ON THE FACTORIZATION OF DISTRIBUTIONS

By Henry Teicher

Purdue University

1. Summary. A family of probability distributions is called "factor-closed" (f.c.) if it is closed under the operation of factorization. The classical binomial family and certain generalizations of it are shown to be f.c. The multinomial family is also f.c. Most families of infinitely divisible distributions are not f.c.

2. Introduction. If $F_1(x)$ and $F_2(x)$ are any two cumulative distribution functions (c.d.f.'s), the convolution (denoted by $*$) of F_1 with F_2 is again a c.d.f. say

$$F = F(x) = F_1 * F_2 = \int_{-\infty}^{\infty} F_1(x - y)\, dF_2(y) = \Pr\{X < x\} \tag{1}$$

where Pr denotes probability measure and X is called the random variable (r.v.) possessing the c.d.f. F. Further, if X_1 and X_2 are independent r.v.'s having c.d.f.'s F_1 and F_2 with corresponding Fourier transforms or characteristic functions (c.f.'s) $\phi_{x_1}(t)$ and $\phi_{x_2}(t)$, then $F = F_1 * F_2$ is the c.d.f. of $X = X_1 + X_2$ having, as is well known, the c.f.

$$\phi_x(t) = \phi_{x_1}(t) \cdot \phi_{x_2}(t) = \int_{\infty}^{\infty} e^{itx}\, dF(x). \tag{2}$$

If one commences with $\phi_x(t)$ or $F(x)$, any such representation as (2) or (1) is termed a *factorization* of $\phi_x(t)$ or $F(x)$ and the components $\phi_{x_i}(t)$ or $F_i(x)$ are called *factors*.

For an arbitrary distribution F, factorization is not unique. That is, $F = F_1 * F_2 = F_1 * F_3$ does not imply $F_2 = F_3$. If F is infinitely divisible, this is no longer possible. Many results concerning factorization, as well as references, are given by Lévy [4], [5].

To avoid trivialities, we presume in what follows that all c.d.f.'s have at least two points of increase and consider two c.f.'s $\phi_1(t)$ and $\phi_2(t)$ as equivalent if for some real α,

$$\phi_1(t) = \exp\{i\alpha t\}\, \phi_2(t).$$

The starting point of this investigation is the following

Definition. A family $\mathcal{S}$ of c.d.f.'s will be said to be *decomposable* ($\mathcal{S}'$) if, for any element F of $\mathcal{S}$, the relationship $F = G_1 * G_2$ implies that G_1 and G_2 are members of the family $\mathcal{S}'$. In particular, if $\mathcal{S} = \mathcal{S}'$, the family $\mathcal{S}$ will be called *factor-closed* (f.c.).[1]

Received 12/14/53.

[1] The class of all c.d.f.'s as well as the family of prime or indecomposable c.d.f.'s (i. e., the only "factors" of $\phi(t)$ are the trivial ones $\exp\{i\alpha t\}$ and $\phi(t)\exp\{-i\alpha t\}$) are trivially f.c.

Thus, Cramér's theorem [1], [2] on the factorization of the normal distribution states that the normal family is f.c. A corresponding result of Raikov [3] avers that the Poisson family is f.c.

For later usage, $P(z)$ is defined to be a *quasi-polynomial* if, for real d_j and r_j and integral $m \geqq 2$, $P(z) = \sum_{j=1}^{m} r_j z^{d_j}$. If, in addition, $m = 2$, $P(z)$ will be termed a *binomial* quasi-polynomial.

