THE MEAN, MEDIAN AND MODE OF UNIMODAL DISTRIBUTIONS: A CHARACTERIZATION

by

Sanjib Basu and Anirban DasGupta*
University of California, Santa Barbara
Purdue University

Technical Report #92-40

Department of Statistics
Purdue University

September 1992

*Research supported by NSF grant DMS 89-230-71 at Purdue University.

THE MEAN, MEDIAN, AND MODE OF UNIMODAL DISTRIBUTIONS : A CHARACTERIZATION

by

Sanjib Basu

University of California, Santa Barbara and

Anirban DasGupta

Purdue University

August 1992

Abstract

For a unimodal distribution on the Real line, the celebrated *mean-median-mode inequality* states that they often occur in an alphabetical (or its reverse) order. Various sufficient conditions for the validity of the inequality are known. This article explicitly characterizes the three dimensional set of means, medians, and modes of unimodal distributions. It is found that the set is pathwise connected but not convex. Some fundamental inequalities among the mean, the median and mode of unimodal distributions are also derived. These inequalities are used : (*i*) to prove nonunimodality of certain distributions, and (*ii*) for obtaining bounds on the median of a unimodal distribution. In a multivariate setting, the generalized notion of *α-unimodality* is used, and characterizations are given for the set of mean vectors, when the mode is fixed, or when it varies in a sphere. In particular, it is found that the set of mean vectors for generalized unimodal distributions with a specified mode and covariance matrix is an exact ellipsoid and this ellipsoid is explicitly described.

Key words : α-unimodality, connected, convex, ellipsoid, mean, mean-median-mode inequality, median, mode, moment problem, sphere, star unimodality, uniform distribution, unimodality.

AMS 1991 subject classifications. 60E05, 62E10, 60E15, 62H05

1 Introduction

Let X be a real valued random variable with cumulative distribution function (cdf) F, and *mean* $\boldsymbol{\mu} = \boldsymbol{\mu}[F] = E_F(X) < \infty$. A number $\mathbf{m} = \mathbf{m}[F]$ is said to be a *median* of X if $P(X \leq \mathbf{m}) \geq \frac{1}{2}$ and $P(X \geq \mathbf{m}) \geq \frac{1}{2}$. A median $\mathbf{m}$, thus defined, always exists, although in general, a random variable X may have several medians. Suppose, furthermore, that X is unimodal about some point $\mathbf{M} = \mathbf{M}[F]$ (called a *mode* of X), i.e., $F(x)$ is convex for $x \in (-\infty, \mathbf{M})$, and concave for $x \in (\mathbf{M}, \infty)$. It is easy to see that a median $\mathbf{m}$ is uniquely defined for a unimodal random variable X.

The well known *mean-median-mode inequality* states that for a unimodal distribution F, often, the mean, median, and mode occur in an alphabetical or reverse alphabetical order, i.e.,

$$\mathbf{M} \leq \mathbf{m} \leq \boldsymbol{\mu} \qquad \text{or} \qquad \mathbf{M} \geq \mathbf{m} \geq \boldsymbol{\mu}. \tag{1}$$

It is well known that this inequality, however, is not always true as can be seen from the following very simple example.

Example 1. Let $X \sim F$, where F is a mixture of three distributions : $F(x) = (\frac{1-\delta}{4})F_1(x) + (\frac{3-\delta}{4})F_2(x) + \frac{\delta}{2}F_3(x)$, where $0 < \delta < 1$, and F_1, F_2, and F_3 are, respectively, cdfs of the Uniform$[-3+\delta, 0]$, Uniform$[0, 1-\delta]$, and the degenerate distribution at 0. Clearly, F is unimodal about $\mathbf{M} = 0$, and $\boldsymbol{\mu} = E_F(X) = 0$. However, the median $\mathbf{m}$ equals $= \frac{(1-\delta)^2}{3-\delta} > 0$, violating both inequalities in (1). Other examples can be found in Dharmadhikari and Joag-dev (1988) ■

Various sufficient conditions for the validity of (1) are given in Groeneveld and Meeden (1977), Runnenburg (1978), van Zwet (1979), and MacGillivray (1981). Dharmadhikari and Joag-Dev (1983, 1988) give a sufficient condition using stochastic ordering of distributions; this encompasses the works of the previous authors. In section 2, we briefly review the literature on the validity of the *mean-median-mode inequality.*

Our approach in this article is different. We consider the class of all unimodal distributions with a fixed variance σ^2,

$$\mathcal{F}_\sigma = \left\{ F \,:\, F \text{ is unimodal and } \operatorname{Var}_F(X) = \sigma^2 \right\}; \tag{2}$$

and look at the three dimensional set

$$\mathcal{M}_\sigma = \{(\boldsymbol{\mu}[F], \mathbf{m}[F], \mathrm{M}[F]) \; : \; F \in \mathcal{F}_\sigma\} \subseteq \Re^3, \tag{3}$$

the collection of all possible triplets of mean, median, and mode for unimodal distributions on the Real line with a given variance. Knowledge about the set $\mathcal{M}_\sigma$ is valuable on its own right; furthermore, it points out the extent to which inequality (1) may get violated, and also may help to single out distributions that cause such violations. The set $\mathcal{M}_\sigma$ turns out to be a connected but nonconvex set; in section 3, we analytically describe the envelope of the set. The derivation of the set $\mathcal{M}_\sigma$ is a nontrivial mathematical exercise. By moment theory techniques we first reduce an appropriate infinite dimensional problem to finite dimensions; some further calculations of a rather difficult nature then yield the exact boundaries of $\mathcal{M}_\sigma$.

It is well known that $|\mathrm{M}[F] - \boldsymbol{\mu}[F]| \leq \sqrt{3}\sigma[F]$ for any unimodal distribution F (cf. Johnson and Rogers (1951)). Using the exact formulae for the boundary of the set $\mathcal{M}_\sigma$, we generalize this inequality to all paired combinations of $\boldsymbol{\mu}[F]$, $\mathbf{m}[F]$, and $\mathrm{M}[F]$ (see corollary 4). The use of these inequalities in : (a) establishing nonunimodality, and (b) for obtaining bounds on $\mathbf{m}[F]$ or $\sigma[F]$ (when F is known to be unimodal), is described in section 3.5. We also give an explicit quantification of the extent to which the mythical inequality in (1) can get violated (Theorem 5).

Section 4 focuses on multidimensional random variables. Here, we use the generalized notion of α-unimodality due to Olshen and Savage (1970), and prove that for an *α-unimodal* distribution, an inequality similar to the one dimensional case holds between the mode $\underset{\sim}{\mathrm{M}}$ and the mean $\underset{\sim}{\boldsymbol{\mu}}$. For a fixed mode $\underset{\sim}{\mathrm{M}}$, Theorem 7 shows that the set of mean vectors $\underset{\sim}{\boldsymbol{\mu}}$ for *α-unimodal* distributions with a specified covariance matrix is an ellipsoid around $\underset{\sim}{\mathrm{M}}$. This ellipsoid is explicitly described. We consider this exact ellipsoidal representation very satisfying. If the covariance matrix is identity and we let $\underset{\sim}{\mathrm{M}}$ vary in a sphere, the mean vectors then form a sphere (see Theorem 8).

The principal achievements of this article are the following :

(i) to explicitly characterize the three dimensional set $\mathcal{M}_\sigma$ of all possible means, medians, and modes of univariate unimodal distributions;

(ii) to use the set $\mathcal{M}_\sigma$ in quantifying the extent and nature of violations of the celebrated *mean-median-mode inequality*;

(*iii*) to use the set $\mathcal{M}_\sigma$ in establishing non-unimodality of certain distributions;

(*iv*) to derive new sharp inequalities relating the mean, median, and mode; and

(*v*) to obtain some neat characterizations in the multivariate case for generalized unimodal distributions.

2 Conditions for validity of the inequality : a review

The search for sufficient conditions under which the *mean-median-mode inequality* (cf. (1)) holds for a continuous unimodal distribution F dates back to Groeneveld and Meeden (1977). They assume F to be absolutely continuous (w.r.t. Lebesgue measure) with density f, and their sufficient condition requires that $f(\mathbf{m}+x) - f(\mathbf{m}-x)$ changes sign once for $x > 0$ and that $f(\mathbf{M}+x) - f(\mathbf{M}-x)$ does not change sign (van Zwet (1979) and MacGillivray (1981) point out that their restriction to nonnegative random variables is superfluous).

van Zwet (1979) shows that a more general sufficient condition is given by (assuming F has a density f)

$$F(\mathbf{m}-x) + F(\mathbf{m}+x) \geq 1 \quad \text{for all } x. \tag{4}$$

The following result is available.

Theorem 1 (van Zwet) *If condition (4) holds, then* $\mu \leq \mathbf{m} \leq \mathbf{M}$. *If, moreover* $\mathbf{m} \neq \mathbf{M}$, *then* $\mu < \mathbf{m} < \mathbf{M}$.

An even more general sufficient condition, based on stochastic ordering, is given in Dharmadhikari and Joag-Dev (1983).

Theorem 2 (Dharmadhikari and Joag-Dev (1988)) *Let X be a unimodal random variable. If $(X-\mathbf{m})^+$ is stochastically larger than $(X-\mathbf{m})^-$, then X has a mode $\mathbf{M}$ satisfying* $\mathbf{M} \leq \mathbf{m} \leq \mu$.

