SHEPPARD'S CORRECTIONS FOR A DISCRETE VARIABLE

CECIL C. CRAIG

In the Annals of Mathematical Statistics,[1] J. R. Abernethy gave a derivation of the corrections to eliminate the systematic errors in the moments of a discrete variable due to grouping. It is the purpose of this note to considerably shorten and simplify the derivation of these corrections by an adoption of a device used by R. A. Fisher (not published so far as I know) in the case of the ordinary Sheppard's corrections.

Let us suppose that m consecutive values of the discrete variable in question are grouped in a frequency class of width k. The m smaller intervals of width k/m go to make up the class width k, the actual points representing the m values of the variable being plotted at the centers of the sub-intervals. Now let us suppose that each of m consecutive boundary points of the sub-intervals is as likely to be chosen as a boundary point of the larger intervals as any other. Then, if x_i is the class mark of the i-th frequency class, for any *true* value, x, of the discrete variable included in this frequency class, we have

$$x_i = x + \epsilon$$

in which x and ϵ are independent variables and ϵ takes on the m values

$$-\frac{m-1}{2}\,k/m,\; -\frac{m-3}{2}\,k/m,\; \cdots,\; \frac{m-3}{2}\,k/m,\; \frac{m-1}{2}\,k/m,$$

with the equal relative frequencies $1/m$.

The moments of x_i are those calculated from the grouped frequency distribution; the problem is to express the average values of the moments of x in terms of the calculated moments and k and m. The use of moment generating functions at once leads to the desired results. Denoting the s-th moment of x_i about any origin by ν'_s, the like moment of x by μ'_s, the respective moment generating functions of the two variables by $M_{x_i}(\vartheta)$ and $M_x(\vartheta)$ respectively, we have at once

$$(1) \qquad M_{x_i}(\vartheta) = M_x(\vartheta) \sum_{\epsilon=-\frac{m-1}{2}k/m}^{\frac{m-1}{2}k/m} \frac{e^{\epsilon\vartheta}}{m},$$

[1] "On the Elimination of Systematic Errors Due to Grouping," vol. IV (1933), pp. 263–277.

in which by definition

$$M_{x_i}(\vartheta) = 1 + \nu_1'\vartheta + \nu_2'\,\vartheta^2/2! + \nu_3'\,\vartheta^3/3! + \cdots,$$
$$M_x(\vartheta) = 1 + \mu_1'\vartheta + \mu_2'\,\vartheta^2/2! + \mu_3'\,\vartheta^3/3! + \cdots.$$

The computation necessary to get the actual corrections consists in the calculation of the coefficients in the formal expansion of

$$M_\epsilon(\vartheta) = \sum_{\epsilon=-\frac{m-1}{2}k/m}^{\frac{m-1}{2}k/m} \frac{e^{\epsilon\vartheta}}{m}, \tag{2}$$

in powers of ϑ and then solving for the μ_s''s in (1).

But the summation indicated in (2) is readily effected by means of the calculus of finite differences. In fact, we get

$$M_\epsilon(\vartheta) = \frac{e^{\frac{m+1}{2}k\vartheta/m} - e^{-\frac{m-1}{2}k\vartheta/m}}{m(e^{k\vartheta/m} - 1)} = \frac{\sinh k\vartheta/2}{m \sinh k\vartheta/2m}. \tag{3}$$

Then (2) becomes

$$M_{x_i}(\vartheta) = M_x(\vartheta)\,\frac{\sinh k\vartheta/2}{m \sinh k\vartheta/2m}. \tag{4}$$

If we let $m \to \infty$ we get the corresponding result for a continuous variable

$$M_{x_i}(\vartheta) = M_x(\vartheta)\,\frac{\sinh k\vartheta/2}{k\vartheta/2} \tag{5}$$

already given by Langdon and Ore,[2] though in a less elegant manner; for in this case, the expression analogous to (1) is immediately seen to be

$$M_{x_i}(\vartheta) = M_x(\vartheta) \int_{-k/2}^{k/2} e^{\epsilon\vartheta}\,\vartheta\epsilon/k\,.$$

Returning to (4), taking the logarithms of both sides, remembering that the logarithm of the moment generating function is the generating function of the semi-invariants of Thiele, we get,

$$\begin{aligned} &\lambda_1\vartheta + \lambda_2\,\vartheta^2/2! + \lambda_3\,\vartheta^3/3! + \cdots \\ &\quad = \bar\lambda_1\vartheta + \bar\lambda_2\,\vartheta^2/2! + \bar\lambda_3\,\vartheta^3/3! + \cdots - \log\frac{k\vartheta/2m \sinh k\vartheta/2}{k\vartheta/2 \sinh k\vartheta/2m}, \end{aligned} \tag{6}$$

in which the $\bar\lambda_r$'s are the calculated semi-invariants and the λ_r's the corrected ones.

---

[2] W. H. Langdon and O. Ore, Semi-invariants and Sheppard's Corrections, Annals of Mathematics, vol. 31 (1930), pp. 230–232.