3. The general binomial family. We define a sequence of independent r.v.'s $\{X_j\}$, for $j = 1, 2, \cdots, n \geqq 2$, by

$$\Pr\{X_j = a_j\} = p_j, \qquad \Pr\{X_j = b_j\} = q_j = 1 - p_j,$$

where $a_j > b_j$ are real and $0 < p_j < 1$ for $j = 1, 2, \cdots, n$. Let

$$c_j = a_j - b_j > 0, \qquad V' = \sum_{j=1}^{n} X_j.$$

The c.f. of V' is

$$\phi_{V'}(t) = \prod_{j=1}^{n} (p_j e^{ita_j} + q_j e^{itb_j}) = \exp\left\{it \sum_{j=1}^{n} b_j\right\} \prod_{j=1}^{n} (p_j e^{itc_j} + q_j).$$

It suffices to consider the equivalent c.f.

$$\phi_V(t) = \prod_{j=1}^{n} (p_j e^{itc_j} + q_j) = A \prod_{j=1}^{n} (e^{itc_j} + \bar{q}_j)$$

where A is a constant and $\bar{q}_j = q_j p_j^{-1} > 0$. As $\phi_V(t)$ depends on the parameters a_j, b_j, p_j, and n, it represents a family of c.f.'s and there exists the corresponding family of c.d.f.'s, say $\mathfrak{I}$, whose explicit form is not required here. This family will be dubbed the *general binomial* family since it constitutes an obvious generalization of the classical binomial distributions connected with coin tossing, etc. It will be shown to be f.c. under certain conditions.

As the c.d.f. of X_j and hence of V is a step function with a finite number of jumps, the same must be true of any factor of the c.d.f. of V. We may therefore confine our attention (in looking for factors of $\phi_V(t)$) to c.f.'s of the form $\phi(t) = \sum_{j=1}^{m} r_j \exp\{itd_j\}$, where r_j is positive, d_j is real, and m is a positive integer $\geqq 2$. That is, we need only consider c.f.'s which are quasi-polynomials in $z = e^{it}$ with positive coefficients.

LEMMA. *If a polynomial with nonnegative coefficients admits a factorization into quasi-polynomials with nonnegative coefficients, it admits a factorization into (ordinary) polynomials having the same coefficients.*

PROOF. Let $P_0(z) = \prod_1^r P_i(z)$, where $P_0(z)$ is an ordinary polynomial with nonnegative coefficients and $P_i(z)$ is a quasi-polynomial for $i = 1, 2, \cdots, r$. Also, let m_i be the smallest exponent of $P_i(z)$ for $i = 0, 1, \cdots, r$. Since $\sum_1^r m_i$ is a nonnegative integer m_0, we have immediately $P_0'(z) = \prod_1^r P_i'(z)$, where $P_i'(z)$ is a quasi-polynomial with $m_i' = 0$ for $i = 0, 1, \cdots, r$. As any exponent appearing on the right side of the above equation must also appear on the left side, the $P_i'(z)$ must be ordinary polynomials. Q.E.D.

The distinguishing characteristic of the family $\mathfrak{I}$ is that $\phi_V(t)$ may be represented, by substituting $z = e^{it}$, as a product of binomial quasi-polynomials. If $\phi_V(t)$ may also be expressed as a product of quasi-polynomials which are not all reducible to binomial quasi-polynomials, it will be established that $\mathfrak{J}$ is not, in general f.c. Consider the identity

$$(z + 3)(z + 4)(z^3 + 8) \equiv (z + 2)(z^4 + 5z^3 + 2z^2 + 4z + 48) \equiv (z + 2)\cdot P_4(z). \tag{3}$$

Now, although $P_4(z)$ is in general reducible to $(z + 3)(z + 4)(z^2 - 2z + 4)$, it is irreducible into ordinary polynomials having nonnegative coefficients. By the lemma it is also irreducible to quasi-polynomials having nonnegative coefficients. If each parenthetic factor in (3) is divided by the sum of its coefficients and z is replaced by e^{it}, the expression on the left side is the c.f. of a member of $\mathfrak{J}$, while that in the middle is a product of two c.f.'s, the second of which is not a member of $\mathfrak{J}$.

