ON THE MIXTURE OF DISTRIBUTIONS*

By Henry Teicher

Purdue University

Summary. If $\mathfrak{F} = \{F\}$ is a family of distribution functions and μ is a measure on a Borel Field of subsets of $\mathfrak{F}$ with $\mu(\mathfrak{F}) = 1$, then $\int F(\cdot)\, d\mu(F)$ is again a distribution function which is called a μ-mixture of $\mathfrak{F}$. In Section 2, convergence questions when either F_n or μ_k (or both) tend to limits are dealt with in the case where $\mathfrak{F}$ is indexed by a finite number of parameters. In Part 3, mixtures of additively closed families are considered and the class of such μ-mixtures is shown to be closed under convolution (Theorem 3). In Section 4, a sufficient as well as necessary conditions are given for a μ-mixture of normal distributions to be normal. In the case of a product-measure mixture, a necessary and sufficient condition is obtained (Theorem 7). Generation of mixtures is discussed in Part 5 and the concluding remarks of Section 6 link the problem of mixtures of Poisson distributions to a moment problem.

1. Introduction. Let $\mathfrak{F} = \{F\}$ be a family of one[1] dimensional cumulative distribution functions (c.d.f.'s), and let $\mathfrak{M} = \{\mu\}$ be a class of measures defined on $\mathfrak{A}$, a Borel Field of subsets of $\mathfrak{F}$, with $\mu(\mathfrak{F}) = 1$, all $\mu \in \mathfrak{M}$. ($\mathfrak{A}$ may be taken to be the smallest sigma-Algebra containing sets $A_{x,y} = \{F \mid F(x) \leqq y, F \in \mathfrak{F}\}$). Then, [9], [11], $\int g(F)\, d\mu(F)$ is defined in the usual manner for measurable mappings g of $\mathfrak{F}$ into the real line. If $g = g_x(F) = F(x)$, this becomes

$$H = H(x) = \int_{\mathfrak{F}} F(x)\, d\mu(F). \tag{1}$$

The resultant distribution function H will be called a "mixture" or more specifically a μ-mixture of $\mathfrak{F}$ providing the "mixing measure" μ does not assign measure one to a particular member of $\mathfrak{F}$. Thus, the term mixture[2], as employed here, signifies a genuine weighted average of c.d.f.'s.

For a stipulated $\mathfrak{F}$, the family $\mathfrak{H} = \mathfrak{H}(\mathfrak{F})$ of mixtures H, swept out as μ varies over $\mathfrak{M}$, will be called the class of $\mathfrak{M}$-mixtures of $\mathfrak{F}$ or (if $\mathfrak{M}$ is definitive in some sense) simply the class of mixtures of $\mathfrak{F}$.

In particular, the family $\mathfrak{F}$ may be indexed by a finite number of parameters $\alpha_1, \alpha_2, \cdots, \alpha_m$ each α_i varying over the real line, that is,

$$\mathfrak{F} = \{F(x; \alpha_1, \alpha_2, \cdots, \alpha_m)\}.$$

Received January 9, 1959; revised October 10, 1959.

* This work was supported in part by an Office of Naval Research Contract.

[1] $\mathfrak{F}$ may also be taken to be a family of r-dimensional c.d.f.'s for any positive integer r.

[2] Actually, (1) corresponds to what Bourbaki [3] calls "integration of measures". However, except for convergence, the questions considered here have no contact with those of [3].

Let $\mathcal{G} = \{G(\alpha_1, \alpha_2, \cdots, \alpha_m)\}$ denote the class of m-dimensional c.d.f.'s and $F(x; \alpha_1, \alpha_2, \cdots, \alpha_m)$ be measurable on $(m + 1)$-dimensional Euclidean space R^{m+1}. Then, $\mathfrak{M}$ may be taken to be the class of Lebesgue-Stieltjes measures $\{\mu_G\}$ on R^m induced by $G \varepsilon \mathcal{G}$ and (1) becomes

$$H(x) = \int_{R^m} F(x; \alpha_1, \alpha_2, \cdots, \alpha_m)\, dG(\alpha_1, \alpha_2, \cdots, \alpha_m). \tag{2}$$

Similarly, $\mathfrak{F}$ may be $\{F(x; \alpha_1, \cdots, \alpha_m)\} = \{F(x; \alpha)\}$ where now $\alpha = (\alpha_1, \cdots, \alpha_m)$ is restricted to R_1^m, some measurable subset of R^m. However, since μ_G will assign zero measure to $R^m - R_1^m$, one may define $F(x; \alpha)$ to be an arbitrary c.d.f. for $\alpha \varepsilon R^m - R_1^m$. Then (1) again takes the form (2), the class $\mathcal{G}$ (or $\mathfrak{M}$) being suitably restricted.

If $\{F(x; \alpha)\}$ is a discrete family whose discontinuity points are independent of $\alpha_1, \cdots, \alpha_m$, then the resultant distribution under mixture on α will be discrete, inheriting the common points of discontinuity. The situation may be otherwise if the discontinuity points vary with α. Thus, if $m = 1$ and $F(x; \alpha)$ has unit saltus at $x = \alpha$, $H(x)$ is continuous if, and only if G is, since $H = G$. In general, x_0 is a discontinuity point of $H(x)$ if and only if the α-set for which $F(x; \alpha)$ is discontinuous at x_0 has positive μ-measure.

On the other hand, if $F(x; \alpha)$ is absolutely continuous for every $\alpha \varepsilon R^m$ then $f(x; \alpha) = \partial/\partial x F(x; \alpha)$ is measurable on R^{m+1}, whence, by Fubini's theorem, the resultant mixture $H(x)$ is absolutely continuous with a density $h(x)$ given by $\int_{R^m} f(x; \alpha)\, dG(\alpha)$. In other words, for an absolutely continuous family $\{F(x; \alpha)\}$, h is a μ_G (or simply G)–mixture of $\{f(x; \alpha)\}$.[3] Conversely, if a probability density function (p.d.f.) $h(x)$ is (merely almost everywhere) a G-mixture of a family of p.d.f.'s $\{f(x; \alpha)\}$, with $f(x; \alpha)$ measurable on R^{m+1}, its c.d.f. $H(x)$ will be a G-mixture of the corresponding family of c.d.f.'s $\{F(x; \alpha)\}$.

It follows directly from Theorem 5 of [13] or the fact that $F(x; \alpha)$ and $G(\alpha)$ determine a joint distribution that if H is a G-mixture of $\mathfrak{F}$ (of the form (2)), then its characteristic function $\varphi(t)$ is a G-mixture[3] of the class $\mathfrak{F}^* = \{\varphi(t; \alpha)\}$ of Fourier transforms of the elements of $\mathfrak{F}$. Similarly, any existing moment of H is a G-mixture of the family of moments (of the same order) of $\mathfrak{F}$ (which exist except perhaps for a set of μ_G-measure zero). If $G(\alpha_1, \cdots, \alpha_m) = \prod_{i=1}^m G_i(\alpha_i)$, the mixture will be termed a "product measure" mixture. Analogously, we may speak of a discrete or absolutely continuous mixture according as $G(\alpha_1, \cdots, \alpha_m)$ (or μ) is a discrete or absolutely continuous c.d.f. (measure). Finally, a finite (countable) mixture is one for which μ is discrete and assigns measure one to a finite (countable) set of points.

A question of importance concerning mixtures is that of unique characterization. That is, for a specific family $\mathfrak{F}$, which distributions H uniquely determine the mixing measure μ. In this connection, we give the following

Definition: A μ-mixture of $\mathfrak{F}$, say H, will be called "identifiable" if, for *any*

[3] Here, we have tacitly extended the terminology "μ-mixture of $\mathfrak{F}$" to cases where the family $\mathfrak{F}$ has as elements functions $f(x; \alpha)$ which, for each $\alpha \varepsilon R^m$, are not c.d.f.'s in the remaining variable.

probability measure μ^*, the relationship $H(x) = \int F(x)\, d\mu(F) = \int F(x)\, d\mu^*(F)$ implies $\mu = \mu^*$. If every member of a class $\mathcal{H}$ of μ-mixtures of $\mathfrak{F}$ is identifiable, $\mathcal{H}$ itself will be called identifiable.

In numerous problems in probability and statistics, one is interested in the distribution of a random variable X but knows only the conditional distributions of X given the values of some auxiliary random variable Y. Then the desired distribution of X is simply a mixture of the known conditional distributions. Similarly, one may know the limit of a sequence of distributions of random variables X_n for all fixed values of other random variables Y_n as well as the limiting distribution of the Y_n (say G), whereas one requires the limiting distribution of the X_n. Under certain conditions, (see especially Theorem 2 and propositions A, B of Section 2) the latter will be a G-mixture of the former.