Notice that Dharmadhikari and Joag-Dev do not assume existence of densities; neither do they assume that X has a unique mode. They further show that Theorem 1, and the sufficient condition of Groeneveld and Meeden follow as corollaries from Theorem 2; so also does the following result :

Corollary 1 (van Zwet) *Let X be a unimodal random variable with density f and cdf F. If $f\left(F^{-1}(t)\right) \leq f\left(F^{-1}(1-t)\right)$ for all $0 < t < \frac{1}{2}$, then $\mu \leq \mathrm{m} \leq \mathrm{M}$. If, moreover, $\mathrm{m} \neq \mathrm{M}$, then $\mu < \mathrm{m} < \mathrm{M}$.*

For various other ramifications of these results, we refer the reader to Groeneveld and Meeden (1977), van Zwet (1979), and Dharmadhikari and Joag-dev (1983, 1988).

3 Mean, median, and mode for univariate unimodal distributions

3.1 Preliminaries

Let X be a unimodal random variable with cdf $F \in \mathcal{F}_\sigma$. The goal is to explicitly characterize the set $\mathcal{M}_\sigma$ of the mean μ, the median m, and the mode M for unimodal distributions with variance σ^2. Let $\mathcal{F}_\sigma^\mu = \{F : F \text{ is unimodal and } E_F(X) = \mu,\ \mathrm{Var}_F(X) = \sigma^2\}$, and let $\mathcal{M}_\sigma^\mu = \{(\mu, \mathrm{m}[F], \mathrm{M}[F]) : F \in \mathcal{F}_\sigma^\mu\} \subseteq \{\mu\} \times \Re^2$. Clearly, $\mathcal{M}_\sigma^\mu = (\mu,\mu,\mu) + \sigma\mathcal{M}_1^0$; further, the set $\mathcal{M}_\sigma$ is the union of the $\mathcal{M}_\sigma^\mu$ sets, i.e., $\mathcal{M}_\sigma = \bigcup_\mu \mathcal{M}_\sigma^\mu$. Hence, from now on, we assume that $\mu = 0$ and $\sigma = 1$.

We first use the known fact that for any $F \in \mathcal{F}_1^0$, the set of modes form the interval $[-\sqrt{3}, \sqrt{3}]$. Determination of the set $\mathcal{M}_1^0$ now goes along the following two principal steps :

(i) Fix $\mathrm{M} \in [-\sqrt{3}, \sqrt{3}]$, and determine

$$\underline{\mathrm{m}} = \inf_{F \in \mathcal{F}_1^{0,M}} \mathrm{m}[F], \qquad \overline{\mathrm{m}} = \sup_{F \in \mathcal{F}_1^{0,M}} \mathrm{m}[F],$$

where $\mathcal{F}_1^{0,M} = \{F \in \mathcal{F}_1^0 : F \text{ is unimodal at } \mathrm{M}\}$.

(ii) For every $\underline{\mathrm{m}} < m < \overline{\mathrm{m}}$, show that $\exists\, F \in \mathcal{F}_1^{0,M}$ such that m is the median of F.

Step (i) determines the boundaries of $\mathcal{M}_1^0$, whereas step (ii) proves that it is path connected (as $\underline{\mathrm{m}} \leq 0 \leq \overline{\mathrm{m}}$ for every fixed M, see Theorem 4 and Figure 1), and hence connected. Practically all the work goes into accomplishing step (i); the proof of step (ii) is easy on comparison.

Proof of step (ii) : Fix $m \in (\underline{\mathbf{m}}, \overline{\mathbf{m}})$. Since the infimum $\underline{\mathbf{m}}$ and the supremum $\overline{\mathbf{m}}$ are, in fact, attained (as we will show while accomplishing step (i)), $\exists$ F_1 and $F_2 \in \mathcal{F}_1^{0,M}$ such that $\mathbf{m}[F_1] = \underline{\mathbf{m}}$ and $\mathbf{m}[F_2] = \overline{\mathbf{m}}$. For $0 \leq \alpha \leq 1$, define the real valued function $g(\alpha) = \alpha F_1(m) + (1-\alpha)F_2(m)$. $g(\alpha)$ is continuous in $\alpha \in [0,1]$; further $g(0) = F_2(m) \leq \frac{1}{2}$ (since $m < \overline{\mathbf{m}}$) and $g(1) = F_1(m) \geq \frac{1}{2}$ (as $m > \underline{\mathbf{m}}$). Hence $\exists \alpha^* \in [0,1]$ such that $g(\alpha^*) = \frac{1}{2}$. Thus m is the median of $F^* = \alpha^* F_1 + (1-\alpha^*)F_2$. Clearly, $E_{F^*}(X) = 0$, $\text{Var}_{F^*}(X) = 1$, and F^* is unimodal about $\mathbf{M}$ (since both F_1 and F_2 are so); hence $F^* \in \mathcal{F}_1^{0,M}$, and step (ii) is obtained ■

3.2 Reduction to mixtures of two uniforms

To find the infimum median $\underline{\mathbf{m}}$ over the family $\mathcal{F}_1^{0,M}$ (step (i) above), we use the following easily proved Lemma (Lemma 2 in Basu and DasGupta (1992)) to show that $\underline{\mathbf{m}}$, in fact, can be described in terms of suprema of probabilities of intervals.

Lemma 1 $$\underline{\mathbf{m}} = \inf_{F \in \mathcal{F}_1^{0,M}} \mathbf{m}[F] = \inf_{\tau \in \Re} \left\{ \tau \; : \; \sup_{F \in \mathcal{F}_1^{0,M}} F(\tau) \geq \frac{1}{2} \right\}$$

Proof : See Basu and DasGupta (1992) ■

Determination of $\underline{\mathbf{m}}$, thus, can be done along the following steps : (i) for $\tau \in \Re$, find $\sup_{F \in \mathcal{F}_1^{0,M}} F(\tau)$; (ii) find $\mathcal{T} = \left\{ \tau : \sup_{F \in \mathcal{F}_1^{0,M}} F(\tau) \geq \frac{1}{2} \right\}$; and (iii) find $\underline{\mathbf{m}} = \inf\{\tau : \tau \in \mathcal{T}\}$. A similar statement holds for $\overline{\mathbf{m}} = \sup_{F \in \mathcal{F}_1^{0,M}} \mathbf{m}[F]$.

The next Theorem shows that, for any $\tau \in \Re$, $\sup_{F \in \mathcal{F}_1^{0,M}} F(\tau)$ and $\inf_{F \in \mathcal{F}_1^{0,M}} F(\tau)$ are, in fact, attained at distributions $F^* \in \mathcal{F}_1^{0,M}$ which are mixtures of at most two uniforms.

Theorem 3 *For any* $\tau \in \Re$,

$$\inf_{F \in \mathcal{F}_1^{0,M}} F(\tau) = \inf_{H \in \mathcal{H}_M} H(\tau) \quad \textit{and} \quad \sup_{F \in \mathcal{F}_1^{0,M}} F(\tau) = \sup_{H \in \mathcal{H}_M} H(\tau),$$

where $\mathcal{H}_M = \left\{ H \in \mathcal{F}_1^{0,M} \; : \; H = (1-p)U_{\eta_1}^M + pU_{\eta_2}^M, \;\; 0 \leq p \leq 1, \;\; \eta_1 \leq \eta_2 \in \Re \right\}$, *and* U_η^M *is the cdf of the* $\text{Uniform}[\min(\mathbf{M}, \mathbf{M}+\eta), \max(\mathbf{M}, \mathbf{M}+\eta)]$ *distribution.*

Remark : From existing moment theory techniques (see, for example, Mulholland and Rogers (1958), and Kemperman (1968)), it follows that the problem of finding extrema of $F(\tau)$ over the class $\mathcal{F}_1^{0,M}$ can be reduced to finding the extrema over mixtures of at most three uniforms. The above Theorem reduces the dimensions further only to mixtures of two uniforms, without which the exact analytic calculations we do in Theorem 4 perhaps would have been impossible.

Proof : We will only describe the reduction for the infimum problem. Let $X \sim F \in \mathcal{F}_1^{0,M}$. By familiar arguments, we can write $X \stackrel{\mathcal{L}}{=} \mathrm{M} + UZ$, where $U \sim$ Uniform[0,1], $Z \sim G$ is a real random variable, and U and Z are independent. Further, $E_F(X) = 0 \Leftrightarrow E_G(Z) = -2\mathrm{M}$ and $\mathrm{Var}_F(X) = 1 \Leftrightarrow E_G(Z^2) = 3(1+\mathrm{M}^2)$. Without loss of generality, we assume $\mathrm{M} \geq 0$ (the conclusion for the case $\mathrm{M} < 0$ follows by symmetry). We have to treat the following two cases separately.

Case I : $\tau \geq \mathrm{M}$. Since $\tau - \mathrm{M} \geq 0$, a straightforward argument shows that $P_F(X \leq \tau) = P_F(UZ \leq \tau - \mathrm{M}) = \int_{-\infty}^{\infty} f(z)dG(z)$, where $f(z)$ equals 1 for $z \in (-\infty, \tau - \mathrm{M}]$, and equals $\frac{\tau - \mathrm{M}}{z}$ for $z \in (\tau - \mathrm{M}, \infty)$. The problem at hand now reduces to finding $\inf \int_{-\infty}^{\infty} f(z)dG(z)$ subject to $E_G(Z) = -2\mathrm{M}$ and $E_G(Z^2) = 3(1+\mathrm{M}^2)$. The assertion of Theorem 3 will follow from general moment theory (see, for instance, Theorem 2.1 and Remark 2.3 in chapter XII in Karlin and Studden (1966)) if we can show that a quadratic $a + bz + cz^2$ such that $f(z) \geq a + bz + cz^2$ for all z can cut f at at most two points. However, this is easy to see. For such a quadratic, it easily follows that (i) c must be ≤ 0, and (ii) $c \neq 0$. Let $h(z) = f(z) - (a + bz + cz^2)$. $h(z) \geq 0$, and strictly convex on each of the subintervals $(-\infty, \tau - \mathrm{M}]$ and $(\tau - \mathrm{M}, \infty)$; thus proving that at each subinterval $h(z)$ can have at most one zero.

