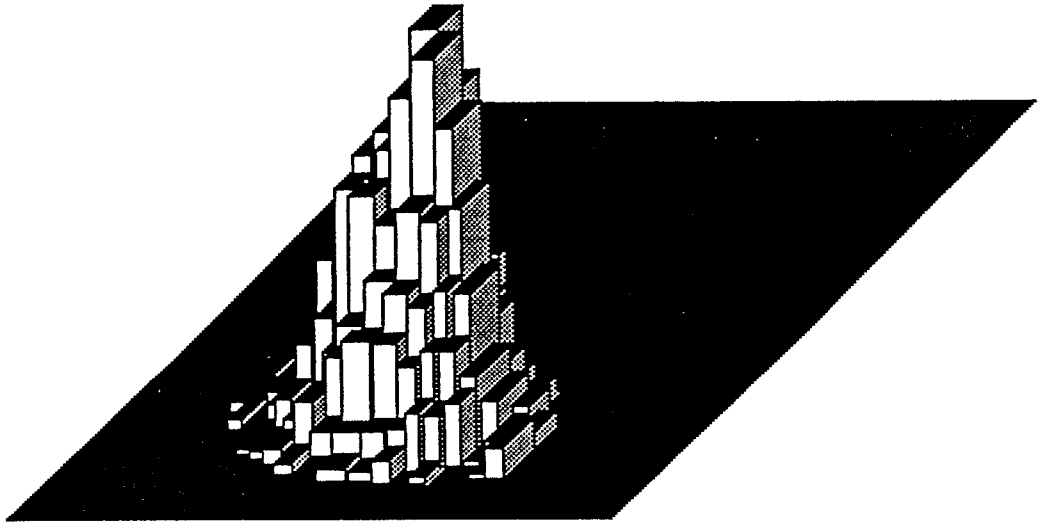

Figure 4.2.47. Semigrid histogram (density as in Figure 4.2.43, sample size = 2000).

Dirchlet sample size = 2000							
Parameters			Theoretical IMSE				
α	β	γ	Regular	Semireg.	Grid	Semigrid	Free
3	3	3	.2018	.1938	.1865	.1612	.1358
5	5	5	.3164	.3072	.2987	.2650	.2279
3	3	10	.4926	.4325	.3792	.3497	.3157
7	7	7	.4363	.4249	.4146	.3698	.3198
8	6	8	.4584	.4448	.4293	.3855	.3350
10	5	7	.4948	.4823	.4653	.4157	.3607
10	5	5	.4888	.4801	.4706	.4139	.3562
4	10	4	.4989	.4866	.4776	.4172	.3549
10	7	7	.5212	.5114	.5011	.4456	.3855
10	5	10	.5548	.5344	.5058	.4577	.4003
10	7	5	.5331	.5268	.5217	.4582	.3951
3	10	3	.5623	.5436	.5282	.4524	.3774
10	3	3	.5626	.5373	.5284	.4483	.3774
10	10	10	.6178	.6028	.5890	.5277	.4575
10	4	4	.4989	.4881	.4776	.4155	.3549
10	3	10	.6270	.5970	.5311	.4843	.4275
4	4	10	.4474	.4099	.3754	.3439	.3068
10	4	10	.5720	.5483	.5087	.4615	.4047
5	5	10	.4481	.4197	.3933	.3584	.3175
5	10	5	.4888	.4783	.4707	.4152	.3562

Table 4.2.5.

Dirchlet sample size = 2000							
Parameters			Theoretical Efficiencies relative to Free Grid				
α	β	γ	Regular	Semireg.	Grid	Semigrid	Free
3	3	3	.6729	.7007	.7282	.8424	1.000
5	5	5	.7203	.7419	.7630	.8600	1.000
7	7	7	.7330	.7526	.7713	.8648	1.000
10	10	10	.7405	.7590	.7767	.8670	1.000
3	3	10	.6409	.7299	.8325	.9028	1.000
10	3	3	.6708	.7024	.7142	.8418	1.000
3	10	3	.6712	.6943	.7145	.8342	1.000
10	3	10	.6818	.7161	.8049	.8827	1.000
4	4	10	.6857	.7485	.8173	.8921	1.000
10	4	4	.7114	.7271	.7431	.8542	1.000
4	10	4	.7114	.7293	.7431	.8507	1.000
10	4	10	.7075	.7381	.7956	.8769	1.000
5	5	10	.7085	.7565	.8073	.8859	1.000
10	5	5	.7287	.7419	.7569	.8606	1.000
5	10	5	.7287	.7447	.7567	.8579	1.000
10	5	10	.7215	.7491	.7914	.8746	1.000
10	5	7	.7290	.7479	.7752	.8677	1.000
10	7	5	.7411	.7500	.7573	.8623	1.000
10	7	7	.7396	.7538	.7693	.8651	1.000
8	6	8	.7308	.7531	.7803	.8690	1.000

