Statistica Sinica **10**(2000), 559-575

ASYMPTOTICALLY OPTIMAL REPRESENTATIVE POINTS OF BIVARIATE RANDOM VECTORS

Yingcai Su

Southwest Missouri State University

Dedicated to the memory of Professor Stamatis Cambanis

Abstract: Optimal representative points (also called principal points) of scalar random variables are known in several cases and asymptotically optimal representative points are known for all densities, as the number of the points increases to infinity. When random quantities are uniformly distributed over a bounded two dimensional region, the centers of regular hexagons as representative points are asymptotically optimal. So far asymptotically optimal representative points of non-uniform multivariate distributions are not reported. Here, we give a method of designing representative points for non-uniform bivariate random vectors and show that the proposed method is asymptotically optimal. Examples of simulations with Gaussian, Pearson Type VII and Laplacian density functions are considered.

Key words and phrases: Asymptotically optimal discretizer, principal points, representative points.

1. Introduction

This paper considers the problem of discretizing a continuous random vector. This has applications in all digital systems, in designing optimal sizes and shapes of clothes or protection masks, and in cluster analysis (see Hartigan (1975)). A general discretizer forms a finite partition of the set of values of a random quantity and chooses a point in each region to minimize the mean square error. These points are called representative points in Fang and Wang (1994) and Fang, Yuan and Bentler (1994), principal points in Flury (1990) and quantization points in engineering. Representative points are coded for processing or transmission in digital systems, and clothes or protection masks of sizes or shapes described by principal points are mass-produced. Optimal scalar discretizers are known for several densities and certain numbers of representative points (Tarpey (1994)). Also, for a large class of densities asymptotically optimal scalar discretizers are known (Cambanis and Gerr (1983)).

When a two-dimensional random vector is uniformly distributed over the unit square, Neumann (1982) showed that an asymptotically optimal discretizer corresponds to the tessellation of regular hexagons with representative points

at their centers. But general procedures of finding optimal or asymptotically optimal representative points of non-uniform distributions in two or higher dimensions are unknown even for Gaussian or Laplacian distributions. A lot of work has been done on designing discretizers to reduce mean square error. Flury (1990) obtained optimal representative points of several different sizes for bivariate normal distributions. When a random vector has an elliptical distribution, Tarpey, Li and Flury (1995) showed that optimal representative points or principal points exist in the linear subspace spanned by principal components of the covariance matrix of the random vector. Fischer and Dicharry (1984) generated discretizers of sizes $8, 16$ and 32 for Gaussian, Laplacian and Gamma distributions by using the Monte Carlo simulation method proposed by Linde, Buzo and Gray (see Fischer and Dicharry (1984) and the references therein). One drawback of simulation algorithms is that errors are random. Fang and Wang (1994) used a number-theoretic (quasi Monte Carlo) method to generate representative points of sizes $3, \ldots, 32$, for two or higher dimensional distributions such as Gaussian and symmetric multivariate Pearson Type VII distributions. For a two dimensional Gaussian distribution and sizes $8, 16$ and 32, Fang and Wang's number-theoretic method yields slightly lower (deterministic) mean square errors than those of simulation algorithms (see Fang and Wang (1994)). The same quasi Monte Carlo method was also considered in Fang, Yuan and Bentler (1994). However, asymptotic properties of the number-theoretic method are not reported.

In this paper, we present a discretizing method for bivariate random vectors with a finite second moment. We first properly partition the set of values of a random vector into disjoint subregions, then decompose each subregion into regular hexagons. The number of subregions grows to infinity with the size of discretizer but at a slower rate, and the number of hexagons in each subregion is determined by an auxiliary positive density. It is shown that centers of these hexagons, as representative points, are asymptotically optimal for a proper choice of the auxiliary density function. Examples with Gaussian, Pearson Type VII and Laplacian distributions are considered.

2. Results and Examples

Let $X = (X_1, X_2)$ be a random vector distributed over a region D in R^2 according to a known probability density function $p(x), x \in D$. For a fixed positive integer N, consider a partition of D into N disjoint subregions $D_{i,N}, i = 1, \ldots, N$: $D = \bigcup_{i=1}^{N} D_{i,N}$, select a representative point $x_{i,N}$ from each region $D_{i,N}$ and define the discretizer function

$$Q_N(x) = \sum_{i=1}^{N} x_{i,N} 1_{D_{i,N}}(x),\ x \in D,$$

where $1_{D_{i,N}}(\cdot)$ denotes the indicator function of the set $D_{i,N}$. The discretizer $Q_N(\cdot)$ maps (or rounds off) every point in $D_{i,N}$ to the representative point $x_{i,N}$. Here, the random vector $X = (X_1, X_2)$ is assumed to have finite second moment: $E\|X\|^2 < \infty$, where $\|x\| = (x_1^2 + x_2^2)^{1/2}$ is the Euclidean norm, and the performance of a discretizer is measured by the mean square error (mse)

$$e^2(Q_N, p) \equiv E\|X - Q_N(X)\|^2 = \int_D \|x - Q_N(x)\|^2 p(x)dx$$
$$= \sum_{i=1}^{N} \int_{D_{i,N}} \|x - x_{i,N}\|^2 p(x)dx. \tag{2.1}$$

A minimizer Q_N^o of $e^2(Q_N, p)$ over all discretizers Q_N with N representative points is called an optimal discretizer.

A simple argument shows that for N fixed representative points $\{x_{i,N}\}_{i=1}^N$, the best possible partition is the Voronoi partition, namely, each $D_{i,N}$ is the set of points closer to $x_{i,N}$ than to any other representative points. Thus, once the best possible representative points are specified, an optimal discretizer Q_N^o is fully determined. However, in two and higher dimensions, it turns out to be too complicated to find the optimal representative points even for simple bivariate densities and small N. Therefore, an asymptotic approach is considered to circumvent the difficulties of fixed sizes. A sequence of discretizers $\{Q_N^*\}$ is asymptotically optimal for a density p if

$$\lim_{N\to\infty} e^2(Q_N^*, p) / \inf_{Q_N} e^2(Q_N, p) = 1, \tag{2.2}$$

where the infinimum is taken over all discretizers with N representative points. Zador (1982) examined the denominator of (2.2) in arbitrary dimensions and obtained its limiting behavior, but the approach was not constructive. In the two dimensional case, this limiting behavior becomes

$$\lim_{N\to\infty} N \inf_{Q_N} e^2(Q_N, p) = \frac{5\sqrt{3}}{54} \{\int_D p^{1/2}(x)dx\}^2. \tag{2.3}$$