On the other hand if $\mathfrak{J}$ is suitably restricted, it is f.c. Let $\mathfrak{J}_{c,2c}$ denote the subfamily of $\mathfrak{J}$ with $c_j = c$ or $2c$ for $j = 1, 2, \cdots, n$. We have then

THEOREM 1. *The family* $\mathfrak{J}_{c,2c}$ *is* f.c. *for any (positive)* c.

PROOF: If $c_j = 1$ or 2 for $j = 1, 2, \cdots, n$, then

$$\psi(z) = \phi_V(t) = A \prod_{j=1}^{n} (z^{c_j} + \bar{q}_j)$$

is the canonical decomposition of $\psi(z)$ into linear and quadratic factors. As $\bar{q}_j > 0$, it is clear that no matter how $\psi(z)$ is factored into ordinary polynomials, these must always be reducible to products of binomial factors with positive coefficients. With the lemma, this proves the theorem for the case $c = 1$. For arbitrary (positive) c, the transformation $y = z^c$ returns one to the case just examined.

COROLLARY 1. *Let* $\mathfrak{J}^p_{c,2c}$ *denote the subfamily of* $\mathfrak{J}_{c,2c}$ *wherein* $p_j = p$ *for* $j = 1, 2, \cdots, n$. *Then* $\mathfrak{J}^p_{c,2c}$ *is* f.c.

PROOF: By Theorem 1, $\mathfrak{J}^p_{c,2c}$ is decomposable $(\mathfrak{J}_{c,2c})$. That $\mathfrak{J}^p_{c,2c}$ is also f.c. follows directly from the fact that (for $c = 1$) all the roots of $\psi(z)$ must be equal to $-\bar{q}$ or $\pm i(\bar{q})^{1/2}$.

From Corollary 1, it follows that the only factors of the classical binomial (Bernoulli) distributions are themselves binomial distributions. It suffices here to choose $a_j \equiv 1$, $b_j \equiv 0$, and $p_j \equiv p$.

One might also define $\mathfrak{J}^p$ as that subfamily of $\mathfrak{J}$ for which $p_j = p$ for $j = 1, 2, \cdots, n$. However, it is simple to show via a counter-example that $\mathfrak{J}^p$ is not f.c.

In generalization of the preceding, we define for any integral $k \geqq 2$ the general k-nomial family of distributions, say, U_k, as follows: Let $\{X_j\}$ be a sequence of independent random variables with

$$\Pr\{X_j = a_{ji}\} = p_{ji}, \qquad i = 1, 2, \cdots, k,$$

$$0 < p_{ji} < 1, \qquad \sum_{i=1}^{k} p_{ji} = 1, \qquad \text{all } j = 1, 2, \cdots, n.$$

There is no loss of generality in supposing $a_{j1} > a_{j2} > \cdots > a_{jk}$ for all j. Then $V' = \sum_1^n X_j$ will be a c.v. having the "general k-nomial distribution." We consider the case $k = 3$.

THEOREM 2. *If a_{ji} forms, for each j, an arithmetic progression whose common difference is independent of j, and if $p_{j2}^2 < 4p_{j1}p_{j3}$ for all j, then U_3 is* f.c.

PROOF. Let $b_{ji} = a_{ji} - a_{j3} > 0$ for $i = 1$ or 2. Then if $b_{j2} = b$, by hypothesis $b_{j1} = 2b$. As earlier, it suffices to consider

$$\phi_V(t) = \prod_{j=1}^n (p_{j1}e^{it2b} + p_{j2}e^{itb} + p_{j3}).$$

But this is the canonical decomposition of a polynomial in $W = e^{itb}$. In view of the positivity of the coefficients, and the lemma, U_3 is necessarily f.c. The conditions $p_{j2}^2 < 4p_{j1}p_{j3}$ preclude trivial decompositions into binomial distributions.

