NOTE ON A METHOD OF SAMPLING

By Carlos E. Dieulefait

National University of Litoral, Argentina

Olds[1] has considered the following problem: *Given a lot of size $m = s + r$ containing s items of a specified kind. Items are drawn without replacement until j of the s items have been drawn. The problem is to determine the probability law of n, the number of drawings which have to be made.* In the present note, we shall consider a certain limiting form for the probability function of n and make some remarks concerning repeated sampling of this type.

If n is the size of a drawing $j \le n \le r + j$ its probability law $P(n)$ is given by:

$$P(n) = \frac{C_{r,n-j}C_{s,j}}{C_{m,n}} \cdot \frac{j}{n} = \frac{\Gamma(s+1)}{\Gamma(j)\Gamma(s-j+1)} C_{r,n-j} \int_0^1 x^{n-1}(1-x)^{m-n}\, dx.$$

The characteristic function of n is

$$\varphi(t, n) = \sum_{n=j}^{r+j} P(n) e^{nt} =$$

$$\frac{\Gamma(s+1)}{\Gamma(j)\Gamma(s-j+1)} \cdot e^{jt} \int_0^1 x^{j-1}(1-x)^{s-j}(1 - x + xe^t)^r\, dx.$$

Differentiating we find

$$(1) \qquad \frac{\varphi'(t, n)}{\varphi(t, n)} = j + re^t \frac{\int_0^1 x^j(1-x)^{s-j}(1-x+xe^t)^{r-1}\, dx}{\int_0^1 x^{j-1}(1-x)^{s-j}(1-x+xe^t)^r\, dx}$$

and hence

$$m_1(n) = \sum_{n=j}^{r+j} P(n)n = [\varphi'(t, n)]_{t=0} = j \cdot \frac{m+1}{s+1}.$$

For the calculation of moments about the mean we take

$$(2) \qquad \varphi(t, n - m_1) = e^{-m_1 t}\varphi(t, n),$$

from which we obtain

$$[\varphi^{(k)}(t, n - m_1)]_{t=0} = \sum_{n=j}^{r+j} P(n)(n - m_1)^k = \mu_k(n).$$

In particular, $\mu_2 = \dfrac{rj(m+1)(s+1-j)}{(s+1)^2(s+2)}$. The values of $m_1(n)$ and $\mu_2(n)$ have already been given by Olds using another method. Putting $\dfrac{rj}{s+1} = \beta$, we have

[1] E. G. Olds, *Annals of Math. Stat.*, Vol. 11 (1940), p. 355.

$$\mu_2 = \beta(1 - \beta) + \frac{r^{(2)}(j, 2)}{(s + 1, 2)}$$

$$\mu_3 = 3(1 - \beta)\mu_2 + \beta(2 - 3\beta + \beta^2) + \frac{r^{(3)}(j, 3)}{(s + 1, 3)}$$

$$\mu_4 = (6 - 4\beta)\mu_3 - (11 + 4\beta + 6\beta^2)\mu_2 - \beta(6 - 11\beta + 6\beta^2 + \beta^3) + \frac{r^{(4)}(j, 4)}{(s + 1, 4)},$$

where $r^{(k)} = r(r - 1) \cdots (r - k + 1)$, $(j, k) = j(j + 1) \cdots (j + k - 1)$.[2]

We can obtain a limiting form for $P(n)$ in the following way:
Since

$$\varphi\left(t, \frac{n - j}{r}\right) = e^{-jt/r} \varphi\left(\frac{t}{r}, n\right)$$

we find

$$\varphi\left(t, \frac{n - j}{r}\right) = \frac{\Gamma(s + 1)}{\Gamma(j)\Gamma(s - j + 1)} \int_0^1 x^{j-1}(1 - x)^{s-j}(1 - x + xe^{t/r})^r \, dx.$$

Therefore

$$\lim_{r \to \infty} \varphi\left(t, \frac{n - j}{r}\right) = \int_0^1 L(x)e^{xt} \, dx, \tag{3}$$

where

$$L(x) = \frac{\Gamma(s + 1)}{\Gamma(j)\Gamma(s - j + 1)} x^{j-1}(1 - x)^{s-j}.$$

The interpretation of (3) is that the distribution $\left\{\frac{n - j}{r};\ P(n)\right\}$ has as its limiting form the distribution $\{x, L(x)\}$ as $r \to \infty$.

Letting $n_1, n_2, \cdots n_w$ be a sample of size w and $\bar{n}$ the mean, $\bar{n} = \frac{1}{w} \sum_{i=1}^{w} n_i$. For the characteristic function of $\bar{n}$ we have

$$\varphi(t, \bar{n}) = \sum_{n_i = j}^{r+j} \prod_{i=1}^{w} P(n_i)e^{n_i t/w} = \prod_{i=1}^{w} \varphi\left(\frac{t}{w}, n\right) = \left[\varphi\left(\frac{t}{w}, n\right)\right]^w$$

and hence

$$\frac{\varphi'(t, \bar{n})}{\varphi(t, \bar{n})} = w \frac{\varphi'\left(\frac{t}{w}, n\right)}{\varphi\left(\frac{t}{w}, n\right)} = \left[\frac{\varphi'(t, n)}{\varphi(t, n)}\right]_{t = t/w}.$$

[2] For an easy symbolical method of calculation cf. C. Dieulefait, *Comptes Rendu*, Vol. 208, p. 145.

