## Inverse Statistical Variates

HALDANE[1] has recently explained a method of using a hæmacytometer in which counting is stopped when a fixed number of cells has been recorded. This sampling technique, with suitable modifications, can be applied to other populations and the theory reveals some interesting relationships.

In ordinary binomial sampling, if the variate $(r)$ is the number of occurrences of a certain event when it is given a fixed number $(N)$ of opportunities of occurring, in each of which the probability of occurrence is $\pi$, then the seminvariant-generating function of $r$ is

$$L_N(t) \equiv \ln\sum_{r=0}^{N} e^{-rt}(NC_r)\pi^r(1-\pi)^{N-r}$$
$$= N \ln(1-\pi+\pi e^{-t}) = NL_1(t).$$

If, however, the number of occurrences is fixed $(R)$, the probability that $n$ opportunities will be required is equal to the probability that the event will occur $R-1$ times in the first $n-1$, compounded with the probability of an occurrence at the $n$th opportunity. Hence it is

$$\frac{(n-1)!}{(R-1)!(n-R)!}\pi^R(1-\pi)^{n-R},$$

(the coefficients being those of the expansion of a binomial of power $-R$), and the seminvariant-generating function of $n$ can be shown to be $-R\ln\left(1-\frac{1-e^t}{\pi}\right)$.

This seminvariant-generating function is simply $RL_1^{-1}(t)$, where $L_1(L_1^{-1}(t)) \equiv t$—a rather striking result. Moreover, the coefficients of variation of $r$ and $n$ are equal when $R = N\pi$. In this case, the value of $R$ in the distribution of $n$ is the mean value of $r$, while the value of $N$ in the distribution of $r$ is the mean value of $n$. In view of these relationships between the two distributions, I suggest that the sampling with a fixed number of occurrences should be termed *inverse binomial sampling*.

The Poisson limit is obtained on writing $\pi = k\mu$, $x = nk$, $X = Nk$, and then making $k$ vanish while $x, X$ and $\mu$ remain finite. The seminvariant-generating functions of $r$ and $x$ become $X\mu(e^{-t}-1)$ and $-R\ln(1+t/\mu)$, which are of the form $Xf(t)$, $Rf^{-1}(t)$, where $f(f^{-1}(t)) \equiv t$; and the coefficients of variation are equal when $R = X\mu$. The first of these seminvariant-generating functions is that of the usual Poisson limit, while the second corresponds to a distribution of Pearson's Class III, in which the variate can take any non-negative value, and such a distribution could therefore be termed the *inverse* of the Poisson distribution. That inverse Poisson sampling with $R = 1$ gives a Class III distribution of this type has already been shown by Marsden and Barratt[2].

A similar relation appears in the theory of the effect of Brownian motion on the times taken by colloid particles to travel a fixed distance under electrophoresis or convection. The general theory, developed by using the properties of seminvariant-generating functions, will be published elsewhere, but a result relevant here is that if a number of particles are released in a plane at time $\eta = 0$ and the distribution of their distances from this plane is Gaussian with variance $\beta\eta$ after a time $\eta$, $\beta$ being a constant—so that, if their mean moves with constant velocity $\omega$, the seminvariant-generating function of their displacements $(x)$ from the original plane is $\eta(-\omega t + \frac{1}{2}\beta t^2)$, which is equal to $\eta F(t)$, say—

then the seminvariant-generating function of the times $(y)$ taken by the individual particles to reach a parallel plane at a distance $\xi$ is

$$\xi F^{-1}(t) = \xi(\omega - \sqrt{\omega^2 + 2\beta t})/\beta.$$

The corresponding distribution function has been given by Schrödinger[3] and Smoluchowski[4], who derived it by a different method. The coefficients of variation of $x$ and $y$ are equal when $\xi = \eta\omega$.

In general, if a variate $x$ has the seminvariant-generating function $\alpha L(t)$, and another variate $y$ the seminvariant-generating function $\beta L^{-1}(t)$, then their means are respectively $\alpha\varkappa_1$ and $\beta/\varkappa_1$ and their variances $\alpha\varkappa_2$ and $\beta\varkappa_2/\varkappa_1^3$, where $\varkappa_1$ and $\varkappa_2$ are the first and second seminvariants of the seminvariant-generating function $L(t)$. The coefficients of variation are equal if $\beta = \alpha\varkappa_1$, so that the mean value of $x$ is $\beta$, while that of $y$ is $\alpha$. Thus, when the seminvariant-generating functions of $x$ and $y$ are so related, and the accuracy is measured by the coefficient of variation, $y/\beta$ is as accurate an estimate of $1/\varkappa_1$ as $x/\alpha$ is of $\varkappa_1$. Further, if the coefficient of variation is small, it is approximately the same for $\beta/y$, which is therefore as good an estimate of $\varkappa_1$ as $x/\alpha$. I suggest that two variates related as $x$ to $y$ should be termed *inverse variates*, and either distribution may be said to be the inverse of the other.

I have been able to prove that the seminvariant-generating functions must be related in this way in the case of the Brownian motion problem, even when the latter is generalized by the removal of the assumption that the particles spread out on either side of their mean in a Gaussian manner. On the other hand, I have not been able to formulate a general sampling rule by which the inverse variate can be obtained from a given group of distributions, so that the definition of inverse variates remains a relation between their seminvariant-generating functions and between their means.

M. C. K. TWEEDIE.

Radium Centre,
North Staffordshire Royal Infirmary,
Stoke-on-Trent.
Jan. 22.

[1] Haldane, *Nature*, **155**, 49 (1945).
[2] See Rutherford, Chadwick and Ellis, "Radiations from Radioactive Substances" (1930), §34c.
[3] Schrödinger, *Phys. Z.*, **16**, 289 (1915).
[4] Smoluchowski, *Phys. Z.*, **16**, 318 (1915).

---

MR. TWEEDIE'S demonstration of the relation between the seminvariant-generating (or cumulant-generating) functions of two sets of related distributions is of great interest. However, the word "inverse" is already used in statistical theory, first in the phrase "inverse probability", and secondly in connexion with the problem of finding a distribution with a given characteristic function. If it is used in yet a third sense, some confusion may possibly arise, although "inversion" is already used in geometry and algebra with different meanings, and without inconvenience. It may be added that Mr. Tweedie's letter does not give an alternative solution of the problem raised by my own, since I am concerned with the distribution, not of $n$, but of $\frac{R-1}{n-1}$, which gives an unbiased estimate of the probability $\pi$.

J. B. S. HALDANE.

Department of Biometry,
University College,
London.