# SCALE MIXING OF SYMMETRIC DISTRIBUTIONS WITH ZERO MEANS[1]

By E. M. L. Beale[2] and C. L. Mallows[3]

*Princeton University*

**0. Summary.** Suppose that a distribution $A$ is a mixture of distributions similar to $B$ but with different scale parameters; or (almost equivalently) that a distribution $F$ is a convolution of a given distribution $G$ with some other distribution. We derive conditions on (i) the moments of $A$ and $F$ and (ii) on the derivatives of $A$ and $F$; these conditions are necessary, but are not sufficient in general. The conditions (ii) are appropriate when $B$ (or $G$) is of Pólya type 3.

**1. Introduction.** Suppose $A(x)$ and $B(x)$ are cumulative distribution functions (c.d.f.'s) on the real line, continuous on the right, and a.e. symmetric about the origin, so that

$$A(x) + A(-x - 0) = 1 = B(x) + B(-x - 0), \qquad -\infty < x < \infty. \tag{1}$$

We write $X_A$ for a random variable (r.v.) having the c.d.f. $A(x)$, and similarly for $X_B$. We shall say that $A(x)$ is a $B$-mixture if there exist r.v.'s $X_A$, $X_B$, and $Y$, where $Y$ is non-negative and independent of $X_B$, such that

$$X_A = X_B Y \tag{2}$$

or equivalently, if there exists a c.d.f. $C(\sigma^2)$ on $[0, \infty)$, continuous on the right, such that

$$A(x) = \int_0^\infty B(x/\sigma)\, dC(\sigma^2), \qquad 0 < x \tag{3}$$

where we interpret $B(x/0)$ as 1 for $x > 0$. It is clear that $A(x)$ is discontinuous at $x = 0$ if $C(0) \neq 0$.

In a closely related situation (see Section 3 below), if $F(x)$ and $G(x)$ are c.d.f.'s on the real line (not necessarily symmetric), we shall say that $F$ is a $G$-convolution if there exist r.v.'s $X_F$, $X_G$, and $Z$ ($Z$ having c.d.f. $H(x)$ and being independent of $X_G$, not necessarily non-negative) such that

$$X_F = X_G + Z. \tag{4}$$

Some general theorems concerning the existence and measurability of functions related to mixtures of distributions were proved by Robbins [3]. Teichroew [6] considered the case where $B(x)$ is the unit Normal c.d.f., and $C(\sigma^2)$ is of Pearson type III.

---

Received September 15, 1958; revised March 10, 1959.

[1] Research at Princeton University sponsored by the Office of Ordnance Research, U. S. Army; Statistical Techniques Research Group, contract No. DA-36-034-ORD 2297.

[2] Now at Admiralty Research Laboratory, Teddington, Middlesex, England.

[3] Now at University College London, England.

Our interest in the mixture problem arose out of some research where it was possible to prove that a certain procedure was optimal whenever an error distribution was Normal with zero mean but arbitrary variance, and also whenever it was a mixture of such distributions. It was thus of interest to determine as far as possible the properties of such mixtures. In general, given $A$ and $B$, we would like to be able to determine whether or not $A$ can be regarded as a $B$-mixture; and similarly for the convolution problem.

Hirschman and Widder [1] investigate (4) at great length, but their results are not in the form we desire; thus for the case where $X_G$ is normal (with mean zero and variance $v$, say) they give two sets of necessary and sufficient conditions for (4) to hold. The first of these ([1] Theorem 12.2) requires a knowledge of $\partial F(x)/\partial v$, and the second ([1] Theorem 12.4) achieves the inversion of (4) by means of an infinite series of derivatives of $F(x)$; the required condition is that the sum of this series be everywhere non-decreasing (i.e. gives a c.d.f.). This last formula has been much used in practice; see e.g. Smart [5].

We assume that the distribution $A$ (or $F$) is completely known; we do not say anything about the statistical problems of testing whether a random sample can reasonably be assumed to come from some B-mixture, and if so of estimating the mixing distribution $C$. Robbins [4] considers this estimation problem. He remarks that it is of considerable importance in other connections, but awaits a satisfactory practical solution.