Case II : $\tau < \mathrm{M}$. The argument for case II is quite similar and we skip the details ∎

Corollary 2 $\inf_{F \in \mathcal{F}_1^{0,M}} \mathbf{m}[F] = \underline{\mathbf{m}} = \inf_{H \in \mathcal{H}_M} \mathbf{m}[H]$ *and* $\sup_{F \in \mathcal{F}_1^{0,M}} \mathbf{m}[F] = \overline{\mathbf{m}} = \sup_{H \in \mathcal{H}_M} \mathbf{m}[H]$.

Proof : Follows trivially from Lemma 1 and Theorem 3 ∎

3.3 Formulae for extremal medians

Our next objective is to determine exact expressions for $\underline{\mathbf{m}}$ and $\overline{\mathbf{m}}$ (in terms of the mode M).

Theorem 4 *For* $\mathrm{M} \geq 0$,

$$\underline{\mathbf{m}} = \min_{F \in \mathcal{F}_1^{0,M}} \mathbf{m}[F] = \frac{\mathrm{M}^3 - 27\mathrm{M} + (\mathrm{M}^2+9)^{3/2}}{27(\mathrm{M}^2-3)}$$

$$\textit{and} \quad \overline{\mathbf{m}} = \max_{F \in \mathcal{F}_1^{0,M}} \mathbf{m}[F] = \begin{cases} \frac{\mathrm{M}^3 - 27\mathrm{M} - (\mathrm{M}^2+9)^{3/2}}{27(\mathrm{M}^2-3)} & \textit{if} \quad 0 \leq \mathrm{M} \leq \sqrt{0.6} \\ \frac{3 - \mathrm{M}\sqrt{6-\mathrm{M}^2}}{\sqrt{6-\mathrm{M}^2} - \mathrm{M}} & \textit{if} \quad \sqrt{0.6} \leq \mathrm{M} \leq \sqrt{3} \end{cases}$$

Proof : Since the proof is entirely technical, we defer it to the appendix ■

The following consequences of Theorem 4 are worth noting separately.

Corollary 3

$$(i) \quad \{\mathbf{m}[F] \;:\; F \textit{ is unimodal at } 0,\; E_F(X) = 0,\; Var_F(X) = 1\} = [-\tfrac{1}{3}, \tfrac{1}{3}].$$

$$(ii) \quad \left\{\mathbf{m}[F] \;:\; F \textit{ is unimodal at } \sqrt{3},\; E_F(X) = 0,\; Var_F(X) = 1\right\} = \{0\}.$$

Proof : Both (*i*) and (*ii*) follow trivially from Theorem 4. An alternative proof of (*ii*) is that the only unimodal F with $E_F(X) = 0$, $\mathrm{Var}_F(X) = 1$ and $\mathrm{M}[F] = \sqrt{3}$ is $F = U[-\sqrt{3}, \sqrt{3}]$ ■

3.4 Subsequent results

In Figure 1, we plot the boundaries of the set $\mathcal{M}_1^0$, i.e., $\underline{\mathbf{m}}$ and $\overline{\mathbf{m}}$ against M $(\mathrm{M} \in [-\sqrt{3}, \sqrt{3}])$. By step (*ii*) of section 3.1, $\mathcal{M}_1^0$ is a connected set. The following corollary to Theorem 4 can be readily seen from Figure 1.

Corollary 4 *For a unimodal distribution* F *with* $E_F(X) = \mu$, median$[F] = \mathbf{m}$, mode$[F] = \mathrm{M}$, *and* $\mathrm{Var}_F(X) = \sigma^2$,

$$(i) \quad \frac{|\mathrm{M} - \mu|}{\sigma} \leq \sqrt{3}$$

$$(ii) \quad \frac{|\mathrm{m} - \mu|}{\sigma} \leq \sqrt{0.6}$$

$$(iii) \quad \frac{|\mathrm{M} - \mathrm{m}|}{\sigma} \leq \sqrt{3}.$$

Moreover, each inequality is attained.

Proof : Result (*i*) is well known. It was first obtained in Johnson and Rogers (1951). Also see Dharmadhikari and Joag-Dev (1988, pp 9).

For (*ii*), w.l.g., we assume $\sigma = 1, \mu = 0$, and $\mathrm{M} \geq 0$. For fixed $\mathrm{M} \geq 0$, let $\underline{\mathrm{m}}\{\mathrm{M}\} = \min_{F \in \mathcal{F}_1^{0,M}} \mathbf{m}[F]$, and $\overline{\mathrm{m}}\{\mathrm{M}\} = \max_{F \in \mathcal{F}_1^{0,M}} \mathbf{m}[F]$, the lower and upper boundary points of $\mathcal{M}_1^0$ corresponding to M. From Theorem 4, it can be proved that $\frac{d}{d\mathrm{M}}\underline{\mathrm{m}}\{\mathrm{M}\} > 0$ for $\mathrm{M} \in [0, \sqrt{3})$, i.e.,

$\underline{\mathrm{m}}\{\mathrm{M}\}$ is $\uparrow$ in M. Also, $\overline{\mathrm{m}}\{\mathrm{M}\}$ is nondecreasing for $\mathrm{M} \in [0, \sqrt{6}]$, and nonincreasing for $\mathrm{M} \in (\sqrt{6}, \sqrt{3}]$. Since $\underline{\mathrm{m}}\{\mathrm{M}\} \leq 0$ and $\overline{\mathrm{m}}\{\mathrm{M}\} \geq 0$, it follows that $|\mathrm{m}| \leq \max\left[\max_{\mathrm{M}\geq 0} \overline{\mathrm{m}}\{\mathrm{M}\}, -\min_{\mathrm{M}\geq 0} \underline{\mathrm{m}}\{\mathrm{M}\}\right]$ $= \max\left[\overline{\mathrm{m}}\{\sqrt{0.6}\}, -\underline{\mathrm{m}}\{0\}\right] = \max[\sqrt{0.6}, \frac{1}{3}] = \sqrt{0.6}$. This completes the proof of (*ii*).

Towards proving (*iii*), let $\phi_1(\mathrm{M}) = |\overline{\mathrm{m}}\{\mathrm{M}\} - \mathrm{M}| = \overline{\mathrm{m}}\{\mathrm{M}\} - \mathrm{M}$ for $0 \leq \mathrm{M} \leq \sqrt{0.6}$, and $= \mathrm{M} - \overline{\mathrm{m}}\{\mathrm{M}\}$ for $\sqrt{0.6} \leq \mathrm{M} \leq \sqrt{3}$. Notice $\phi_1'(\mathrm{M}) \leq 0$ for $\mathrm{M} \in [0, \sqrt{0.6}]$ and ≥ 0 otherwise; thus $\max \phi_1(\mathrm{M}) = \max\left[\phi_1(0), \phi_1(\sqrt{3})\right] = \phi_1(\sqrt{3}) = \sqrt{3}$. Similarly, $\phi_2(\mathrm{M}) = |\underline{\mathrm{m}}\{\mathrm{M}\} - \mathrm{M}| = \mathrm{M} - \underline{\mathrm{m}}\{\mathrm{M}\}$ is nondecreasing in M, thus $\max \phi_2(\mathrm{M}) = \phi_2(\sqrt{3}) = \sqrt{3}$. This proves (*iii*) and completes the proof of the corollary ■

Figure 2 is a three-dimensional plot of the set $\mathcal{M}_{\sigma=1} = \{(\boldsymbol{\mu}[F], \mathrm{m}[F], \mathrm{M}[F]) : F \text{ is unimodal and } \mathrm{Var}_F(X) = \sigma^2 = 1\}$. As we mentioned before, $\mathcal{M}_{\sigma=1}$ is simply the three dimensional set obtained by translating the origin of the set $\mathcal{M}_1^0$ along the vector $(\boldsymbol{\mu}, \boldsymbol{\mu}, \boldsymbol{\mu})$.

Corollary 5 *The set $\mathcal{M}_\sigma$ is connected but not convex.*

<u>Proof :</u> Nonconvexity is trivial ($\mathcal{M}_1^0$ is not convex). We will show that $\mathcal{M}_\sigma$ is pathwise connected. Note that, for every $\mu \in \Re$, the point $(\mu, \mu, \mu) \in \mathcal{M}_\sigma$ (Because there is a symmetric unimodal distribution with mean $= \mu$, median $= \mu$, mode $= \mu$, and variance $= \sigma^2$). The proof now follows from the fact that each $\boldsymbol{\mu}$-section of $\mathcal{M}_\sigma$ is path connected (see section 3.1) ■

In Example 1 and Figure 1, we have observed that there are unimodal distributions for which the alphabetical ordering of mean, median and mode does not hold. It is of natural interest to quantify the amount of maximum possible deviation from this ordering. Towards this end, let $I = \{(\boldsymbol{\mu}, \mathrm{m}, \mathrm{M}) \in \Re^3 :$ inequality (1) or its reverse holds $\}$, and let $\mathcal{S}_\sigma = \mathcal{M}_\sigma \cap I$, the subset of $\mathcal{M}_\sigma$ where inequality (1) holds. For any $\underset{\sim}{\theta} = (\boldsymbol{\mu}, \mathrm{m}, \mathrm{M}) \in \mathcal{M}_\sigma$, let $d(\underset{\sim}{\theta}, \mathcal{S}_\sigma) = \inf_{\underset{\sim}{\eta} \in \mathcal{S}_\sigma} ||\underset{\sim}{\theta} - \underset{\sim}{\eta}||$ be the L^2-distance between the point $\underset{\sim}{\theta}$ and the set $\mathcal{S}_\sigma$. The quantity $d(\mathcal{M}_\sigma, \mathcal{S}_\sigma) = \sup_{\underset{\sim}{\theta} \in \mathcal{M}_\sigma} d(\underset{\sim}{\theta}, \mathcal{S}_\sigma)$ is a reasonable quantification of the maximum possible deviation from the mythical ordering.

