A SYMMETRIC METHOD OF OBTAINING UNBIASED ESTIMATES AND EXPECTED VALUES

BY PAUL L. DRESSEL

Michigan State College, East Lansing, Michigan

The problem of finding the relationship between moment functions of a sample and moment functions of the population from which the sample was obtained has, of necessity, received much attention. The problem has two parts: first, to find the expected value of a given sample moment function; second, to find the estimate of a given population moment function. Thus, if m_i represent the ith central moment of a sample and μ_i represent the ith central moment of the population, the first part of the problem requires that we find the mean value of m_i for all possible samples of a given size and express it in term of the μ_i's. The second part requires that we find a function of the m_i's such that the mean value, taken for all possible samples of a given size, be a given μ_i. For the case $i = 4$ we have the well known results:

$$E[m_4] = \frac{(n-1)(n^2-3n+3)}{n^3}\mu_4 + \frac{3(n-1)(2n-3)}{n^3}\mu_2^2,$$

$$E^{-1}[\mu_4] = \frac{n^2(n^2-2n+3)}{n^{(4)}}m_4 - \frac{3n^2(2n-3)}{n^{(4)}}m_2^2.$$

These results are based on the assumption of an infinite population. In spite of the inverse relationship existing between estimates and expected value, the expressions above show no simple relationship. This lack of simplicity of relationship between estimate and expected value is directly traceable to the fact that such results are usually obtained for infinite populations. When results are obtained for finite populations a symmetry is found to exist which reduces to a single problem the two parts stated above. Since this should be evident to anyone upon reflection, the main purpose of the present paper may be considered as that of indicating one method of demonstrating the result stated above as well as showing relationship of this method to material appearing in previously published papers.

Consider a finite population consisting of N items $x_1 \cdots x_N$ and samples of n items taken from that population, the sampling being done without replacement. We shall utilize the power product notation of P. S. Dwyer [1; p. 13]

$$(1) \qquad (q_1 \cdots q_r) = \sum_{i_1 \neq i_2 \neq \cdots \neq i_r}^{n} x_{i_1}^{q_1} x_{i_2}^{q_2} \cdots x_{i_r}^{q_r}$$

to represent a power product formed for the sample and

$$(2) \qquad [q_1 \cdots q_r] = \sum_{i_1 \neq i_2 \neq \cdots \neq i_r}^{N} x_{i_1}^{q_1} x_{i_2}^{q_2} \cdots x_{i_r}^{q_r}$$

to represent like power products formed for the population. An arbitrary moment function of weight r of the sample is indicated by

$$(3) \qquad \Sigma a_{q_1^{\pi_1} \cdots q_t^{\pi_t}} \frac{r!}{(q_1!)^{\pi_1} \cdots (q_t!)^{\pi_t} \pi_1! \cdots \pi_t!} (q_1)^{\pi_1} \cdots (q_t)^{\pi_t}$$

and likewise a moment function of the population is indicated by

$$(4) \qquad \Sigma A_{q_1^{\pi_1} \cdots q_t^{\pi_t}} \frac{r!}{(q_1!)^{\pi_1} \cdots (q_t!)^{\pi_t} \pi_1! \cdots \pi_t!} [q_1]^{\pi_1} \cdots [q_t]^{\pi_t}$$

where the summation extends over all partitions of r.

It now is convenient to express each of the expressions (3) and (4) in terms of power products. We shall utilize for this purpose an expansion theorem which is the converse of a theorem stated by Dwyer, [1; p. 34] and [2; pp. 37–39], which can be proved in a similar fashion.

