NOTE ON THE DISTRIBUTIONS OF THE STANDARD DEVIATIONS AND SECOND MOMENTS OF SAMPLES FROM A GRAM-CHARLIER POPULATION

By G. A. Baker

T. N. Thiele in his "Theory of Observations" makes the following statement with regard to the distributions of the higher half-invariants in samples of n: "Not even for μ_2 have I discovered the general law of errors."[1] The purpose of this paper is to shed some light on the distribution of μ_2 and to give the distribution of second moments about a fixed point when the sampled population can be represented by a Gram-Charlier series.

The distribution of the second moments about a fixed point of samples is given in complete generality. It is known that if the sampled population is normal there is a simple relation between the distribution of the standard deviations of samples of n and the distribution of the second moments of the samples about the mean of the population. It was thought that such a relation might exist in case the sampled population could be represented by a Gram-Charlier series. Such is not the case. Again, it was thought that by obtaining the distribution of the standard deviations for samples of 2, 3, 4, $\cdots$ it might be possible to deduce empirically a general law of distribution. This proved an unfruitful line of investigation but required so much labor that the results should be reported to save others time and energy.

First, suppose that a population may be represented as

$$f(x) = a_0\varphi_0(x) + a_3\varphi_3(x) + a_5\varphi_5(x) + \cdots \tag{1}$$

where

$$\varphi_i(x) = \frac{d^i(e^{-\frac{1}{2}x^2})}{dx^i}.$$

Then applying Theorem II of the author's paper on "Random Sampling from Non-Homogeneous Populations"[2] we deduce at once the following theorem.

Theorem I. The distribution of the second moments about the origin of (1) of samples of n drawn at random from a population represented by (1) is precisely the same as the distribution of the second moments about the same point of samples of n drawn from a population represented by the first term of (1), that is a normal population, and is proportional to $x^{\frac{n-2}{2}} e^{-\frac{1}{2}x}$ (loc. cit.)

[1] Thiele, T. N., "The Theory of Observations," reprinted in the *Annals of Mathematical Statistics*, Vol. 2, No., 2, May, 1931, p. 208.

[2] *Metron*, Vol. 8, No. 3, Feb. 28, 1930.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Mathematical Statistics.* www.jstor.org

This is not so surprising as it may seem at first if it is remembered that the odd subscript terms of a Gram-Charlier series slice off frequencies on one side of the mean of $a_0\varphi_0(x)$ and add them onto the other side in the same manner.

If we suppose that a population is given as

$$f(x) = a_0\varphi_0(x) + a_3\varphi_3(x) + a_4\varphi_4(x) + \cdots \tag{2}$$

in the same manner we get the following theorem.

THEOREM II. The distribution of the second moments measured from the origin of (2) of samples of n drawn at random from (2) will be a combination of distributions of the type of Theorem I with only even subscript terms contributing anything. The variations in the component distributions will consist of differences in the constant factors and the exponent of x, the estimate of the second moment. The lowest exponent will be $\frac{n-2}{2}$.

For instance, if

$$f(x) = a_0\varphi_0(x) + a_3\varphi_3(x) + a_4\varphi_4(x) \tag{3}$$

and $n = 2$, the estimates of the second moment will be distributed as proportional to

$$e^{-\frac{1}{2}x}\left[(a_0+3)^2 - 12a_4(a_0+3)x + (36a_4^2 + 6a_0a_4 + 18a_4)\frac{x^2}{2!} - 36a_4^2\frac{x^3}{3!} + 9a_4^2\frac{x^4}{4!}\right].$$

Thus, it can be said that we know the distribution of the second moments of samples about a fixed point if the sampled population is of the Gram-Charlier type in the sense that given the number of terms necessary for an adequate representation and the number in the samples we can write down the desired distribution. However, this is not a simple matter. Further, if some relation existed between the distributions of the second moments about a fixed point and the standard deviations of the samples we would know the latter distribution also. Such a relation is not apparent for samples of 2 and 3.