For $t = 0$ we have $m_1(\bar{n}) = m_1(n)$. But:

$$\frac{d^\alpha}{dt^\alpha}\frac{\varphi'(t, \bar{n})}{\varphi(t, \bar{n})} = \frac{1}{w^\alpha}\left[\frac{d^\alpha}{dt^\alpha}\frac{\varphi'(t, n)}{\varphi(t, n)}\right]_{t=t/w}$$

Then for $t = 0$ we arrive at

$$\mu_{\alpha+1}(\bar{n}) = \frac{\mu_{\alpha+1}(n)}{w^\alpha}.$$

For $\alpha = 1$, we have

$$\mu_2(\bar{n}) = \frac{\mu_2(n)}{w}$$

and this leads us to

$$\sigma_{\bar{n}} = \sqrt{\frac{rj(m+1)(s+1-j)}{w(s+1)^2(s+2)}}.$$

By the Tchebycheff theorem we obtain

$$P(|\,\bar{n} - m_1(n)\,| < l\sigma_{\bar{n}}) > 1 - \frac{1}{l^2}.$$

We can take l and w as large as we please; then we have the following stochastic limit

$$\lim_{w\to\infty} \bar{n} = m_1(n).$$

Now, we have

$$\Phi_w(t) = \varphi\left(t, \frac{\bar{n} - m_1}{\sigma_{\bar{n}}}\right) = e^{-m_1 t/\sigma_{\bar{n}}}\left[\varphi\left(\frac{t}{\sigma_{\bar{n}} w}, n\right)\right]^w$$

and

$$\frac{\Phi_w'(t)}{\Phi_w(t)} = -\frac{m_1}{\sigma_{\bar{n}}} + \frac{1}{\sigma_{\bar{n}}}\left[\frac{\varphi'(t, n)}{\varphi(t, n)}\right]_{t=t/\sigma_{\bar{n}}w}.$$

Remembering (1) we readily obtain

$$\frac{\Phi_w'(t)}{\Phi_w(t)} = \frac{\dfrac{rt}{\sigma_{\bar{n}}^2 w}\left[-\dfrac{rj^2}{(s+1)^2} + \dfrac{(r-1)j(j+1)}{(s+1)(s+2)} + \dfrac{j}{s+1}\right] + \cdots}{1 + \dfrac{rt}{\sigma_{\bar{n}} w}\dfrac{j}{s+1} + \cdots}$$

$$= \frac{t + \cdots}{1 + \dfrac{rt}{\sigma_{\bar{n}} w}\dfrac{j}{s+1} + \cdots}.$$

Thus, we find

$$\lim_{w\to\infty} \frac{\Phi_w'(t)}{\Phi_w(t)} = t.$$

This result implies that the distribution $\left\{\frac{\bar{n}_w - m_1(\bar{n})}{\sigma_{\bar{n}}}; P(\bar{n})\right\}$ has the limiting normal distribution $\left\{x, \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\right\}$, as $w \to \infty$.

A SEQUENCE OF DISCRETE VARIABLES EXHIBITING CORRELATION DUE TO COMMON ELEMENTS

By Carl H. Fischer

University of Michigan

1. **Introduction.** Studies of correlation due to common elements have been made more or less sporadically over the past thirty years in attempts to throw more light on the meaning of correlation. Numerous examples may be cited. One of the earliest was a study by Kapteyn [1] in which he showed that two sums, each of n elements drawn from a normal population with k elements in common, had a correlation coefficient of k/n. This was considerably generalized by the writer [3] who considered sums of different numbers of elements drawn from quite arbitrary continuous distributions. The work was extended to include sequences of three or more such sums. Antedating this latter paper, Rietz [2] has devised various urn schemata in one of which pairs of drawings of s balls each were produced with t balls held in common. The coefficient of correlation between the numbers of white balls in each of the pairs of drawings was found to be t/s.

Fairly recently some interest has been shown in this subject in connection with the study of heredity; hence it appeared that it might be of value to present the following study by elementary methods of a sequence of discrete variables in which each member is linked to the adjacent members by various specified numbers of common elements.

2. **Two variables.** A pair of discrete variables is defined as follows: The first, x, is equal to the number of white balls in a set of s_1 balls drawn one at a time from an urn which is so maintained that the probability of drawing a white ball is always a constant, p. The second, y, is equal to the number of white balls in a second set of s_2 balls formed by drawing t_{12} balls at random from the s_1 balls of the first set plus $s_2 - t_{12}$ balls drawn directly from the urn. The numbers s_1 and s_2 may or may not be equal.

Evidently the marginal distribution of x follows the Bernoulli law and is given by $\binom{s_1}{x} q^{s_1 - x} p^x$.[1] The first step in finding $P(x, y:t_{12})$, the bivariate distribution

---

[1] By $\binom{a}{b}$ is meant the number of combinations of a items taken b at a time. It shall be understood that $\binom{a}{b} = 0$ if $b < 0$ or $b > a$.