The Annals of Statistics
1978, Vol. 6, No. 5, 1159-1164

HOW BROAD IS THE CLASS OF NORMAL SCALE MIXTURES?

BY BRADLEY EFRON[1] AND RICHARD A. OLSHEN[2]

Stanford University and University of California, San Diego

We study the class of scale mixtures of normal distributions with mean zero. Given that the cdf $F(x)$ of such a mixture is fixed at two points, say $F(x_1) = \alpha_1$, $F(x_2) = \alpha_2$, we answer the question of how widely $F(x_3)$ can vary at some third point x_3. A brief final section mentions extensions of our theorem.

1. Introduction and summary. Andrews et. al. [1] restrict their well-known study of robust estimators to error distributions which are mixtures of scaled normal variates with mean zero. A typical such mixture has (right continuous) cdf in the family $\mathscr{F}$,

$$\mathscr{F}: F(x) = \int_{[0,\infty]} \Phi(hx)\mu(dx) \tag{1.1}$$

where Φ is the standard normal cdf, and μ is some *mixing distribution*. For example, if μ itself is taken to be standard normal, then F is the Cauchy cdf. It will be convenient in what follows to let μ put mass points at ∞ and 0, corresponding respectively to mixture components of F with all of their mass at zero or all of their mass placed symmetrically at $\pm\infty$. Formula (1.1) is valid for x finite and nonzero. Some interesting properties of $\mathscr{F}$ are given in [5] and [2].

Normal scale mixtures are attractive for Monte Carlo studies because they are easy to work with, [2] and [7]. It is natural to ask how broad is the class of such distributions. The specific form of the question considered here is as follows: fix[3] two values of x, $0 < x_1 < x_2$, and two percentile values, $\frac{1}{2} < \alpha_1 < \alpha_2$, and let $\mathscr{F}(x_1, x_2; \alpha_1, \alpha_2)$ indicate those members of $\mathscr{F}$ satisfying

$$F(x_1) = \alpha_1, \qquad F(x_2) = \alpha_2. \tag{1.2}$$

Then at a third value of x, say x_3, what are the maximum and minimum of $F(x_3)$ attainable in $\mathscr{F}(x_1, x_2; \alpha_1, \alpha_2)$? In other words, if we fix two percentiles of F, how widely can we vary a third percentile within the class of normal scale mixtures?

Figure 1 illustrates the answer to this question for the class $\mathscr{F}(0.253, 2.400; .6, .9)$. The standard Cauchy is in this class. For x_3 between x_1 and x_2 the set

Received December 1976; revised December 1977.

[1] Research supported in part by National Science Foundation Grant MPS74-21416.

[2] Research supported in part by National Science Foundation Grant MCS76-08314.

[3] Notice that all cdf's (1.1) represent distributions symmetric about 0 so that the conditions $0 < x_1$, $\frac{1}{2} < \alpha$, are not really restrictive. By rescaling we could always take $x_1 = 1$, say, but this would be inconvenient for comparison with standard distributions like the Cauchy.

AMS 1970 *subject classifications.* Primary 62E10, 62G05.

Key words and phrases. Normal scale mixtures, Tchebycheff systems.