W.l.g., let $\sigma = 1$. Instead of looking at the 3-dimensional sets $\mathcal{M}_{\sigma=1}$ and $\mathcal{S}_{\sigma=1}$, we first look at a $\boldsymbol{\mu}$-section of them, in particular, the $\boldsymbol{\mu} = 0$ section. Recall, $\mathcal{M}_1^0$ was our notation for the $\boldsymbol{\mu} = 0$ section of $\mathcal{M}_{\sigma=1}$. Let $I^0 = \{(\boldsymbol{\mu}, \mathrm{m}, \mathrm{M}) \in I : \boldsymbol{\mu} = 0\}$, and $\mathcal{S}_1^0 = \mathcal{M}_1^0 \cap I^0$. The set I^0 is the dashed area in Figure 1.

Theorem 5 $\quad d\,(\mathcal{M}_1^0, \mathcal{S}_1^0) \;=\; 0.294931.$

Proof : For brevity, we will consider the case $\mathrm{M} \geq 0$ (the proof for $\mathrm{M} < 0$ is similar). From Figure 1, it follows that $d\,(\mathcal{M}_1^0, \mathcal{S}_1^0) = \sup_{\underset{\sim}{\alpha} \in \mathcal{M}_1^0} \inf_{\underset{\sim}{\beta} \in \mathcal{S}_1^0} ||\underset{\sim}{\alpha} - \underset{\sim}{\beta}\,||$ is given by the maximum of

[A] the distance of the farthest point on the the upper boundary $\overline{\mathrm{m}}\{\mathrm{M}\}$ of $\mathcal{M}_1^0$ from the 45° line $\mathbf{m} = \mathbf{M}$ (for $0 \leq \mathrm{M} \leq \sqrt{0.6}$), and

[B] $\sup_{\mathrm{M} \geq 0} \{$ min of (i) and $(ii)\}$, where for each fixed $\mathrm{M} \geq 0$

(i) distance of $\underline{\mathbf{m}}\{\mathrm{M}\}$ from the horizontal axis, and

(ii) distance of $\underline{\mathbf{m}}\{\mathrm{M}\}$ from the 45° line $\mathbf{m} = \mathrm{M}$.

Towards [A], for each fixed M, the distance between the point $(\overline{\mathrm{m}}\{\mathrm{M}\}, \mathrm{M})$ and the 45° line is the same as the distance between the two points $(\overline{\mathrm{m}}\{\mathrm{M}\}, \mathrm{M})$, and $([\overline{\mathrm{m}}\{\mathrm{M}\} + \mathrm{M}]/2, [\overline{\mathrm{m}}\{\mathrm{M}\} + \mathrm{M}]/2)$, which equals $\sqrt{0.5}|\mathrm{M} - \overline{\mathrm{m}}\{\mathrm{M}\}|$. Since $|\mathrm{M} - \overline{\mathrm{m}}\{\mathrm{M}\}|$ is nonincreasing in $\mathrm{M} \in [0, \sqrt{0.6}]$, the maximum attains at $\mathrm{M} = 0$, and the maximum distance $= 0.235702$.

In [B], (ii) is similar to [A]; the distance is given by (for each fixed M) the distance between the points $(\underline{\mathrm{m}}\{\mathrm{M}\}, \mathrm{M})$ and $([\underline{\mathrm{m}}\{\mathrm{M}\} + \mathrm{M}]/2, [\underline{\mathrm{m}}\{\mathrm{M}\} + \mathrm{M}]/2)$. Hence, for each fixed M, minimum of (i) and (ii) $=$ min $\left[-\underline{\mathrm{m}}\{\mathrm{M}\}, \sqrt{0.5}|\mathrm{M} - \underline{\mathrm{m}}\{\mathrm{M}\}|\right] = \sqrt{0.5}|\mathrm{M} - \underline{\mathrm{m}}\{\mathrm{M}\}|$ if $0 \leq \mathrm{M} \leq 0.122164$, and $= -\underline{\mathrm{m}}\{\mathrm{M}\}$ if $0.122164 \leq \mathrm{M} \leq \sqrt{3}$ (after some simplifications). But $|\mathrm{M} - \underline{\mathrm{m}}\{\mathrm{M}\}|$ and $\underline{\mathrm{m}}\{\mathrm{M}\}$ are both nondecreasing in M; thus the supremum over M is attained at $\mathrm{M} = 0.122164$ and the maximum distance $= 0.294931$. This proves the Theorem ∎

From Theorem 5, it follows trivially that the L^2-distance between the two 3-dimensional sets $\mathcal{M}_\sigma$ and $\mathcal{S}_\sigma$ satisfies $d(\mathcal{M}_\sigma, \mathcal{S}_\sigma) \leq 0.294931\,\sigma$. We strongly suspect that, in fact, equality holds, i.e.,

Conjecture 1 $\quad d(\mathcal{M}_\sigma, \mathcal{S}_\sigma) \;=\; 0.294931\,\sigma,$

though we were not able to prove this conjecture analytically.

3.5 Examples

Our objective in this section is to point out the possible directions of use of our obtained results. For clarity, we look at two simple examples, rather than considering complex (probably more realistic) examples. The first example shows how inequality (*ii*) in corollary 4 can be used to prove nonunimodality of certain distributions. The second example outlines a method for obtaining useful bounds on the median of a fixed distribution when the actual median may be hard to obtain.

<u>Example 1 :</u> Let F be a mixture of two point masses and a uniform : $F = p_1\delta_{\{-n\}} + p_2 U[-n,1] + p_3\delta_{\{1\}}$, $p_i \geq 0$, $p_1 + p_2 + p_3 = 1$, and the p_i's are so chosen that $E_F(X) = 0$ ($\delta_{\{\theta\}}$ denotes the degenerate distribution at θ). F is clearly bimodal with modes at $-n$ and 1, the two endpoints of its support. Let $\widetilde{X}_k$ denote the sample median of a sample of $(2k-1)$ observations from F and let $\widetilde{F}_k$ denote the distribution of $\widetilde{X}_k$. It is known that median of $\widetilde{F}_k$ = median of F, i.e., $\mathbf{m}[\widetilde{F}_k] = \mathbf{m}[F]$ (see Reiss (1989)). For small values of k, say $k = 2$ or 10, though the form of $\widetilde{F}_k$ can easily be written down, proving unimodality or nonunimodality of $\widetilde{F}_k$ seems nontrivial. However, violation of the inequality in part (*ii*) of corollary 4 $\left(V^* = 0.6\text{Var}(\widetilde{X}_k) - \left[\mathbf{m}[F] - E(\widetilde{X}_k)\right]^2 < 0\right)$ is a sufficient condition for nonunimodality of the distribution $\widetilde{F}_k$. Evaluation of $E(\widetilde{X}_k)$ and $\text{Var}(\widetilde{X}_k)$ needs some numerical work, but is much more straightforward compared to a direct check of unimodality. For example, for $n = 1.1$ and p_3 close to 0.5 (but < 0.5), V^* turns out to be < 0 for all $k \leq 39$, thus proving that $\widetilde{X}_k$ has a nonunimodal distribution for $k \leq 39$.

<u>Example 2 :</u> The inequality of corollary 4, namely $(\text{median} - \text{mean})^2 \leq 0.6\,\text{Variance}$, can also be used to obtain useful bounds on mean or median or variance, depending on which two of the three are easier to obtain.

As a simple verification of how it works, let $X \sim \exp(\lambda)$ with $\mu = E(X) = \frac{1}{\lambda}$. Further, $F(x) = 1 - e^{-\lambda x}$, $x > 0$, and solving $F(x) = \frac{1}{2}$ gives median $\mathbf{m} = \frac{\log(2)}{\lambda}$. From the above inequality, we obtain $\text{Var}(X) \geq \frac{1}{0.6}(\mu - \mathbf{m})^2 = \frac{0.5114213}{\lambda^2}$, which is approximately half of $\text{Var}(X) = \frac{1}{\lambda^2}$.

For another simple example, let $X \sim F = \text{Beta}(\alpha, \beta)$. Thus, $\mu = \frac{\alpha}{\alpha+\beta}$ and $\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$. Moreover, if $\alpha > 1$, and $\beta > 1$, then F is unimodal at $\text{M} = \frac{\alpha-1}{\alpha+\beta-2}$; if $\alpha \leq 1, \beta \geq 1$ ($\alpha \geq 1, \beta \leq 1$), F is unimodal at $\text{M} = 0$ ($\text{M} = 1$). But evaluation of the median $\mathbf{m}$ of F requires solving an equation involving incomplete Beta integrals. Corollary 4, however, gives the following useful bounds on $\mathbf{m}$ (for $\alpha \not< 1$ or $\beta \not< 1$).

$$\max\left\{\mu - \sqrt{0.6}\sigma, \text{M} - \sqrt{3}\sigma\right\} \leq \mathbf{m} \leq \min\left\{\mu + \sqrt{0.6}\sigma, \text{M} + \sqrt{3}\sigma\right\} \tag{5}$$

Table 1: Bounds on the median **m** of Beta(α, β) distribution

		β			
α		0.5	1.0	1.5	2.0
1.0	Bounds	(0.484,0.898)	(0.500,0.724)	(0.197,0.454)	(0.151,0.408)
	m	0.750	0.500	0.370	0.293
1.5	Bounds	(0.567,0.944)	(0.546,0.803)	(0.306,0.694)	(0.248,0.609)
	m	0.837	0.630	0.500	0.414
2.0	Bounds	(0.634,0.966)	(0.592,0.849)	(0.391,0.752)	(0.327,0.673)
	m	0.879	0.707	0.586	0.5
3.0	Bounds	(0.729,0.985)	(0.665,0.900)	(0.511,0.822)	(0.445,0.755)
	m	0.921	0.794	0.693	0.614
4.0	Bounds	(0.785,0.993)	(0.717,0.926)	(0.592,0.863)	(0.529,0.805)
	m	0.941	0.841	0.756	0.686

These bounds can easily be evaluated without any numerical work. Table 1 shows the bounds obtained from (5) along with the actual values of the median **m** for different combinations of α and β.