Table 4.2.6.

The greatest improvements in IMSE made possible by adaptive meshes were observed for mixed normal distributions in which the variances differed widely. It was already noted that a 20% improvement can be expected for mixed normal distributions having the same variance as well as for unimodal normal and Dirichlet distributions. An improvement in efficiency of over 90% may be obtained when the variances of bimodal normals differ, for example when $\sigma_1 = \sigma_2 = .2$, $\sigma_3 = \sigma_4 = 3$ (see Tables 4.2.3 and 4.2.4).

5. Application

Scott (1986) provided data on paired cholesterol and triglyceride levels in 320 patients with heart disease. It is assumed that these levels have a bivariate elliptical normal distribution with mean vector and covariance matrix estimated from the sample. Histograms were constructed using the four mesh types discussed in previous sections. All of the histograms suggest a bimodal structure in which the modes are separated by less than 2 σ so that the bimodality is not represented by separate peaks but rather by an elongation of the cells in the y-direction.

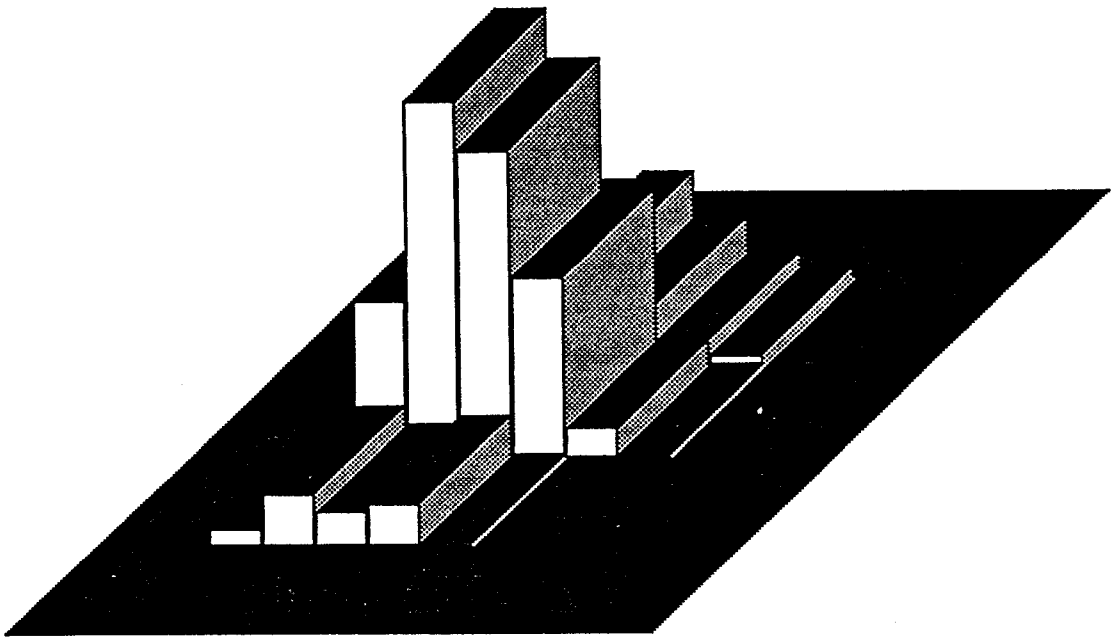

Figure 5.1. Regular histogram based upon bivariate elliptical normal data from 320 cardiology patients; normal model as in equation 3.4.1 with p = 1, $\rho_1 = .226$, $\mu_1 = 2.162$, $\sigma_1 = .4301$, $\mu_2 = 1.794$, $\sigma_2 = 1.019$.

