ON THE EVALUATION OF MOMENTS OF DISTINCT UNITS IN A SAMPLE

By P. K. PATHAK

Indian Statistical Institute

SUMMARY. In this paper, exact expressions for the moments of distinct units that appear in a sample, are derived under any sampling scheme. The importance of such a problem arises, e.g., when we select a simple random sample (with replacement) from a finite population and require the variance of the average of distinct units selected. It has been shown by Basu (1958) that this average is a better estimator of the population mean than the usual overall average.

1. Preliminaries

In this section we give a lemma which will be used in the next section.

Lemma : *The coefficient* $C_m(n)$ *of*

$$Z_1^{\alpha_1} Z_2^{\alpha_2} \ldots Z_m^{\alpha_m} \text{ (where } m \leqslant N, \alpha_i\text{'s} > 0 \text{ and } \sum_{i=1}^{m} \alpha_i = n)$$

in the expansion of $(Z_1+Z_2+\ldots+Z_N)^n$,

*is given by**

$$C_m(n) = m^n - \binom{m}{1}(m-1)^n + \ldots (-)^{m-1}\binom{m}{m-1}1^n. \qquad \ldots (1.1)$$

In terms of the 'differences of zeros', $C_m(n)$ can be represented as

$$C_m(n) = \Delta^m O^n = \Delta^m x^n | x = 0, \qquad \ldots (1.2)$$

where Δ is the difference operator with unit increments. We shall be using freely these two expressions (1.1) and (1.2) for $C_m(n)$, whichever will be convenient to us in subsequent sections.

Corollary 1.1 : Coefficient of $Z_{i_1}^{\alpha_1} Z_{i_2}^{\alpha_2} \ldots Z_{i_m}^{\alpha_m}$ (where $i_1, i_2, \ldots, i_m$ are any m different integers chosen out of $1, 2, \ldots, N$; α_i's > 0 and $\sum_{i=1}^{m} \alpha_i = n$) in the expansion of $(Z_1+Z_2+\ldots+Z_N)^n$, is given by $C_m(n)$.

Corollary 1.2 :

$$C_m(n) = m\,[C_m(n-1) + C_{m-1}(n-1)]. \qquad \ldots (1.3)$$

---

*Note that $C_m(n) = 0$ for $m > n$.

Corollary 1.3 : For all positive integral values of n and N, we have

$$N^n = \sum_{m=1}^{N} C_m(n) \binom{N}{m} \qquad \dots \quad (1.4)$$

2. Moments of distinct units

Positive order. Consider a population containing N units and a sampling scheme S for which

p_i = probability of inclusion of the i-th unit in the sample; ($i = 1, \dots, N$)

q_i = probability of exclusion of the i-th unit from the sample;

p_{ij} = probability of inclusion of the i-th and j-th units in the sample;

q_{ij} = probability of exclusion of the i-th and j-th units from the sample, etc.

We shall denote by ν, the number of distinct units that appear in a sample. It is obvious that

$$\nu = Z_1 + Z_2 + \dots + Z_N, \qquad \dots \quad (2.1)$$

where $$Z_i = \begin{cases} 1 & \text{if the } i\text{-th unit is included in the sample;} \\ 0 & \text{otherwise.} \end{cases}$$

Now, by definition, if n is any positive integer, the n-th order moment of ν is given by

$$E(\nu^n) = E(Z_1 + Z_2 + \dots + Z_N)^n$$

$$= E\left[\sum_{m=1}^{N} \Sigma_1 \Sigma_2 Z_1^{\alpha_1} \dots Z_m^{\alpha_m}\right], \qquad \dots \quad (2.2)$$

where Σ_1 denotes the summation over $\binom{N}{m}$ combinations of mZ's chosen out of $Z_1, Z_2, \dots, Z_N$ and Σ_2 denotes the summation over all products of the type

$$Z_1^{\alpha_1} Z_2^{\alpha_2} \dots Z_m^{\alpha_m} \quad (\alpha_i\text{'s} > 0; \ \sum_{i=1}^{m} \alpha_i = n).$$

Obviously, $$E(\Sigma_2 Z_1^{\alpha_1} Z_2^{\alpha_2} \dots Z_m^{\alpha_m}) = p_{12\cdots m} C_m(n),$$

and therefore, $$E(\nu^n) = \sum_{m=1}^{N} C_m(n) \Sigma_1 p_{12\cdots m}, \qquad \dots \quad (2.3)$$

and

$$E(\nu^n) = \sum_{m=1}^{N} \binom{N}{m} C_m(n) p_{12\cdots m}$$

when $p_{i_1, i_2, \dots, i_m} = p_{12\cdots m}$ for every set of m distinct units $i_1, i_2, \dots, i_m$.