This converse theorem follows:

If any isobaric sum of products of power sums indicated by

$$(5) \qquad \Sigma A_{q_1^{\pi_1} \cdots q_t^{\pi_t}} \frac{r!}{(q_1!)^{\pi_1} \cdots (q_t!)^{\pi_t} \pi_1! \cdots \pi_t!} [q_1]^{\pi_1} [q_2]^{\pi_2} \cdots [q_t]^{\pi_t}$$

be expanded in terms of power products in a form indicated by

$$(6) \qquad \Sigma B_{p_1^{\pi_1} \cdots p_s^{\pi_s}} \frac{r!}{(p_1!)^{\pi_1} \cdots (p_s!)^{\pi_s} \pi_1! \cdots \pi_s!} [p_1^{\pi_1} \cdots p_s^{\pi_s}]$$

then the coefficient B_r of the power sum $[r]$ is given by

$$(7) \qquad B_r = \Sigma \frac{r!}{(p_1!)^{\pi_1} \cdots (p_s!)^{\pi_s} \pi_1! \cdots \pi_s!} A_{p_1^{\pi_1} \cdots p_s^{\pi_s}}$$

and the coefficient $B_{r_1 \cdots r_m}$ of $[r_1 r_2 \cdots r_m]$ is

$$(8) \qquad B_{r_1 r_2 \cdots r_m} = \overline{B_{r_1} B_{r_2} \cdots B_{r_m}}$$

where the barred product indicates a symbolic multiplication by suffixing of subscripts.

This is exemplified by

$$\begin{aligned} B_{32} = \overline{B_3 B_2} &= \overline{(A_3 + 3A_{21} + A_{111})(A_2 + A_{11})} \\ &= A_{32} + A_{311} + 3A_{221} + 4A_{2111} + A_{11111}. \end{aligned}$$

Using this theorem the moment functions (3) and (4) are easily expanded in terms of power products. In this latter form the expected value of the sample moment function is easily found by utilizing the fact that

$$E\left(\frac{(q_1 \cdots q_s)}{n^{(s)}}\right) = \frac{[q_1 \cdots q_s]}{N^{(s)}}.$$

Now if the expected value of the sample moment function be equated to the population moment function (both being in power product form) we obtain a set of equations connecting the coefficients of a sample moment function and a population moment function. Since either the coefficients of the sample moment function or those of the population moment function may be assigned and the others solved for, this set of equations enables one to solve two problems. First, we may find unbiased estimates—moment functions of the sample such that their expected value is some preassigned population moment function. Second, we may find expected values—moment functions of the population such that they are expected values of some preassigned sample moment function. From the symmetry of this set of equations, we shall see that any result obtained from the system has, through the symmetry, a dual role.

The foregoing discussion may be clarified by an example. Let $A_2[2] + A_{11}[1]^2$ be the population moment function. In terms of power products this becomes $(A_2 + A_{11})[2] + A_{11}[11]$. The sample moment function $a_2(2) + a_{11}(1)^2$ becomes in terms of power products $(a_2 + a_{11})(2) + a_{11}(11)$ and its expected value is

$$\frac{n}{N}(a_2 + a_{11})[2] + \frac{n^{(2)}}{N^{(2)}} a_{11}[11].$$

By equating this to the population moment function above we obtain

$$n^{(2)}a_{11} = N^{(2)}A_{11},$$

$$n(a_2 + a_{11}) = N(A_2 + A_{11}),$$

and the symmetry of the system is apparent.
If

$$\rho_i = \frac{n^{(i)}}{N^{(i)}}, \qquad \tau_i = \frac{N^{(i)}}{n^{(i)}} = \frac{1}{\rho_i},$$

the solutions of the system are

$$\begin{aligned} a_{11} &= \tau_2 A_{11}, & A_{11} &= \rho_2 a_{11}, \\ a_2 &= \tau_1 A_2 + (\tau_1 - \tau_2)A_{11}, & A_2 &= \rho_1 a_2 + (\rho_1 - \rho_2)a_{11}. \end{aligned} \tag{9}$$