In Section 2 we derive necessary and sufficient conditions for the existence of some $A$ that is a $B$-mixture having given moments through order $2r$. In Section 3 we examine the relation between the mixture problem and the convolution problem. In Section 4 we obtain a necessary condition for a given $A$ to be a $B$-mixture (or for a given $F$ to be a $G$-convolution) in terms of the frequency functions $A'(x)$ (or $F'(x)$) and their derivatives; the validity of these conditions depends on certain properties of the derivatives of $B$ (or $G$), related to the theory of Pólya types; this relation is explored in Section 5.

**2. Conditions on moments.** From (2) we have immediately that if $A$ is a $B$-mixture, then

$$E(X_A^{2r}) = E(X_B^{2r})E(Y^{2r}) \tag{5}$$

and the l.h.s. exists if and only if each of the factors on the r.h.s. is finite. Since $Y^2$ is to be a r.v. on $[0, \infty]$, its moments must satisfy certain inequalities, the simplest of which is the obvious one $E(Y^4) \geqq \{E(Y^2)\}^2$. Hence we obtain necessary relations between the moments of $A$ and $B$; the simplest is

$$\mu_4(A)/\mu_2(A)^2 \geqq \mu_4(B)/\mu_2(B)^2. \tag{6}$$

so that the kurtosis of a mixture is never less than the kurtosis of a single component. Conversely, these relations are sufficient for the existence of some distribution $A(x)$ that is a $B$-mixture, having the given moments.

The result that a mixture of Normal distributions (with zero means) is necessarily leptokurtic (unless it reduces to a single Normal distribution) seems to be widely known, though apparently unpublished. It is worth bearing in mind when considering the argument that practical error distributions "must" tend to Normality because the error is the sum of many independent components. It is arguable that many error distributions are mixtures of distributions with a common mean but different variances, and can therefore be expected to be leptokurtic.

**3. The distribution of** $\ln |X_A|$. Another simple line of approach to the mixture problem is to consider the distribution of $\ln |X_A|$. Before we can do this we must consider the probability that $X_A = 0$, since $\ln X_A$ is then undefined. Writing $A_0 = \Pr \{X_A \neq 0\}$ and similarly for $B_0$ and $C_0$, we have immediately from (2) that

$$A_0 = B_0 C_0 . \tag{7}$$

Also from (2), conditioned that none of the r.v.'s are zero (i.e. $X_A \neq 0$), we have

$$\ln |X_A| = \ln |X_B| + \ln Y \tag{8}$$

which is exactly (4). Thus we have transformed the mixture problem into the convolution problem. If we define the conditional characteristic functions of $\ln |X_A|$ and $\ln |X_B|$ by

$$\varphi_A(t) = \frac{2}{A_0} \int_{0+}^{\infty} x^{it}\, dA(x), \qquad \varphi_B(t) = \frac{2}{B_0} \int_{0+}^{\infty} x^{it}\, dB(x), \tag{9}$$

then we have from (7) and (8)

THEOREM 1: *A necessary and sufficient condition for A to be a B-mixture is that* $A_0 \leqq B_0$ *and* $\varphi_A(t)/\varphi_B(t)$ *is the ch.fn. of some distribution on* $(-\infty, \infty)$.

In a sense this is the complete answer to the problem, but unfortunately the criterion is not in general easy to apply. In some circumstances a numerical approach based on (8) may be effective. An approach via the moments of $\ln |X_A|$ and $\ln |X_B|$ (similar to that in the previous section) will yield a series of necessary conditions.

**4. Conditions on the frequency function.** We now consider criteria based on derivatives of the c.d.f.'s. Let us assume that $B(x)$ is four times differentiable everywhere, and that $b = B'(x) > 0$ for all $x$. It will follow that any $B$-mixture $A(x)$ is four times differentiable everywhere except perhaps at $x = 0$, and that $a(x) = A'(x) > 0$ wherever this exists.

Now $A$ is assumed to be a mixture of distributions with zero mean and varying scale parameter $\sigma$; so that part of the distribution $A$ near $x = 0$ will consist primarily of those components with small $\sigma$, while the part with $|x|$ large will consist primarily of components with large $\sigma$. We may expect to find a necessary condition for $A$ to be a $B$-mixture based on this fact, and the simplest such condition seems to be the following:

*Conjecture*: If one assumed that only one component contributed to $a(x)$ for a particular $x$, and one estimated the value of $\sigma$ for this component from the values of $a(x)$ and $a'(x)$, then for any $A$ that is a $B$-mixture, the value of $\sigma$ so defined is a non-decreasing function of $|x|$.