4 Multivariate Unimodality

Our objective in this section is to generalize some of the results of section 3 in the setting of multivariate unimodal distributions. Unlike in one dimension, there are several definitions of unimodality in higher dimensions such as *star unimodality, block unimodality, central convex unimodality, and log concavity* (and more; see Dharmadhikari and Joag-Dev (1988)). We will restrict ourselves to the generalized notion of *α-unimodality*, introduced by Olshen and Savage (1970), of which *star-unimodality* is a special case.

Definition 1 (Olshen and Savage) *A p-dimensional random vector $\underset{\sim}{X}$ is said to have an* α-unimodal ($\alpha > 0$) *distribution about* $\underset{\sim}{\mathrm{M}}$ *if, for every bounded, nonnegative, Borel-measurable function g on* $\Re^p$, $t^\alpha E\left[g(t(\underset{\sim}{X} - \underset{\sim}{\mathrm{M}}))\right]$ *is nondecreasing in* $t \in (0, \infty)$.

The definition given by Olshen and Savage is, in fact, more general; it applies for $\underset{\sim}{X}$ taking values in any p-dimensional vector space. It can be seen that ordinary unimodality on $\Re$ is equivalent to definition 1 with $p = 1$ and $\alpha = 1$. In general, definition 1 corresponds to the property of starshaped level sets for the distribution of $\underset{\sim}{X}$.

Theorem 6 (Olshen and Savage) *A p-dimensional random vector $\underset{\sim}{X}$ is* α-unimodal *(about $\underset{\sim}{M}$) if and only if $\underset{\sim}{X} \overset{\mathcal{L}}{=} U^{1/\alpha}\underset{\sim}{Z} + \underset{\sim}{M}$ where $U \sim$ Uniform$[0,1]$ and $\underset{\sim}{Z}$ is a p-dimensional random vector independent of U.*

A p-dimensional and p-unimodal (about $\underset{\sim}{M}$) random vector $\underset{\sim}{X}$ is called *star unimodal* (about $\underset{\sim}{M}$). Generally, *star unimodality* is defined through a natural extension of the idea of one-dimensional unimodality, and the above definition follows as an equivalent version. The concept of *star unimodality* has recently been successfully used in inference problems : see DasGupta, Ghosh and Zen (1991).

Since the use of a median in the multivariate case is less pervasive, we focus our efforts on extending part (i) of corollary 4 given in section 3.4. Let us make it more precise.

Let $\Theta_{M,\Sigma}$ denote the set of all possible mean vectors $\underset{\sim}{\mu}$ corresponding to p-dimensional *α-unimodal* random vectors $\underset{\sim}{X}$ with mode$[\underset{\sim}{X}] = \underset{\sim}{M}$ and covariance matrix $D(\underset{\sim}{X}) = \Sigma$, i.e.,

$$\Theta_{M,\Sigma} \;=\; \left\{ \underset{\sim}{\mu} = E(\underset{\sim}{X}) \;:\; \underset{\sim}{X} \text{ is } \alpha\text{-}unimodal \text{ about } \underset{\sim}{M} \text{ and } D(\underset{\sim}{X}) = \Sigma \right\} \qquad (6)$$

We have the following neat ellipsoidal representation of $\Theta_{M,\Sigma}$.

Theorem 7 *Assume $|\Sigma| \neq 0$. Then*

$$\Theta_{M,\Sigma} \;=\; \left\{ \underset{\sim}{\mu} \;:\; (\underset{\sim}{\mu} - \underset{\sim}{M})^T\Sigma^{-1}(\underset{\sim}{\mu} - \underset{\sim}{M}) \;\leq\; \alpha(\alpha+2) \right\}.$$

Proof : W.l.g, we take $\underset{\sim}{M} = \underset{\sim}{0}$. Further, if $\underset{\sim}{X}$ is *α-unimodal* (we use *α-unimodal* to imply *α-unimodal* about $\underset{\sim}{0}$) with $E(\underset{\sim}{X}) = \underset{\sim}{\mu}$ and $D(\underset{\sim}{X}) = \Sigma$, then for $A_{p\times p}$ nonsingular, $\underset{\sim}{Y} = A\underset{\sim}{X}$ has $E(\underset{\sim}{Y}) = A\underset{\sim}{\mu}$ and $D(\underset{\sim}{Y}) = A\Sigma A^T$. It follows from Theorem 6 that $\underset{\sim}{Y}$ is also *α-unimodal.* Taking $A = \Sigma^{-1/2}$ shows that it is enough to prove the result for $\Sigma = I$.

The proof will be divided into the following steps

(i) $\underset{\sim}{\mu} \in \Theta_{0,I} \;\Rightarrow\; \underset{\sim}{\mu}^T\underset{\sim}{\mu} \leq \alpha(\alpha+2) = \alpha^*$ (say).

(ii) $\underset{\sim}{\mu} \in \Theta_{0,I} \Rightarrow P\underset{\sim}{\mu} \in \Theta_{0,I}$ for any orthogonal matrix P.

(iii) For every scalar γ satisfying $-\sqrt{\alpha^*} \leq \gamma \leq \sqrt{\alpha^*}$, $\underset{\sim}{\mu} = (\gamma, 0, \ldots, 0)^T \in \Theta_{0,I}$.

We claim that these three steps show $\Theta_{0,I} = \left\{\underset{\sim}{\mu} : \underset{\sim}{\mu}^T \underset{\sim}{\mu} \leq \alpha^*\right\}$ and proves the Theorem. That $\Theta_{0,I} \subseteq \left\{\underset{\sim}{\mu} : \underset{\sim}{\mu}^T \underset{\sim}{\mu} \leq \alpha^*\right\}$ follows from step (i). Next, for any $\underset{\sim}{\mu}$ with $\underset{\sim}{\mu}^T \underset{\sim}{\mu} = \gamma^2 \leq \alpha^*$, $\exists\, P$ orthogonal such that $P\underset{\sim}{\mu} = (\gamma, 0, \ldots, 0)^T$. By (iii), $P\underset{\sim}{\mu} \in \Theta_{0,I}$, and hence by step (ii), $\underset{\sim}{\mu} \in \Theta_{0,I}$.

Step (ii) is an easy consequence of Theorem 6. For proving (i), let $\underset{\sim}{X}$ be *α-unimodal* with $D(\underset{\sim}{X}) = I$. By Theorem 6, $\underset{\sim}{X} \stackrel{\mathcal{L}}{=} V\underset{\sim}{Z}$ where V is a scalar random variable, $V \stackrel{\mathcal{L}}{=} U^{1/\alpha}$ with $U \sim$ Uniform$[0,1]$, and V and $\underset{\sim}{Z}$ are independent. Note that $\mu_v = E(V) = \frac{\alpha}{\alpha+1}$ and $\sigma_v^2 = \text{Var}(V) = \frac{\alpha}{(\alpha+2)(\alpha+1)^2}$; thus $\alpha^* = \mu_v^2/\sigma_v^2$. Now,

$$I = D(\underset{\sim}{X}) = D(V\underset{\sim}{Z}) = \sigma_v^2 E(\underset{\sim}{Z}\,\underset{\sim}{Z}^T) + \mu_v^2 D(\underset{\sim}{Z}) \geq \sigma_v^2 E(\underset{\sim}{Z}\,\underset{\sim}{Z}^T). \tag{7}$$

On the other hand,

$$\begin{aligned} \underset{\sim}{\mu}\,\underset{\sim}{\mu}^T = E(\underset{\sim}{X}\,\underset{\sim}{X}^T) - D(\underset{\sim}{X}) &= E(V^2)E(\underset{\sim}{Z}\,\underset{\sim}{Z}^T) - I \\ &\leq \left[\frac{E(V^2)}{\sigma_v^2} - 1\right] I \qquad \text{(from (7))} \\ &= \frac{\mu_v^2}{\sigma_v^2} I \end{aligned}$$

Thus $\underset{\sim}{\mu}^T \underset{\sim}{\mu} \leq \mu_v^2/\sigma_v^2$, which proves (i).

To prove (iii), fix $-\sqrt{\alpha^*} \leq \gamma \leq \sqrt{\alpha^*}$. Consider p independent scalar random variables $Z_1, Z_2, \ldots, Z_p$ such that $E(Z_1) = \frac{\gamma}{\mu_v}$, $E(Z_1^2) = \frac{1+\gamma^2}{\sigma_v^2+\mu_v^2}$, and for $i = 2, \ldots, p$, $E(Z_i) = 0$, $E(Z_i^2) = \frac{1}{\sigma_v^2+\mu_v^2}$. Take $\underset{\sim}{Z} = (Z_1, \ldots, Z_p)^T$ and now pick a scalar random variable V independent of $\underset{\sim}{Z}$ such that $V \stackrel{\mathcal{L}}{=} U^{1/\alpha}$ with $U \sim$ Uniform$[0,1]$. $\underset{\sim}{X} = V\underset{\sim}{Z}$ is clearly *α-unimodal*, moreover, it can be seen that $D(\underset{\sim}{X}) = I$ and $E(\underset{\sim}{X}) = (\gamma, 0, \ldots, 0)^T$. The proof of the Theorem is therefore complete ■

<u>Remark :</u> Notice the ellipsoids of Theorem 7 are nested, i.e., the same ellipsoid is obtained if $D(\underset{\sim}{X}) \leq \Sigma$.