In a similar manner if we use moment functions of weight 3 we begin with

$$A_3[3] + 3A_{21}[2][1] + A_{111}[1]^3,$$

$$a_3(3) + 3a_{21}(2)(1) + a_{111}(1)^3,$$

and obtain the system of equations

$$n^{(3)}a_{111} = N^{(3)}A_{111}$$

$$n^{(2)}(a_{21} + a_{111}) = N^{(2)}(A_{21} + A_{111})$$

$$n(a_3 + 3a_{21} + a_{111}) = N(A_3 + 3A_{21} + A_{111})$$

with solutions

$$
\begin{aligned}
A_{111} &= \rho_3 a_{111}, \\
A_{21} &= \rho_2 a_{21} + (\rho_2 - \rho_3) a_{111}, \qquad (10) \\
A_3 &= \rho_1 a_3 + 3(\rho_1 - \rho_2) a_{21} + (\rho_1 - 3\rho_2 + 2\rho_3) a_{111}.
\end{aligned}
$$

The solutions for the a's in terms of the A's are obtainable from the given results in an obvious manner.

If we use the Carver functions [3; p. 104]

$$
\begin{aligned}
&P_1 = \rho_1, && P_{11} = \rho_2 \quad \cdots && P_{1^i} = \rho_i, \\
&P_2 = \rho_1 - \rho_2, && P_{21} = \rho_2 - \rho_3 \quad \cdots && \\
&P_3 = \rho_1 - 3\rho_2 + 2\rho_3, && P_{22} = \rho_2 - 2\rho_3 + \rho_4 \cdots && \\
&P_4 = \rho_1 - 7\rho_2 + 12\rho_3 - 6\rho_4, && &&
\end{aligned}
$$

or in general

$$
P_r = \sum_{t=1}^{r} \rho_t \sum (-1)^{t-1} \frac{r!(t-1)!}{(p_1!)^{\pi_1} \cdots (p_s!)^{\pi_s} \pi_1! \cdots \pi_s!} \qquad (11)
$$

and

$$
P_{r_1 r_2 \cdots r_s} = \overline{\overline{P_{r_1} P_{r_2} \cdots P_{r_s}}}
$$

where the double barred product indicates a symbolic multiplication by addition of subscripts exemplified by

$$
\begin{aligned}
P_{32} = \overline{\overline{P_3 P_2}} &= (\rho_1 - 3\rho_2 + 2\rho_3)(\rho_1 - \rho_2) \\
&= \rho_2 - 4\rho_3 + 5\rho_4 - 2\rho_5;
\end{aligned}
$$

the results (9) and (10) may be written

$$
\begin{aligned}
A_{11} &= P_{11} a_{11}, & A_3 &= P_1 a_3 + 3P_2 a_{21} + P_3 a_{111}, \\
A_2 &= P_1 a_2 + P_2 a_{11}, & A_{21} &= P_{11} a_{21} + P_{21} a_{111}, \\
& & A_{111} &= P_{111} a_{111}.
\end{aligned}
$$

Similarly for weight 4 we obtain

$$
\begin{aligned}
A_4 &= P_1 a_4 + 4P_2 a_{31} + 3P_2 a_{22} + 6P_3 a_{211} + P_4 a_{1111}, \\
A_{31} &= P_{11} a_{31} + 3P_{21} a_{211} + P_{31} a_{1111}, \\
A_{22} &= P_{11} a_{22} + 2P_{21} a_{211} + P_{22} a_{1111}, \\
A_{211} &= P_{111} a_{211} + P_{211} a_{1111}, \\
A_{1111} &= P_{1111} a_{1111}.
\end{aligned}
$$

In general

$$(12)\qquad A_r = \Sigma P_{\pi_1+\pi_2+\cdots+\pi_s} \frac{r!}{(p_1!)^{\pi_1}(p_2!)^{\pi_2}\cdots(p_s!)^{\pi_s}\pi_1!\cdots\pi_s!} a_{p_1^{\pi_1}p_2^{\pi_2}\cdots p_s^{\pi_s}},$$

and

$$(13)\qquad A_{r_1r_2\cdots r_m} = \overline{A_{r_1}A_{r_2}\cdots A_{r_m}},$$

where as before the barred product indicates a symbolic multiplication by suffixing of subscripts.