From (2.2) and (2.3), it follows that a sequence of discretizers $\{Q_N^*\}$ is asymptotically optimal if and only if

$$\lim_{N\to\infty} N e^2(Q_N^*, p) = \frac{5\sqrt{3}}{54} \{\int_D p^{1/2}(x)dx\}^2. \tag{2.4}$$

In the following, without loss of generality, we state our procedure for bivariate random vectors distributed in a bounded region D. If D is unbounded,

by the assumption of finite second moment, one can always choose a bounded region D_N such as a circle or square, so that

$$N \int_{D \setminus D_N} \|x\|^2 p(x) dx \to 0.$$

Then one works with the bounded domain D_N. Asymptotically, the contribution to the distortion error from the unbounded region $D \setminus D_N$ is of a higher order than N^{-1}. More naturally, one can select the region D_N so that the value of $Prob\,(X \in D \setminus D_N) = \int_{D \setminus D_N} p(x) dx$ is relatively small.

Specifically, we proceed as follows.

Partition of the Region D. For a positive density function $g(x)$ on D and each fixed N, partition D into $m(N)$ subregions $\Pi_{k,N}, k = 1, \ldots, m(N)$, in such a way that the following conditions are satisfied.

Condition 1. $m(N) \to \infty$ and $m(N)/N \to 0$ as $N \to \infty$.

Condition 2. For all N,

$$m(N) \min_{1 \le k \le m(N)} \int_{\Pi_{k,N}} g(x) dx \ge c,$$

where c is a positive constant.

Condition 3. Each $\Pi_{k,N}$, $k = 1, \ldots, m(N)$, contains $n_{k,N} = \lceil N \int_{\Pi_{k,N}} g(x) dx \rceil$ full regular hexagons with radius $d_{k,N}$, where

$$d_{k,N}^2 = \frac{2}{3\sqrt{3}} \mid \Pi_{k,N} \mid / (N \int_{\Pi_{k,N}} g(x) dx), \tag{2.5}$$

$\lceil \cdot \rceil$ denotes the integer part, and $\mid A \mid$ stands for the area of the region A.

Condition 4. For $k = 1, \ldots, m(N)$, let $H_{k,N}$ be the union of the hexagons in Condition 3 and $T_{k,N}$ be the set of their centers; then

$$\sup_{x \in \Pi_{k,N}} \|x - T_{k,N}\| \le a d_{k,N}$$

for some constant a, where $\|x - A\| = \inf_{y \in A} \|x - y\|$ is the distance of the point x to the set A.

According to Conditions 1-4, each subregion $\Pi_{k,N}$ is essentially tessellated into a number (equal to $\lceil N \int_{\Pi_{k,N}} g(x) dx \rceil$) of regular hexagons. In consideration of computing the integral $\int_{\Pi_{k,N}} g(x) dx$, one should choose $\Pi_{k,N}$ such that $g(x)$ varies little in $\Pi_{k,N}$ and can be integrated easily. Condition 1 implies that the number of subregions tends to infinity with N but at a slower rate, for instance,

$m(N) = \log(N)$, $\sqrt{N}$, etc. Condition 2 excludes trivial subregions, and is fulfilled for instance, if for every k, $\int_{\Pi_{k,N}} g(x)dx = 1/m(N)$. Condition 3 says no matter what particular shape each $\Pi_{k,N}$ takes, it should contain about $n_{k,N}$ full regular hexagons each with radius $d_{k,N}$ satisfying the compatibility condition

$$\begin{aligned} n_{k,N} \times \{area\ of\ each\ hexagon\} &= n_{k,N}(3\sqrt{3}/2)d_{k,N}^2 \\ &= |\ \Pi_{k,N}\ |\ \frac{\lceil N \int_{\Pi_{k,N}} g(x)dx \rceil}{N \int_{\Pi_{k,N}} g(x)dx}\ \cong |\ \Pi_{k,N}\ | \end{aligned} \tag{2.6}$$

within a negligible margin. If there is enough space for only one array of hexagons, the centers of these hexagons are uniformly spaced. Condition 4 excludes partitions with subregions having elongated sides, and is satisfied, for example, by squares, circles, rectangles whose sides are not too long or too short, etc.

Design of the Discretizer. Consider the discretizer $Q_{N',g}(\cdot)$ with representative points $\bigcup_{k=1}^{m(N)} T_{k,N}$, each of which represents a corresponding hexagon, where the second subscript indicates the dependence of the discretizer on the density function g. For the values of X falling in the boundary areas $\Pi_{k,N} \setminus H_{k,N}$, choose the closest points in $T_{k,N}$ as their representative points. This discretizer has a total number of $N' = n_{1,N} + \cdots + n_{m(N),N} \leq N$ representative points, which is generally smaller than N. In fact, by Condition 3,

$$N - N' = \sum_{k=1}^{m(N)} \{N \int_{\Pi_{k,N}} g(x)dx - \lceil N \int_{\Pi_{k,N}} g(x)dx \rceil\} \leq m(N). \tag{2.7}$$

To obtain a discretizer $Q_{N,g}(\cdot)$ with N representative points, one can select $N-N'$ additional points in the boundary region $\bigcup_{k=1}^{m(N)}(\Pi_{k,N} \setminus H_{k,N})$, appropriately, and have them represent their nearest neighborhood areas in the boundary region.

In view of the partition of D and the design of discretizer above, it is clear that in each region $H_{k,N}$ we employ regular hexagonal partition with the size or the number of hexagons determined by the density function g. The hexagons are the nearest neighborhoods of the representative points (centers of hexagons). Thus, they form a Voronoi partition with exceptions near the boundary. There is some flexibility in the boundary regions. The discretizers $Q_{N',g}(\cdot)$ and $Q_{N,g}(\cdot)$ actually provide two different ways to deal with the boundary regions. It is shown that $Q_{N',g}(\cdot)$ is asymptotically optimal, even though it generally uses fewer points than $Q_{N,g}(\cdot)$. Since, clearly, $e^2(Q_{N,g}(\cdot), p) \leq e^2(Q_{N',g}(\cdot), p)$, the results for $Q_{N',g}(\cdot)$ also hold for $Q_{N,g}(\cdot)$ and are stated in the following.