Corollary 6 *Let $\underset{\sim}{X}$ be a p-dimensional random vector,* star unimodal *about* $\underset{\sim}{\mathrm{M}}$, *with* $E(\underset{\sim}{X}) = \underset{\sim}{\mu}$, $D(\underset{\sim}{X}) = \Sigma$. *Then*

$$(\underset{\sim}{\mu} - \underset{\sim}{\mathrm{M}})^T \Sigma^{-1} (\underset{\sim}{\mu} - \underset{\sim}{\mathrm{M}}) \leq p(p+2).$$

Proof : Immediate from Theorem 7 ■

Thus, according to Theorem 7, the means $\underset{\sim}{\mu}$ of *α-unimodal* distributions lie in an ellipsoid with center at the mode $\underset{\sim}{M}$, and the axes of the ellipsoid are multiples of the eigenvectors of the dispersion matrix Σ.

In the above, we assume that the mode $\underset{\sim}{M}$ is fixed at a certain point in $\Re^p$. Next, we look at the set of mean vectors $\underset{\sim}{\mu}$ of *α-unimodal* distributions, when the mode $\underset{\sim}{M}$ is also allowed to vary in a sphere in $\Re^p$.

Theorem 8 *Let $\Theta_{M,\Sigma}$ be as defined before, and let*

$$\Theta_I = \bigcup_{\underset{\sim}{M} \in \Omega} \Theta_{M,I} = \left\{ \underset{\sim}{\mu} = E(\underset{\sim}{X}) : D(\underset{\sim}{X}) = I, \text{ and } \underset{\sim}{X} \text{ is } \alpha\text{-unimodal about } \underset{\sim}{M} \text{ with } \underset{\sim}{M} \in \Omega \right\},$$

where the Ω is the sphere $\left\{ \underset{\sim}{M} : (\underset{\sim}{M} - \underset{\sim}{M}_0)^T (\underset{\sim}{M} - \underset{\sim}{M}_0) \leq \beta^2 \right\}$.

Then the set of mean vectors $\underset{\sim}{\mu}$ is again a sphere, and equals

$$\Theta_I = \left\{ \underset{\sim}{\mu} : (\underset{\sim}{\mu} - \underset{\sim}{M}_0)^T (\underset{\sim}{\mu} - \underset{\sim}{M}_0) \leq (\sqrt{\alpha^*} + \beta)^2 \right\}.$$

Proof : The proof is notationally slightly complex, but is indeed nothing more than a proof of the fact that the Minkowski sum of two spheres is again a sphere.

W.l.g., we take $\underset{\sim}{M}_0 = \underset{\sim}{0}$. For a random vector $\underset{\sim}{X}$, *α-unimodal* about $\underset{\sim}{M}$ with $E(\underset{\sim}{X}) = \underset{\sim}{\mu}$ and $D(\underset{\sim}{X}) = I$, by Theorem 7, we have $(\underset{\sim}{\mu} - \underset{\sim}{M})^T (\underset{\sim}{\mu} - \underset{\sim}{M}) \leq \alpha^*$. If, moreover, $\underset{\sim}{M} \in \Omega$, then $||\underset{\sim}{\mu}|| \leq ||\underset{\sim}{\mu} - \underset{\sim}{M}|| + ||\underset{\sim}{M}|| \leq \sqrt{\alpha^*} + \beta$; this shows that $\Theta_I \subseteq \left\{ \underset{\sim}{\mu} : \underset{\sim}{\mu}^T \underset{\sim}{\mu} \leq (\sqrt{\alpha^*} + \beta)^2 \right\}$.

For proving that Θ_I equals the entire sphere, pick any $\underset{\sim}{\mu}$ with $||\underset{\sim}{\mu}|| = \gamma \leq \sqrt{\alpha^*} + \beta$. Find an orthogonal matrix P such that $P\underset{\sim}{\mu} = (\gamma, 0, \ldots, 0)^T$, and next find $0 \leq \gamma_1 \leq \sqrt{\alpha^*}$ and $0 \leq \gamma_2 \leq \beta$ such that $\gamma = \gamma_1 + \gamma_2$. Clearly, $\underset{\sim}{M}^* = (\gamma_2, 0, \ldots, 0)^T \in \Omega$, and for such an $\underset{\sim}{M}^*$, $\exists$ a random vector $\underset{\sim}{X}$, *α-unimodal* about $\underset{\sim}{M}^*$, with $D(\underset{\sim}{X}) = I$ and $E(\underset{\sim}{X}) = \underset{\sim}{\mu}^*$ such that $\underset{\sim}{\mu}^* - \underset{\sim}{M}^* = (\gamma_1, 0, \ldots, 0)^T$. Thus, $\underset{\sim}{\mu}^* = (\underset{\sim}{\mu}^* - \underset{\sim}{M}^*) + \underset{\sim}{M}^*$, and hence $\underset{\sim}{\mu}^* \in \Theta_I$. Now, $\underset{\sim}{\mu} = (P^T \underset{\sim}{\mu}^* - P^T \underset{\sim}{M}^*) + P^T \underset{\sim}{M}^*$, and $P^T \underset{\sim}{\mu}^* - P^T \underset{\sim}{M}^* \in \Theta_{P^T M^*, I}$; $P^T \underset{\sim}{M}^* \in \Omega$, hence $\underset{\sim}{\mu} \in \Theta_I$. This completes the proof ■

5 Summary and discussions

Past work in this area was directed towards finding more and more general sufficient conditions for the validity of the *mean-median-mode inequality.* In contrast, we concentrate our attention to exact characterizations of the three dimensional set of means, medians and modes. Derivation of the set involves some novel use of moment theory, and, most surprisingly, we were able to analytically describe this set. Furthermore, this exact description enabled us to derive some fundamental inequalities among the mean, median and mode of unimodal random variables. We were also able to specify the region where the *inequality* gets violated, and found the maximum extent to which a unimodal random variable may violate the *mean-median-mode inequality.* In the multivariate case, we obtain a very pleasant ellipsoidal characterization of the set of means of *α-unimodal* distributions with a fixed mode $\underset{\sim}{\mathrm{M}}$. When the mode $\underset{\sim}{\mathrm{M}}$ is also allowed to vary in a sphere, we prove that the mean vectors form a larger sphere.

Several open questions remain. We conjectured that the maximum deviation from the *inequality* is 0.294931σ and we proved it for each fixed mean μ, but were unable to prove it in its full generality. In Theorem 8 we assume $\Sigma = I$, and that the mode $\underset{\sim}{\mathrm{M}}$ is in a sphere. The case of a general Σ and when the mode varies in other nicely structured convex sets (such as an ellipsoid, or a rectangle) are hard geometric problems, and again, open for exploration. We believe the geometric results in DasGupta and Studden (1988) are of probable relevance here.

Appendix

Proof of Theorem 4 : The derivation of $\underline{\mathbf{m}}$ and $\overline{\mathbf{m}}$ looks intimidating, but it is really straightforward on patient verifications. First note that, by corollary 2, we only need consider distributions of the form $H = (1-p)U_{\eta_1}^M + pU_{\eta_2}^M \quad (\eta_1 \leq \eta_2)$ satisfying $E_H(Z) = -2\mathrm{M}$ and $E_H(Z^2) = 3(1+\mathrm{M}^2)$. Without loss of generality, we assume $\mathrm{M} \geq 0$. Since $E_H(Z) = -2\mathrm{M} \leq \mathrm{M}$, it follows that η_1 must be ≤ 0.

Case I $[\eta_1 \leq 0,\ \eta_2 \geq 0]$: For notational simplicity, we write $\eta_1 = -2a,\ \eta_2 = 2b$, where $a, b \geq 0$. Thus, the distribution H is of the form

$$H \ = \ (1-p)\,U[-2a+\mathrm{M},\,\mathrm{M}] \ + \ p\,U[\mathrm{M},\,\mathrm{M}+2b]. \tag{8}$$

Now, $E_H(Z) = -2\mathrm{M} \Rightarrow p = \frac{a-\mathrm{M}}{a+b}$, $p \geq 0 \Leftrightarrow a \geq \mathrm{M}$. Further, $E_H(Z^2) = 3(1+\mathrm{M}^2)$ allows us to write

$$a = \frac{4\mathrm{M}b + 3(1+\mathrm{M}^2)}{4(b+\mathrm{M})} \quad \text{or} \quad b = \frac{3(1+\mathrm{M}^2) - 4a\mathrm{M}}{4(a-\mathrm{M})} \tag{9}$$

(we will use either expression as necessary). For ease of calculations, we further subdivide *Case I* to the following two subcases.

<u>*Subcase A* $[\,\mathbf{m}[H] > \mathrm{M}\,]$:</u> From (8), $\mathbf{m}[H] > \mathrm{M} \Leftrightarrow p > \frac{1}{2} \Leftrightarrow b < -\mathrm{M} + \frac{\sqrt{3-\mathrm{M}^2}}{2}$ on simplification (using (9)). The condition $b \geq 0$ requires $-\mathrm{M} + \frac{\sqrt{3-\mathrm{M}^2}}{2} \geq 0 \Leftrightarrow \mathrm{M}^2 \leq 0.6$; hence, *Subcase A* is possible only if $0 \leq \mathrm{M} \leq \sqrt{0.6}$.

Since H has no jumps, we have $1 - H(\mathbf{m}) = \frac{1}{2}$. Solving for $\mathbf{m}$ (and use of (9)) gives $\mathbf{m} = 2b + \mathrm{M} - b\,\frac{2\mathrm{M}b + b^2 + 0.75(1+\mathrm{M}^2)}{0.75 - 0.25\mathrm{M}^2} = \psi_1(b)$ (say). $b^* = \frac{-4\mathrm{M} + \sqrt{\mathrm{M}^2+9}}{6}$ is the only nonnegative solution to $\psi_1'(b) = 0$. Moreover, $\psi_1''(b^*) < 0$, and b^* satisfies the required boundary conditions. Hence, $\max_{Case\ I-A} \mathbf{m}[H] = \psi_1(b^*) = \frac{\mathrm{M}^3 - 27\mathrm{M} - (\mathrm{M}^2+9)^{1.5}}{27(\mathrm{M}^2-3)}$.