Let us investigate the correlation surfaces of the means and standard deviations of samples of 2 and 3 drawn at random from a population represented by the first few terms of a Gram-Charlier series after the method of Dr. A. T. Craig.[3] The distributions of the standard deviations can then be obtained immediately by integration.

Suppose that

$$f(x) = a_0\varphi_0(x) + a_3\varphi_3(x) + a_4\varphi_4(x) \tag{4}$$

[3] *Annals of Mathematical Statistics*, Vol. 3, No. 2, May, 1932, pp. 126–140.

and that we are considering samples of 2. The probability of the concurrence of x_1 and x_2 is

$$f(x_1)f(x_2) \tag{5}$$

and

$$\begin{aligned} x_1 &= -s + x \\ x_2 &= s + x \end{aligned} \tag{6}$$

where s is the standard deviation and x is the mean of a sample of 2. By means of (6), (5) becomes

$$\begin{aligned} &e^{-(s^2+x^2)}[a_0^2 + a_0a_3(-6s^2x - 2x^3 + 6x) \\ &+ a_0a_4(2s^4 + 12s^2x^2 - 12s^2 - 12x^2 + 6) \\ &+ a_3^2(-s^6 + 3s^4x^2 + 6s^4 - 3s^2x^4 - 9s^2 + 9x^2 - 6x^4 + x^6) \\ &+ a_3a_4(2s^6 - 6s^4x^3 - 6s^4x + 6s^2x^5 - 12s^2x^3 + 18s^2x - 2x^7 \\ &\qquad + 18x^5 - 42x^3 + 18x) \\ &+ a_4^2(s^8 - 4s^6x^2 - 12s^6 + 6s^4x^4 + 12s^4x^2 + 42s^4 - 4s^2x^6 \\ &\qquad + 12s^2x^4 - 36s^2x^2 - 36s^2 + x^8 - 12x^6 + 42x^4 - 36x^2 + 9)]. \end{aligned} \tag{7}$$

To find the distribution of s we must integrate from $-\infty$ to ∞ with respect to x. Thus, (8) is obtained.

$$\begin{aligned} &\sqrt{\pi}\, e^{-s^2}\Big[a_0^2 + a_0a_4(2s^4 - 6s^2) + a_3^2\Big(-s^6 + \frac{15}{2}s^4 - \frac{45}{4}s^2 + \frac{15}{8}\Big) \\ &\qquad + 2a_3a_4s^6 + a_4^2\Big(s^8 - 14s^6 + \frac{105}{2}s^4 - \frac{105}{2}s^2 + \frac{105}{2}\Big)\Big]. \end{aligned} \tag{8}$$

If we retain only two terms of (3), i.e. use

$$f(x) = a_0\varphi_0(x) + a_3\varphi_3(x) \tag{9}$$

and consider samples of 3 we obtain as the correlation surface of x and s

$$\begin{aligned} &\frac{18\pi}{\sqrt{3}}\, s e^{-\frac{1}{2}(3x^2+3s^2)}\Big[a_0^3 - \frac{a_0^2a_3}{4}(-40x^3 + 24xs^2 - 24x) \\ &\quad + \frac{a_0a_3^2}{64}(-84s^6 + 525x^2s^4 - 2752x^4s^2 \\ &\qquad + 576s^4 - 1008x^2s^2 - 288s^2 - 5586x^6 + 270x^4 - 1728x^2) \\ &\quad + \frac{a_3^3}{64}(28s^6 - 6189x^2s^4 - 28x^4s^2 - 629x^6 + 288s^4 + 1344x^4 \\ &\qquad + 4608x^2s^2 - 288s^2 + 729x^2)\Big]. \end{aligned} \tag{10}$$

The distribution of s can be obtained as before. The processes involved in obtaining (7) and (10) are so complicated that the general rule for writing the distribution of s is not apparent. Also, the relation of the distributions of s to the corresponding distributions of the second moments about a fixed point is not apparent.

In summary, the general distributions of the second moments about a fixed point of samples from a population represented by a definite number of terms of a Gram-Charlier series and the distributions of the standard deviations of samples of 2 and 3 from the same type of population are given and compared. No apparent relation exists between them.