If in

$$(14)\qquad \Sigma a_{q_1^{\pi_1}\cdots q_t^{\pi_t}} \frac{r!}{(q_1!)^{\pi_1}\cdots(q_t!)^{\pi_t}\pi_1!\cdots\pi_t!}(q_1)^{\pi_1}\cdots(q_t)^{\pi_t}$$

$$(15)\qquad a_{q_1^{\pi_1}\cdots q_t^{\pi_t}} = \frac{(-1)^{\pi_1+\pi_2+\cdots+\pi_t}(\pi_1+\pi_2+\cdots+\pi_t-1)!}{n^{\pi_1+\pi_2+\cdots+\pi_t}}$$

the moment function of the sample which is thereby represented is the Thiele seminvariant l_r of the sample. If the A's are solved for by means of the appropriate set of equations the expected value of l_r is found. Thus we find

$$(16)\qquad \begin{aligned} E[l_2] &= \frac{N^2 n^{(2)}}{N^{(2)} n^2}\lambda_2, \\ E[l_3] &= \frac{N^3 n^{(3)}}{N^{(3)} n^3}\lambda_3, \\ E[l_4] &= \frac{N^4 n^{(4)}}{N^{(4)} n^4}\lambda_4 + \frac{n^{(2)} N^2}{n^4 N^{(4)}}(n-N)(Nn-6)\kappa_4, \\ E[l_2^2] &= \frac{N^4 n^{(4)}}{N^{(4)} n^4}\lambda_2^2 - \frac{N^2 n^{(2)}}{N^{(4)} n^4}(n-N)(nN-n-N-1)\kappa_4, \\ E[l_5] &= \frac{N^5 n^{(5)}}{N^{(5)} n^5}\lambda^5 + \frac{5N^3 n^{(3)}}{N^{(5)} n^5}(n-N)(Nn-12)\kappa_5, \\ E[l_3 l_2] &= \frac{N^5 n^{(5)}}{N^{(5)} n^5}\lambda_3\lambda_2 - \frac{N^3 n^{(3)}}{N^{(5)} n^5}(n-N)(Nn-n-N-5)\kappa_5, \end{aligned}$$

where the κ system of seminvariants used here is defined by

$$(17)\qquad \begin{aligned} \kappa_{2r} &= \frac{1}{2}\sum_{i=0}^{2r}(-1)^i\binom{2r}{i}\mu_i\mu_{2r-i}, \\ \kappa_{2r+1} &= \sum_{i=0}^{r}(-1)^{i+r}\binom{2r}{i+r}\frac{2i+1}{i+r+1}\mu_{r-i}\mu_{r+i+1}. \end{aligned}$$

By virtue of the symmetry noted earlier it follows that the estimates of the Thiele seminvariants and products of these seminvariants of weight ≤ 5 are obtainable from the last results by replacing E by E^{-1} (estimate of), κ_i by k_i l_i by λ_i, and N by n. In this manner we find that L_4, the estimate of λ_4 is

$$(18) \qquad L_4 = E^{-1}[\lambda_4] = \frac{n^4 N^{(4)}}{n^{(4)} N^4} l_4 + \frac{N^{(2)} n^2}{N^4 n^{(4)}} (N - n)(Nn - 6)k_4.$$

It is of some interest to note in the results (16) above that in those expected values or estimates which contain more than one term the factor $N - n$ occurs in the second term. This, and the form of other coefficients involved in the terms, shows that as the sample size approaches the population size the sample seminvariants approach the population seminvariants. Another characteristic of such results as those given in (16) is that infinite sampling formulas are easily obtainable therefrom. Thus if in L_4 given in (18) $N \to \infty$, we find

$$\begin{aligned} L_4 &= \frac{n^4}{n^{(4)}} l_4 + \frac{n^3}{n^{(4)}} k_4 \\ &= \frac{n^3(n+1)}{n^{(4)}} m_4 - \frac{3n^3(n-1)}{n^{(4)}} m_2^2, \end{aligned}$$

the first of these forms checking the result given by Dressel [4; p. 45] and the second form being identical with that given by Fisher [5].

