ON THE MEAN CONSERVING PROPERTY

BY C. RADHAKRISHNA RAO
(Statistical Laboratory, Calcutta)

Received November 2, 1945
(Communicated by Dr. U. Sivaraman Nair, F.A.SC.)

1. INTRODUCTION

IT is the object of this paper to investigate into the general forms of the distribution laws which possess the mean conserving property and arrive at new frequency curves useful for graduation purposes and in tests of significance connected with means in random samples.

The mean conserving property may be defined as follows. Let a variate x be characterised by the probability differential

$$df = \phi\,(x, \lambda, \mu, \ldots.)\, dx \tag{1·1}$$

$\lambda, \mu, \ldots.$ being constants which may be called the parameters in the distribution law of x. Let $x_1, x_2, \ldots.$ be independent variates from distribution laws of the type (1·1) defined by the sets of parameters given as rows of the matrix

$$\begin{pmatrix} \lambda_1 & \mu_1 & \ldots. \\ \lambda_2 & \mu_2 & \ldots. \\ .. & .. & \ldots. \end{pmatrix} \tag{1·2}$$

The type (1·1) is said to possess the mean conserving property denoted by

$$\mathrm{M}^{\alpha,\, \beta,\, \nu, \ldots.}_{\xi,\, \eta,\, \zeta, \ldots.} \tag{1·53}$$

where $\alpha, \beta, \gamma, \ldots.$ and $\xi, \eta, \zeta, \ldots.$ are the sets of parameters which have to be kept fixed and can be varied in the distribution laws of $x_1, x_2, \ldots.$ if their mean follows the distribution law of the same type as in (1·1) but with the set of parameters

$$(\lambda, \mu, \ldots.), \tag{1·4}$$

where $\lambda, \mu, \ldots.$ are functions of the number of x's and the elements of the matrix (1·2).

A similar definition holds for the sum conserving property defined by

$$\mathrm{S}^{\alpha,\, \beta,\, \gamma, \ldots.}_{\xi,\, \eta,\, \zeta, \ldots.} \tag{1·5}$$

Both (1·3) and (1·5) can be made identical by the introduction of a new constant as a multiplier of x, and the property (1·3) or (1·5) will be referred to as M.

2. Properties of Distribution Laws Satisfying M

From the definition of M we derive the relations

$$\prod_{r=1}^{n} c\,(t,\ \lambda_r,\ \mu_r,\ldots.)= c\,(t,\ \lambda,\ \mu,\ldots.) \tag{2·1}$$

$$\sum_{r=1}^{n} k_s\,(\lambda_r,\ \mu_r,\ldots\ldots)= k_s\,(\lambda,\ \mu,\ldots.) \tag{2·2}$$

where $c\,(t)$ is the characteristic function corresponding to the distribution law of x and k_s, the sth semi-invariant of x.

These simple properties enable us to discover readily whether any distribution law satisfies M when a study of its moments and semi-invariants are made. Let us consider the Bessel function populations defined by

$$T_0\, e^{-\frac{c}{b}x}\ |\,x\,|^m \left.\begin{matrix} \Pi\ I_m\,(x) \\ k_m\,(x) \end{matrix}\right\} dx \tag{2·3}$$

where the upper function is taken when $|\,c\,| > 1$ and the lower when $|\,c\,| < 1$. The moment generating function, in either case, is given by

$$\{f\,(t,\,b,\,c)\}^{m+\frac{1}{2}}= \{(1-c^2)/(1-c+tb)\}^{m+\frac{1}{2}} \tag{2·4}$$