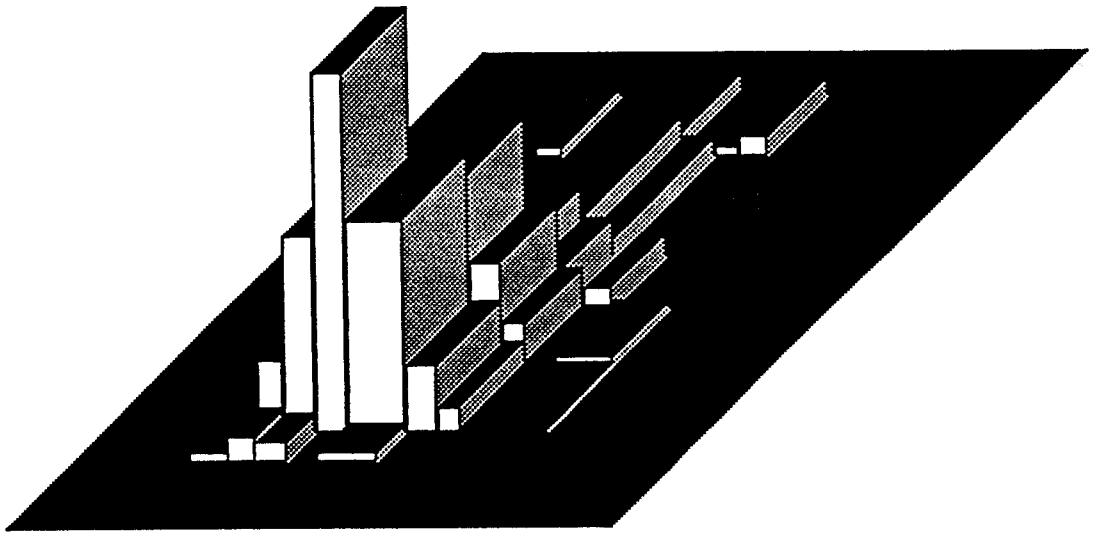

Figure 5.2. Grid histogram based upon bivariate elliptical normal data from 320 cardiology patients; normal model as in equation 3.4.1 with p = 1, ρ_1 = .226, μ_1 = 2.162, σ_1 = .4301, μ_2 = 1.794, σ_2 = 1.019.

A similar structure is apparent in the theoretical cases in which the modes are separated by 2 σ. It is impossible using our methods to distinguish between the form of distributions when the modes are between 2.2 and 2.7 σ apart with such small sample sizes. For example, we were not able to demonstrate separate peaks with our histograms at a sample size of 10,000 when the modes were separated by 2.4 σ even though the true functional form had two clearly defined peaks. See Figures 5.3 and 5.4.