The results exhibited above for finite sampling may lead to a mistaken idea about the simplicity of the results. Simplicity decreases rapidly as the weight increases. Thus for weight 6 we find

$$(19) \qquad \begin{aligned} E[l_6] = {} & \frac{N^6 n^{(6)}}{N^{(6)} n^6} \lambda_6 + \frac{2N^4 n^{(4)}}{N^{(6)} n^6} (n - N)(Nn - 20)[8\mu_6 - 15\mu_4\mu_2 + 10\mu_3^2 - 45\mu_2^3] \\ & + \frac{N^3 n^{(3)}}{N^{(6)} n^6} (n - N)[Nn(n + N) - 12nN + 60] \\ & \qquad \cdot [11\mu_6 + 105\mu_4\mu_2 - 50\,\mu_3^2 + 60\mu_2^3] \\ & - \left\{ \frac{4N^2 n^{(2)}}{N^{(6)} n^6} (n - N)[Nn(N^2 + nN + n^2) - 14nN(N + n) + 71Nn - 120] \right. \\ & \left. + \frac{10Nn^{(3)}}{N^{(4)} n^5} (n - N) + \frac{6n^{(2)}}{N^{(4)} n^5} (n - N)(N + n - 5) - \frac{2n^{(2)}}{N^{(3)} n^5} (n - N) \right\} \kappa_6. \end{aligned}$$

Again by letting $N \to \infty$ infinite sampling results are obtained. Much of this last result vanishes in that case.

It has been demonstrated that the κ system of seminvariants are invariant under estimation in the case of infinite sampling [4; p. 53]. It is therefore of some interest to note that this system also possesses the property for finite sampling without replacement. The proof of this is quite simple. Denote the estimate of κ_i by K_i and the fundamental relations are

$$K_{2r} = \frac{n^2}{n^{(2)}} k_{2r}, \qquad K_{2r+1} = \frac{n^3}{n^{(3)}} k_{2r+1}.$$

These expressions hold for any n and hence for a population of N. Let K'_{2r} and K'_{2r+1} denote functions corresponding to K_{2r} and K_{2r+1} but with population moments replacing sample moments and we have

$$K'_{2r} = \frac{N^2}{N^{(2)}} \kappa_{2r}, \qquad K'_{2r+1} = \frac{N^3}{N^{(3)}} \kappa_{2r+1}.$$

Since the power product mode of formulation of K_{2r} and K_{2r+1} insures that

$$E[K_{2r}] = K'_{2r}, \qquad E[K_{2r+1}] = K'_{2r+1}$$

it follows that

$$E[K_{2r}] = E\left[\frac{n^2}{n^{(2)}} k_{2r}\right] = K'_{2r} = \frac{N^2}{N^{(2)}} \kappa_{2r},$$

or

$$E[k_{2r}] = \frac{n^{(2)} N^2}{n^2 N^{(2)}} \kappa_{2r}.$$

Similarly

$$E[k_{2r+1}] = \frac{n^{(3)} N^3}{n^3 N^{(3)}} \kappa_{2r+1},$$

thus establishing the theorem stated above.

REFERENCES

[1] P. S. Dwyer, "Combined expansions of products of symmetric power sums and sums of symmetric power products with applications to sampling," Part I, *Annals of Math. Stat.*, Vol. 9 (1938), pp. 1–47. Part II, Vol. 9, (1938) pp. 97–132.

[2] P. S. Dwyer, "Moments of any rational integral isobaric sample moment function," *Annals of Math. Stat.*, Vol. 8 (1937), pp. 21–65.

[3] H. C. Carver, "Fundamentals of the theory of sampling," *Annals of Math. Stat.*, Vol. 1 (1930), pp. 101–121; 260–274.

[4] P. L. Dressel, "Seminvariants and their estimates," *Annals of Math. Stat.*, Vol. 11 (1940), pp. 33–57.

[5] R. A. Fisher, "Moments and product moments of sampling distributions," *Proc. Lond. Math. Soc.*, Vol. 2 (30), (1929), pp. 199–238.