It is surprising, therefore, that, except for the special case $F(x; \alpha) = F(x - \alpha)$ of convolution, general properties of mixtures have received relatively little attention. A treatment of mixtures appears in [13] and specific mixture problems are dealt with in [8] and [14]. The compound Poisson distributions (see e.g. [8]) are precisely mixtures of Poisson distributions which are necessarily (by a prior remark) discrete distributions with jumps at the non-negative integers.

Mixtures of distributions are of interest for reasons other than those already cited. For example, in the course of determining limit distributions of sums of interchangeable random variables [2] mixtures of normal distributions are encountered and this provides one of several motivations for a study of such creatures.

2. Convergence of mixtures. If $G(\alpha)$, $G_k(\alpha)$, $k = 1, 2, \cdots$, is a sequence of c.d.f.'s such that $G_k(\alpha)$ converges to $G(\alpha)$ on all continuity intervals of G (equivalently, $\lim_{k\to\infty} \int f(\alpha)\, dG_k(\alpha) = \int f(\alpha)\, dG(\alpha)$ for every bounded continuous function $f(\alpha)$), we write $G_k \Rightarrow G$.

Let $F(x; \alpha)$, $F_n(x; \alpha)$, $n = 1, 2, \cdots$, be a sequence of families of c.d.f.'s (all functions of x, α are supposed measurable on R^{m+1}) such that $F_n(x; \alpha) \Rightarrow F(x; \alpha)$, all $\alpha \,\varepsilon\, R^m$; similarly, let $G_k(\alpha) \Rightarrow G(\alpha)$ and define H_{nk} (H) to be a μ_{G_k} (μ_G)-mixture of $\{F_n(x; \alpha)\}$ $(\{F(x; \alpha)\})$, that is,

$$H_{nk}(x) = \int F_n(x; \alpha)\, dG_k(\alpha), \qquad H(x) = \int F(x; \alpha)\, dG(\alpha).$$

In case $G_k \equiv G$ $(F_n \equiv F)$, we write $H_{n\cdot}$ $(H_{\cdot k})$ for H_{nk}. As indicated in Section 1, it is the convergence of the diagonal sequence $H_{nn} \Rightarrow H$ that is of special interest in probability and statistics. However, it seems pertinent to cite more general results. (A related but different convergence question is treated in [17a].)

We first consider the simple cases of $H_{n\cdot}$ and $H_{\cdot k}$. It follows from the dominated convergence theorem and a remark of the preceding section concerning the relationship between discontinuities of H and those of the family $\mathfrak{F}$ that $H_{n\cdot} \Rightarrow H$. On the other hand, $H_{\cdot k}$ need not converge to H as the following example shows:

Take $m = 1$ and let $G_k(\alpha)$ be a step function with jumps of $k^{2j}e^{-k^2}/j!$ at the points $\alpha = -k + j/k$, $j = 0, 1, 2, \cdots$. By a classical result,

$$\lim_{k\to\infty} G_k(\alpha) = (1/\sqrt{2\pi}) \int_{-\infty}^{\alpha} e^{-y^2/2}\, dy.$$

Choose $\mathfrak{F}$ so that $F(x; \alpha) = F_1(x)$ or $F_2(x)$ as α is rational or irrational. Clearly $H_{.k} \equiv F_1$, while $H = F_2$. Thus if $F_1 \not\equiv F_2$, $H_{.k} \not\Rightarrow H$. The source of trouble is thereby indicated, leading to

THEOREM 1: *If, for each continuity point x_0 of $H(x)$,*

$$\mu_G\{\alpha \mid F(x_0 ; \alpha) \text{ is discontinuous}\} = 0, \text{ then } H_{.k} \Rightarrow H.$$

PROOF: The theorem is an immediate consequence of an extension of the Helly-Bray theorem, which, in turn, follows directly from known results. For example, Theorem 2.1 of [1] (see also [4]) insures that if, for a sequence of random vectors X_k defined on a probability space,

$$G_k(\alpha) = P\{X_k < \alpha\} \Rightarrow G(\alpha) = P\{X < \alpha\},$$

then $F_k(\alpha) = P\{h(X_k) < \alpha\} \Rightarrow P\{h(X) < \alpha\} = F(\alpha)$, provided only that the set of discontinuities of the measurable function h has μ_G-measure zero. But if h is also bounded, the rth moments of $h(X_k)$ converge to the rth moments of $h(X)$. For $r = 1$, this shows, for any bounded measurable function $h(\alpha)$ whose discontinuity set has μ_G-measure zero, that $G_k \Rightarrow G$ implies

$$\lim_{k\to\infty} \int_{R^m} h(\alpha)\, dG_k(\alpha) = \lim_{k\to\infty} \int_{R^1} y\, dF_k(y) = \int_{R^1} y\, dF(y) = \int_{R^m} h(\alpha)\, dG(\alpha).$$

Applying this result to $h(\alpha) = F(x_0 ; \alpha)$, the theorem follows.

In the double sequence case, it follows along the lines of [10, Theorem 26, p. 284] that, if the total variation $V[G_k - G] \to 0$, then $H_{nk} \Rightarrow H$ $(n, k \to \infty)$. But, as in the example, $G_k \Rightarrow G$ is compatible with $V[G_k - G] \equiv 2$.

Now if $\mathfrak{B}$ denotes the class of Borel sets of R^m and $G(\alpha)$, $G_k(\alpha)$ are absolutely continuous with densities $g(\alpha)$, $g_k(\alpha)$, $k = 1, 2, \cdots$, such that

$$g_k(\alpha) \to g(\alpha)$$

pointwise, then

$$V[G_k - G] = 2 \sup_{B\varepsilon\mathfrak{B}} \mid \mu_{G_k}(B) - \mu_G(B) \mid = 2 \sup_{B\varepsilon\mathfrak{B}} \left| \int_B g_k(\alpha)\, d\alpha - \int_B g(\alpha)\, d\alpha \right| \to 0$$

since [15] $\lim_{k\to\infty} \int_B g_k(\alpha)\, d\alpha = \int_B g(\alpha)\, d\alpha$ uniformly in B. Consequently,

A. If $F_n(x; \alpha) \Rightarrow F(x; \alpha)$ and $g_k(\alpha) = G_k'(\alpha) \to G'(\alpha) = g(\alpha)$ for all $\alpha \varepsilon R^m$, then $H_{nk} \Rightarrow H$.

B. Let $G(\alpha)$, $G_k(\alpha)$ be discrete c.d.f.'s with $G_k(\alpha) \Rightarrow G(\alpha)$, $F_n(x; \alpha) \Rightarrow F(x; \alpha)$ all α. If, for every point α' of positive mass of G, the mass of G_k at α' converges to that of G, then $H_{nk} \Rightarrow H$.

Proposition B follows in the same fashion as A since a slight extension of Scheffé's theorem [15] yields an analogue for discrete c.d.f.'s under the stated assumption.

We note that the additional proviso is automatically insured by prior assumptions if the discontinuity points of G have no finite limit point and are not themselves limit points of discontinuities of $\{G_k\}$.

Next, the rather stringent condition $V[G_k - G] \to 0$ may be replaced as follows:

THEOREM 2: *Let $G_k(\alpha) \Rightarrow G(\alpha)$, and let R_1^m be a measurable set of R^m with $\mu_G\{R_1^m\} = 1$. In order that $H_{nk} \Rightarrow H (n, k \to \infty)$ it is sufficient that for each continuity point x_0 of $H(x)$*

(i) $\mu_G[\alpha \mid F(x_0 ; \alpha)$ *is discontinuous*$] = 0$

(ii) $\lim_{n\to\infty} F_n(x_0 ; \alpha) = F(x_0 ; \alpha)$ *uniformly in $S \cdot R_1^m$ for every closed bounded α-rectangle S of R^m.*

PROOF: For arbitrary $\epsilon > 0$, choose the "continuity rectangle" A such that $\mu_G(A) > 1 - \epsilon$ and let $\bar{A}$ denote its closure. Then if $B = R^m - A \cdot R_1^m$,

$$| H_{nk}(x_0) - H(x_0) | \leqq \left| \int [F_n(x_0 ; \alpha) - F(x_0 ; \alpha)] \, dG_k(\alpha) \right|$$

$$+ \left| \int F(x_0 ; \alpha) \, dG_k(\alpha) - \int F(x_0 ; \alpha) \, dG(\alpha) \right| \leqq \int_{\bar{A} R_1^m} | F_n(x_0 ; \alpha) - F(x_0 ; \alpha) | \, dG_k(\alpha)$$

$$+ 2 \int_B dG_k(\alpha) + \left| \int F(x_0 ; \alpha) \, dG_k(\alpha) - \int F(x_0 ; \alpha) \, dG(\alpha) \right| \leqq \epsilon + 4\epsilon + \epsilon = 6\epsilon$$

for sufficiently large n and k by Theorem 1 and (ii).