Theorem. *Under Conditions* 1-4, *the following are true.*

i) *If the density functions* $p(x)$ *and* $g(x)$ *are continuous and the function* $p(x)/g(x)$ *is Riemann integrable, then*

$$\lim_{N\to\infty} Ne^2(Q_{N,g}(\cdot), p) = \frac{5\sqrt{3}}{54}\int_D \frac{p(x)}{g(x)}dx. \tag{2.8}$$

ii) *If* $p(x)$ *is continuous and* $p^{1/2}(x)$ *is Riemann integrable, then the sequence* $\{Q_{N,g^o}(\cdot)\}$*, where*

$$g^o(x) = p^{1/2}(x)/\int_D p^{1/2}(u)du, \tag{2.9}$$

is asymptotically optimal.

Remarks. It is clear that when the random vector X is uniformly distributed over a bounded region D, g^o is uniform and thus an asymptotically optimal discretizer corresponds to the regular hexagonal tessellation of D plus some boundary pieces. However, it is also not clear how to choose the number $m(N)$ for small N, though asymptotically the only requirement is that $m(N)/N \to 0$. For an unbounded region D and a small size N of discretization, the errors come from two different sources: the error of truncation of D and the discretizing error. It is also not clear how to truncate D in the best possible way for a small N. Asymptotically, the truncation error of D is negligible. The density function $g^o(x)$ in (2.9) is called the asymptotically optimal discretizing density and $Q_{N,g^o}(\cdot)$ achieves the asymptotically optimal constant $5\sqrt{3}\{\int_D \sqrt{p(x)}dx\}^2/54$.

Next we consider two spherically symmetric distributions, Gaussian and Pearson Type VII, and a diagonally symmetric distribution, the Laplacian. All three examples have an unbounded region $D = R^2$. We truncate D by making $N\int_{D\setminus D_N} \|x\|^2 p(x)dx$ or $Prob(X \in D \setminus D_N) = \int_{D\setminus D_N} p(x)dx$ relatively small.

For each of these cases, we take $N = 8, 16, 32, 64$ and 128 and run 15,000 simulations. We partition the plane into circularly symmetric strips for Gaussian and Pearson Type VII distributions, and diagonally symmetric strips for the Laplacian distribution, for convenience of computations. The numbers of points in these strips are taken to be the numbers of hexagons one would conveniently place in the corresponding regions. These numbers of points then in turn determine the sizes of the strips through the density $g^o(x)$. The *mse* values of the discretizer $Q_{N,g^o}(\cdot)$ are compared with those of the Monte Carlo method (Fischer and Dicharry (1984)), those of Fang and Wang (1994), and the asymptotically optimal mean square error, namely, the *asymptotically optimal constant* $\div N$. We found that for $N = 8, 16$ and 32, $Q_{N,g^o}(\cdot)$ can achieve a similar performance as that of the Monte Carlo algorithm and that of Fang and Wang's method, and for $N = 64$ and 128, $Q_{N,g^o}(\cdot)$ matches the asymptotically optimal performance, as illustrated in the following. Note the comparison of our results with those of

Fang and Wang are more reliable than with those of the Monte Carlo method, due to the error of the Monte Carlo simulation.

Example 1. *Gaussian distribution.* Here, the density function of the random vector is

$$p(x) = (2\pi\sigma^2)^{-1} \exp\{-\|x\|^2/(2\sigma^2)\}, \quad x \in R^2,$$

and the asymptotically optimal discretizing density is

$$g^o(x) = (4\pi\sigma^2)^{-1} \exp\{-\|x\|^2/(4\sigma^2)\}, \quad x \in R^2.$$

For this distribution, we partition R^2 into $m+1$ circular disks centered at the origin. Their radii $0 = r_0 < r_1 < \cdots < r_m < r_{m+1} = \infty$ are determined by

$$\int_{r_{i-1}<\|x\|<r_i} g^o(x)dx = a_i, \quad i = 1, \ldots, m+1,$$

where $a_i > 0$ and $\sum_{i=1}^{m+1} a_i = 1$, which yields

$$r_i = 2\sigma\{-\log(1 - \sum_{k=1}^{i} a_k)\}^{1/2}, \quad i = 1, \ldots, m+1. \tag{2.10}$$

We work with the finite region $D_N = \{x : \|x\| \le r_m\}$, where r_m is determined by properly choosing the value of a_{m+1}. It is straightforward to verify that

$$N \int_{D\setminus D_N} \|x\|^2 p(x)dx = 2\sigma^2 N(1 - 2\log(a_{m+1}))a_{m+1}^2.$$

We choose a number of discretizing sizes, $N = 8, 16, 32, 64, 128$, and for convenience, take $\sigma = 1$; then the corresponding asymptotically optimal constant is 4.030665.

When $N = 8$, we choose $m = 1$. The eight points are the centers of eight regular hexagons and are located symmetrically about the origin in the circle $\|x\| < r_1$, as shown in Figure 1 (a). The radius r_1 is determined by (2.10) with $a_1 = 1 - a_2$. The radius d of each hexagon satisfies the following compatibility condition

$$\begin{aligned} &\textit{number of points} \times \textit{area of of each hexagon} \\ &= \textit{number of points} \times \tfrac{3\sqrt{3}}{2} d^2 \approx \textit{area of the circle } \|x\| < r_1 \ (= \pi r_1^2), \end{aligned}$$

namely, $d \approx .38878 r_1$. We tried a number of values for a_2 and found that a small *mse* is obtained if $a_2 = .32$. The corresponding *mse* is .388509 while the Monte Carlo algorithm gives a value .406 and Fang and Wang's method gives .400536.

When $N = 16$, we take $m = 2$ and put four points in the first circle $\|x\| < r_1$ and twelve points along the edge of a hexagon in the disk $\|x\| > r_1$. The radius r_1 is as in (2.10) with $a_1 = 4/16 = .25$. We tried a number of values for a_3

and found that $a_3 = .31$ gives a relatively small mse ($a_2 = .44$). The design of representative points is shown in Figure 1 (b) and has $mse = .226$. This is about one percent higher than the .216 of the Monte Carlo algorithm and the .213588 of Fang and Wang's method. We found that for this design of discretizer, if $a_3 \in [.37, .5]$, the corresponding mse is higher than .226 by about 8×10^{-3}.