Negative order. To derive the negative moments of ν of any order under any sampling scheme, we assume that $\nu \geq 1$*, i.e.,

$$q_{12 \ldots N} = 0$$

and define $\quad u_i = 1 - Z_i \quad (i = 1, 2, \ldots, N).$

Therefore, if t is any positive integer, the negative moment of ν of order t is given by

$$E\left(\frac{1}{\nu^t}\right) = E\left[\frac{1}{(N-u_1-u_2-\ldots-u_N)^t}\right] = \frac{1}{N^t} E\left[1 - \frac{\Sigma u_i}{N}\right]^{-t}. \quad \ldots \quad (2.4)$$

Since $0 < \sum_{i=1}^{N} u_i \leq (N-1)$, the infinite expansion is possible. Now let

$$(1-x)^{-t} = 1 + \sum_{r=1}^{\infty} A_r x^r,$$

so that $$E\left(\frac{1}{\nu^t}\right) = \frac{1}{N^t} E\left[1 + \sum_{r=1}^{\infty} \frac{A_r}{N^r}(u_1+u_2+\ldots+u_N)^r\right]. \quad \ldots \quad (2.5)$$

Since the infinite series $1 + \sum_{r=1}^{\infty} \frac{A_r}{N^r}(u_1+u_2+\ldots+u_N)^r$ is bounded above by the absolutely convergent series

$$1 + \sum_{r=1}^{\infty} A_r \left(\frac{N-1}{N}\right)^r,$$

it, therefore, follows that

$$E\left(\frac{1}{\nu^t}\right) = \frac{1}{N^t}\left[1 + \sum_{r=1}^{\infty} \frac{A_r}{N^r} E(u_1+\ldots+u_N)^r\right]. \quad \ldots \quad (2.6)$$

But it is apparent from (2.3) that

$$E(u_1+u_2+\ldots+u_N)^r = E\left[\sum_{m=1}^{N} \Sigma_1 \Sigma_2 \, u_1^{\alpha_1} \ldots u_m^{\alpha_m}\right]$$

$$= \sum_{m=1}^{(N-1)} (\Sigma_1 q_{12 \ldots m}) \Delta^m x^r | x = 0.$$

---

*It is evident that this assumption is indeed necessary, otherwise no negative moment of ν exists. In this paper, we restrict ourselves to those sampling schemes for which $P(\nu \geq 1) = 1$.

The N-th term vanishes by assumption $q_{12 \dots N} = 0$. Putting this in (2.6), we obtain

$$E\left(\frac{1}{\nu^t}\right) = \frac{1}{N^t}\left[1+\sum_{r=1}^{\infty}\frac{A_r}{N^r}\sum_{m=1}^{(N-1)}(\Sigma_1 q_{12 \dots m})\Delta^m x^r|_{x=0}\right]$$

$$= \frac{1}{N^t}\left[1+\sum_{m=1}^{(N-1)}(\Sigma_1 q_{12 \dots m})\Delta^m \sum_{r=1}^{\infty}\frac{A_r}{N^r}x^r|_{x=0}\right]$$

$$= \frac{1}{N^t}\left[1+\sum_{m=1}^{(N-1)}(\Sigma_1 q_{12 \dots m})\Delta^m\left(1-\frac{x}{N}\right)^{-t}|_{x=0}\right],$$

which on expansion gives

$$E\left(\frac{1}{\nu}\right)=\frac{1}{N^t}+\sum_{m=1}^{(N-1)}(\Sigma_1 q_{12\cdots m})\left[\frac{1}{(N-m)^t}-\frac{\binom{m}{1}}{(N-m+1)^t}+\dots(-)^m\frac{\binom{m}{m}}{N^t}\right]. \quad \dots (2.7)$$

In case, $q_{i_1, i_2, \dots, i_m} = q_{12\cdots m}$ for every set $(i_1, i_2, \dots, i_m)$ of m distinct units, this reduces to

$$=\frac{1}{N^t}+\sum_{m=1}^{(N-1)}\binom{N}{m}q_{12\cdots m}\left[\frac{1}{(N-m)^t}-\frac{\binom{m}{1}}{(N-m+1)^t}+\dots(-)^m\frac{\binom{m}{m}}{N^t}\right].$$

Corollary 2.1 : Putting $t = 1$ in the above result, we get

$$E\left(\frac{1}{\nu}\right)=\left[\frac{1}{N}+\sum_{m=1}^{(N-1)}(\Sigma_1 q_{12\cdots m})\left\{\frac{1}{(N-m)}-\frac{\binom{m}{1}}{(N-m+1)}+\cdots(-)^m\frac{\binom{m}{m}}{N}\right\}\right].$$

Since $$\frac{1}{(N-m)}-\frac{\binom{m}{1}}{(N-m+1)}+\cdots(-)^m\frac{\binom{m}{m}}{N}=\frac{m\,!}{N(N-1)\dots(N-m)},$$

$$\therefore\ E\left(\frac{1}{\nu}\right)=\frac{1}{N}\left[1+\frac{1}{(N-1)}\Sigma_1 q_1+\frac{1.2}{(N-1)(N-2)}\Sigma_1 q_{12}+\cdots+\frac{1.2\dots(N-1)}{(N-1)\dots 2.1}\Sigma_1 q_{12\cdots N-1}\right]. \quad \dots (2.8)$$