3. Mixtures of additively closed families. We recall [18] that a family $\mathfrak{F} = \{F(x; \alpha)\} = \{F(x; \alpha_1, \cdots, \alpha_m)\}$ where α_j varies over an additive abelian semi-group D_j', $j = 1, 2, \cdots, m$, is called "additively closed" if for every admissible α, β,

$$F(x; \alpha) * F(x; \beta) = F(x; \alpha + \beta) \tag{3}$$

where, as usual, $*$ denotes the convolution operation. The families of normal, Poisson, binomial and many other distributions are encompassed within this definition.

Suppose that D_j denotes either R^1 or some measurable subset of R^1 that is an additive Abelian semi-group, $j = 1, 2, \cdots, m$, and that μ_G assigns measure one to $D = D_1 \times D_2 \times \cdots \times D_m$. Then, under (3), $\int_D F(x; \alpha) \, dG(\alpha)$ is a mixture of the additively closed family $\mathfrak{F} = \{F(x; \alpha)\}$.

THEOREM 3: *Let H_i be a G_i-mixture of the additively closed family $\mathfrak{F}$, $i = 1, 2$. Then the convolution $H_1 * H_2$ is a $(G_1 * G_2)$-mixture of $\mathfrak{F}$. Conversely, if for some $r \geqq 1$ and all G_1, G_2 having exactly r points of positive mass, the convolution of a G_1-mixture of $\mathfrak{F}$ with a G_2-mixture of $\mathfrak{F}$ is a $(G_1 * G_2)$-mixture of $\mathfrak{F}$, then $\mathfrak{F}$ is additively closed.*

PROOF: Let $H = H_1 * H_2$, $G = G_1 * G_2$ and denote by $\varphi(t)$, $\varphi_1(t)$, $\varphi_2(t)$ and $\varphi(t; \alpha)$ the characteristic functions (c.f.'s) respectively of H, H_1, H_2 and $F(x; \alpha)$. Since, as remarked earlier, $\varphi_i(t)$ is a G_i-mixture of $\{\varphi(t; \alpha)\}$, $i = 1, 2$, we have

$$\begin{aligned}\varphi(t) &= \varphi_1(t)\cdot\varphi_2(t) = \int_D \varphi(t;\alpha)\,dG_1(\alpha)\cdot\int_D \varphi(t;\beta)\,dG_2(\beta)\\ &= \int_D\int_D \varphi(t;\alpha+\beta)\,dG_1(\alpha)\,dG_2(\beta)\\ &= \int_D\int_D \varphi(t;\gamma)\,dG_1(\gamma-\beta)\,dG_2(\beta)\\ &= \int_D \varphi(t;\gamma)\,dG(\gamma),\end{aligned}$$

employing (3) and Theorem 5 of [13]. In view of the one-to-one correspondence between c.d.f.'s and c.f.'s, this implies that H is a G-mixture of $\mathfrak{F} = \{F(x;\alpha)\}$.

In proving the converse, we suppose $r = 2$ (and the distributions discrete) for brevity's sake. By hypothesis,

$$\int_{R^m} \varphi(t;\alpha)\,dG_1(\alpha)\cdot\int_{R^m} \varphi(t;\beta)\,dG_2(\beta) = \int_{R^m} \varphi(t;\gamma)\,dG(\gamma),$$

where we may choose

$$\mu_{G_1}(\alpha_0) = 1 - \frac{1}{n} = 1 - \mu_{G_1}(\alpha_1)$$

$$\mu_{G_2}(\beta_0) = 1 - \frac{1}{n} = 1 - \mu_{G_2}(\beta_1).$$

Since $G = G_1 * G_2$, $\varphi(t;\alpha_i+\beta_j)$, $i, j = 0, 1$, belongs to the class of transforms of $\mathfrak{F}$, i.e., the domain D of α is an Abelian semi-group and

$$\begin{aligned}&\left[\frac{n-1}{n}\varphi(t;\alpha_0) + \frac{1}{n}\varphi(t;\alpha_1)\right]\left[\frac{n-1}{n}\varphi(t;\beta_0) + \frac{1}{n}\varphi(t;\beta_1)\right]\\ &\quad= \left(\frac{n-1}{n}\right)^2\varphi(t;\alpha_0+\beta_0) + \frac{n-1}{n^2}[\varphi(t;\alpha_0+\beta_1) + \varphi(t;\alpha_1+\beta_0)]\\ &\qquad+ \frac{1}{n^2}\varphi(t;\alpha_1+\beta_1).\end{aligned}$$

Letting $n \to \infty$, we see that $\varphi(t;\alpha_0)\cdot\varphi(t;\beta_0) = \varphi(t;\alpha_0+\beta_0)$. The conclusion now follows from the fact that α_0 and β_0 are arbitrary points of R^m (or some measurable sub-region D thereof).

COROLLARY 1: *An infinitely divisible mixing* (G) *of an additively closed family* ($\mathfrak{F}$) *yields an infinitely divisible mixture* (H).

COROLLARY 2: *The convolution of two compound Poisson distributions* (*see Section* 1) *is again a compound Poisson distribution whose mixing* c.d.f. *is the convolution of the two given mixing* c.d.f.'s.

This was proved by Feller [8] and follows from Theorem 3 by taking $m = 1$, $D_1 = [0, \infty)$, $\varphi(t;\alpha) = \exp\{\alpha(e^{it} - 1)\}$. For this same choice of $\varphi(t;\alpha)$, Corollary 1 appears in [11a]. Similarly, for $m = 1$,

$$D_1 = [0, \infty), \varphi(t; \alpha) = \exp\{-\alpha \mid t \mid^\beta\}, 0 < \beta \leqq 2,$$

we have

Corollary 3: *The convolution of two mixtures of symmetric stable distributions of fixed exponent β is again a mixture of the same type with mixing* c.d.f. *the convolution of the given mixing* c.d.f.'s.

Let $m = 2$, $D_1 = R^1$,

$$D_2 = [0, \infty), \varphi(t; \alpha_1, \alpha_2) = \varphi(t; \theta, \sigma^2) = \exp\{i\theta t - \sigma^2 t^2/2\}.$$

By an extension of terminology, a mixture of normal distributions (on both parameters) might be called compound normal. Then Corollary 2 remains valid if everywhere therein the word "Poisson" is replaced by the word "normal".

In [18], it was shown that, except for a pathological case (arising when α varies in a continuum and which may be excluded by a slight additional assumption), if $\mathfrak{F} = \{F(x; \alpha)\}$ is additively closed, then $\varphi(t; \alpha)$ is of the form

$$\prod_{i=1}^m, [f_i(t)]^{\alpha_i};$$

specifically, for $m = 1$,

$$\varphi(t; \alpha) = [\varphi(t)]^\alpha, \qquad \alpha \geqq 0, \tag{4}$$

where $\varphi(t)$ is a c.f. independent of α. In order to avoid detailing the conditions, let us say that $\mathfrak{F} \varepsilon C_1'$ if (4) holds. Most of the classical one-parameter families of distributions belong to C_1'.

We pose the question whether a G-mixture of $\mathfrak{F}$, with $\mathfrak{F} \varepsilon C_1'$, may itself be an element of $\mathfrak{F}$. If $\mathfrak{F}$ is the additively closed family of unitary distributions, i.e. $\varphi(t; \alpha) = [e^{it}]^\alpha$, all real α, then the very definitions of c.f. and mixture show that any non degenerate c.d.f., H, is a mixture (in fact an H-mixture) of $\mathfrak{F}$ and hence not an element of $\mathfrak{F}$. The following theorem shows this situation to prevail under considerably less trivial circumstances.

Theorem 4: *Take $m = 1$ and let $\mathfrak{F} = \{F(x; \alpha)\} \varepsilon C_1'$. If*

(i) *$\varphi(t; \alpha)$ is real-valued (for real t) and $\lim_{t \to t_0} \varphi(t) = 0$ for t_0 finite or infinite or*

(ii) *$\{F(x; \alpha)\}$ has finite second moments and non-zero first moments then no $G(\alpha)$-mixture of $\mathfrak{F}$ belongs to $\mathfrak{F}$.*

Proof: If a G-mixture of $\mathfrak{F}$ is an element of $\mathfrak{F}$, say with c.f. $\Psi(t) = [\varphi(t)]^\gamma$, $\gamma \geqq 0$, we have

$$\int_0^\infty [\varphi(t)]^\alpha \, dG(\alpha) = \Psi(t) = [\varphi(t)]^\gamma \tag{4.1}$$

or

$$1 = \int_0^{\gamma-} [\varphi(t)]^{\alpha-\gamma} \, dG(\alpha) + \int_\gamma^\infty [\varphi(t)]^{\alpha-\gamma} \, dG(\alpha) \tag{4.2}$$

In case (i), if $G(\gamma +) = 0$, letting $t \to t_0$ in (4.2), we reach the contradiction $1 = 0$. Since $G(\gamma +) - G(\gamma -) = 1$ is precluded, $G(\gamma -) = p > 0$ and we may choose ϵ, $0 < \epsilon < \gamma$ such that $G(\gamma - \epsilon) > 0$. Also, if t_0 is finite, we may take it to be the smallest (positive) zero of $\varphi(t)$ in which case $\varphi(t) > 0$ for $|t| < t_0$ whether t_0 is finite or infinite. Thus, for $|t| \leqq t_0$, from (4.2)

$$1 \geqq \int_0^{\gamma-\epsilon} [\varphi(t)]^{\alpha-\gamma}\, dG(\alpha) \geqq [\varphi(t)]^{-\epsilon} G(\gamma - \epsilon)$$

which is clearly impossible for t sufficiently close to t_0.