Since

$$\prod_{r=1}^{n} \{f\,(t,\,b,\,c)\}^{m_r+\frac{1}{2}}= \{f\,(t,\,b,\,c)\}^{mn+n/2} \tag{2·4}$$

where

$$mn= m_1+m_2+\cdots+m_n \tag{2·5}$$

it follows that the sum of n variates from populations of the type (2·3) defined by

$$\begin{pmatrix} b & c & m_1 \\ b & c & m_2 \\ \cdot\cdot & \cdot\cdot & \cdot\cdot \\ b & c & m_n \end{pmatrix} \tag{2·6}$$

follows the same type defined by

$$(b \quad c \quad n\,(m+\tfrac{1}{2})-\tfrac{1}{2}) \tag{2·7}$$

For the Bessel function population defined by

$$T_0= x^{\frac{m}{2}}\ e^{-dx}\ I_m\,(q\sqrt{x})\ dx \tag{2·8}$$

the sth semi-invariant is given by

$$k_s = (s-1)!\left\{\frac{(m+1)}{a^s} + \frac{q^2}{4}\frac{s}{a^{s+1}}\right\}. \tag{2·10}$$

Hence the sth semi-invariant of $nz = x_1 + x_2 + \ldots\ldots + x_n$ where x_i follows the law (2·8) with the parameters α, m_i, q_i is given by

$$k_s = (s-1)!\left\{\frac{m(n+1)}{a^s} + \frac{nq^2}{4}\frac{s}{a^{s+1}}\right\} \tag{2·10}$$

where $\qquad mn = m_1 + m_2 + \ldots. + m_n$

and $\qquad nq^2 = q_1{}^2 + q_2{}^2 + \ldots. + q_n{}^2$

which shows that (2·8) satisfies $M^{\alpha}{}_{m,q}$, the set of parameters

$$\begin{pmatrix} \alpha & m_1 & q_1 \\ \cdot\cdot & \cdot\cdot & \cdot\cdot \\ \alpha & m_n & q_n \end{pmatrix} \tag{2·11}$$

giving the set

$$(\alpha \qquad n(m+1)-1 \qquad \sqrt{n}q) \tag{2·12}$$

for the distribution of the sum. This result has been obtained by Bose (1937) when m's and q's are the same for all the variables.

If $k_s(r)$, $r = 1, 2, \ldots. n$ are the semi-invariants of the variates $x_1, x_2, \ldots. x_n$, the distribution laws of which satisfy M, then the semi-invariants of the variate $Z = x_1 + x_2 + \ldots. + x_n$ satisfy (2·2). Hence we get the result that *the distribution law of the sum or the mean of any number of independent variates, whose distribution laws satisfy M, also satisfies M.* From this it follows that the distribution law

$$df = c\, e^{-\alpha_i x}\, x^{n\lambda-1} \sum_{m=0}^{\infty} \frac{m!\, A_m Z^m}{\Gamma(n\lambda+m)} \tag{2·13}$$

$$n\lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_n$$

$$A_m = \frac{d^m}{da^m}\left[\prod_{j=1}^{m} \{1-(\alpha_i-\alpha_j)\, a\}^{-\lambda_j}\right]_{a=0}$$

derived by the author (1942) as the distribution of the sum of n different gamma variates following the laws

$$c\, e^{-\alpha_r x}\, x^{\lambda_r - 1}\, dx \tag{2·14}$$

$$r = 1, 2, \ldots., n$$

satisfies M.

It is well known that the distribution law

$$\frac{a^\lambda}{\Gamma(\lambda)}\, e^{-ax}\, x^{\lambda-1}\, dx \tag{2·15}$$

satisfies M^{α}_{λ}. Let the variates $x_1, x_2, \ldots, x_n$ have the probability densities

$$f(x_r, a_r, b_r, \ldots) \qquad r = 1, 2, \ldots n \tag{2·16}$$

with the corresponding cumulant functions

$$S(\beta, a_r, b_r, \ldots) \tag{2·17}$$

If it is known that the distribution of $Z = x_1 + x_2 + \ldots + x$ is (2·15) with the cumulant function $-\lambda \log(1 - i\beta/\alpha)$ then by hypothesis we have

$$\sum_{r=1}^{n} S(\beta, a_r, b_r, \ldots) = -\lambda \log(1 - i\beta/\alpha) \tag{2·18}$$