But since

$$\log \frac{\sinh x}{x} = \sum_{s=1}^{\infty} (-1)^{s+1} \frac{B_s}{2s\,(2s)!} (2x)^{2s}$$

we have on setting:

$$(7) \qquad -\log \frac{k\vartheta/2m \sinh k\vartheta/2}{k\vartheta/2 \sinh k\vartheta/2m} = a_0 + a_1\vartheta + a_2\,\vartheta^2/2! + a_3\,\vartheta^3/3! + \cdots,$$

$$a_0 = 0, \qquad a_{2s+1} = 0, \qquad s = 0, 1, 2, \cdots$$

$$(8) \qquad a_{2s} = \frac{(-1)^s B_s k^{2s}}{2s}\left(1 - \frac{1}{m^{2s}}\right), \qquad s = 1, 2, 3, \cdots.$$

Obviously these a's are the "Sheppard's" corrections for the semi-invariants. We have generally

$$\lambda_{2s+1} = \bar{\lambda}_{2s+1}, \qquad s = 0, 1, 2, \cdots.$$

$$\lambda_{2s} = \bar{\lambda}_{2s} + (-1)^s \frac{B_s k^{2s}}{2s}\left(1 - \frac{1}{m^{2s}}\right).$$

In particular

$$\lambda_2 = \bar{\lambda}_2 - \left(1 - \frac{1}{m^2}\right) k^2/12 \qquad \lambda_6 = \bar{\lambda}_6 - \left(1 - \frac{1}{m^6}\right) k^6/252$$

$$\lambda_4 = \bar{\lambda}_4 + \left(1 - \frac{1}{m^4}\right) k^4/120 \qquad \lambda_8 = \bar{\lambda}_8 + \left(1 - \frac{1}{m^8}\right) k^8/240.$$

For $m \to \infty$, these give of course the results reached by Langdon and Ore.[3]

To get the corrections for the moments let us set

$$\frac{m \sinh k\vartheta/2m}{\sinh k\vartheta/2} = \alpha_0 + \alpha_1\vartheta + \alpha_2\,\vartheta^2/2! + \alpha_3\,\vartheta^3/3! + \cdots.$$

From (7) and (8)

$$\alpha_0 = 1, \qquad \alpha_{2n+1} = 0, \qquad n = 0, 1, 2, \cdots$$

$$(9) \qquad \alpha_{2n} = \sum \frac{(2n)!\, a_2^r a_4^s a_6^t \cdots}{(2!)^r (4!)^s (6!)^t \cdots r!\,s!\,t! \cdots}$$

the summation extending over all positive, integral values of $r, s, t, \cdots$ for which,

$$r + 2s + 3t + \cdots = n.$$

[3] Loc. cit.

Then finally we have the formula,

$$\mu'_n = \sum_{s=0}^{\left[\frac{n}{2}\right]} \binom{n}{2s} \alpha_{2s} \nu'_{n-2s}, \tag{10}$$

for the corrected moments.

Writing out the first four α's, we have for the first eight moments about the mean

$$\mu_1 = \nu_1 = 0$$

$$\mu_2 = \nu_2 - (1 - 1/m^2)\, k^2/12$$

$$\mu_3 = \nu_3$$

$$\mu_4 = \nu_4 - (1 - 1/m^2)\, \nu_2 k^2/2 + (1 - 1/m^2)(7 - 3/m^2)\, k^4/240.$$

$$\mu_5 = \nu_5 - 5(1 - 1/m^2)\nu_3 k^2/6.$$

$$\mu_6 = \nu_6 - 5(1 - 1/m^2)\, \nu_4 k^2/4 + (1 - 1/m^2)(7 - 3/m^2)\, \nu_2 k^4/16$$
$$- (1 - 1/m^2)(31 - 18/m^2 + 3/m^4)\, k^6/1344$$

$$\mu_7 = \nu_7 - 7(1 - 1/m^2)\, \nu_5 k^2/4 + 7(1 - 1/m^2)(7 - 3/m^2)\, \nu_3 k^4/48$$

$$\mu_8 = \nu_8 - 7(1 - 1/m^2)\, \nu_6 k^2/3 + 7(1 - 1/m^2)(7 - 3/m^2)\, \nu_4 k^4/24$$
$$- (1 - 1/m^2)(31 - 18/m^2 + 3/m^4))\, \nu_2 k^6/48$$
$$+ (1 - 1/m^2)(381 - 239/m^2 + 55/m^4 - 5/m^6)\, k^8/11520.$$

The final term in μ_{2n} as given above is α_{2n}.

The above method is readily extended to the case of two or more variables. We will illustrate the procedure by getting the results likely to be required for two variables. As before we suppose that m consecutive values of x are grouped in a frequency class of width k, and we shall similarly suppose that n values of y are grouped in a frequency class of width l. And arguing as before we write now

$$x_i = x + \epsilon$$

$$y_i = y + \eta$$

in which ϵ and η are independent of x and y and of each other.