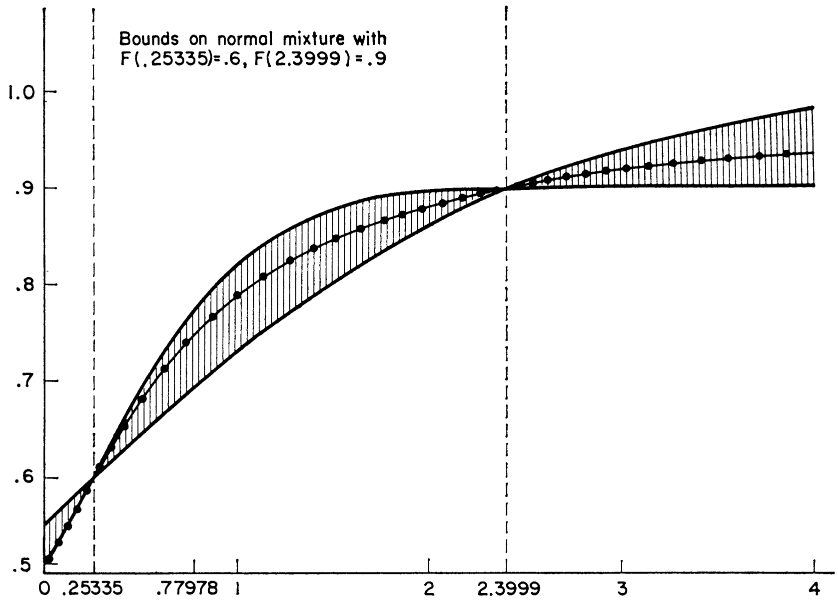

FIG. 1. Bounds on normal mixture cdf's with $F(0.25335) = .6$, $F(2.3999) = .9$. The standard Cauchy is in this class. The upper and lower boundaries correspond to normal mixtures with one point of support in $(0, \infty)$ and one point at either 0 or ∞. The dotted line indicates the Cauchy cdf.

of attainable percentile values forms a "banana" containing the Cauchy cdf. The boundaries of the banana are themselves cdf's. The upper one, between x_1 and x_2, is a two-point mixture putting mass .802 at $h = .796$ and mass .198 at $h = 0$. The lower one puts mass .891 at $h = .506$ and mass .109 at $h = \infty$. It turns out that this simple form of solution holds for any choice of $x_1, x_2, \alpha_1, \alpha_2$.

THEOREM. *There exists a unique F^* in $\mathscr{F}(x_1, x_2; \alpha_1, \alpha_2)$ with mixing distribution μ putting mass on one point h^* in $(0, \infty)$ and on $h = \infty$, such that*

$$(1.3) \qquad \begin{aligned} F^*(x_3) &= \max_{\mathscr{F}(x_1,x_2;\alpha_1,\alpha_2)} F(x_3) \quad \textit{for all} \quad x_3 \in (x_1, x_2) \qquad \textit{and} \\ F^*(x_3) &= \min_{\mathscr{F}(x_1,x_2;\alpha_1,\alpha_2)} F(x_3) \quad \textit{for all positive} \quad x_3 \notin (x_1, x_2)\,. \end{aligned}$$

Likewise there exists a unique F^{**} in $(x_1, x_2; \alpha_1, \alpha_2)$ with μ putting mass at one point h^{**} in $(0, \infty)$ and on $h = 0$, such that

$$(1.4) \qquad \begin{aligned} F^{**}(x_3) &= \min_{\mathscr{F}(x_1,x_2;\alpha_1,\alpha_2)} F(x_3) \quad \textit{for all} \quad x_3 \in (x_1, x_2) \qquad \textit{and} \\ F^{**}(x_3) &= \max_{\mathscr{F}(x_1,x_2;\alpha_1,\alpha_2)} F(x_3) \quad \textit{for all positive} \quad x_3 \notin (x_1, x_2)\,. \end{aligned}$$

The proof of the theorem, which is closely related to the Tchebycheff system considerations of [3] and (implicitly of) [4], is presented in Section 2[4]. Notice

[4] Indeed, a finite set of mean 0, normal densities with different variances is a Tchebycheff system on $[0, \infty)$, but the set of their indefinite integrals is not. It is the latter set with which we are concerned.

that the theorem makes it easy to find the upper and lower bounds. The mixing distribution corresponding to F^* puts mass μ^* at h^* and mass $1 - \mu^*$ at 0, where μ^* and h^* are obtained by numerically finding a solution to the two equations

$$\mu^*\Phi(x_i \cdot h^*) + (1 - \mu^*) \cdot .5 = \alpha_i , \qquad i = 1, 2 . \tag{1.5}$$

Likewise, the equations

$$\mu^{**}\Phi(x_i \cdot h^{**}) + (1 - \mu^{**}) = \alpha_i , \qquad i = 1, 2 \tag{1.6}$$

give F^{**}.