Notice that, for each fixed $\mathrm{M} \geq 0$, this *Subcase* restricts the median $\mathbf{m}[H]$ to be $> \mathrm{M}$, whereas *Subcase I-B* and *Case II* below only allows $\mathbf{m}[H] \leq \mathrm{M}$. Thus, for those M where *Case I-A* is possible (i.e., for $0 \leq \mathrm{M} \leq \sqrt{0.6}$), $\overline{\mathbf{m}} = \max_{Case\ I-A} \mathbf{m}[H]$. For the same reason (since other feasible cases allow $\mathbf{m}[H] \leq \mathrm{M}$), we refrain from evaluating $\inf_{Case\ I-A} \mathbf{m}[H]$ □

<u>*Subcase B* $[\,\mathbf{m}[H] \leq \mathrm{M}\,]$:</u> For ease of calculations, we formulate this case in terms of 'a'. The domain of 'a' is bounded by : (i) $p \geq 0 \Leftrightarrow a \geq \mathrm{M}$, (ii) $b \geq 0 \Leftrightarrow a \leq \frac{3(1+\mathrm{M}^2)}{4\mathrm{M}} = A_1$, and (iii) $\mathbf{m}[H] \leq \mathrm{M} \Leftrightarrow a \leq \frac{2\mathrm{M}+\sqrt{3-\mathrm{M}^2}}{2} = A_2$. Thus, $a \in [\mathrm{M}, \min(A_1, A_2)]$, and $\min(A_1, A_2) = A_2$ if $0 \leq \mathrm{M}^2 \leq 0.6$, and $= A_1$ otherwise.

Solving $H(\mathbf{m}) = \frac{1}{2}$ gives $\mathbf{m} = \mathrm{M} - 2a + a\,\frac{4a^2 - 8a\mathrm{M} + 3 + 3\mathrm{M}^2}{3 - \mathrm{M}^2} = \psi_2(a)$ (say). $\psi_2'(a) = 0$ has only one solution $a^* = \frac{4\mathrm{M} + \sqrt{\mathrm{M}^2+9}}{6}$ in the domain of 'a'; further, $\psi_2''(a^*) > 0$, i.e., a^* is a local minima.

Towards determining $\max \mathbf{m}[h]$, note that we are really only interested in the case $\mathrm{M}^2 > 0.6$ (see *Case I-A*), which implies $a \in [\mathrm{M},\ A_1]$. Clearly, $\max \mathbf{m}[H]$ is attained at the boundaries of the domain of 'a'. Thus, for $\sqrt{0.6} < \mathrm{M} \leq \sqrt{3}$, $\max_{Case\ I-B} \mathbf{m}[H] = \max\{\psi_2(\mathrm{M}), \psi_2(A_1)\} = \psi_2(A_1) = \frac{(3-\mathrm{M}^2)^2}{16\mathrm{M}^3}$.

$\min \psi_2(a)$ is clearly attained at $a = a^*$, and $\min_{Case\ I-B} \mathbf{m}[H] = \psi_2(a^*) = \frac{\mathrm{M}^3 - 27\mathrm{M} + (\mathrm{M}^2+9)^{1.5}}{27(\mathrm{M}^2-3)}$. It can be shown that $\min_{Case\ I-B} \mathbf{m}[H] < 0$ for $0 \leq \mathrm{M} < \sqrt{3}$, and $= 0$ at $\mathrm{M} = \sqrt{3}$ □

<u>*Case II* $[\,\eta_1 \leq 0,\ \eta_2 \leq 0\,]$:</u> Again, for notational convenience, we write $\eta_1 = -2a,\ \eta_2 = -2b$,

where $a \geq b \geq 0$. Thus,

$$H = (1-p)\,U[-2a+\mathrm{M},\,\mathrm{M}] \;+\; p\,U[-2b+\mathrm{M},\,\mathrm{M}], \tag{10}$$

and from the conditions $E_H(Z) = -2\mathrm{M}$ and $E_H(Z^2) = 3(1+\mathrm{M}^2)$, we have

$$p = \frac{a-\mathrm{M}}{a-b} \quad \text{and} \quad b = \frac{4a\mathrm{M} - 3(1+\mathrm{M}^2)}{4(a-\mathrm{M})}. \tag{11}$$

Now, $0 \leq p \leq 1 \Rightarrow a \geq \mathrm{M} \geq b \geq 0$, and $b \geq 0 \Leftrightarrow a \geq \frac{3(1+\mathrm{M}^2)}{4\mathrm{M}} = A_1^*$ (from (11)). Thus, $a \geq \max(\mathrm{M}, A_1^*) = A_1^*$. Depending on the position of $\mathbf{m}[H]$, we subdivide *Case II* into following subcases.

<u>*Subcase A* $[\,\mathbf{m}[H] \leq -2b+\mathrm{M}\,]$:</u> From (10), $\mathbf{m}[H] \leq -2b+\mathrm{M} \Leftrightarrow (1-p)\frac{2a-2b}{2a} \geq \frac{1}{2} \Leftrightarrow a \leq \frac{\mathrm{M}+\sqrt{6-\mathrm{M}^2}}{2} = A_2^*$. Thus, $A_1^* \leq a \leq A_2^*$, however, for $\mathrm{M}^2 < 0.6$, $A_2^* < A_1^*$; hence this subcase is possible only if $0.6 \leq \mathrm{M}^2 \leq 3$.

As before, solving $H(\mathbf{m}) = \frac{1}{2}$ gives $\mathbf{m} = \mathrm{M} - 2a - a\,\frac{4a^2-8a\mathrm{M}+3+3\mathrm{M}^2}{3-\mathrm{M}^2} = \psi_3(a)$ (say). $\psi_3(a)$ turns out to be increasing for $a \in [A_1^*, A_2^*]$, $(\psi_3'(a) > 0)$; hence $\max_{Case\ II-A} \mathbf{m}[H] = \psi_3[A_2^*] = \frac{3-\mathrm{M}\sqrt{6-\mathrm{M}^2}}{\sqrt{6-\mathrm{M}^2}-\mathrm{M}}$ and $\min_{Case\ II-A} \mathbf{m}[H] = \psi_3[A_1^*] = \frac{(\mathrm{M}^2-3)^2}{16\mathrm{M}^3} \geq 0$ □

<u>*Subcase B* $[\,\mathbf{m}[H] \geq -2b+\mathrm{M}\,]$:</u> $\mathbf{m}[H] \geq -2b+\mathrm{M} \Rightarrow a \geq A_2^*$; thus $a \geq \max[A_1^*, A_2^*] = A_1^*$ if $\mathrm{M}^2 \leq 0.6$, and $= A_2^*$ for $\mathrm{M}^2 \geq 0.6$. Following similar steps as in *Case II-B*, we get $\mathbf{m} = \mathrm{M} - a\,\frac{3+3\mathrm{M}^2-4a\mathrm{M}}{4a\mathrm{M}-4a^2-\mathrm{M}^2+3} = \psi_4(a)$ with $\psi_4'(a) < 0\ \ \forall a$. Thus, $\max_{Case\ II-B} \mathbf{m}[H] = \psi_4[A_2^*] = \max_{Case\ II-A} \mathbf{m}[H]$ (if $\mathrm{M}^2 \geq 0.6$) and $\min_{Case\ II-B} \mathbf{m}[H] = \lim_{a\to\infty} \psi_4(a) = 0$ □

Combining all the cases, we know from *Case I-A* that for $0 \leq \mathrm{M} \leq \sqrt{0.6}$, $\overline{\mathbf{m}} = \max_{Case\ I-A} \mathbf{m}[H]$. For $\sqrt{0.6} < \mathrm{M} \leq \sqrt{3}$, $\overline{\mathbf{m}} = \max\left\{\max_{Case\ I-B} \mathbf{m}[H],\ \max_{Case\ II} \mathbf{m}[H]\right\} = \max_{Case\ II} \mathbf{m}[H]$. For $\underline{\mathbf{m}}$, we note that $\min_{Case\ II-A} \mathbf{m}[H]$ and $\min_{Case\ II-B} \mathbf{m}[H]$ are both ≥ 0, whereas $\min_{Case\ I-B} \mathbf{m}[H] \leq 0$. The proof of Theorem 4 is therefore complete ■

<u>Acknowledgement</u> We have benefitted from valuable discussions with Johannes Kemperman, Friedrich Pukelsheim, Yosi Rinott, and Willem van Zwet. The first author is indebted to Rajeeva Karandikar, Mathew Penrose, and Srikanth Iyer for numerous helpful discussions during this work. They all have our sincerest appreciation.

References

[1] Basu, S. and DasGupta, A. (1992), "Bayesian analysis with distribution bands : the role of the loss function", Technical Report, **208**, Department of Statistics and Applied Probability, University of California, Santa Barbara.

[2] DasGupta, A., Ghosh, J.K. and Zen, M. (1991) "A new general method for constructing confidence sets in arbitrary dimensions, with applications" To appear in *Ann. Statist.*.

[3] DasGupta, A. and Studden, W.J. (1988), "Robust Bayesian Analysis and optimal experimental designs in normal linear models with many parameters - I" Technical Report, **88-14**, Department of Statistics, Purdue University.

[4] Dharmadhikari, S.W. and Joag-dev, K. (1983), "Mean, median, mode III", *Staist. Neerl.*, **37**, pp. 165-168.

