NOTE ON THE DISTRIBUTION OF MEANS OF SAMPLES OF N DRAWN FROM A TYPE A POPULATION

By

CECIL C. CRAIG
National Research Fellow

Recently in this journal, Dr. George A. Baker has found "the distribution of the means of samples drawn at random from a population represented by a Gram-Charlier series."[1] It is the purpose of this note to call attention to the fact that by the use of the semi-invariant notation Dr. Baker's results may be reached in very many fewer steps.

Let the parent population be represented by

$$f(x) = \phi(x)\left[1 + \frac{a_3}{\sigma_x^3} H_3\left(\frac{x}{\sigma_x}\right) + \frac{a_4}{\sigma_x^4} H_4\left(\frac{x}{\sigma_x}\right) + \cdots + \frac{a_k}{\sigma_x^k} H_k\left(\frac{x}{\sigma_x}\right)\right] \tag{1}$$

in which

$$\phi(x) = \frac{1}{\sigma_x\sqrt{2\pi}}\, e^{-\frac{x^2}{2\sigma_x^2}} \tag{2}$$

[1]Vol. 1, No. 3 (Aug., 1930), pp. 199-204.

the origin for x being chosen at the mean, and

(3) $$H_k(t)e^{-\frac{t^2}{2}} = D_t^k\left(e^{-\frac{t^2}{2}}\right).$$

We shall first find the distribution function of $z = x_1 + x_2 + \cdots + x_N$ in which x_i, $i = 1, 2, \cdots N$, has the frequency function $f(x)$. Let us assume the frequency function of z is given by

(4) $$F(z) = \phi(z)\left[1 + \frac{A_3}{\sigma_z^3} H_3\left(\frac{z}{\sigma_z}\right) + \frac{A_4}{\sigma_z^4} H_4\left(\frac{z}{\sigma_z}\right) + \cdots + \frac{A_l}{\sigma_z^l} H_l\left(\frac{z}{\sigma_z}\right)\right]$$

Then the semi-invariants of $f(x)$, $\lambda_1, \lambda_2, \cdots \lambda_k$ are defined by the formal identity in t:

(5) $$e^{\lambda_1 t + \frac{1}{2!}\lambda_2 t^2 + \frac{1}{3!}\lambda_3 t^3 + \cdots} = \int_{-\infty}^{\infty} dx\, f(x) e^{xt} \quad (\lambda_1 = 0 \text{ in this case})$$

and on integration, using (3), we get at once on the right:

$$e^{\lambda_2 \frac{t^2}{2}}\left[1 - a_3 t^3 + a_4 t^4 + \cdots (-1)^k a_k t^k\right]$$

Similarly for the semi-invariants $L_1, L_2, L_3 \cdots$ of $F(z)$ we have

(6) $$e^{L_1 t + \frac{1}{2!}L_2 t^2 + \frac{1}{3!}L_3 t^3 + \cdots} = e^{L_2 \frac{t^2}{2}}\left[1 - A_3 t^3 + A_4 t^4 - \cdots + (-1)^l A_l t^l + \cdots\right]$$

But because of the well-known fact that $L_r = N\lambda_r$ this gives

$$1 - A_3 t^3 + A_4 t^4 - \cdots - (-1)^l A_l t^l$$
$$= \left[1 - a_3 t^3 + a_4 t^4 - \cdots (-1)^k a_k t^k \right]^N$$

an identity in t. Thus

$$(7)\quad A_r = \sum \frac{N!}{\nu_3!\,\nu_4! \cdots \nu_k!\,(N-\nu_3-\nu_4 \cdots \nu_k)!}\, a_3^{\nu_3} a_4^{\nu_4} \cdots a_k^{\nu_k}$$

the summation including all terms for which

$$3\nu_3 + 4\nu_4 + \cdots + k\nu_k = r$$

Remembering that $\sigma_z = \sqrt{L_2} = \sqrt{N}\,\sigma_x$, we have on substitution in (4) the expression for $F(z)$ since only a finite number of A_r's (depending on N) are different from zero.

To get the distribution of $z' = \frac{x_1 + x_2 + \cdots + x_N}{N}$ only involves the appropriate change of unit.

Cecil C. Craig

Stanford University.