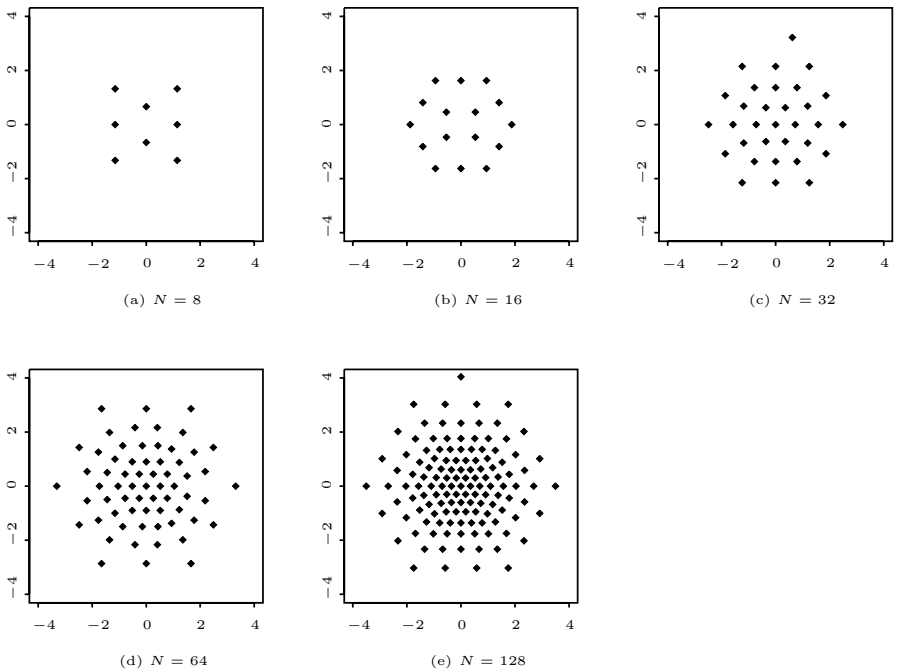

Figure 1.

When $N = 32$ we select $m = 3$, with 7 points in the circle $\|x\| < r_1$, 12 in the disk $r_1 < \|x\| < r_2$ and 13 in the disk $\|x\| > r_2$, where $r_i, i = 1, 2$ are determined by (2.10) with $a_1 = 7/32 = .21875$ and $a_2 = 12/32 = .375$. We choose $a_4 = .128$ ($a_3 = 1 - a_1 - a_2 - a_4 = .27825$) which gives a smaller value of mse than other choices. Those 7 points in the circle $\|x\| < r_1$ are at the centers of 7 regular hexagons; each of them has a radius $d_1 = .4156r_1$. The 12 points in the disk $r_1 < \|x\| < r_2$ and 12 of the points in $\|x\| > r_2$ are placed along the edge of a hexagon, and the last point in $\|x\| > r_2$ is in the first quadrant, as shown in Figure 1 (c). The corresponding value of mse is .117405 compared with .1138 for the Monte Carlo algorithm, .113112 for Fang and Wang's method and the asymptotically optimal $mse = .1259583$. For this design of discretizer, we also tried other values of a_4 and found that when $a_4 \in [.121, .134]$ the corresponding values of mse are higher than .1174 by a margin of 9×10^{-5}.

As shown in Figure 1 (d), when $N = 64$ we take $m = 4$, with 19 points in the circle $\|x\| < r_1$, 17 in the disk $r_1 < \|x\| < r_2$, 16 in $r_2 < \|x\| < r_3$ and finally 12 in the region $\|x\| > r_3$, where $r_i, i = 1, 2, 3$ are as in (2.10) with $a_1 = 19/64 = .296875$, $a_2 = 17/64 = .265625$ and $a_3 = 16/64 = .25$. We take a_5 to be .028 and thus $a_4 = 1 - a_1 - a_2 - a_3 - a_5 = .1595$. The 19 points in the circle $\|x\| < r_1$ are at the centers of 19 regular hexagons, each of which has a radius $d_1 = .252r_1$. The rest of the points are placed along the edge of a hexagon in the corresponding disks. The mse is .062686 while the asymptotically optimal mse = .06297914. For this design of representative points, we tried other values for a_5 and found that when $a_5 \in [.018, .046]$, the corresponding values of mse are higher than .062686 by a margin of 4×10^{-4}.

Table 1. Values of mean square errors for the Gaussian distribution.

Sizes N	8	16	32	64	128
Monte Carlo method	.406	.216	.1138	$\sim$	$\sim$
Fang and Wang's method	.4005	.2136	.1131	$\sim$	$\sim$
Asymp. optimal const. $\div N$	.5038	.252	.1259	.062979	.03149
$Q_{N,g^o}(\cdot)$	.3886	.226	.1174	.062686	.03143

When $N = 128$, we use $m = 6$, with 19 points in the circle $\|x\| < r_1$, 18 in the disk $r_1 < \|x\| < r_2$, 24 in each of the disks $r_i < \|x\| < r_{i+1}, i = 2, 3, 4$, 18 in $r_5 < \|x\| < r_6$, and finally one point in the region $\|x\| > r_6$, as shown in Figure 1 (e), where $r_i, i = 1, \ldots, 6$ satisfy (2.10) with $a_1 = 19/128 = .1484375$, $a_2 = 18/128 = .140625$, $a_3 = a_4 = a_5 = 24/128 = .1875$. Here, we use $a_7 = .031$ ($a_6 = .1174375$). The 19 points are the centers of 19 regular hexagons with radius $d_1 = .252r_1$. All other points are located along the edge of a hexagon in the corresponding regions. The mse for this design is .031429 compared with the value .03149 for the asymptotically optimal mse. For this discretizer, we found that if the value of a_7 is in $[.024, .039]$, the corresponding values of mse are higher than .031429 by a margin of 9×10^{-5}. The available values of mean square errors for all methods referred to above are in Table 1, with $\sim$ indicating "not available".

Example 2. *Pearson Type VII distribution.* Here we consider the density function

$$p(x) = \frac{\Gamma(\lambda)}{\Gamma(\lambda - s/2)}(\pi\mu)^{-s/2}(1 + \|x\|^2/\mu)^{-\lambda}, \quad x \in R^2.$$

This was treated by Fang and Wang (1994) for $s = 2$, $\mu = 4$ and $\lambda = 15$. For these parameter values, the asymptotically optimal discretizing density is

$$g^o(x) = \frac{13}{8\pi}(1 + \|x\|^2/4)^{-15/2}, \quad x \in R^2,$$

and the asymptotically optimal constant is .6678025.