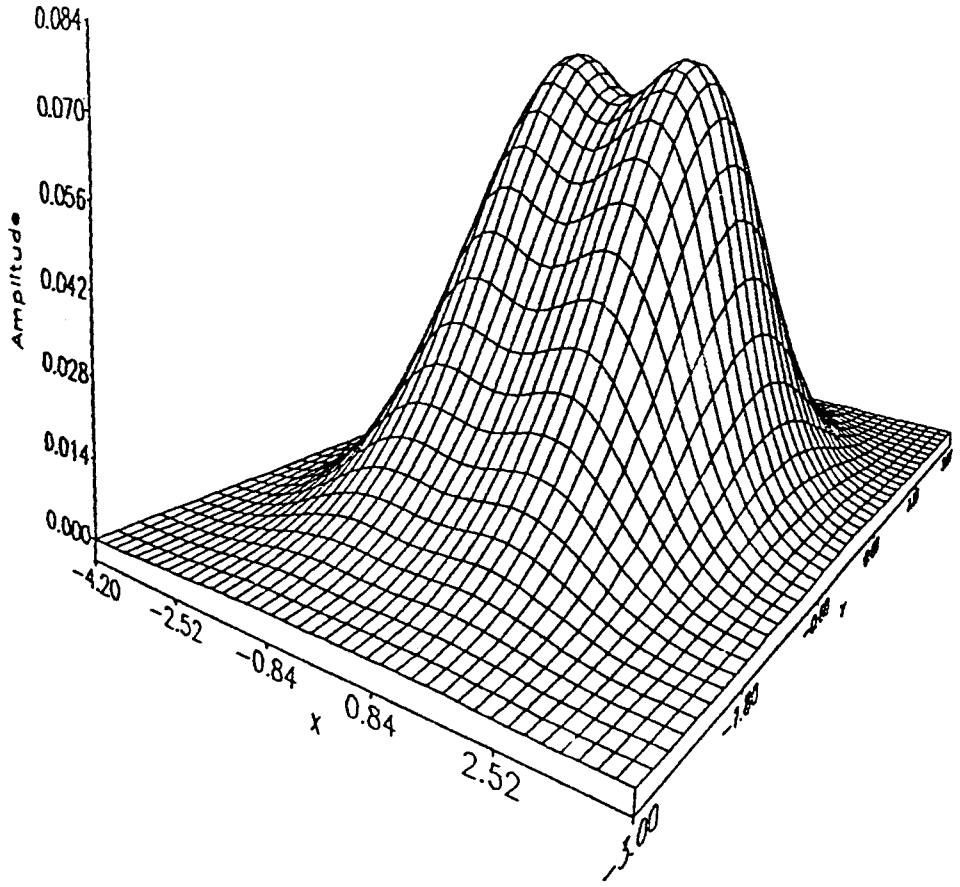

Figure 5.3. Bivariate normal density 3.4.1 with p = 1, $\rho_1 = \rho_2 = 0$, $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = 1$, $\mu_1 = -1.2$, $\mu_2 = 0$, $\mu_3 = 1.2$, $\mu_4 = 0$.

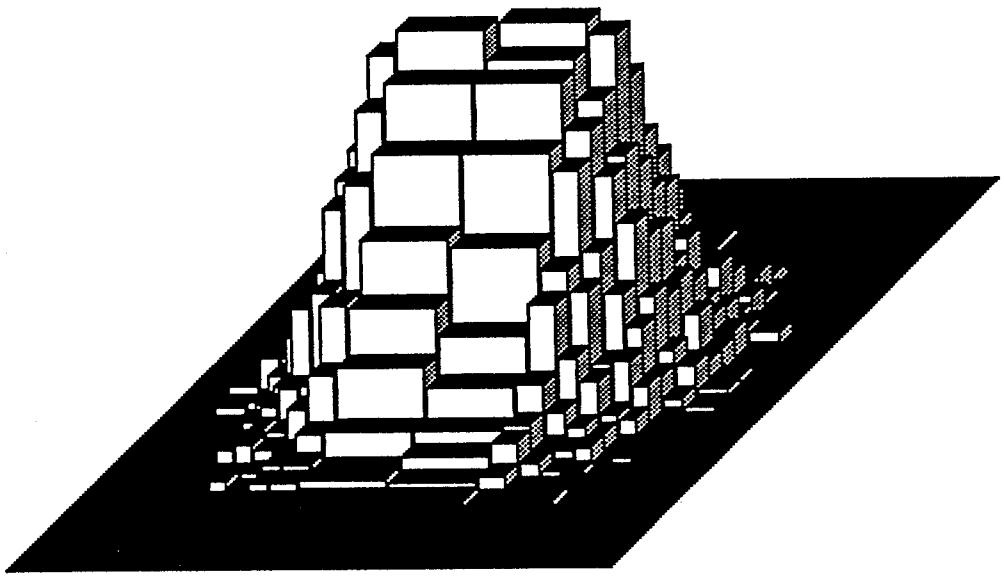

Figure 5.4. Semiregular histogram (density as in Figure 5.1 with sample size of 10,000). The Grid and Semigrid histograms constructed from the same density and sample size were similar in appearance and did not show bimodality.

6. Discussion

Some general observations may be made regarding the implications of our research. Specifically, we should indicate under what conditions variable-dimension meshes are preferable to fixed-dimension meshes. Whenever the density consists of distinct sections whose forms are substantially different, such as is the case for a mixed bivariate normal density having unequal variances, we strongly recommend the use of variable-dimension meshes. In the case of densities whose forms approach that of the circular normal and where the choice of mesh will not result in substantial decreases in error, the investment of time and effort relative to improvement of estimate should be considered. Although the Grid mesh will not produce as efficient an estimator as the Semigrid it, nevertheless, may be preferable to the Semigrid because of its ease of implementation and symmetry.