or

$$\sum_{r=1}^{n} F(\beta, a_r, b_r, \ldots) = -\lambda \tag{2·19}$$

where $F(\beta) = S(\beta)/\log(1 - i\beta/\alpha)$. Differentiating (2·19) with respect to β we get

$$\Sigma F'(\beta,) a_r, b_r, \ldots) = 0 \tag{2·20}$$

If this holds for all sets of $a_r, b_r, \ldots$ then we get by setting them equal values for all r, that

$$n F'(\beta, a, b, \ldots) = 0 \tag{2·21}$$

or

$$S(\beta, a, b, \ldots) = c \log(1 - i\beta/\alpha) \tag{2·22}$$

where c must necessarily be negative if the right-hand side is to represent a cumulant function. Hence $f_r(x_r, a_r, b_r, \ldots)$ is of the gamma type. So we get the result that *if the sum of* n *independent variates drawn from* n *different populations of the same type (the mathematical form remaining the same and the parameters may be varying) follows the gnmma type distribution law, then the above original populations also belong to the gamma type.* As a result of this we get the necessary and sufficient condition for the sum of n into pendent observations from a population to follow the gamma type is that the population itself is of the gamma type. Also it easily follows that *if the sum of two variates of which one follows the gamma type, is distributed in the gamma type, then the other variate also follows the gamma type.*

3. Differential Equation Satisfied by the Characteristic Function

When all the n variates are drawn from the same population, we have, if $c(t)$ represents the characteristic function of the distribution in the population satisfying M,

$$\{c(t, \lambda', \mu', \ldots)\}^n = c(t, \lambda, \mu, \ldots). \tag{3·1}$$

Taking logarithms and representing log $c(t)$ by $\psi(t)$ we get

$$n\,\psi(t, \lambda', \mu', \ldots.) = \psi(t, \lambda, \mu, \ldots.). \tag{3·2}$$

Starting from $\psi(t, \lambda', \mu', \ldots.)$ we can form the differential equation satisfied by ψ under some analytical conditions by eliminating the constants $\lambda', \mu', \ldots.$ The order of the differential equation is, in general, equal to the number of constants eliminated. Since $\psi(t, \lambda, \mu, \ldots.)$ also should satisfy this equation, we require that ψ and $n\psi$ should both satisfy the differential equation for ψ. If the differential equation satisfied by ψ is

$$\mathrm{D}(\psi, \psi' \ldots.) = 0 \tag{3·3}$$

then

$$\mathrm{D}(n\psi, n\psi', \ldots.) = 0 \tag{3·4}$$

which shows that D must be homogeneous in $\psi, \psi', \psi'', \ldots.$ the homogeneity of $\mathrm{D}(x, y, \ldots.)$ being defined as

$$\mathrm{D}(ax, ay, \ldots.) = f(a)\,\mathrm{D}(x, y, \ldots.) \tag{3·5}$$

The differential equation (3·4) may be denoted by $\mathrm{D_H} = 0$. Hence we get the results that *the semi-invariant generating function corresponding to a distribution law satisfying M satisfies a homogeneous differential equation homogeneity being defined as in* (3·5).

From the above differential equation $\mathrm{D_H} = 0$, we can derive the differential equation satisfied by $c(t)$ by making the substitutions

$$\psi = \log c, \quad \psi' = c'/c \text{ etc.} \tag{3·6}$$

If the differential equation $\mathrm{D_H} = 0$ arising out of the probability differential $\phi(x)\,dx$ is homogeneous then the differential equation arising out of the probability differential $\chi(a)\,e^{ax}\,\phi(x)\,dx$ is also homogeneous which shows that the property M is conserved by the multiplication of the distributive law by an exponential factor.