Of course our theorem does not answer the question of whether or not the class of normal mixtures is broad enough for any particular Monte Carlo study. It does give some useful information to help with the answer. For instance, any normal mixture agreeing with the Cauchy at the 60th and 90th percentiles cannot have $F(x) > .77$ at the Cauchy 75th percentile. Extensions of the theorem to the case of more than two restrictions $F(x_i) = \alpha_i$ are easy to make, but will only be sketched here.

2. Proof of the theorem. Fix $0 < x_1 < x_2 < \infty$. As before, let $\mathscr{F} = \mathscr{F}(x_1, x_2; \alpha_1, \alpha_2)$ be the class of scale mixtures of normal distributions for which $E_\mu(\Phi(hx_1)) \equiv \int_{[0,\infty]} \Phi(hx_1)\mu(dh) = \alpha_1$ and $E_\mu(\Phi(hx_2)) = \alpha_2$. Suppose that $\mathscr{F}$ is not empty. Fix x_3, $x_1 < x_3 < x_2$. Let $X = \{x_\mu\}$ be the subset of R^3 with coordinates $(E_\mu(\Phi(hx_3)), E_\mu(\Phi(hx_1)), E_\mu(\Phi(hx_2)))$ as μ ranges over all probabilities on the Borel subsets of $[0, \infty]$. As μ varies over the subset of probabilities which consists of point masses, $\{x_\mu\}$ traces a curve $\mathscr{C}$ in R^3. Because the Borel probabilities on $[0, \infty]$ are all mixtures of these point masses, X is the convex hull of $\mathscr{C}$. What Lemma 2 below says geometrically is that all the points of $\mathscr{C}$ are extreme points of its convex hull.

LEMMA 1. *X is closed, convex, and bounded.*

PROOF. The set of all probabilities on $[0, \infty]$ is (weak*) compact, and for every x, $\mu \to E_\mu(\Phi(x))$ is weak* continuous. This shows that X is closed. Convexity and boundedness are obvious. □

It follows from Lemma 3 below that X has a nonempty interior and is strictly convex (that is, X has no flat spots).

LEMMA 2. *The extreme points of X correspond to all probabilities μ degenerate at a single point.*

PROOF. Fix a point $x_\mu \in X$, where μ is not a point mass. Then write $\mu = \alpha\nu + \bar{\alpha}\lambda$ $(0 < \alpha < 1$, $\bar{\alpha} = 1 - \alpha$, ν and λ probabilities), such that $s \leqq t$ for each $s \in$ support (ν), $t \in$ support (λ). Then $x_\mu = \alpha x_\nu + \bar{\alpha} x_\lambda$ and by the construction of ν and λ, $x_\nu \neq x_\lambda$. Therefore x_μ is not an extreme point of X. It follows that the extreme points of X are of the form $(\Phi(h_0 x_3), \Phi(h_0 x_1), \Phi(h_0 x_2))$ for some h_0,

$0 \leqq h_0 \leqq \infty$. It is obvious that the points corresponding to $h_0 = 0, \infty$, namely $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ and $(1, 1, 1)$, are extreme. Concentrate on h_0 for which $0 < h_0 < \infty$.

We now construct a plane which is tangent to X at precisely the point $(\Phi(h_0 x_3), \Phi(h_0 x_1), \Phi(h_0 x_2))$ of $\mathscr{C}$. The existence of such a plane implies that $(\Phi(h_0 x_3), \Phi(h_0 x_1), \Phi(h_0 x_2))$ is an extreme point of X. Thus, study $L(a, c, h) = L(h) = a\Phi(hx_3) + \Phi(hx_1) - c$ for $a, c \neq 0$. If $L(h_0) = 0$, then $a\Phi(h_0 x_3) + \Phi(h_0 x_1) = c$. If $L'(h_0) = 0$, then $ax_3\varphi(h_0 x_3) + x_1\varphi(h_0 x_1) = 0$. For both L and L' to be 0 at h_0, $a = -x_1/x_3(\varphi(h_0 x_1))/(\varphi(h_0 x_3))$ and $c = \Phi(h_0 x_1) - x_1/x_3(\varphi(h_0 x_1))/(\varphi(h_0 x_3))\Phi(h_0 x_3)$. Now, in an abuse of notation, coordinatize R^3 as (z_3, z_1, z_2) and study

$$L(z_3, z_1, z_2) - z_1 - \frac{x_1}{x_3}\frac{\varphi(h_0 x_1)}{\varphi(h_0 x_3)} z_3 + \frac{x_1}{x_3}\frac{\varphi(h_0 x_1)}{\varphi(h_0 x_3)} \Phi(h_0 x_3) - \Phi(h_0 x_1) .$$