In case (ii), since the c.d.f. of $\Psi(t)$ has finite second moment and the first and second moments of $F(x; \alpha)$ are $-i\alpha\varphi'(0)$ and

$$-\{\alpha^2[\varphi'(0)]^2 + \alpha[\varphi''(0) - (\varphi'(0))^2]\},$$

it follows that G has its first two moments finite whence differentiation under the integral sign in (4.1) is permissible. Thus, for $\varphi'(t) \neq 0$,

$$[\Psi'(t)]/[\varphi'(t)] = \int \alpha[\varphi(t)]^{\alpha-1} dG(\alpha)$$

and for $\varphi(t) \cdot \varphi'(t) \neq 0$,

$$\text{(4.3)} \quad \frac{\varphi'\Psi'' - \Psi'\varphi''}{[\varphi']^3} + \frac{\Psi'}{\varphi\varphi'} - \left(\frac{\Psi'}{\varphi'}\right)^2 = \int \alpha^2[\varphi(t)]^{\alpha-2}\, dG - \left(\int \alpha[\varphi(t)]^{\alpha-1}\, dG(\alpha)\right)^2.$$

When $\Psi = \varphi^\gamma$, (4.3) becomes

$$\text{(4.4)} \quad \gamma^2\varphi^{\gamma-2}(t) - [\gamma\varphi^{\gamma-1}(t)]^2 = \int \alpha^2[\varphi(t)]^{\alpha-2}\, dG(\alpha) - \left(\int \alpha[\varphi(t)]^{\alpha-1}\, dG(\alpha)\right)^2$$

for all t such that $\varphi(t) \cdot \varphi'(t) \neq 0$.

Since $\varphi'(0) \neq 0$, we may substitute $t = 0$ directly in (4.4) obtaining

$$0 = \int [\alpha - \int \alpha\, dG(\alpha)]^2\, dG(\alpha)$$

which implies that G is a unitary distribution.

In the case $\varphi'(0) = 0$ but $\int \alpha^2\, dG(\alpha) < \infty$, there exists an interval $[0, \epsilon)$ in which $\varphi'(t) \not\equiv 0$ since the contrary would imply $\varphi(t) \equiv 1$ in $[0, \epsilon)$, hence $\varphi(t) \equiv 1 \equiv \varphi(t; \alpha)$ and μ_G degenerate. Consequently, (4.4) holds for a sequence of t-values approaching zero and, by continuity, at zero also, again yielding the prior contradiction.

Note that Theorem 4 does not preclude $\int [\varphi(t)]^\alpha\, dG(\alpha) = \varphi(at)$ for some real a; here $G(\alpha) = G_a(\alpha)$. Taking $\Psi(t) = \varphi(at)$ in (4.3), we see that under (ii), $a \geqq 1$ (with equality rendering G degenerate). Example 1 of section 5 illustrates this possibility.

COROLLARY: *No mixture of symmetric stable distributions with fixed exponent β, $(0 < \beta \leqq 2)$ is a symmetric stable distribution with exponent β.*

On the other hand, Wintner [20] has shown that any symmetric stable distribution of exponent β $(0 < \beta < 2)$ is a "mixture" of symmetric stable distributions of some fixed larger exponent γ (with a non-finite "mixing measure").

When $m = 2$, there is much greater latitude for a G-mixture of $\{F(x; \alpha)\}$ since in the representation of $\varphi(t; \alpha)$, $f_1(t)$ and $f_2(t)$ need not both be c.f.'s and even when they are, α_1 or α_2 may assume both positive and negative values. The next section deals with this case when $\mathfrak{F}$ is the two-parameter family of normal distributions.

4. Mixtures of normal distributions. On occasion, the underlying population (distribution) of interest to the statistician is not prescribed to be normal but rather is generated by selecting one of a collection of alternative normal distributions according to some probability mechanism or scheme. If the resulting mixture of normal distributions is itself normal, many classical results may be utilized.

Consider, therefore, mixtures of the two-parameter family of normal distributions and under what circumstances, i.e. for what measures μ, such mixtures may themselves be normal.

Define $\Phi(x) = (1/\sqrt{2\pi}) \int_{-\infty}^{x} e^{-y^2/2}\, dy$ and $\Phi(x; \theta, \sigma^2) = \Phi(x - \theta)/\sigma$, where $\theta \in R^1$ and $\sigma \in (0, \infty)$. The question arises whether "degenerate normal distributions" (viz., $\sigma = 0$ which is interpreted as $\Phi(x; \theta, 0) = 0$, $x \leqq \theta$ and $\Phi(x; \theta, 0) = 1$, $x > \theta$) should be mixed; these will be banned because if μ assigns measure one to $\{\sigma^2 = 0\}$ an arbitrary distribution may be thereby obtained.

If $G(\theta, \sigma^2)$ is a c.d.f. which is zero on the lower half and boundary ($\sigma^2 = 0$) of the (θ, σ^2) plane and $\mu(\theta, \sigma^2) = \mu_G$ is the corresponding measure on the Borel sets of R^2, let

$$H(x) = \int_{R^2} \Phi(x; \theta, \sigma^2)\, dG(\theta, \sigma^2) = \int_{R^2} \Phi\left(\frac{x - \theta}{\sigma}\right) d\mu. \tag{5}$$

The class $\mathfrak{H}$ of mixtures (5) of normal distributions is by no means identifiable (see section 1 for definition); this will be apparent momentarily if it is not already so. On the other hand, the class $\mathfrak{H}_m$ (respectively, $\mathfrak{H}_v$) of mixtures on means only (respectively, on variances only) is identifiable.

For $\mathfrak{H}_m$ may also be characterized as the class of c.d.f.'s containing a fixed normal factor with mean zero and specified variance, say unity. Since the normal c.f. is non-vanishing, $\Phi * G_1 = \Phi * G_2$ implies $G_1 = G_2$ which is therefore tantamount to the identifiability of $\mathfrak{H}_m$. (Of course, $\Phi(x; 0, \sigma_1^2) * G_1(x) = \Phi(x; 0, \sigma_2^2) * G_2(x)$ has solutions $G_1 \neq G_2$ if $\sigma_1 \neq \sigma_2$, but this is not the issue).

If $H \in \mathfrak{H}_v$, we may suppose without loss of generality that its mean is zero, so that its c.f. is $\int_0^\infty \exp\{-t^2\sigma^2/2\}\, dG(\sigma^2)$, whence the identifiability of $\mathfrak{H}_v$ is an immediate consequence of the uniqueness theorem for Laplace transforms.

In returning to a consideration of $\mathfrak{H}$, we note that by integrating (5) over the regions $\theta < x$ and $\theta \geqq x$, $\mu\{\theta < x\} < 2H(x)$, $\mu\{\theta \geqq x\} < 2[1 - H(x)]$; that is, the "tails" of the distribution of means (θ) are dominated by those of H.[4]

Since θ and σ^2 are Euclidean "random variables", a conditional distribution of θ given σ^2 exists. Denote it by $G_{\sigma^2}(\theta)$ and let $\phi_{\sigma^2}(t)$ be the corresponding c.f.;

[4] If the distribution of means is degenerate, equality may hold in the second relationship for some x.

also, define $G_1(\sigma'^2) = \mu\{\sigma^2 \mid \sigma^2 < \sigma'^2\}$. Then (5) may be rewritten as

$$(6)\quad \begin{aligned} H(x) &= \int_0^\infty \int_{-\infty}^\infty \Phi\left(\frac{x-\theta}{\sigma}\right) dG_{\sigma^2}(\theta)\, dG_1(\sigma^2) \\ &= \int_0^\infty [\Phi(x; 0, \sigma^2) * G_{\sigma^2}(x)]\, dG_1(\sigma^2). \end{aligned}$$

Since H is a G_1-mixture of the bracketed family of c.d.f.'s its c.f., say $\varphi(t)$, is given by

$$(7)\quad \varphi(t) = \int_0^\infty e^{-\sigma^2 t^2/2} \cdot \varphi_{\sigma^2}(t)\, dG_1(\sigma^2).$$

Equating the real parts of (7), we note that, if H is a symmetric c.d.f., i.e., $\varphi(t)$ is real-valued (e.g. normal) and a μ_1-mixture of $\{\Phi(x; \theta, \sigma^2)\}$, it is also a μ-mixture of $\{\Phi(x; \theta, \sigma^2)\}$, where μ is such that the c.d.f.'s $G_{\sigma^2}(\theta)$ are symmetric.