We choose a number of discretizing sizes, $N = 8, 16, 32, 64$ and 128, and partition R^2 into $m+1$ disks centered at origin. Their radii $0 = r_0 < r_1 < \cdots < r_{m+1} = \infty$ are determined by

$$\int_{r_{i-1}<\|x\|<r_i} g^o(x)dx = a_i, \quad i = 1, \ldots, m+1,$$

where $a_i > 0$ and $\sum_{i=1}^{m+1} a_i = 1$, which means

$$r_i = 2\{(1 - \sum_{k=1}^{i} a_k)^{-2/13} - 1\}^{1/2}, \quad i = 1, \ldots, m+1. \tag{2.11}$$

Since this distribution is also circularly symmetric about the origin, though it does not have independent components, we design the discretizing points as in Example 1 for the sizes $N = 8, 16, 32, 64$ and 128, as shown in Figure 2 (a)-(e), respectively. Here we truncate the infinite region R^2 by choosing a relatively small a_{m+1}.

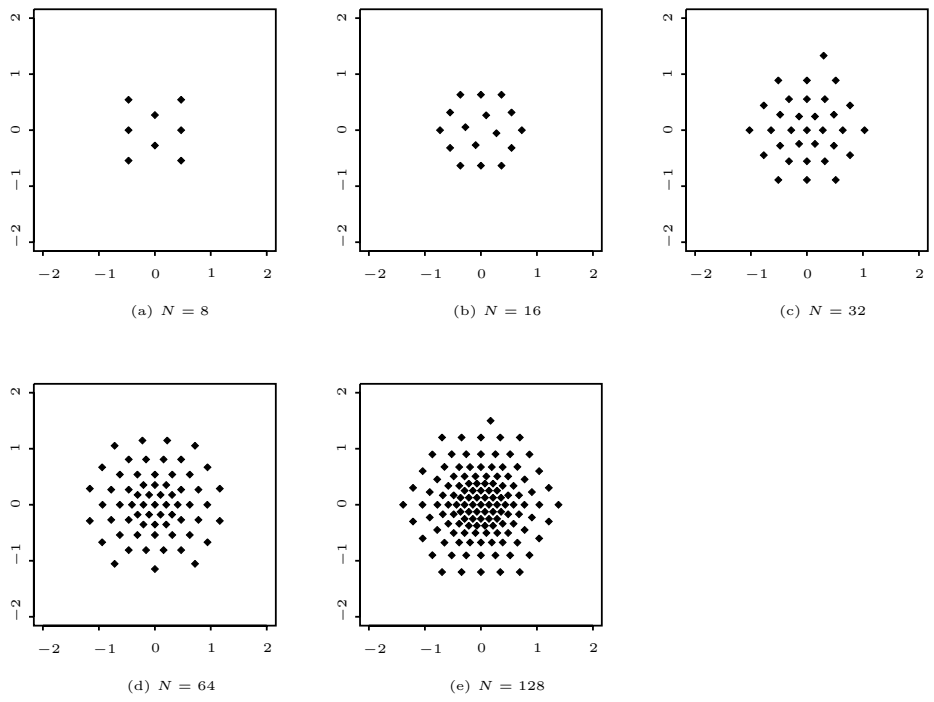

Figure 2.

When $N = 8$ we choose $m = 1$. The radius r_1 is determined by (2.11) with $a_1 = 1 - a_2$. The eight points in the circle $\|x\| < r_1$ are the centers of eight regular hexagons which are located symmetrically about the origin, as shown in Figure 2 (a). The radius d_1 of each hexagon is $d_1 = .4156r_1$. We tried a number of values for a_2 and found that a small mse is obtained if $a_2 = .4$. The mean square error is found to be .062596 while Fang and Wang's method gives a value of .064588.

When $N = 16$ we take $m = 2$. The radius r_1 is as in (2.11) with $a_1 = 4/16 = .25$. We tried a number of values for a_3 ($a_2 = 1 - 4/16 - a_3$) and found that a good choice is $a_3 = .35$ or $a_2 = .4$. The 4 points in the circle $\|x\| < r_1$ are placed at the centers of 4 regular hexagons, and 12 in the second region $r_1 < \|x\| < r_2$ are placed along the edge of a hexagon. The design of discretizer has a value of $mse = .037053$ compared with .035172 for Fang and Wang's method. We also found that if a_3 falls in $[.26, .42]$, the corresponding values of mse are higher than .37053 by about 2×10^{-3}.

When $N = 32$ we select $m = 3$, with 7 points in the circle $\|x\| < r_1$ and 12 in the disk $r_1 < \|x\| < r_2$, where $r_i, i = 1, 2$ are determined by (2.11) with $a_1 = 7/32 = .21875$ and $a_2 = 12/32 = .375$. The remaining 13 points are placed in the disk $\|x\| > r_2$, with 12 on the edge of a hexagon and one in the first quadrant. We choose $a_4 = .145$ or $a_3 = 1 - 7/32 - 12/32 - a_4 = .17625$, which gives a smaller value of mse than other choices. The corresponding value of mse is .018995 compared with .018572 for Fang and Wang's method. For other choices of a_4, we found that if a_4 is chosen from the interval $[.133, .158]$, the corresponding values of mse are higher than .018995 by about 5×10^{-5}.

Table 2. Values of mean square errors for the Pearson Type VII distribution.

Sizes N	8	16	32	64	128
Fang and Wang's method	.0616	.0352	.0186	$\sim$	$\sim$
Asymp. optimal const. $\div N$	.0835	.0417	.0209	.0104	.005217
$Q_{N,g^o}(\cdot)$	.0626	.03705	.0189	.0102	.005174

When $N = 64$ we take a relatively small $m = 3$, with 19 points in the circle $\|x\| < r_1$, 30 in the disk $r_1 < \|x\| < r_2$, and 15 in $\|x\| > r_2$, where $r_i, i = 1, 2$ are as in (2.11) with $a_1 = 19/64 = .296875$ and $a_2 = 30/64 = .46875$. The 19 points in the circle $\|x\| < r_1$ are at the centers of 19 regular hexagons, each of which has a radius $d_1 = .252r_1$, the 30 points in the second region are placed along the edges of two hexagons and the 15 points in the third region are along the edge of a hexagon, as shown in Figure 2 (d). We use $a_4 = .08$ ($a_3 = .1543375$). The corresponding mse is .010218 while the asymptotically optimal $mse = .010434$.

We found that if $a_4 \in [.07, .09]$, the mse is higher than .010218 by a margin of 4×10^{-4}.