We shall now consider some distribution laws obtained by inversion from $\mathrm{D_H} = 0$. The simplest case is when the order of $\mathrm{D_H} = 0$ is one, in which case the differential equation becomes

$$\frac{\psi'}{\psi} = f(t) \text{ (an arbitray function)} \tag{3·7}$$

which gives the solution

$$\psi = \lambda\, e^{\phi(t)} \text{ where } \phi(t) = \int f(t)\,\mathrm{d}t \tag{3·8}$$

and $c(t) = e^{\psi}$. If $e^{\phi(t)}$ admits an expansion in series we get

$$c(t) = e^{\lambda\left(a_0 + a_1 \frac{it}{1!} + a_2 \frac{(it)^2}{2!} + \cdots\right)}$$

which shows the $\lambda a_1, \lambda a_2, \ldots.$ are the semi-variants of the distribution. In particular if $a_1, a_2, \ldots.$ are the semi-invariants for any distribution law then $na_1, na_2, \ldots.$ are the semi-invariants for the sum of n independent observations from the above distribution. This gives the result that *the distribution law of the sum or mean of a number of observations from any distribution law with finite semi-invariants satisfies M.* The functional form of the distribution law may change with n but may be capable of being represented by a general type of function. Thus we get a huge class of distribution laws satisfying M.

4. Measures of Departure from M

Given the probability density $\phi_1(x, \lambda, \mu, \ldots.)$ of a variate x, we can, in general, replace the constants $\lambda, \mu, \ldots.$ by an equivalent number of semi-invariants of suitable orders so that $\phi_1(x, \lambda, \mu, \ldots.)$ may be written as $\phi(x, k_1, k_2, \ldots.)$ where $k_1, k_2, \ldots.$ are the first, second, etc., semi-invariants. Let the cumulant generating function be $k(t, k_1, k_2, \ldots.)$. Then the cumulant generating function of the mean of n observations is $nk(t, k_1, k_2, \ldots.)$. Let $\mathrm{R}(t, n, k_1, k_2, \ldots.)$ be defined by

$$n\,k\left(\frac{t}{n}, k_1, k_2 \ldots.\right) = k(t, k_1, k_2/n, \ldots.) + \mathrm{R}(t, n, k_1, \ldots.). \tag{4·1}$$

When the distribution law satisfies M, R vanishes. If not, it can be written, when it admits expansion, as

$$\mathrm{R}(t, n, k_1, \ldots.) = \sum_{p=s}^{\infty} \frac{a_p}{p!} t^p \tag{4·2}$$

where s depends on the number of constants involved in the distribution law of x. Taking the exponentials in (4·1) we get

$$e^{nk\left(\frac{t}{n}, k_1, k_2, \cdots\right)} = e^{k(t, k_1, k_2/n, \cdots)} \left(1 + \sum_{p=s}^{\infty} \frac{b_p}{p!} t^p\right). \tag{4·3}$$

On taking the integral transform we get the probability density $\mathrm{S}(z)$ of the mean as

$$\mathrm{S}(z) = \phi(z, k_1, k_2/n, \cdots) + \sum_{p=s}^{\infty} \frac{b_p}{p!} \frac{d^p}{dz^p} \phi(z). \tag{4·4}$$

The expression consists of two portions. The second part vanishes when M holds and measures the departure from M when M does not hold. The considerations of replacing $\mathrm{S}(z)$ by $\phi(z, k_1, k_2/n, \ldots.)$ depend upon the magnitude of this measure. It is proposed to study the effect of the departure from M when n increases and also to consider the effect of replacing $\phi(z, k_1, k_2/n)$ by the normal approximation.