We turn directly to the case $H(x) = \Phi(x; \theta_0, \sigma_0^2)$ for fixed $\theta_0, \sigma_0 > 0$. The fact that σ^2 then has a bounded spectrum (see Theorem 6), together with the domination of the tails of the θ distribution by those of the normal (see the paragraph preceding (6)), insure that $G(\theta, \sigma^2)$ has finite moments of all orders.

If $\Phi(x; \theta_0, \sigma_0^2)$ is a $\bar{G}$-mixture of $\{\Phi(x; \theta, \sigma^2)\}$, then, replacing x by $\sigma_0 x + \theta_0$,

$$(8)\quad \Phi(x) = \int_{R^2} \Phi\left[\frac{x - (\theta - \theta_0)/(\sigma_0)}{\sigma/\sigma_0}\right] d\bar{G}(\theta, \sigma^2) = \int_{R^2} \Phi\left(\frac{x-\theta}{\sigma}\right) dG(\theta, \sigma^2),$$

and we therefore suppose without loss of generality that $\theta_0 = 0, \sigma_0 = 1$. Taking (bilateral) Laplace transforms in (8), we have

$$(9)\quad e^{s^2/2} = \int_{R^2} e^{(\sigma^2 s^2/2) - \theta s}\, dG(\theta, \sigma^2).$$

Multiplying (9) by $\exp\{\beta s^2/2\}$, replacing s by $s(1+\beta)^{-\frac{1}{2}}$ and changing integration variables shows that, if $\Phi(x)$ is a G-mixture of $\{\Phi(x; \theta, \sigma^2)\}$, it is likewise a G_β-mixture of $\{\Phi(x; \theta, \sigma^2)\}$, where $G_\beta(\theta, \sigma^2) = G(\theta\sqrt{1+\beta}, \sigma^2(1+\beta) - \beta)$, $\beta \geqq 0$. (Note that since the mass of G is contained in the strip $0 \leqq \sigma^2 \leqq 1$ (Theorem 6), the mass of G_β is constrained to lie in the strip $\beta/(1+\beta) \leqq \sigma^2 \leqq 1$.) Thus, contrary to the Compound Poisson case, specification of H by no means determines the mixing measure μ.

Now the representation (6), with $H(x) = \Phi(x)$, elicits the obvious "solutions" $G_{\sigma^2}(\theta) = \Phi(\theta; 0, 1 - \sigma^2)$, $G_1(\sigma^2) =$ arbitrary c.d.f. on (0, 1). On the other hand, the following easily proved

LEMMA: $\Psi(x) = (1+d)\Phi(x; \theta_1, \sigma_1^2) - d\Phi(x; \theta_2, \sigma_2^2)$, $d > 0$, *is a* c.d.f. *if and only if* $\sigma_2 < \sigma_1$, and

$$d^{-1} \geqq \frac{\sigma_1}{\sigma_2} \exp\left\{\frac{1}{2}\frac{(\theta_2 - \theta_1)^2}{\sigma_1^2 - \sigma_2^2}\right\} - 1,$$

shows that

$$G_1(\sigma^2) = \begin{cases} 0, \sigma^2 \leqq \sigma_1^2 \\ (1+d)^{-1}, \sigma_1^2 < \sigma^2 \leqq \sigma_2^2 \\ 1, \sigma^2 > \sigma_2^2 \end{cases}$$

$$G_{\sigma_1^2}(\theta) = (1+d)\Phi(\theta; 0, 1-\sigma_1^2) - d\Phi(\theta; \theta_2, \sigma_2^2 - \sigma_1^2)$$

$$G_{\sigma_2^2}(\theta) = \begin{matrix} 0, \theta \leqq \theta_2 \\ 1, \theta > \theta_2 \end{matrix}$$

also constitute solutions when d, θ_2, σ_1, σ_2 are as prescribed. We have thus proved (with the exception of the parenthetical statement which follows from Theorem 6)

THEOREM 5: *Suppose (as the conclusion requires) that* $\mu\{\sigma^2 \mid \sigma^2 > \sigma_0^2\} = 0$. *Then a sufficient but unnecessary condition that a* μ*-mixture of normal distributions be normal with mean* θ_0 *and variance* σ_0^2 *is that the conditional distribution of* θ *given* σ^2 *be normal with mean* θ_0 *and variance* $\sigma_0^2 - \sigma^2$ *for all values of* σ^2 *for which it is defined.*

Naturally, (8) imposes contraints on the distributions of means θ and variances σ^2 and we now proceed to establish some of these.

THEOREM 6: *In order that a* μ*-mixture of normal distributions be normal with mean* $\theta_0 = 0$ *and variance* $\sigma_0^2 = 1$, *it is necessary that*

(i) $\mu\{\sigma^2 \mid \sigma^2 > 1\} = 0 = \mu\{\theta, \sigma^2 \mid \sigma^2 = 1, \theta \neq 0\}$. *Hence it may be supposed that* $\mu\{\sigma^2 \mid \sigma^2 \geqq 1\} = 0$.

(ii)[5] $\mu\left\{\theta, \sigma^2 \middle| \frac{\theta}{1-\sigma^2} > C\right\} \cdot \mu\left\{\theta, \sigma^2 \middle| \frac{\theta}{1-\sigma^2} < -C\right\} > 0$, *all* $C > 0$,

(iii) $\int_{R^2} \exp\left\{\frac{\theta^2}{2(1-\sigma^2)}\right\} d\mu = \infty$,

(iv) $\mu\{\theta \mid |\theta| < e^{-\frac{1}{2}}\} \cdot \mu\{\theta, \sigma^2 \mid \theta^2 \leqq \sigma^2 \log_e 1/\sigma^2\} > 0$,

(v) the θ-spectrum of μ not be confined to a subset of numbers in arithmetic progression; further, for all integers m (all real b) and all integers $n \geqq 1$,

$$\mu\left\{\theta \mid \bar{\gamma} \in \bigcup_{j=0}^{n-1} \left[\frac{8j+1}{8n}, \frac{8j+3}{8n}\right]\right\} < 1,$$

where $\bar{\gamma}$ signifies the fractional part of γ and either $\gamma = \theta - m/n$ or $\gamma = b\theta$.

PROOF: Rewrite (9) as

$$(6.1) \qquad 1 = \int_{R^2} e^{[(\sigma^2-1)s^2/2]-\theta s}\, d\mu.$$

Suppose now that for some $\epsilon > 0$, $B_\epsilon = \{\sigma^2 \mid \sigma^2 \geqq 1 + \epsilon\}$ has positive μ-measure. Then for sufficiently large C, so does $A = B_\epsilon \cdot \{\theta \mid |\theta| \leqq C\}$, whence, for s real, (6.1) implies

$$1 \geqq \int_A e^{[(\sigma^2-1)s^2/2]-\theta s}\, d\mu \geqq e^{\epsilon s^2/2 - C|s|}\mu\{A\}.$$

[5] The writer cordially thanks his colleague Prof. Michael Golomb for helpful conversations relating to an early version of (ii).

For sufficiently large s, this is manifestly impossible. Thus $\mu\{B_\epsilon\} = 0$, all $\epsilon > 0$, which implies the first equality in (i); the second follows in similar fashion from (6.1). Further, if μ_G assigns measure $p_0 > 0$ to the point $\theta = 0$, $\sigma^2 = 1$, subtracting $p_0\Phi(x)$ from both sides of (8) and dividing by $1 - p_0$, a new related mixture μ^* is obtained for which $p_0 = 0$. Generality is clearly maintained in supposing $\mu = \mu^*$.

To prove (ii), let $W = \{\theta, \sigma^2 \mid 0 < \sigma^2 < 1\}$ and observe from (6.1) and (i) that for all real s

$$(6.2)\qquad 1 = \int_W e^{-(1-\sigma^2)s^2/2-\theta s}\, d\mu = \int_W \exp\left\{-\frac{(1-\sigma^2)}{2}\left[s + \frac{\theta}{1-\sigma^2}\right]^2\right\} d\mu_1,$$

where $d\mu_1 = \exp\{\theta^2/2(1-\sigma^2)\}\, d\mu$. If, now, for some $C > 0$,

$$\mu\{\theta, \sigma^2 \mid \theta/(1-\sigma^2) > C\} = 0,$$

then $s_2 < s_1 < -C$ would imply

$$\int_W \exp\left\{\frac{-(1-\sigma^2)}{2}\left[s_2 + \frac{\theta}{1-\sigma^2}\right]^2\right\} d\mu_1 < \int_W \exp\left\{\frac{-(1-\sigma^2)}{2}\left[s_1 + \frac{\theta}{1-\sigma^2}\right]^2\right\} d\mu_1$$

in violation of (6.2). The remaining part of (ii) is analogous.

If (iii) did not obtain, it would be legitimate to let $s \to \infty$ within the second integral of (6.2) and conclude that $1 = 0$.