As shown in Figure 2 (e), when $N = 128$ we use $m = 5$, with 37 points in the circle $\|x\| < r_1$, 42 in the disk $r_1 < \|x\| < r_2$, 24 in $r_2 < \|x\| < r_3$ and 25 in the region $\|x\| > r_3$, where $r_i, i = 1, 2, 3$ satisfy (2.11) with $a_1 = 37/128 = .2890625$, $a_2 = 42/128 = .328125$ and $a_3 = 24/128 = .1875$. The 37 points are the centers of 37 regular hexagons with radius $d_1 = .18r_1$. The 42 points in the second region are located along the edges of two hexagons and the remaining points are also along the edges of hexagons in the corresponding regions. Here, we use $a_5 = .07$ ($a_4 = .1207242$). The mse for this design is .005174 compared with the value .005217 for the asymptotically optimal mse. We found that if $a_5 \in [.064, .087]$, the values of mse are higher than .005174 by about 3×10^{-5}. The available values of mean square errors for the methods referred to above are listed in Table 2, with $\sim$ indicating "not available".

Example 3. *Laplacian distribution.* In this example, the random vector is assumed to have a density

$$p(x) = \frac{\lambda\mu}{4} \exp\{-\lambda \mid x_1 \mid -\mu \mid x_2 \mid)\}, \ x \in R^2.$$

The corresponding optimal discretizing density is

$$g^o(x) = \frac{\lambda\mu}{16} \exp\{-\frac{1}{2}(\lambda \mid x_1 \mid +\mu \mid x_2 \mid)\}, \ x \in R^2,$$

and the asymptotically optimal constant is $160/(9\sqrt{3}\lambda\mu)$.

This distribution with $\lambda = \mu = \sqrt{2}$ is treated in Fischer and Dicharry (1984) by use of the Monte Carlo method. Here, we also take $\lambda = \mu = \sqrt{2}$, which yields the value 5.132032 for the asymptotically optimal constant. We select a number of discretizing sizes, $N = 8, 16, 32, 64$ and 128. In a similar fashion to the previous example, for a fixed size N, we partition R^2 into $m + 1$ diamond shaped disks $\rho_{i-1} < \sqrt{2}(\mid x_1 \mid + \mid x_2 \mid) < \rho_i, i = 1, \ldots, m + 1$, centered at the origin. Their radii $0 = \rho_0 < \rho_1 < \cdots < \rho_{m+1} = \infty$ are determined by

$$\int_{\rho_{i-1} < \sqrt{2}(|x_1| + |x_2|) < \rho_i} g^o(x)dx = a_i, \ \ i = 1, \ldots, m + 1,$$

where $a_i > 0$ and satisfy $\sum_{i=1}^{m+1} a_i = 1$. After some calculations, we find

$$1 - (1 + \rho_i/2) \exp\{-\rho_i/2\} = \sum_{k=1}^{i} a_k, \ \ i = 1, \ldots, m + 1. \tag{2.12}$$

When $N = 8$, we choose $m = 1$, and the radius ρ_1 is determined by (2.12) with $a_1 = 1 - a_2$. The $N = 8$ points in the region $\sqrt{2}(|\ x_1\ | + |\ x_2\ |) < \rho_1$ are the centers of eight regular hexagons and are located symmetrically about the origin, as shown in Figure 3 (a). The radius d_1 of each hexagon is $d_1 = .219\rho_1$ which ensures the following compatibility condition

$$\textit{number of points} \times \textit{area of each hexagon} = \textit{number of points} \times (3\sqrt{3}d_1^2/2)$$

$$\approx \textit{area of the diamond } \sqrt{2}(|\ x_1\ | + |\ x_2\ |) \le \rho_1\ (= \rho_1^2).$$

For this design, we tried a number of values of a_1 and found that a small mean square error $mse = .454649$ is achieved at $a_1 = .58$, or $a_2 = .42$, while the Monte Carlo algorithm gives the value .47.

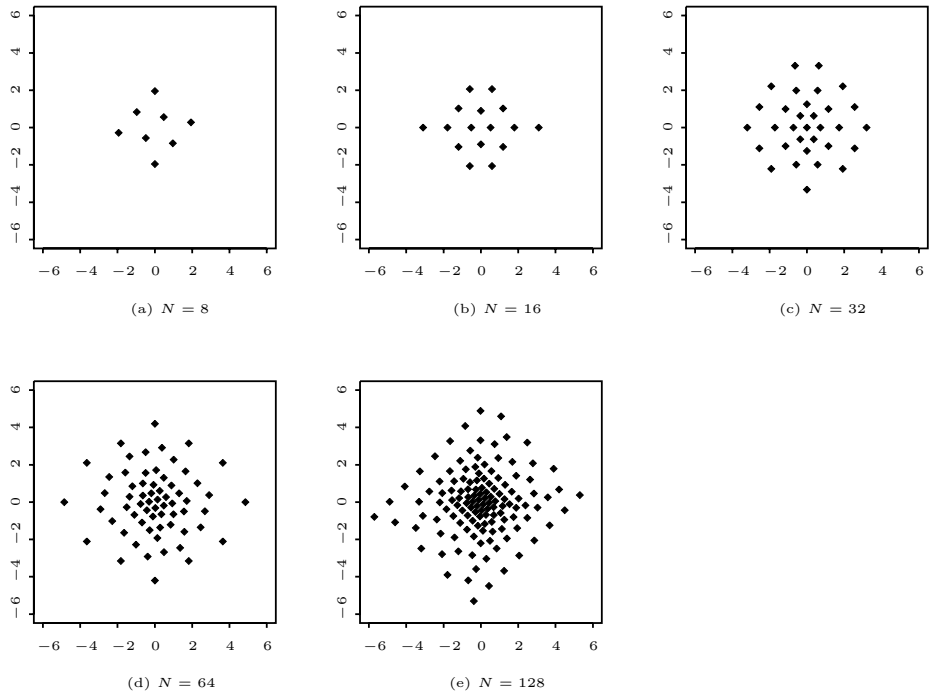

Figure 3.

When $N = 16$ we take $m = 2$, with 4 points in the region $\sqrt{2}(|\ x_1\ | + |\ x_2\ |)$ $< \rho_1$ and being the centers of the regular hexagons with $d_1 = .31\rho_1$. The ρ_1 is determined by (2.12) with $a_1 = 4/16 = .25$. Ten of the 12 remaining points are placed along the next diamond disk and two of them are placed on the horizontal

axis, as shown in Figure 3 (b). The a_2 is chosen to be .4 and the corresponding $mse = .259273$ compares with .264 for the Monte Carlo algorithm.