The value of $\sigma$ described is defined by the equation

$$(x/\sigma)b'(x/\sigma)/b(x/\sigma) = xa'(x)/a(x). \tag{10}$$

If this equation holds for more than one $\sigma$, we could make the estimate unique by agreeing to take the smallest value satisfying (10). But we can hardly expect the conjecture to be true unless $xb'(x)/b(x)$ is a strictly monotone function of $x$. This is equivalent to the condition that the distribution $B$ is strictly of Pólya type 2 (monotone likelihood ratio) with respect to the parameter $\sigma$, as defined by Karlin [2].

It turns out (see Section 5) that the conjecture is correct if $B$ is also of Pólya type 3 with respect to $\sigma$. Although it is possible to construct symmetrical distributions that are not Pólya type 3 with respect to $\sigma$, almost all the principal cases occurring in statistical practice—such as the Normal, double-exponential, Cauchy, rectangular, triangular—are of this type.

In terms of the distribution of $\ln |X_A|$ and $\ln |X_B|$, the conjecture asserts that if $F$ is a $G$-convolution, and writing $f(x) = F'(x)$, $g(x) = G'(x)$, then the value of $\mu$ defined by

$$g'(x - \mu)/g(x - \mu) = f'(x)/f(x) \tag{11}$$

is a non-decreasing function of $x$. In the following, we shall work in terms of the convolution problem. We shall write

$$R_j(x) = g^{(j)}(x)/g(x) \tag{12}$$

so that

$$R_1' = dR_1/dx = R_2 - R_1^2. \tag{13}$$

THEOREM 2: *If for all* $x$, (i) $g(x) > 0$, (ii) $dR_1/dx < 0$, (iii) $R_2(x)$ *is a convex function of* $R_1(x)$, *then* $\mu$, *defined by* (11), *is a non-decreasing function of* $x$.

*Conversely, given* (i) *and* (ii), *if* $\mu$ *is non-decreasing for all G-convolutions, then* (iii) *must hold.*

The statement of the theorem for the mixture problem, with $\sigma$ defined by (10), is the same as this with the $R$'s defined as

$$R_1(x) = 1 + x\frac{b'(x)}{b(x)}, \qquad R_2(x) = 1 + 3x\frac{b'(x)}{b(x)} + x^2\frac{b''(x)}{b(x)}. \tag{14}$$

PROOF: From (11),

$$R_1(x - \mu) = f'(x)/f(x) \tag{15}$$

$$= \frac{\int R_1(x - m)g(x - m)\, dH(m)}{\int g(x - m)\, dH(m)}, \tag{16}$$

which shows that $\mu$ exists (by (ii)). Multiplying (15) by $f(x)$ and differentiating with respect to $x$, we find

$$f'(x)R_1(x - \mu) + f(x)R_1'(x - \mu)(1 - d\mu/dx) = f''(x)$$

so that (using (13) and (15))

$$\begin{aligned} -f(x)R_1'(x - \mu)\, d\mu/dx &= f''(x) - f(x)R_2(x - \mu) \\ &= \int \{R_2(x - m) - R_2(x - \mu)\}g(x - m)\, dH(m). \end{aligned} \tag{17}$$

Now $f(x) > 0$, and $R_1'(x - \mu) < 0$ by (ii); further, (iii) implies that for each $x - \mu$ there exists some number $k$ (independent of $m$) such that for all $m$

$$R_2(x - m) - R_2(x - \mu) \geqq k\{R_1(x - m) - R_1(x - \mu)\}. \tag{18}$$

But from (16)

$$\int \{R_1(x - m) - R_1(x - \mu)\}g(x - m)\, dH(m) = 0 \tag{19}$$

so that the r.h.s. of (17) is $\geqq 0$, and $d\mu/dx \geqq 0$ as required.