$\{L = 0\}$ is a plane passing through $(\Phi(h_0 x_3), \Phi(h_0 x_1), \Phi(h_0 x_2))$. To show that the stated point is extreme in X it suffices to show $\{L(\Phi(hx_3), \Phi(hx_1), \Phi(hx_2)), h \neq h_0\}$ are all of the same sign. Put

$$f(h) = L(\Phi(hx_3), \Phi(hx_1), \Phi(hx_2))$$

and observe that: (i) f is continuous on $[0, \infty]$; (ii) $f(h_0) = 0 = f'(h_0)$; and (iii) f' is negative for $h < h_0$, and positive for $h > h_0$. Thus, f strictly decreases until h_0, is 0 there, and strictly increases thereafter. In short, $(\Phi(h_0 x_3), \Phi(h_0 x_1), \Phi(h_0 x_2))$ is extreme. □

LEMMA 3. *Let* $0 < w_1 < w_2 < \cdots < w_n < \infty$, *and* $a_1, \cdots, a_n$ *be nonzero constants. Then*

$$g(h) \equiv \sum_1^n a_i \Phi(hw_i) = \text{constant}$$

has at most n solutions in $(0, \infty)$, *and* $g'(h) = 0$ *has at most* $n - 1$ *solutions in* $(0, \infty)$.

PROOF. The assertion regarding g' follows from a reparameterization of Problem 75, page 48 of [6], and the statement concerning g follows easily. □

Let $Q(h) = a_1\Phi(hx_1) + a_2\Phi(hx_2) + a_3\Phi(hx_3)$ for an arbitrary constant c; let r be the number of distinct real roots of $Q(h) = c$ in $[0, \infty]$. Because of the possibility of roots at 0 and ∞, Lemma 3 implies that $r \leqq 5$. The next lemma shows that sometimes more can be said. Define r' as the number of roots in $(0, \infty)$ plus one-half the number of roots at 0 and ∞.

LEMMA 4. *If* $Q(h) \geqq c$ *for all h in* $[0, \infty]$, *then* $r' \leqq \frac{3}{2}$. *Thus* $Q(h) = c$ *has at most two nonnegative roots, one of which is at* 0 *or* ∞.

PROOF. Lemma 3 implies $Q(h) = 0$ has at most 2 distinct roots in $(0, \infty)$. Since $Q(h) \geqq c$, whenever $Q(h) = c$ for $h \in (0, \infty)$ it follows that $Q'(h) = 0$.

In addition, if $h_0 < h_1$ are 2 roots of $Q(h) = c$, then there exists an h^*: $Q'(h^*) = 0$, $h_0 < h^* < h_1$. Therefore, $(r + 1) + (r - \gamma) \leqq 2$, where γ is the number of roots of $Q(h) = c$ at 0 and ∞. Rewording, $r' + \gamma/2 - 1 + r' + \gamma/2 - \gamma \leqq 2$; $2r' \leqq 3$; $r' \leqq \frac{3}{2}$. □

LEMMA 5. *Let*

$$f(x) = a_0 + \sum_{i=1}^{L} a_i \Phi(h_i x)$$
$$g(x) = b_0 + \sum_{j=1}^{L} b_j \Phi(k_j x) .$$

Suppose

(a) *there are* $0 < x_1 < x_2 < \cdots < x_{2L} < \infty$ *for which* $f(x_i) = g(x_i)$, *and that*
(b) $f'(x_i) = g'(x_i)$ *for some* i.

Then $f \equiv g$.