As remarked in a more general context in Section 1, (8) implies a corresponding relationship for densities (here multiplied by $\sqrt{2\pi}$), namely

$$(6.3)\qquad e^{-x^2/2} = \int_{R^2} \frac{1}{\sigma} e^{-(x-\theta)^2/2\sigma^2}\, d\mu,$$

which, evaluated at $x = 0$, becomes

$$(6.4)\qquad 1 = \int_{R^2} \frac{1}{\sigma} e^{-\theta^2/2\sigma^2}\, d\mu.$$

The integrand of (6.4) cannot be less than one on a set of μ-measure one which is equivalent to $\mu\{\theta, \sigma^2 \mid \theta^2 \leqq \sigma^2 \log_e \sigma^{-2}\} > 0$. The remaining portion of (iv) follows by noting that the first part implies $\mu\{\theta \mid |\theta| \leqq e^{-\frac{1}{2}}\} > 0$, and consequently $\mu\{\theta \mid |\theta| < e^{-\frac{1}{2}}\} > 0$, since the negation of the latter would entail $\mu\{\theta, \sigma^2 \mid |\theta| = e^{-\frac{1}{2}}, \sigma^2 = e^{-1}\} = 1$, which is easily seen to be incompatible with (6.3). (iv) also follows from $\int [\Phi[(x+\theta)/\sigma] + \Phi[(x-\theta)/\sigma] - 2\Phi(x)]\, d\mu = 0$.

Next, set $s = it$ in (9), obtaining

$$(6.5)\qquad e^{-t^2/2} = \int_{R^2} e^{(-\sigma^2 t^2/2)+it\theta}\, dG(\theta, \sigma^2).$$

If, in violation of the first statement of (v), the θ spectrum is concentrated at points $a + kb$ where $a, b \neq 0$ are real ($b = 0$ requires $a = 0$; this case is ruled out by the corollary to Theorem 7) and k varies over some subset of the integers, take $t = 2\pi\, n/b$ in (6.5) obtaining

$$(6.6) \qquad 1 = \int e^{2(1-\sigma^2)\pi^2 n^2 b^{-2}} \cos 2\pi n\, a/b\, d\mu$$

for all integral values of n. If a/b is rational, say $a/b = m_1/m_2$ where m_1 and m_2 are relatively prime integers, (6.6) is contradicted by choosing $n = m_2$ and $n = 2m_2$. On the other hand, if a/b is irrational, n may be selected such that the fractional part of na/b lies in $(\frac{1}{4}, \frac{3}{4})$ thereby rendering the integrand of (6.6) negative.

To demonstrate the second part of (v), take $t = 2\pi\, n$ in (6.5) obtaining

$$1 = \int \exp\{2\pi^2 n^2(1-\sigma^2) + 2n\pi i(\theta - m/n)\}\, d\mu$$
$$= \int \exp\{2\pi^2 n^2(1-\sigma^2) + \pi i(2nj + 2k + \epsilon)\}\, d\mu,$$

where j, k, n, are integers and $\epsilon = \epsilon(\theta)$ lies in $[\frac{1}{4}, \frac{3}{4}]$. Since the real part of the integrand is non-positive on a set of measure one, there is a gross contradiction. If $\gamma = b\theta$, set $t = 2\pi\, nb$ after which the argument is the same. Q. E. D.

For any c.d.f. G satisfying (8), let $\varphi_G(t, u)$ denote the corresponding c.f. In view of the domination of the tails of the θ distribution by those of $\Phi(x)$, $\varphi_G(Z, 0)$ is defined and convergent for all complex Z. Also, since σ^2 has a bounded spectrum, $\varphi_G(0, w)$ is an entire function of w. But then (see e.g. Theorem 2 of [19]) $\varphi_G(Z, w)$ is jointly analytic in Z and w. Thus, from (6.5) we see that $\varphi_G(Z, iZ^2/2) = e^{-Z^2/2}$ for all complex Z but this is insufficient to characterize $\varphi_{G(Z,w)}$.

In the case of product measure, the class of measures μ satisfying (5) is given by

THEOREM 7: *A product-measure mixture of normal distributions is normal with mean* θ_0 *and variance* σ_0^2 *if and only if for some* σ_1^2 *in* $(0, \sigma_0^2)$, $G_{\sigma_1^2}(\theta) = \Phi(\theta; \theta_0, \sigma_0^2 - \sigma_1^2)$ *and* $G_1(\sigma^2)$ *is degenerate at* σ_1^2.

PROOF: Sufficiency is obvious and subsumed in Theorem 5. To prove necessity note that, since $G_{\sigma^2}(\theta)$ is constant with respect to σ^2, (6) simplifies to

$$\Phi(x; \theta_0, \sigma_0^2) = G_{\sigma^2}(x) * \textstyle\int_0^\infty \Phi(x; 0, \sigma^2)\, dG_1(\sigma^2).$$

By the theorem of Cramér–Lévy this requires that both factors be normal with moments which add to θ_0, σ_0^2. By the identifiability of $\mathcal{K}_v$, the second factor is normal if and only if G_1 is degenerate.

COROLLARY: *A mixture of normal distributions with identical means cannot be normal*[6].

It follows immediately from (ii) or (iii) of Theorem 6 that a finite mixture of normal distributions cannot be normal (recall that finite here signifies at least two). Furthermore, as a direct consequence of Theorem 7, a countable product-

[6] This corollary and the fact that G_1 is degenerate also flow from Theorem 4. The former is also implicit in a theorem of [2].

measure mixture of normal distributions is non-normal. It seems intuitively plausible that no countable mixture of normal distributions $\{\Phi(x;\theta_j,\sigma_j^2)\}$ is normal and if the variances σ_j^2 have a minimum, this is indeed the case. This follows from a somewhat more general proposition.

Suppose that μ is such that for some real θ_0 and $\sigma_0 > 0$, $\mu\{\sigma^2 \mid \sigma^2 < \sigma_0^2\} = 0$ while

$$\mu\left\{\theta, \sigma^2 \,\middle|\, \begin{matrix}\theta = \theta_0 \\ \sigma^2 = \sigma_0^2\end{matrix}\right\} = p_0 > 0.$$

Then from (6.5) for any μ satisfying (8)

$$e^{-t^2/2} = p_0 e^{-\sigma_0^2 t^2/2 + it\theta_0} + \int_S e^{-\sigma^2 t^2/2 + it\theta}\, d\mu$$

where $S = \left\{\theta, \sigma^2 \,\middle|\, \begin{matrix}\sigma^2 = \sigma_0^2, \theta \neq \theta_0 \\ \text{or } \sigma^2 > \sigma_0^2\end{matrix}\right\}$.

Thus, $e^{-(1-\sigma_0^2)t^2/2} = p_0 e^{it\theta_0} + \int_S e^{-(\sigma^2-\sigma_0^2)t^2/2+it\theta}\, d\mu.$

Since both terms on the right hand side are (to within constants) c.f.'s, this would imply that a continuous c.d.f., namely, $\Phi(x; 0, 1 - \sigma_0^2)$ was a mixture of a discrete and some other distribution, which is patently false. Thus, if a μ-mixture of $\{\Phi(x;\theta,\sigma^2)\}$ is normal, the mixing measure cannot be as supposed here. In particular, if the infimum of the variances is attained, a countable mixture of normal c.d.f.'s is non-normal. If it is not attained, we may suppose that a subsequence of variances approaches zero, whence from (iv) of Theorem 6, it follows that the same conclusion holds if zero is not a value or a point of accumulation of $\{\theta_j\}$.

According to (v) of Theorem 6, a countable mixture of normal c.d.f.'s cannot be normal if the means are a set of numbers in arithmetic progression (or a subset there of); the case of only finitely many different means can be disposed of by a number-theoretic argument but the question of an arbitrary countable mixture remains open.

Given any bounded sequence $\{\sigma_j^2\}$ of (distinct) positive real numbers, say in (0, 1) and arbitrary positive ϵ, there exist sequences $\{\theta_j\}$, $\{c_j\}$ with $c_j > 0$, $\sum_{j=1}^\infty c_j = 1$ such that

$$\sup_x \left| \sum_{j=1}^\infty c_j \Phi(x;\theta_j,\sigma_j^2) - \Phi(x) \right| < \epsilon.$$

This statement follows from Theorems 1 and 5 and shows that a countable mixture of normal distribution can be arbitrarily close to a normal distribution.

The known relationship

$$\int_0^\infty \left[\frac{1}{2\sqrt{\pi\alpha}} e^{-x^2/4\alpha}\right] e^{-\alpha}\, d\alpha = \frac{1}{2} e^{-|x|}$$

reveals that an exponential-mixture of normal distributions (with identical means) has the so-called Laplace distribution.