[5] Dharmadhikari, S.W. and Joag-dev, K. (1988), *Unimodality, Convexity, and Applications*, Academic Press: New York.

[6] Groeneveld, R.A. and Meeden, G. (1977), "The mode, median and mean inequality", *Amer. Statistician*, **33**, pp. 120-121.

[7] Johnson, N.L. and Rogers, C.A. (1951), "The moment problem for unimodal distributions", *Ann. Math. Statist.*, **22**, pp. 432-439.

[8] Karlin, S. and Studden, W.J. (1966), *Tchebycheff systems : with applications in analysis and statistics*, John Wiley & Sons: New York.

[9] Kemperman, J.H.B. (1968), "The general moment problem, a geometric approach", *Ann. Math. Staist.*, **39**, pp. 93-122.

[10] MacGillivray, H.L. (1981), "The mean, median, mode inequality and skewness for a class of densities", *Austral. J. Statist.*, **23**, pp. 247-250.

[11] Mulholland, H.P. and Rogers, C.A. (1958), "Representation theorems for distribution functions", *Proc. London Math. Soc.*, **8**, pp. 177-223.

[12] Olshen, R.A. and Savage, L.J. (1970), "A generalized unimodality", *J. Appl. Probability*, **7**, pp. 21-34.

[13] Reiss, R.-D. (1989), *Approximate Distributions of Order Statistics*, Springer-Verlag.

[14] Runnenburg, J. Th. (1978), "Mean, median, mode", *Statist. Neerl.*, **32**, pp. 73-80.

[15] van Zwet, W.R. (1979), "Mean, median, mode, II", *Statist. Neerl.*, **33**, pp. 1-5.

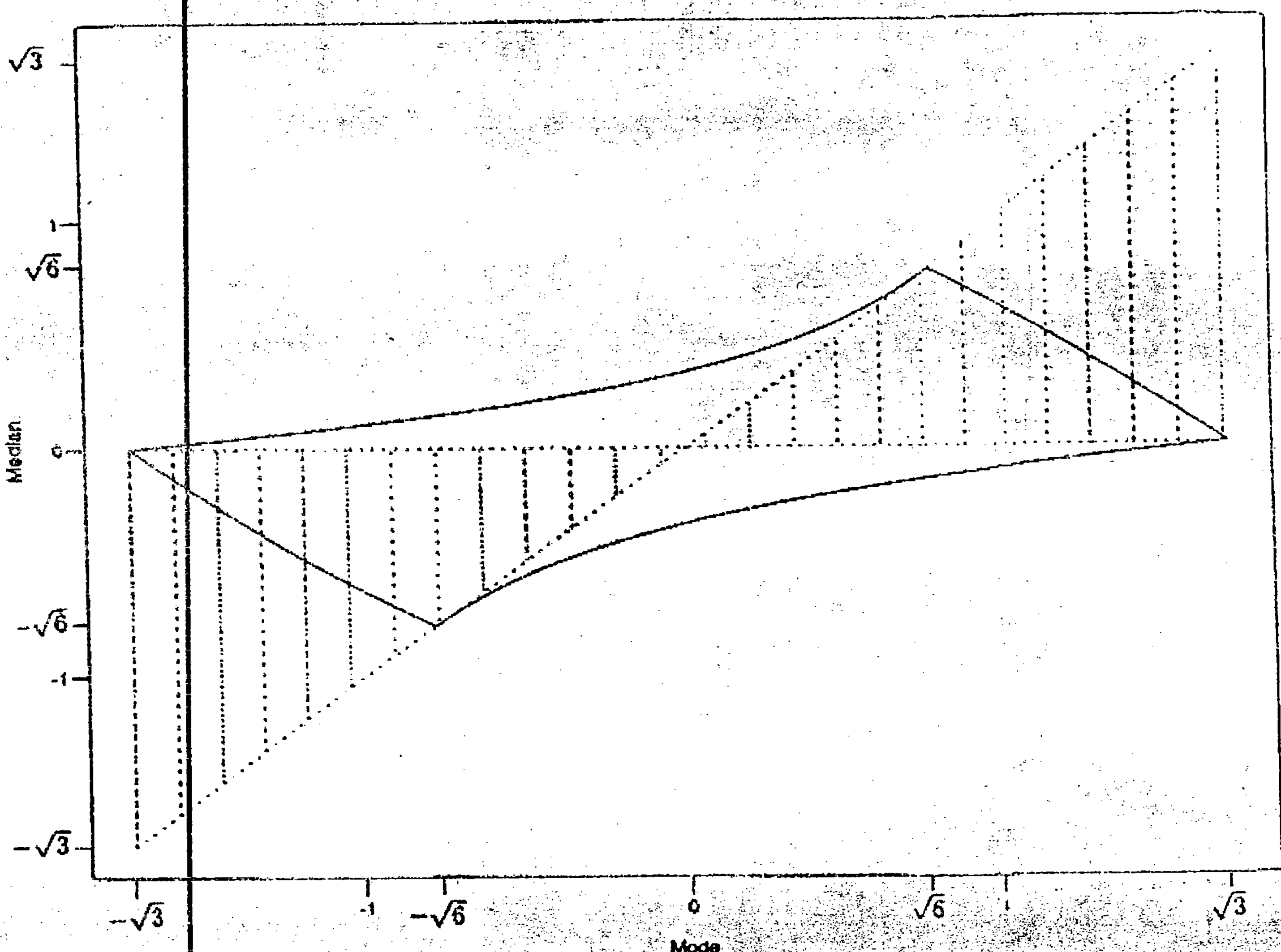

Figure 1: The region bounded by solid lines is the set $\mathcal{M}_1^0$: the set of modes and medians of univariate unimodal distributions when the mean is fixed at 0, and the variance is fixed at 1. The dashed area is the set I^0, where the *mean-median-mode inequality* holds.

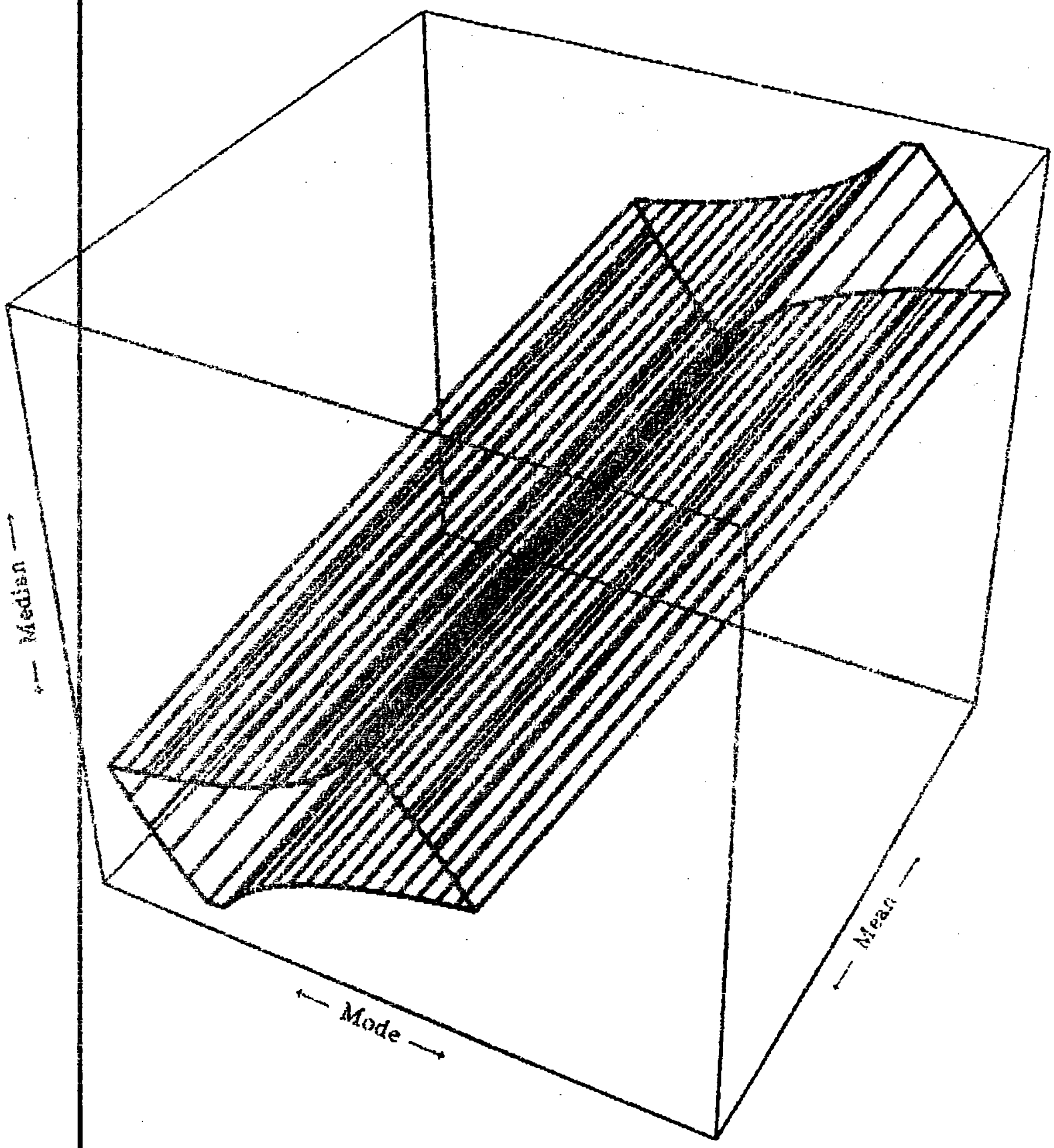

Figure 2: The set $\mathcal{M}_{\sigma=1}$: the set of modes, means and medians of univariate unimodal distributions