PROOF. It follows from (a) that there exist $y_1, \cdots, y_{2L-1}$, $x_i < y_i < x_{i+1}$, for which $f'(y_i) = g'(y_i)$. (Notice that $f' = (f - a_0)'$, $g = (g - b)'$.) Because also $f'(x_i) = g'(x_i)$ for some i, Lemma 3 implies that $f - a_0 \equiv g - b_0$. Therefore, (a) implies that $f \equiv g$. □

The following argument rephrases remarks of Harris ([3], page 529). Let $\tilde{x}^*$ be a boundary point of X. There is a supporting hyperplane at $\tilde{x}^*$ which also contains the extreme points of X of which $\tilde{x}^*$ is a convex combination. Thus, there are numbers $\alpha_1, \alpha_2, \alpha_3$ and a constant c for which $\sum \alpha_i \tilde{x}_i^* = c$ and $\sum \alpha_i \tilde{x}_{\mu i} \geqq c$ for all $x_\mu \in X$. In particular, $\alpha_i \Phi(hx_1) + \alpha_2 \Phi(hx_2) + \alpha_3 \Phi(hx_3) \geqq c$ for all $h \in [0, \infty]$ with equality holding for those extreme points in the supporting hyperplane containing $\tilde{x}^*$.

In view of Lemmas 3, 4, and 5 and the foregoing remarks, for all x_3, $x_1 < x_3 < x_2$,

$$\max_{\mathscr{F}} F(x_3)$$

is achieved by a unique distribution—call it F^*—corresponding to a distribution μ supported by $r' \leqq \frac{3}{2}$ points, where r' is computed as before. In general, the number of points in the support of μ is simply the number of times the supporting hyperplane containing $\tilde{x}^*$ touches X.

An argument like the last one shows that there is a unique F^{**} which for all x_3, $x_1 < x_3 < x_2$, achieves the

$$\min_{\mathscr{F}} F(x_3) .$$

Because F^* can be seen to minimize $F'(x_2)$ over $\mathscr{F}$, the previous arguments and Lemma 5 serve to show that F^* minimizes $F(x_3)$ over $\mathscr{F}$ for $x_3 > x_2$. In short, there is one "maximizer" inside (x_1, x_2) which is a "minimizer" outside. And there is one "minimizer" inside (x_1, x_2) which is a "maximizer" outside. The minimizing μ inside puts positive mass on $\{h = \infty\}$ whereas μ of the maximizer inside puts positive mass on $\{h = 0\}$. That μ is unique can be seen to follow from various facts, the simplest being the cited Problem 75, page 48 of [6].

3. Extensions. The arguments of Section 2 extend readily to the case of $k > 2$ fixed percentiles. And the obvious analogue to the theorem is true. Thus, there are two unique extremal distributions from among those satisfying the constraints: they respectively minimize, maximize, minimize $\cdots$ and maximize, minimize, maximize $\cdots$; moreover, the analogue to r' does not exceed $(k + 1)/2$.

REFERENCES

[1] ANDREWS, D. F., BICKEL, P. J., HAMPEL, F. R., HUBER, P. J., ROGERS, W. H., and TUKEY, J. W. (1972). *Robust Estimates of Location.* Princeton Univ. Press.

[2] ANDREWS, D. and MALLOWS, C. (1974). Scale mixtures of normal distributions. *J. Royal Statist. Soc. Ser. B* **36** 99–102.

[3] HARRIS, B. (1959). Determining bounds on integrals with applications to cataloging problems. *Ann. Math. Statist.* **30** 521–548.

[4] KARLIN, S. and STUDDEN, W. J. (1966). *Tchebycheff Systems: With Applications in Analysis and Statistics.* Interscience, New York.

[5] KELKER, D. (1971). Infinite divisibility and variance mixtures of the normal distribution. *Ann. Math. Statist.* **42** 802–808.

[6] PÓLYA, G. and SZEGÖ, G. (1925). *Aufgaben und Lehrsätze aus der Analysis*, Band II. Springer, Berlin.

[7] RELLES, D. (1970). Variance reduction techniques for Monte Carlo sampling from Student distributions. *Technometrics* **12** 499–515.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA, SAN DIEGO
LA JOLLA, CALIFORNIA 92093