5. Series in Orthogonal Polynomials

Let $M(\alpha)$ be the *m.g.f.* of a variate whose distribution law satisfies M and $f(\alpha)$ any arbitrary function. If

$$G_1 = M_1(\alpha) \sum_{\gamma=0}^{\infty} \frac{a_r}{r!} \{f(\alpha)\}^r \tag{5·1}$$

$$G_2 = M_2(\alpha) \sum_{\gamma=0}^{\infty} \frac{b_r}{r!} \{f(\alpha)\}^r \tag{5·2}$$

are the *m.g.f.*'s of x_1 and x_2 following the distribution laws

$$\phi(x, \lambda_1, \mu_1, \ldots)\, dx \tag{5·3}$$

$$\phi(x, \lambda_2, \mu_2, \ldots)\, dx \tag{5·4}$$

then the *m.g.f.* of $z = x_1 + x_2$ is

$$G = G_1 G_2 = M_1 M_2 \sum_{l=0}^{\infty} \sum_{l=r+s} \frac{a_r\, b_s}{r!\; s!} f^l \tag{5·5}$$

$$= M_3 \sum_{l=0}^{\infty} \frac{c_l}{l!} f^l.$$

The functional form of G will be same as that of G_1 and G_2 if the same holds for c_l and a_r and b_r *i.e.*, a_r satisfies the recurrence relation

$$\sum_{r+s=l} \frac{a_r\, b_s}{r!\; s!} = \frac{c_l}{l!} \tag{5·6}$$

where b_s and c_l are of the same form as a_r differing only in the parameters involved in them. If (5·6) holds then by successive applications we can show that

$$G = G_1\, G_2 \ldots G_n \tag{5·7}$$

has the same functional form as G_i. By a suitable selection of $M(\alpha)$ and $f(\alpha)$ we can get several distribution laws satisfying M.

Let $M(\alpha) = (1-\alpha)^{-p}$ and $f(\alpha) = \alpha/(1-\alpha)$

Since $(1-\alpha)^{-p}\, \alpha^r/(1-\alpha)^r$

$$= \frac{(-1)^r\, \alpha^r}{p\,(p+1)\cdots(p+r-1)} \frac{d^r}{d\alpha^r} (1-\alpha)^{-p}$$

i.e., the *m.g.f.* corresponding to

$$\frac{\Gamma(p)}{\Gamma(p+r)} L_r(x, p)\, \phi(x) \tag{5·8}$$

where $\phi(x) = e^{-x}\, x^{p-1}/\Gamma(p)$ and $L_r(x, p) = \left(\frac{-d}{dx}\right)^r x^r\, \phi(x)$ we see that

$(1-\alpha)^{-p}\left(a_0 + \frac{a_1}{1!} f(l) + \cdots\right)$ is the *m.g.f.* of

$$\gamma(x) = \phi(x) \sum_{\gamma=0}^{\infty} \frac{a_r}{r!} \frac{\Gamma(p)}{\Gamma(p+r)} L_r(x, p) \tag{5·9}$$

where a_r satisfies (5·6). This is a series in Lagurre's polynomials satisfying M.

Let $M(\alpha) = e^{\alpha^2\sigma^2}$, $f(\alpha) = \alpha$. We get that

$$e^{\alpha^2\sigma^2} \sum_{r=0}^{\infty} \frac{a_r}{r!} \alpha^r \tag{5·10}$$

is the *m.g.f.* of

$$\psi(x) = \text{const.}\, e^{-\frac{x^2}{2\sigma^2}} \left(a_0 + \frac{a_1}{1!} H_1 + \frac{a_2}{2!} H_2 + \cdots \right), \tag{5·11}$$

where $$H_p = e^{\frac{x^2}{2\sigma^2}} \frac{d^p}{dx^p} e^{-\frac{x^2}{2\sigma^2}},$$

which is a series in Hermite polynomials satisfying M if a_r satisfies (5·6). By suitable selections of $M(\alpha)$ and $f(\alpha)$ we can obtain the development of probability functions satisfying M in a series of Bessel functions (Neumann's expansion) involving $J_n(x)$, Hypergeometric and other suitable functions. These are omitted here as they are not of direct interest in graduation or tests of significance.