5. Generation of mixtures. If (X, α) have a joint distribution in R^{m+1} then

$H(x) = \int F(x \mid \alpha)\, dG(\alpha)$ and (dually) $G(\alpha) = \int K(\alpha \mid x)\, dH(x)$. Moreover, if for some measure ν on R^m (independent of x), $dK/d\nu \equiv k(\alpha \mid x)$ exists, then the Radon-Nikodym derivative $dG/d\nu \equiv g(\alpha) = \int_{-\infty}^{\infty} k(\alpha \mid x)\, dH(x)$ likewise exists and H is representable in the form

$$(10) \qquad H(x) = \int_{R^m} \frac{\int_{-\infty}^{x} k(\alpha \mid y)\, dH(y)}{g(\alpha)} g(\alpha)\, d\nu = \int_{R^m} \frac{\int_{-\infty}^{x} k(\alpha \mid y)\, dH(y)}{\int_{-\infty}^{x} k(\alpha \mid y)\, dH(y)} d\mu.$$

Since H is a c.d.f., $d\mu = g d\nu$ represents a probability measure. If, in addition, $h(x) \equiv dH/d\lambda$ exists for some linear measure λ, $dF/d\lambda \equiv f(x \mid \alpha) = k(\alpha \mid x)h(x)/g(\alpha)$ whence

$$(11) \qquad h(x) = \int \frac{k(\alpha \mid x)h(x)}{g(\alpha)} d\mu = \int f(x \mid \alpha)\, d\mu.$$

When $m = 1$, α concentrates on 0, 1, 2, $\cdots$ and ν is counting measure, the preceding reduces to

$$(12) \qquad h(x) = \sum_{j=0}^{\infty} g_j f(x \mid j) \quad \text{with} \quad f(x \mid j) = k(j \mid x)h(x)g_j^{-1}$$

and (dually) $g_j = \int k(j \mid x)h(x)\, d\lambda$ with $k(j \mid x) = g_j f(x \mid j)[h(x)]^{-1}$.

In particular, if $k(\alpha \mid x) = 1$ for $x \epsilon I_\alpha = (a_\alpha, a_{\alpha+1})$ and zero otherwise where $a_0 = -\infty$, $\lim_{\alpha \to \infty} a_\alpha = +\infty$ and $a_\alpha < a_{\alpha+1}$, $\alpha = 0, 1, 2, \cdots$ then $g(\alpha) = H(a_{\alpha+1}) - H(a_\alpha)$ and (10) exhibits any non-degenerate c.d.f. H as a finite or countable mixture of c.d.f.'s H_α formed by truncating the distribution H outside I_α, If $a_N = +\infty$ for some finite integer $N > 0$, the mixture is necessarily finite. This method of splitting apart and splicing together a c.d.f. is somewhat artificial but other choices of $k(\alpha \mid x)$ yield more interesting mixtures.

A slightly altered formulation leading to (12) (or (11)) starts with the selection of non-negative measurable functions $g_j(x)$ and constants $a_j \geqq 0$ for which $g(x) = \sum_{j=0}^{\infty} a_j g_j(x)$ is positive and finite on a set S of positive Lebesgue measure. If $h(x)$ is any p.d.f. with spectrum S and such that $0 < \int_S g_j(x)h(x)[g(x)]^{-1}\, dx = b_j < \infty$, $j = 0, 1, 2, \cdots$, then h is a c_j-mixture of $\{f_j(x)\}$ where $c_j = a_j b_j$ and $f_j(x) = g_j(x)h(x)[b_j g(x)]^{-1}$, $x \in S$ and zero elsewhere. Note that $k(j \mid x) = a_j g_j(x)[g(x)]^{-1}$.

An interesting particularization arises from taking $g_j(x) = |x|^j$, $a_0 > 0$. Thus, if $g(x) = \sum_{j=0}^{\infty} a_j |x|^j$ converges for $x \in S$ and h is as just indicated, the prior conclusion holds with $f_j(x) = (|x|^j h(x))/(b_j g(x))$, $x \in S$ and zero elsewhere.

Example 1: Gamma distribution as a negative binomial mixture of commonly scaled but differently exponented Gamma c.d.f.'s.

Let $a > 1$, $\lambda > 0$ and choose $g(x) = e^{(a-1)x}$, $h(x) = e^{-x}x^{\lambda-1}[\Gamma(\lambda)]^{-1}$ on $S = (0, \infty)$. Then $b_j = a^{-(\lambda+j)}\Gamma(\lambda + j)[\Gamma(\lambda)]^{-1}$, $c_j = \binom{-\lambda}{j} a^{-\lambda}(a^{-1} - 1)^j$ and

$$f_j(x) = \frac{x^{\lambda+j-1} e^{-ax}}{a^{-(\lambda+j)}\, \Gamma(\lambda + j)} \text{ on } S.$$

In the earlier notation, $k(j \mid x) = 1/j! e^{-(a-1)x}[(a-1)x]^j$. This example appears in [14] with $\lambda = n/2$ and a change of scale.

Example 2[7]: Select $g(x) = \sigma \exp \{(x^2/2)(1-\sigma^2)/\sigma^2\}$, $0 < \sigma^2 < 1$ and $h(x) = \Phi'(x)$. Then

$$b_j = \frac{(2j)!}{2^j j!}\sigma^{2j}, \qquad a_{2j} = \frac{\sigma}{j!\,2^j}((1-\sigma^2)/\sigma^2)^j, \qquad a_{2j+1} = 0,$$

whence

$$\Phi'(x) = \sum_{j=0}^{\infty} 2^{-2j}\binom{2j}{j}\sigma(1-\sigma^2)^j\left[\frac{2^j j!(2\pi)^{-\frac{1}{2}}}{(2j)!\sigma^2 j + 1}x^{2j}e^{-x^2/2\sigma^2}\right] = \sum_{j=0}^{\infty} c_j f_j(x),$$

expressing the normal density function as a discrete mixture of the bracketed densities $f_j(x)$.

The c.f. (see appendix) of $f_j(x)$ is

$$\varphi_j(t) = \frac{H_{2j}(t)}{H_{2j}(0)}e^{-\sigma^2 t^2/2},$$

where

$$H_{2j}(t) = e^{t^2/2}(-1)^{2j}\frac{d^{2j}}{dt^{2j}}e^{-t^2/2} = \sum_{i=0}^{j}(-1)^i\binom{2j}{2i}[1\cdot 3\cdots(2i-1)]\,t^{2(j-i)}$$

is one version of the Hermite polynomial of order $2j$. The preceding bracketed expression is defined to be one for $i = 0$.

The restatement of the mixture in terms of c.f.'s yields an exponential expansion in terms of even degree Hermite polynomials, viz.,

$$\sigma^{-1}\exp\left\{-t^2/2\left(\frac{1-\sigma^2}{\sigma^2}\right)\right\} = \sum_{j=0}^{\infty}\frac{(-1)^j}{j!}\left(\frac{1-\sigma^2}{2}\right)^j H_{2j}(t),$$

which is virtually that of [12, p. 580].

Example 3: $\Phi(x)$ as a more elaborate mixture of normal distributions.

Let $\{Q_i(x)\}$ be a sequence of positive definite quadratic forms with $\sum_{i=0}^{\infty} Q_i(x) \equiv \infty$. Take $k(j \mid x) = \exp\{-\frac{1}{2}\sum_{i=0}^{j} Q_i(x)\} - \exp\{-\frac{1}{2}\sum_{i=0}^{j+1} Q_i(x)\} > 0$, $j = 0, 1, \cdots$ and define constants $c_j, \theta_j, s_j, j \geqq 0$ by $\Phi'(x)\exp\{-\frac{1}{2}\sum_{i=0}^{j} Q_i(x)\} = c_j/s_j \Phi'((x-\theta_j)/s_j)$. It follows that $g_j = c_j - c_{j+1} > 0$, and letting $d_j = c_{j+1}/(c_j - c_{j+1})$,

$$F(x \mid j) = (1 + d_j)\Phi(x; \theta_j, s_j^2) - d_j\Phi(x; \theta_{j+1}, s_{j+1}^2).$$

Now $\{s_j\}$ is decreasing and positive and if σ_j is sufficiently small, $F(x \mid j)$ contains the factor $\Phi(x; 0, \sigma_j^2)$ and is thus representable as

$$F(x \mid j) = \Phi(x; 0, \sigma_j^2) * [(1 + d_j)\Phi(x; \theta_j, s_j^2 - \sigma_j^2) - d_j\Phi(x; \theta_{j+1}, s_{j+1}^2 - \sigma_j^2)]. \tag{13}$$

[7] The referee has pointed out that example 2 may be obtained directly from example 1. He has also suggested a reformulation of the main idea of this section, leading to greater cohesion.

If the σ_j^2 are distinct, the bracketed term may be regarded as the value at $\sigma^2 = \sigma_j^2$ of a conditional distribution function $G_{\sigma^2}(x)$ and the mixture $\sum_{j=0}^{\infty} g_j F(x \mid j)$ has the structure (6), viz.,

$$\Phi(x) = \sum_{j=0}^{\infty} g_j[\Phi(x/\sigma_j) * G_{\sigma_j^2}(x)]$$

with $G_1(\sigma^2)$ discrete.