The factor-closedness of U_3 cannot be extended even to the case where the a_{ji} are in arithmetic progression but the difference depends on j. It suffices to note the counter-example

$$(z^6 + 30z^3 + 6859/27)(z^2 + 2z + 6)(z^2 + 3z + 6)$$
$$\equiv (z^2 + 5z + 19/3)(z^4 + 25/3z^2 + 2/3z + 38)(z^4 + 10/3z^2 + z + 38).$$

4. The multinomial distribution. The factorization problem, as well as (1) and (2), extend readily to the m-dimensional case, that is, to m random variables or to a single vector random variable with m components. Where X, $F(x)$, and $\phi(t)$ were written previously, we need only substitute $(X_1, \cdots, X_m)$, $F(x_1, \cdots, x_m)$, and $\phi(t_1, \cdots, t_m)$. Cramér has shown [2] that the family of multivariate normal distributions is f.c.

We consider the classical multinomial distribution with n independent repetitions of an experiment whose m mutually exclusive and exhaustive outcomes $A_1, \cdots A_m$ have occurrence probabilities $p_1, \cdots p_m$. If X_j is the r.v. denoting the number of occurrences of A_j in the n trials, then

$$\Pr\{(X_1 = x_1), \quad \cdots, \quad (X_m = x_m)\} = \left(n!\Big/\prod_{i=1}^m x_i!\right) p_1^{x_1}p_2^{x_2} \cdots p_m^{x_m},$$

where $\sum_1^m x_i = n$ and $\sum_1^m p_i = 1$. Here

$$\phi(t_1, t_2, \cdots, t_m) = [p_1e^{it_1} + p_2e^{it_2} + \cdots p_me^{it_m}]^n.$$

Let $z_j = e^{it_j}$ and $\psi(z_1, \cdots, z_m) = \phi(t_1, \cdots, t_m)$. As before, the only possible factors of ψ are of the form

$$\sum_{x_1} \cdots \sum_{x_m} q_{x_1,x_2,\cdots,x_m} z_1^{x_1}z_2^{x_2} \cdots z_m^{x_m} = \psi_1(z_1, z_2, \cdots, z_m). \tag{4}$$

Again there is no loss of generality in supposing the x_i to be nonnegative integers, that is, that ψ_1 is a polynomial rather than a quasi-polynomial. We now prove

THEOREM 3: *The family of* (*classical*) *multinomial distributions is* f.c.

PROOF: Analogous to (2), we have, where ψ_i is of the form (4) for $i = 1$ or 2,

$$(p_1z_1 + p_2z_2 + \cdots + p_mz_m)^n = \psi_1 \cdot \psi_2$$

Since the irreducible factor $(p_1z_1 + \cdots + p_mz_m)$ is an n-fold factor of $\psi_1 \cdot \psi_2$, it must be an n_1-fold factor of ψ_1 and an n_2-fold factor of ψ_2, with $n_1 + n_2 = n$, that is,

$$\psi_i = \left(\sum_{i=1}^{m} p_i z_i\right)^{n_i} Q_i(z_1, \cdots, z_m), \qquad i = 1, 2.$$

Clearly, Q_i = constant = 1, since $\psi(1, 1, \cdots, 1) = \phi(0, \cdots, 0) = 1$. Finally, $0 < n_j < n$ if degenerate c.f.'s and c.d.f's are precluded, as earlier.

Slight generalizations of Theorem 3 are possible. The writer has proved that the family of (correlated) multivariate Poisson distributions is f.c., but this will not be given here.

5. Infinitely divisible families. Returning to the unidimensional case, $F(x)$ is called *infinitely divisible* (i.d.) if $[\phi(t)]^{1/n}$ is a c.f. for every positive integer n. Khintchine's form [6] of Lévy's formula [4] gives as a necessary and sufficient condition for $F(x)$ to be i.d. that

$$\log \phi(t) = i\gamma t + \int_{-\infty}^{\infty} \left(e^{itu} - 1 - \frac{itu}{1 + u^2}\right)\left(\frac{1 + u^2}{u^2}\right) dG(u) \tag{5}$$

where γ is real and $G(u)$ is bounded, monotonic nondecreasing and can be normalized so that $G(u^+) = G(u)$, $G(-\infty) = 0$, and $G(+\infty) = B$. Furthermore, the normalized representation is unique.