In the case of a highly elliptical density whose major axis is not approximately parallel to one of the coordinate axes, no improvement will result from the use of variable dimension meshes. In order to obtain better estimates in this case, a change of coordinates should be made so that the major axis of the density is parallel to one of the coordinate axes, and only then should one of the more efficient meshes be constructed. Further research in mesh orientation is recommended.

Since the histogram estimator, even one based upon variable-dimension mesh, may not adequately represent bimodal densities in which the distance between the modes is small, another type of estimator such as a variety of the Rosenblatt kernel may be preferred. Again, the importance of accuracy and the level of sophistication must be balanced in each application.

REFERENCES

Apostol, T. M. (1964). *Mathematical Analysis.* Addison-Wesley, London.

Bean, S. J., Tsokos, C. P. (1980). Developements in Nonparametric Density Estimation. *Internat. Statist. Ref.,* 48, 267.

Breiman, L., Meisel, W., Purcell, E. (1977). Variable Kernel Estimates of Multivariate Densities. *Technometrics,* 19, 135.

Čencov, N. N. (1962). Evaluation of an Unknown Density Distribution from Observations. *Soviet Math.,* 3, 1559.

David, H. A. (1981). *Order Statistics.* Wiley, New York.

Epanechnikov, V. A. (1969). Nonparametric Estimator of a Multivariate Probability Density. *Theor. Prob. Appl.,* 14, 153.

Freedman, D., Diaconis, P. (1981). On the Histogram as a Density Estimator: L_2 Theory. *Z. Wahrsch. Verw. Gebiete,* 57, 453.

Freedman, D., Diaconis, P. (1981). On the Maximum Deviation Between the Histogram and the Underlying Density. *Z. Wahrsch. Verw. Gebiete,* 58, 139.

Hand, D. J. (1981). *Discrimination and Classification.* Wiley, New York.

IMSL (1984). International Mathematical and Statistical Libraries. Houston, Texas.

Kendall, M. G., Stuart, A. (1969). *The Advanced Theory of Statistics, Vol. I, II, III.* Griffin, London.

Nezames, D. D., Scott, D. W. (1980). Some Results for Estimating Bivariate Densities using Kernel, Orthogonal Series, and Penalized Likelihood Methods. Doctoral Dissertation, Rice University, Houston, Texas.

Olmstead, J. M. H. (1956). *Real Variables.* Appleton-Century-Crofts, New York.

Parzen, E. (1962). On Estimation of a Probability Density Function and Mode. *Ann. Math. Statist.*, 33, 1065.

Rosenblatt, M. (1965). Remarks on Some Nonparametric Estimates of a Density Function. *Ann. Math. Statist.*, 27, 832.

Scott, D. W. (1976). Nonparametric Probability Density Estimation by Optimization Theoretic Techniques. Doctoral Dissertation, Rice University, Houston, Texas.

Scott, D. W. (1979). On Optimal and Data-Based Histograms. *Biometrika*, 66, 605.

Scott, D. W. (1982). Optimal Meshes for Histograms Using Variable-Width Bins. Paper presented to the annual meeting of the American Statistical Association.

Scott, D. W. (1985). A Note on Choice of Bivariate Histogram Bin Shape. Unpublished manuscript.

Scott, D. W. (1986). Histogram Proof a la Frequency Polygon. Unpublished manuscript.

Tapia, R. A., Thompson, J. R. (1978). *Nonparametric Probability Density Estimation.* John Hopkins University Press, Baltimore.

Tarter, M. E., Kronmal, R. A. (1970). On Multivariate Density Estimates Based on Orthogonal Expansions. *Ann. Math. Statist.*, 41, 718.

Tarter, M. E., Kronmal, R. A. (1976). An Introduction to the Implementation and Theory of Nonparametric Density Estimation. *Amer. Statist.*, 30, 105.

Terrell, G. R., Scott, D. W. (1983). Variable Windows Density Estimates. Unpublished manuscript.

Terrell, G. R. (1983). Bivariate Histograms. Unpublished manuscript.

Terrell, G. R. (1984). Orienting Regular Rectangular Histograms. Unpublished manuscript.

Terrell, G. R. (1986). Multivariate Histograms. Unpublished manuscript.