In particular, if $Q_i(x) \equiv (x^2 - 2x - 2ln\ a)$ where $0 < a < e^{-\frac{1}{2}}$, calculations yield $\theta_j = j/(j + 1)$, $s_j^2 = 1/(j + 1)$ and $d_j = q_j/(1 + q_j)$ where $q_j = a((j + 1)/(j + 2))^{\frac{1}{2}} \exp\{\frac{1}{2} - 1/(2(j + 1)(j + 2))\}$. Finally, if $\sigma_j^2 \leqq (1 - a^2e)/(j(1 - a^2e) + 2 - a^2e)$, the bracketed quantity in (13) will be a c.d.f.

6. Remarks on the compound poisson distribution. It seems of interest to note that the problem of mixtures of Poisson distributions is intimately linked to the moment problem. Since, as mentioned earlier,

$$H(x) = \int_0^{\infty} \sum_{j<x} \frac{\alpha^j e^{-\alpha}}{j!}\, dG(\alpha)$$

is a discrete c.d.f. with saltuses at the non-negative integers, it is completely characterized by the probabilities

$$p_j = \int_0^{\infty} \frac{\alpha^j e^{-\alpha}}{j!}\, dG(\alpha), \qquad j = 0, 1, 2, \cdots$$

Let $G^*(\alpha) = 1/p_0 \int_0^{\alpha} e^{-y}\, dG(y)$. Then G^* is a c.d.f. and

$$\frac{j!\,p_j}{p_0} = \int_0^{\infty} \alpha^j\, dG^*(\alpha), \qquad j = 0, 1, 2, \cdots.$$

Consequently, in order that a discrete distribution characterized by mass p_j at j ($j = 0, 1, 2, \cdots$) be a mixture of Poisson distributions, it is necessary that the sequence $\{j!p_j/p_0\}$ be a moment sequence for the Stieltjes moment problem. In particular, it is necessary that the determinants ($i, j = 0, 1, \cdots n$) $\Delta_n = \mid (i + j)!p_{i+j} \mid, \Delta_n' = \mid (i + j + 1)!p_{i+j+1} \mid$ be non-negative for every non-negative integer n. Conversely, if $\{j!p_j/p_0\}$ is a moment sequence on $(0, \infty)$ and the corresponding distribution $G^*(\alpha)$ is such that $\int_0^{\infty} e^{\alpha}\, dG^*(\alpha) = 1/p_0 < \infty$, then $\sum_{j<x} p_j$ is a compound Poisson distribution with mixing c.d.f. $G(\alpha) = p_0 \int_0^{\alpha} e^u\, dG^*(u)$.

It is easy to see (and pointed out in [8]) that the class of compound Poisson distributions is identifiable. Thus, a mixture of Poisson distributions cannot be Poisson.

Appendix

We show by induction that the c.f. of $f_j(x)$ of example 2 of Section 5 is

$$\varphi_j(t) = \frac{H_{2j}(\sigma t)}{H_{2j}(0)}\, e^{-\sigma^2 t^2/2},$$

where

$$H_{2j}(t) = e^{t^2/2}(-1)^{2j}\frac{d^{2j}}{dt^{2j}}e^{-t^2/2} = \sum_{i=0}^{j}(-1)^i\binom{2j}{2i}[1.3....(2i-1)]t^{2(j-i)}$$

is a version of the Hermite polynomial of order $2j$. The preceding bracketed expression is defined to be 1 for $i = 0$.

This is evident for $j = 0$ and (supposing $\sigma = 1$ for simplicity) follows, for $j = 1$ from

$$T\left[\frac{x^2e^{-x^2/2}}{\sqrt{2\pi}}\right] = T\left[\frac{e^{-x^2/2}}{\sqrt{2\pi}} + \frac{d^2}{dx^2}\frac{e^{-x^2/2}}{\sqrt{2\pi}}\right] = e^{-t^2/2} - t^2e^{-t^2/2},$$

where T denotes the operation of Fourier transformation. In general, it suffices to verify that for $j \geqq 2$,

$$T\left[\frac{x^{2j}}{\sqrt{2\pi}}e^{-x^2/2}\right] = (-1)^jH_{2j}(t)e^{-t^2/2}. \tag{14}$$

Now

$$-t^2T[x^{2j}e^{-x^2/2}] = T\left[\frac{d^2}{dx^2}x^{2j}e^{-x^2/2}\right]$$
$$= T[x^{2j+2}e^{-x^2/2} - (4j+1)x^{2j}e^{-x^2/2} + 2j(2j-1)x^{2j-2}e^{-x^2/2}].$$

Hence, if (14) holds for $j - 1$ and $j (j \geqq 1)$, it holds for $j + 1$, since

$$T[x^{2j+2}e^{-x^2/2}] = (-1)^{j+1}e^{-t^2/2}\{(t^2 - 4j - 1)H_{2j}(t) - 2j(2j-1)H_{2j-2}(t)\}$$
$$= (-1)^{j+1}H_{2j+2}(t)e^{-t^2/2}$$

in view of the recursion relation (verified by direct substitution)

$$(t^2 - 4j - 1)H_{2j}(t) - 2j(2j-1)H_{2j-2}(t) = H_{2j+2}(t), j \geqq 1.$$

REFERENCES

[1] PATRICK BILLINGSLEY, "Invariance principle for dependent random variables", *Trans. Amer. Math. Soc.*, Vol. 83 (1956), pp. 250–268.

[2] J. BLUM, H. CHERNOFF, M. ROSENBLATT AND H. TEICHER, "Central limit theorems for interchangeable processes," *Can. Jour. Math.*, Vol. X (1958), pp. 222–229.

[3] N. BOURBAKI, "Elements de mathématique, Livre vi, Intégration," Chapter 5, *Intégration des mesures*, Hermann, Paris, 1957.

[4] HERMAN CHERNOFF, "Large sample theory: Parametric case," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 1–22.

[5] H. CRAMÉR, "Über eine Eigenschaft der normalen Verteilungsfunktion," *Math. Zeit.* Vol. 41 (1936), pp. 405–414.

[6] JEAN DIEUDONNÉ, "Sur la convergence des suites de mesures de Radon," *Anais da Academia Brasiliera de Ciencias*, Vol. 23 (1951), pp. 21–38.

[7] DANIEL DUGUÉ, "Arithmétique des lois de probabilités," *Memorial des Sciences Mathématiques*, fascicule CXXXVII, Gauthier-Villars, Paris, 1957.

[8] W. FELLER, "On a general class of contagious distributions," *Ann. Math. Stat.* Vol. 14 (1943), pp. 389–399.

[9] BRUNO DE FINETTI, "La prévision, ses lois logiques, ses sources subjectives", *Annales de l'Institut Henri Poincaré*, Vol. 17 (1937), pp. 1–68.

[10] LAWRENCE M. GRAVES, *The Theory of Functions of Real Variables*, McGraw-Hill, New York, 1946.

[11] EDWIN L. HEWITT AND LEONARD J. SAVAGE, "Symmetric measures on cartesian products," *Trans. Amer. Math. Soc.*, Vol. 80 (1955), pp. 470–501.

[11a] E. CANSADO MACEDA, "On the compound and generalized Poisson distributions," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 414–16.

[11b] PER OTTESTED, "On certain compound frequency distributions," *Skand. Akt.*, Vol. 27 (1944), pp. 32–42.

[12] HARRY POLLARD, "Distribution functions containing a gaussian factor," *Proc. Amer. Math. Soc.*, Vol. 4 (1953), pp. 578–582.

[13] HERBERT ROBBINS, "Mixture of distributions," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 360–369.

[14] HERBERT E. ROBBINS AND E. J. G. PITMAN, "Application of the method of mixtures to quadratic forms in normal variates," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 552–560.

[15] HENRY SCHEFFÉ, "A useful convergence theorem for probability distributions," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 434–438.

[16] H. M. SCHWARTZ, "Sequences of Stieltjes integrals, III," *Duke Math. Journal*, Vol. 10 (1943), pp. 595–610.

[17] J. A. SHOHAT AND J. D. TAMARKIN, *The Problem of Moments*, Amer. Math. Soc., New York, 1943.

[17a] GEORGE P. STECK, "Limit theorems for conditional distributions," *Univ. of Calif. Pub. in Stat.*, Vol. 2, No. 12, pp. 237–84.

[18] HENRY TEICHER, "On the convolution of distributions," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 775–778.

[19] HENRY TEICHER, "On the convergence of projected distributions," *Ann. Inst. of Stat. Math.*, Vol. 9, No. 2 (1958), pp. 79–86.

[20] AUREL WINTNER, "Stratifications of Cauchy's 'stable' transcendents and of Mittag-Leffler's entire functions," *Amer. J. of Math.*, Vol. 80 (1958), pp. 111–124.