If $G(u)$ is a step function with only a single jump point, that is

$$\text{(a)} \quad G(u) = \begin{cases} 0, & u < 0, \\ \sigma^2, & u \geqq 0, \end{cases} \qquad \text{or (b)} \quad G(u) = \begin{cases} 0, & u < c \neq 0, \\ \frac{1}{2}\sigma^2, & u \geqq c, \end{cases}$$

then (a) yields the normal family of distributions while those in (b) are closely related to the Poisson family. If $c = 1$ and $\gamma = \frac{1}{2}\sigma^2$, then (b) is the Poisson family.

Suppose that $G(u)$ has n discontinuities $a_1 < a_2 < \cdots < a_n$ with saltuses $b_1, b_2, \cdots, b_n$. Then $G(u) = G_n(u) = G_n(u; a_1 \cdots a_n; b_1 \cdots b_n)$ has a corresponding i.d. c.f. $\phi_n(t; a_1 \cdots a_n; b_1 \cdots b_n)$ and c.d.f. $F_n(x; a_1 \cdots a_n; b_1 \cdots b_n)$.

For any fixed $n = 1, 2, \cdots$, consider the family

$$\mathfrak{F}_n = \{F_n(x; a_1 \cdots a_n; b_1 \cdots b_n)\}.$$

If $G_n(u)$ is a step-function, denote the corresponding family of c.d.f.'s by $\mathfrak{F}'_n$.

THEOREM 4. *For any fixed* $n = 2, 3, \cdots$, *the* i.d. *families* $\mathfrak{F}_n$ *and* $\mathfrak{F}'_n$ *are not* f.c.

PROOF. Let $b = \sum_{i=1}^{n-1} b_i$. Define

$$G_n^+(u) = \begin{cases} \dfrac{G_n(u) - b}{B - b}, & u \geqq a_n; \\ 0, & u < a_n; \end{cases} \qquad G_n^-(u) = \begin{cases} \dfrac{G_n(u)}{b}, & u < a_n, \\ 1, & u \geqq a_n. \end{cases}$$

Further, let $\phi_n^+(t)$ be given by (5) with $G_n^+(u)$ replacing $G(u)$, and define $\phi_n^-(t)$ analogously. Clearly, $G_n(u) = (B - b)G_n^+(u) + bG_n^-(u)$, whence

$$\phi_n(t; a_1 \cdots a_n; b_1 \cdots b_n) = \phi_n^+(t) \cdot \phi_n^-(t).$$

Since $G_n^+(u)$ has only one saltus, the c.d.f. corresponding to $\phi_n^+(t)$, say $F_n^+(x)$ belongs to either $\mathfrak{F}_1$ or $\mathfrak{F}_1'$. Thus, $\mathfrak{F}_n$ or $\mathfrak{F}_n'$ is not f.c. for $n \geqq 2$. In particular, if G_n is a step function, $F_n^+(x)$ is a normal or (almost) Poisson c.d.f.

REFERENCES

[1] H. Cramér, "Über eine Eigenschaft der normalen verteilungsfunktion," *Math. Z.*, Vol. 41 (1936), pp. 405–414.

[2] H. Cramér, *Random Variables and Probability Distributions*, Cambridge Tracts in Mathematics, Cambridge, 1937.

[3] D. Raikov, "On the composition of Poisson laws," *C.R. Acad. Sci.* URSS, Vol. 14 (1937), pp. 8–11.

[4] P. Lévy, *Theorie de l'Addition des Variables Aléatoires*, Paris, 1937.

[5] P. Lévy, "L'arithmétique des lois des probabilités," *J. Math. Pures Appl.*, Vol. 17 (1928), pp. 17–39.

[6] A. Khintchine, "Contribution à l'arithmétique des lois de distribution," *Bull. Math. Univ. Moscou*, Vol. 1 (1937), pp. 6–17.