Conversely, suppose (iii) is false. We shall construct a $G$-convolution which has $d\mu/dx < 0$ at $x = 0$. By our assumption, there exist $m_1$, $m_2$, $\mu$ (with $m_1 < \mu < m_2$) such that

$$\tfrac{1}{2}\{R_1(-m_1) + R_1(-m_2)\} = R_1(-\mu), \tag{20}$$

$$\tfrac{1}{2}\{R_2(-m_1) + R_2(-m_2)\} < R_2(-\mu). \tag{21}$$

Now choose $H(m)$ so that $dH(m)/dm = 0$ except at $m = m_1$ and $m_2$, with

$$dH(m_i) = \frac{1}{g(-m_i)}\left\{\frac{1}{g(-m_1)} + \frac{1}{g(-m_2)}\right\}^{-1}, \qquad i = 1, 2. \tag{22}$$

Then by (20), (16) is satisfied (for $x = 0$), and by (21), the r.h.s. of (17) is $<0$, so that $d\mu/dx < 0$.

Theorem 2 provides us with a necessary condition for $F$ to be a $G$-convolution (or for $A$ to be a $B$-mixture); namely, the $\mu$ (or $\sigma$) defined by (11) (or (10)) must be a non-decreasing function of $x$. Unfortunately it will not provide a sufficient condition unless $R_2$ is a linear function of $R_1$; and this is impossible over the whole range of $x$. ($R_2$ can be a piecewise linear function of $R_1$ if we allow $d^3g/dx^3$ to be discontinuous.) However, relaxing condition (i) of the theorem, we can obtain distributions for which $R_2$ is a linear function of $R_1$ wherever $g(x) > 0$; two such distributions for the mixtures problem are the rectangular and the triangular. It is easy to verify that a necessary and sufficient condition for $A$ to be a mixture of rectangular distributions is that $A$ be unimodal; and that necessary and sufficient conditions for $A$ to be a mixture of triangular distributions are that $A$ should have a derivative $a(x)$ everywhere except possibly at $x = 0$, while $a_+'(x)$ exists and is non-positive and non-decreasing for all $x > 0$. If $b(x)$ (or $g(x)$) $> 0$ for all $x$, then, for example, no distribution $A$ for which the estimate of $\sigma$ is constant for all $x > x_0$, but takes a different value for some smaller value of $x$, can be a $B$-mixture.

If $B$ is a Normal distribution (for which the conditions of Theorem 2 are satisfied), the result can be utilized in the form of a log/square plot, in which $\log a(x)$ is plotted as a function of $x^2$. It is easy to see that the slope of this curve is inversely proportional to the estimate of $\sigma$; so the theorem shows that a necessary condition for the given distribution $A(x)$ to be a mixture of Normal distributions is that the log/square plot be convex.

**5. Relation to theory of Pólya types.** The conditions required in Theorem 2 can be expressed in terms of the determinants

$$\Delta_n = \left| \frac{\partial^i}{\partial x^i} \frac{\partial^j}{\partial \mu^j} p(x, \mu) \right|_{i,j=0}^{n-1} \tag{23}$$

for $n = 1, 2, 3$ with $p(x, \mu) = g(x - \mu)$. We shall prove

THEOREM 3: *The conditions* (i), (ii), (iii) *of Theorem* 2 *are equivalent to*

$$\Delta_1 > 0, \qquad \Delta_2 > 0, \qquad \Delta_3 \geqq 0.$$

PROOF: It is easy to see that the signs of these determinants are unaffected by a monotonic increasing transformation of either the independent variable $x$ or the parameter $\mu$; so that a proof of the theorem for the convolution problem will imply the corresponding result for the mixtures problem also. In the following, the argument of all the functions involved is $x - \mu$.