In case, $q_{i_1, i_2, \dots, i_m} = q_{12\cdots m}$ for every set $(i_1, i_2, \dots, i_m)$ of m distinct units, (2.8) reduces to

$$E\left(\frac{1}{\nu}\right)=\frac{1}{N}+\sum_{m=1}^{(N-1)}\frac{q_{12\cdots m}}{(N-m)} \quad \dots (2.8a)$$

Particular cases : (a) *Simple random sampling (with replacement).* In this sampling scheme

$$q_{12\cdots m} = \frac{(N-m)^n}{N^n},$$

where n is the size of the sample.

$$\therefore \quad E\left(\frac{1}{\nu}\right) = \frac{1}{N} + \sum_{m=1}^{(N-1)} \frac{(N-m)^n}{N^n(N-m)} = \frac{1^{n-1}+2^{n-1}+\ldots+N^{n-1}}{N^n}. \quad \ldots \quad (2.8b)$$

In terms of Bernoulli numbers, (2.8b) is given by (Davis, 1935, p. 188)

$$E\left(\frac{1}{\nu}\right) = \frac{1}{n} + \frac{1}{2N} + \frac{1}{n} \sum_{s=1}^{< \frac{n-1}{2}} (-)^{s-1} \binom{n}{2s} \frac{B_s}{N^{2s}}, \quad \ldots \quad (2.9)$$

where B_s is the s-th Bernoulli number.

For large N, this gives us a very convenient method for computing $E\left(\frac{1}{\nu}\right)$.

(b) *Simple random sampling (without replacement).*

Here,

$$q_{12\cdots m} = \begin{cases} \dfrac{\binom{N-m}{n}}{\binom{N}{n}} & \text{for } m < N-n\,; \\ 0 & \text{otherwise} \end{cases}$$

$$\therefore E\left(\frac{1}{\nu}\right) = \left[\frac{1}{N} + \frac{\binom{N-1}{n}}{(N-1)\binom{N}{n}} + \ldots + \frac{\binom{n}{n}}{n\binom{N}{n}}\right]$$

$$= \frac{1}{\binom{N}{n}}\left[\frac{1}{n} + 1 + \frac{(n+1)}{1.2} + \frac{(n+1)(n+2)}{1.2.3} + \ldots \frac{(n+1)(n+2)\ldots(N-2)(N-1)}{1.2.3\ \ldots\ (N-n)}\right].$$

Combining the terms one by one, we get

$$E\left(\frac{1}{\nu}\right) = \frac{1}{\binom{N}{n}} \frac{(n+1)(n+2)\ldots N}{n\cdot 1.2.3\ldots(N-n)} = \frac{1}{n},$$

which is in agreement with the process of sampling.

Corollary 2.2 : For any integer t ($\neq 0$), it can be shown in a similar manner that

$$E(\nu^t) = N^t + \sum_{m=1}^{(N-1)} (\Sigma_1 q_{12\cdots m})\Delta^m (N-x)^t|_{=0}$$

$$= N^t + \sum_{m=1}^{N-1} (\Sigma_1 q_{12\cdots m})\left[(N-m)^t - \binom{m}{1}(N-m+1)^t + \ldots (-)^m \binom{m}{m} N^t\right]. \quad \ldots \quad (2.10)$$

The following theorem giving the expected values of certain functions of ν, is an obvious generalization of (2.10).

Theorem : *If $f(Z)$ is any function of Z defined in the domain, $0 < Z \leqslant N$, for which the infinite expansion in powers of Z is possible, and if the expectation can be taken term by term, then*

$$E[f(\nu)] = f(N) + \sum_{m=1}^{(N-1)} (\Sigma_1 q_{12\cdots m})[f(N-m) - \tbinom{m}{1} f(N-m+1) + \ldots (-)^m f(N)]. \quad \ldots \quad (2.11)$$

Proof: Express $f(Z)$ in the form

$$f(Z) = \sum_{r=-\infty}^{\infty} A_r Z^r.$$

By assumption
$$E[f(\nu)] = E\left[\sum_{r=-\infty}^{\infty} A_r \nu^r\right] = \sum_{r=-\infty}^{\infty} A_r E(\nu^r).$$

Putting the value of $E(\nu^r)$ from (2.10), we get

$$E[f(\nu)] = \sum_{r=-\infty}^{\infty} \left\{ A_r [N^r + \sum_{m=1}^{(N-1)} (\Sigma_1 q_{12\cdots m}) \Delta^m (N-x)^r \big|_{x=0}] \right\}$$

$$= f(N) + \sum_{m=1}^{(N-1)} (\Sigma_1 q_{12\cdots m}) \Delta^m f(N-x) \big|_{x=0},$$

which on expansion gives (2.11).

Acknowledgement

I am indebted to Dr. D. Basu for his keen interest and generous guidance in the preparation of the paper.

References

Basu, D. (1958) : On sampling with and without replacement. *Sankhyā*, 20, 287-294.

Davis, H. T. (1935) : *Tables of the Higher Mathematical Function*, Vol. II, The Principia Press, Inc., Bloomington, Indiana.

Paper received : October, 1960.