Table 3. Values of mean square errors for the Laplacian distribution.

Sizes N	8	16	32	64	128
Monte Carlo method	.47	.264	.1434	$\sim$	$\sim$
Asymp. optimal const. $\div N$	.6415	.3208	.1604	.0802	.040094
$Q_{N,g^o}(\cdot)$	.4547	.2593	.1503	.0806	.040716

When $N = 32$ we select $m = 3$, with 9 points in the region $\sqrt{2}(\mid x_1 \mid + \mid x_2 \mid) < \rho_1$, 10 in the disk $\rho_1 < \sqrt{2}(\mid x_1 \mid + \mid x_2 \mid) < \rho_2$ and 13 in $\sqrt{2}(\mid x_1 \mid + \mid x_2 \mid) > \rho_2$, where $\rho_i, i = 1, 2$ are determined by (2.12) with $a_1 = 9/32 = .28125$ and $a_2 = 10/32 = .3125$. Those 9 points in the first region are the centers of 9 regular hexagons, each of which has a radius $d_1 = .2\rho_1$. The 10 points in the second region correspond to the centers of 10 hexagons, as shown in Figure 3 (c). When a_4 is .164, the value of mse is .150285 compared with .1434 for the Monte Carlo algorithm. Here, the asymptotically optimal mse is .160375. We tried other values for a_4 and found that when $a_4 \in [.158, .17]$, the corresponding values of mse are above .150285 by about 7×10^{-5}.

When $N = 64$ we take $m = 4$ with 14 points in the region $\sqrt{2}(|x_1|+|x_2|) < \rho_1$, 18 in the band $\rho_1 < \sqrt{2}(|x_1|+|x_2|) < \rho_2$, and 20 in $\rho_2 < \sqrt{2}(\mid x_1 \mid + \mid x_2 \mid) < \rho_3$, where $\rho_i, i = 1, 2, 3$ are as in (2.12) with $a_1 = 14/64 = .21875$, $a_2 = 18/64 = .28125$ and $a_3 = .3125$. The 12 remaining points are placed approximately along a hexagon in the region $\sqrt{2}(\mid x_1 \mid + \mid x_2 \mid) > \rho_3$ with two points on the horizontal axis. The 14 points in the first region correspond to the centers of 14 hexagons in a square, after rotation by $\pi/4$ to form a diamond. The next 18 points are rotated (by $\pi/4$) versions of centers of 18 hexagons along the edge of a square. The 20 points in the third region are chosen similarly, as shown in Figure 3 (d). We take $a_5 = .062$. The corresponding mse is .080581 while the asymptotically optimal $mse = .080188$. We found that when $a_5 \in [.055, .069]$, the corresponding values of mse are larger than .080581 by about 2×10^{-4}.

As shown in Figure 3 (e), when $N = 128$, we use $m = 5$ with 22 points in the first region $\sqrt{2}(\mid x_1 \mid + \mid x_2 \mid) < \rho_1$, 24 in the diamond disk $\rho_1 < \sqrt{2}(\mid x_1 \mid + \mid x_2 \mid) < \rho_2$, 27 in $\rho_2 < \sqrt{2}(\mid x_1 \mid + \mid x_2 \mid) < \rho_3$, 27 in $\rho_3 < \sqrt{2}(\mid x_1 \mid + \mid x_2 \mid) < \rho_4$, and 28 in $\rho_4 < \sqrt{2}(\mid x_1 \mid + \mid x_2 \mid) < \rho_5$, where $\rho_i, i = 1, \ldots, 5$, satisfy (2.12) with $a_1 = 22/128 = .171875$, $a_2 = 24/128 = .1875$, $a_3 = 27/128 = .2109375$ and $a_4 = 27/128 = .2109375$. The points in the corresponding region are formed in a similar way as in the case of $N = 64$. When a_6 ($a_5 = .14375$) is chosen to be .075, the mse for this design is .040716, compared with the value .040094 for the asymptotically optimal mse. We found that when $a_6 \in [.068, .086]$, the values of

mse are larger than .040716 by about 8×10^{-5}. The available values of mean square errors for all methods referred to above are shown in Table 3.

Comments. The procedure of designing representative points presented in the paper is asymptotically optimal for every bivariate random vector with a finite second moment. The main idea is to take the centers of regular hexagons in a region as representative points, where the sizes of these regular hexagons are determined by a properly selected density. But for generalization to higher dimensions, other methods such as the number-theoretic and Monte Carlo methods seem more flexible.

3. Proof of the Theorem

Since $e^2(Q_{N',g}(\cdot), p) \geq e^2(Q_{N,g}(\cdot), p)$, it suffices to prove the theorem for $Q_{N',g}(\cdot)$.

The mean square error of $Q_{N',g}(\cdot)$ can be written as

$$e^2(Q_{N',g}(\cdot), p) = \sum_{k=1}^{m(N)} \int_{\Pi_{k,N}} \|x - Q_{N'}(x)\|^2 p(x) dx.$$

By the Mean Value Theorem of integrals, we can pull out the density function $p(x)$ from the integral, and then split $\Pi_{k,N}$ into $H_{k,N}$ and $\Pi_{k,N} \setminus H_{k,N}$, namely,

$$\begin{aligned} e^2(Q_{N',g}(\cdot), p) &= \sum_{k=1}^{m(N)} p(x_k^*) \int_{H_{k,N}} \|x - Q_{N'}(x)\|^2 dx \\ &+ \sum_{k=1}^{m(N)} p(x_k^*) \int_{\Pi_{k,N} \setminus H_{k,N}} \|x - Q_{N'}(x)\|^2 dx \equiv \delta_N^2 + \epsilon_N^2 \end{aligned} \tag{3.1}$$

where $x_k^* \in \Pi_{k,N}$.