For $n = 1$, (23) gives $g > 0$, which is (i). For $n = 2$, (23) gives

$$\begin{vmatrix} g & -g' \\ g' & -g'' \end{vmatrix} > 0, \quad \text{i.e.} \quad \begin{vmatrix} 1 & R_1 \\ R_1 & R_2 \end{vmatrix} < 0, \tag{24}$$

i.e. by (13), $R_1' < 0$, which is (ii). Now

$$d^2R_2/dR_1^2 = (R_1')^{-3}(R_1'R_2'' - R_1''R_2') \tag{25}$$

so that condition (iii) is equivalent to

$$\begin{vmatrix} 1 & 0 & 0 \\ R_1 & R_1' & R_1'' \\ R_2 & R_2' & R_2'' \end{vmatrix} \leqq 0 \tag{26}$$

By differentiation we have successively

$$\begin{aligned} g' &= gR_1, & g'' &= g'R_1 + gR_1', & g''' &= g''R_1 + 2g'R_1' + gR_1'', \\ g'' &= gR_2, & g''' &= g'R_2 + gR_2', & g'''' &= g''R_2 + 2g'R_2' + gR_2''. \end{aligned} \tag{27}$$

Hence manipulating the determinant in (26) according to the scheme

$$\begin{aligned} (\text{col } 3)' &= g(\text{col } 3) + 2g'(\text{col } 2) + g''(\text{col } 1) \\ (\text{col } 2)' &= g(\text{col } 2) + g'(\text{col } 1) \end{aligned} \tag{28}$$

we obtain

$$(29)\qquad \begin{vmatrix} g & g' & g'' \\ g' & g'' & g''' \\ g'' & g''' & g'''' \end{vmatrix} \leqq 0$$

which is equivalent to $\Delta_3 \geqq 0$.

Karlin [2] calls a family of distributions

$$(30)\qquad P(x, \mu) = \beta(\mu) \int_{-\infty}^{x} p(x, \mu)\, d\lambda(x)$$

of Pólya type $m$ (strictly of Pólya type $m$) if the determinants

$$(31)\qquad D_n = | p(x_i, \mu_j)|_{i,j=1}^{n}$$

are $\geqq 0$ ($> 0$) for $n = 1, 2, \cdots, m$, for all

$$(32)\qquad x_1 < x_2 < \cdots < x_n, \qquad \mu_1 < \mu_2 < \cdots < \mu_n .$$

Karlin shows that Pólya $m$ implies $\Delta_n \geqq 0$ ($n = 1, \cdots, m$), while $\Delta_n > 0$ ($n = 1, \cdots, m$) implies strict Pólya $m$.

We are indebted to the referee for the following remarks. One can derive only $\Delta_n \geqq 0$ when assuming strict Pólya $m$, with $\Delta_n > 0$ for almost all $x$ and $\mu$. It is true however that if $p(x, \mu) = p(x - \mu)$ (as is the case in the present problem), then the equivalence is correct. This last result is quite deep and is not published in the literature. Most strict Pólya type distributions satisfy $\Delta_n > 0$ everywhere, but there may be isolated points where equality takes place.

Thus our conditions (i) and (ii) are equivalent to strict Pólya 2, and (iii) is implied by Pólya 3.

Karlin [2] remarks that if $\Delta_n \geqq 0$ ($n = 1, \cdots, m$) with strict inequality almost everywhere, then under a certain weak assumption the convolution of $G(x)$ with a Normal distribution of arbitrarily small variance $\sigma^2$ will be strictly Pólya $m$, and hence (taking the limit as $\sigma^2$ tends to zero) $G$ will be Pólya $m$. In such cases Theorem 2 can still be applied, provided that, whenever (11) does not define $\mu$ uniquely, $\mu$ is taken as the appropriate limit as $\sigma^2$ tends to zero.

The authors are grateful to the referee for his suggestions for improving the presentation of the paper, and for clarifying the situation in the last section.

## REFERENCES

[1] I. I. Hirschman and D. V. Widder, *The Convolution Transform*, Princeton University Press, Princeton, 1955.

[2] Samuel Karlin, "Pólya type distributions, II", *Ann. Math. Stat.* Vol. 28 (1957), pp. 281–308.

[3] Herbert Robbins, "Mixture of distributions," *Ann. Math. Stat.* Vol. 19 (1948), pp. 360–369.

[4] Herbert Robbins, "An empirical Bayes approach to statistics," *Third Berkeley Symposium on Mathematical Statistics and Probability* Vol. 1 (1956), pp. 157–163.

[5] W. M. Smart, *Stellar Dynamics*, Cambridge University Press, 1938.

[6] D. Teichroew, "The mixture of normal distributions with different variances," *Ann. Math. Stat.* Vol. 28 (1957), pp. 510–512.