6. Some Special Series

A series of the form

$$c\, e^{-ax} x^m \Sigma \frac{a_r}{r!} x^r \tag{6·1}$$

is of special interest for the gamma type distribution occurs as a generating function. We shall investigate into the nature of a_r so that (6·1) satisfies M. If x and y follow the type (6·1) with the parameters

$$\begin{pmatrix} a & \lambda & m_1 & p_1 & \cdots \\ a & \lambda & m_2 & p_2 & \cdots \end{pmatrix} \tag{6·2}$$

then the distribution of $z = x + y$ is given by

$$c'\, e^{-az}\, dz \frac{d}{dz} \iint_{\Omega} \Sigma\Sigma \frac{a_r b_s}{r!\, s!} x^{m_1+\lambda r} y^{m_2+\lambda s}\, dx\, dy \tag{6·3}$$

where the integral is over the domain Ω defined by $x > 0$, $y > 0$ and $x + y \leqslant z$. This becomes apart from const. $e^{-az} dz$,

$$\frac{d}{dz} \Sigma\Sigma \frac{a_r}{r!\, \Gamma(m_1 + \lambda r + 1)} \frac{b_s}{s!\, \Gamma(m_2 + \lambda s + 1)} \times \frac{z^{m_1+m_2+\lambda(r+s)+2}}{\Gamma(m_1 + m_2 + \lambda(r+s) + 3)}$$

$$= \Sigma\Sigma \frac{a_r' b_s'}{r!\, s!} \frac{z^{m_1+m_2+\lambda(r+s)+1}}{\Gamma(m_1 + m_2 + \lambda(r+s) + 2)} \tag{6·4}$$

where

$$a_r' = a_r / \Gamma(m_1 + \lambda r + 1)$$

$$b_s' = b_s / \Gamma(m_2 + \lambda s + 1)$$

The distribution of Z now becomes

$$c'' e^{-az}\, dz \sum_{l=0}^{\infty} \frac{A_l'}{l!} \frac{z^{m_1+m_2+\lambda l+1}}{\Gamma(m_1+m_2+\lambda l+1)}$$

$$= c = e^{-az}\, dz \sum_{l=0}^{\infty} \frac{A_l}{l!} z^{m+\lambda l}$$

where $m = m_1 + m_2 + 1$ and

$$A_l\, \Gamma(m_1+m_2+\lambda l+1) = A_l' = l!\, \sum_{r+s=l}\sum \frac{a_r'}{r!}\frac{b_s'}{s!}$$

which shows that (6·1) satisfies M if a'_r satisfies (5·6). The method of proof can be extended to the sum of n variates. Some particular forms of a'_r give rise to important distribution laws.

(*a*) $a'_r = q^r$ then (6·1) satisfies $M^{\alpha,\lambda}_{m,q}$

(*b*) $a'_r = \Gamma\left(\frac{m+1}{\lambda} + r\right)$ satisfies $M^{\alpha,\lambda,q}_{m}$

(*c*) $a'_r = \Gamma(p+r)\, q^r$ satisfies $M^{\alpha,\lambda,q}_{p,m}$

(*d*) $a'_r = \sum_{\gamma_1+\gamma_2+\ldots=r} \Gamma(p_1+r_1)\, q_1^{r_1}\, \Gamma(p_2+r_2)\, q^{r_2}$ satisfies $M^{\alpha,\lambda,q_1,q_2,\ldots}_{m,p_1,p_2,\ldots}$

The solutions for (*a*), (*b*), (*c*) and (*d*), in the special case $\lambda = 1$, become

(*a*) $c\, e^{-dx}\, x^{\frac{m}{2}}\, I_m(q\sqrt{x})$

(*b*) $c\, e^{-ax}\, x^m\, I_m(qx)$

(*c*) $c\, e^{-ax}\, x^m\, {}_1F_1(p, m+1, qx)$

(*d*) The distribution (2·13) of Section 2.

References

Bose, R. C. .. "On the distribution of the means of samples drawn from a Bessel function population," *Sankhya*, 1937, 3, (3), 262–64

Rao, C. R. .. "On the sum of '*n*' observations from different gamma type populations," *Science and Culture*, 1942, 7, 614–15.