Use the fact that the integral of $\|x\|^2$ over a regular hexagon centered at the origin with radius d equals $(5\sqrt{3}/8)d^4$. Then $\int_{\Pi_{k,N}} g(x)dx = \mid \Pi_{k,N} \mid g(x_k^{**})$, where $x_k^{**} \in \Pi_{k,N}$, implies

$$\begin{aligned} \delta_N^2 &= \sum_{k=1}^{m(N)} p(x_k^*) n_{k,N} \frac{5\sqrt{3}}{8} d_{k,N}^4 \\ &= \frac{5\sqrt{3}}{8} \sum_{k=1}^{m(N)} p(x_k^*) n_{k,N} \{\frac{2}{3\sqrt{3}} \mid \Pi_{k,N} \mid /(N \int_{\Pi_{k,N}} g)\}^2 \\ &= \frac{5\sqrt{3}}{54N} \sum_{k=1}^{m(N)} \frac{p(x_k^*)}{g(x_k^{**})} \mid \Pi_{k,N} \mid \frac{n_{k,N}}{N \int_{\Pi_{k,N}} g} \end{aligned}$$

$$= \frac{5\sqrt{3}}{54N} \sum_{k=1}^{m(N)} \frac{p(x_k^*)}{g(x_k^{**})} \mid \Pi_{k,N} \mid \{1 - \frac{N \int_{\Pi_{k,N}} g - \lceil N \int_{\Pi_{k,N}} g \rceil}{N \int_{\Pi_{k,N}} g}\}$$

$$= \frac{5\sqrt{3}}{54N} \{ \sum_{k=1}^{m(N)} \frac{p(x_k^*)}{g(x_k^{**})} \mid \Pi_{k,N} \mid - \sum_{k=1}^{m(N)} \frac{p(x_k^*)}{g(x_k^{**})} \mid \Pi_{k,N} \mid \frac{N \int_{\Pi_{k,N}} g - \lceil N \int_{\Pi_{k,N}} g \rceil}{N \int_{\Pi_{k,N}} g}\}$$

$$\equiv \delta_{1,N}^2 - \delta_{2,N}^2. \tag{3.2}$$

Riemann integration of the function p/g yields

$$\lim_{N\to\infty} N\delta_{1,N}^2 = \frac{5\sqrt{3}}{54} \int_D \frac{p(x)}{g(x)} dx. \tag{3.3}$$

Using Condition 2 and noticing that $0 \le N \int_{\Pi_{k,N}} g - \lceil N \int_{\Pi_{k,N}} g \rceil \le 1$, we have

$$N\delta_{2,N}^2 \le \frac{5\sqrt{3}}{54c} \frac{m(N)}{N} \sum_{k=1}^{m(N)} \frac{p(x_k^*)}{g(x_k^{**})} \mid \Pi_{k,N} \mid .$$

Thus

$$\limsup_{N\to\infty} N\delta_{2,N}^2 \le \frac{5\sqrt{3}}{54c} \int_D \frac{p}{g}(x) dx \lim_{N\to\infty} \frac{m(N)}{N} = 0. \tag{3.4}$$

For the second term ϵ_N^2 in (3.1), observe that the area of each hexagon in $H_{k,N}$ is $(3\sqrt{3}/2)d_{k,N}^2$ and, by (2.5),

$$\mid H_{k,N} \mid = n_{k,N} \frac{3\sqrt{3}}{2} d_{k,N}^2 = \frac{\lceil N \int_{\Pi_{k,N}} g \rceil}{N \int_{\Pi_{k,N}} g} \mid \Pi_{k,N} \mid .$$

From Conditions 4 and 2, we obtain

$$\epsilon_N^2 \le a^2 \sum_{k=1}^{m(N)} p(x_k^*) d_{k,N}^2 \mid \Pi_{k,N} \setminus H_{k,N} \mid$$

$$= \frac{2a^2}{3\sqrt{3}} \sum_{k=1}^{m(N)} p(x_k^*) \frac{\mid \Pi_{k,N} \mid}{N \int_{\Pi_{k,N}} g} (\mid \Pi_{k,N} \mid - \frac{\lceil N \int_{\Pi_{k,N}} g \rceil}{N \int_{\Pi_{k,N}} g} \mid \Pi_{k,N} \mid)$$

$$= \frac{2a^2}{3\sqrt{3}} \frac{1}{N} \sum_{k=1}^{m(N)} \frac{p(x_k^*)}{g(x_k^{**})} \mid \Pi_{k,N} \mid \frac{(N \int_{\Pi_{k,N}} g - \lceil N \int_{\Pi_{k,N}} g \rceil)}{N \int_{\Pi_{k,N}} g}$$

$$= const.\ \delta_{2,N}^2, \tag{3.5}$$

which, together with (3.4), yields

$$\limsup_{N\to\infty} N\epsilon_N^2 = 0. \tag{3.6}$$

Result i) follows from (3.1)–(3.6), and result ii) follows from (2.4) and result i).

Acknowledgements

This paper is dedicated to the late Professor Stamatis Cambanis. The work presented in the paper reflects many early ideas and suggestions of Professor Cambanis. The author would also like to thank an anonymous referee, an Associate Editor, and the present Editor (Professor Ching-Shui Cheng) for their careful reading of the manuscript and suggestions that helped to improve the presentation of the paper. This work is supported in part by 1998 Summer Faculty Fellowship of Southwest Missouri State University.

References

Cambanis, S. and Gerr, N. L. (1983). A simple class of asymptotically optimal quantizers. *IEEE Trans. Inform. Theory* **IT-29**, 664-676.

Fang, K.-T. and Wang, Y. (1994). Number-theoretic methods in statistics. *Monographs Statist. Appl. Probab.* **51**. Chapman & Hall, London.

Fang, K.-T., Yuan, K.-H. and Bentler, P. M. (1994). Applications of number-theoretic methods to quantizers of elliptically contoured distributions. In *Multivariate Analysis and Its Applications*, 237-251. IMS Lecture Notes-Monograph Series.

Fischer, T. R. and Dicharry, R. M. (1984). Vector quantizer design for memoryless Gaussian, Gamma, and Laplacian Sources. *IEEE Trans. Commun.* **COM-32**, 1065-1069.

Flury, B. A. (1990). Principal points. *Biometrika* **77**, 33-41.

Hartigan, J. A. (1975). *Clustering Algorithms*. Wiley, New York.

Newmann, D. J. (1982). The hexagon theorem. *IEEE Trans. Inform. Theory* **IT-28**, 137-139.

Tarpey, T. (1994). Two principal points of symmetric, strongly unimodal distributions. *Statist. Probab. Lett.* **20**, 253-257.

Tarpey, T., Li, L. and Flury, B. D. (1995). Principal points and self-consistent points of elliptical distributions. *Ann. Statist.* **23**, 103-112.

Zador, P. L. (1982). Asymptotically quantization error of continuous signals and the quantization dimension. *IEEE Trans. Inform. Theory* **IT-28**, 139-149.

Department of Mathematics, Southwest Missouri State University, 901 S. National Ave., Springfield, MO 65804, U.S.A.

E-mail: yis780f@mail.smsu.edu

(Received May 1998; accepted June 1999)