The moment generating function of two variables is defined by the identity in ϑ and ω:

$$M_{x,y}(\vartheta, \omega) = 1 + (\mu'_{10}\vartheta + \mu'_{01}\omega) + \frac{1}{2!}(\mu'_{20}\vartheta^2 + 2\mu'_{11}\vartheta\omega + \mu'_{02}\omega^2) + \cdots$$

$$= 1 + (\mu'_{10}\vartheta + \mu'_{01}\omega) + \frac{1}{2!}(\mu'_{10}\vartheta + \mu'_{01}\omega)^{(2)} + \frac{1}{3!}(\mu'_{10}\vartheta + \mu'_{01}\omega)^{(3)} + \cdots,$$

in which the manner of expansion of $(\mu'_{10}\vartheta + \mu'_{01}\omega)^{(r)}$ is evident.

Then from the properties of moment generating functions, we have

$$(11)\qquad \begin{aligned} M_{x_i, x_j}(\vartheta, \omega) &= M_{x,y}(\vartheta, \omega) \sum_{\epsilon=-\frac{m-1}{2}k/m}^{\frac{m-1}{2}k/m} \sum_{\eta=-\frac{n-1}{2}l/n}^{\frac{n-1}{2}l/n} \frac{e^{\epsilon\vartheta+\eta\omega}}{mn} \\ &= M_{x,y}(\vartheta, \omega) \frac{\sinh k\vartheta/2}{m \sinh k\vartheta/2m} \frac{\sinh l\omega/2}{n \sinh l\omega/2n}. \end{aligned}$$

As in the case of a single variable it will be simpler first to get the corrections for the semi-invariants. The logarithm of the moment generating function is the generating function of the semi-invariants; thus

$$\log M_{x,y}(\vartheta, \omega) = (\lambda_{10}\vartheta + \lambda_{10}\omega) + \frac{1}{2!}(\lambda_{10}\vartheta + \lambda_{01}\omega)^{(2)} + \frac{1}{3!}(\lambda_{10}\vartheta + \lambda_{01}\omega)^{(3)} + \cdots,$$

in which

$$(\lambda_{10}\vartheta + \lambda_{01}\omega)^{(3)} = \lambda_{30}\vartheta^3 + 3\lambda_{21}\vartheta^2\omega + 3\lambda_{12}\vartheta\omega^2 + \lambda_{03}\omega^3,$$

etc.

We write (see (7)),

$$(12)\qquad \begin{aligned} \log \frac{m \sinh k\vartheta/2m}{\sinh k\vartheta/2} &= a_2\,\vartheta^2/2! + a_4\,\vartheta^4/4! + \cdots \\ \log \frac{n \sinh l\omega/2n}{\sinh l\omega/2} &= b_2\,\omega^2/2! + b_4\,\omega^4/4! + \cdots, \end{aligned}$$

with

$$a_{2r} = \frac{(-1)^r B_r k^{2r}}{2r}(1 - 1/m^{2r})$$

$$b_{2s} = \frac{(-1)^s B_s l^{2s}}{2s}(1 - 1/n^{2s}).$$

Then from (11) we have

$$(13)\qquad \begin{aligned} (\lambda_{10}\vartheta + \lambda_{01}\omega)^{(2s+1)} &= (\bar{\lambda}_{10}\vartheta + \bar{\lambda}_{01}\omega)^{(2s+1)}, \qquad s = 0, 1, 2, \cdots \\ (\lambda_{10}\vartheta + \lambda_{01}\omega)^{(2s)} &= (\bar{\lambda}_{10}\vartheta + \bar{\lambda}_{01}\omega)^{(2s)} + a_{2s}\vartheta^{2s} + b_{2s}\omega^{2s}, \qquad s = 1, 2, 3, \cdots, \end{aligned}$$

in which, of course, $\bar{\lambda}_{rs}$ is a calculated semi-invariant and λ_{rs} a corrected one. We read off

$$\lambda_{rs} = \bar{\lambda}_{rs}, \qquad rs \neq 0,$$

as already shown by Wold in the case of continuous variables,[4]

$$\lambda_{2s+1,0} = \bar{\lambda}_{2s+1,0}, \qquad \lambda_{0,2s+1} = \bar{\lambda}_{0,2s+1}.$$

[4] Herman Wold: Sheppard's Correction Formulae in Several Variables: Skandinavisk Aktuarietidskrift, vol. XVII (1934), pp. 248–255.

The values of $\lambda_{2s,0}$ are the same as those for λ_{2s} given above and those for $\lambda_{0,2s}$ are obtained from these merely by replacing in them m and k by n and l. And it is quite obvious that for any number of variables the only semi-invariants to be corrected are those in which a single figure of the index is different from zero and is moreover even. For such semi-invariants the corrections are naturally those derived for a single variable.

Now to derive the corrections for the moments, we write

$$\frac{m \sinh k\vartheta/2m}{\sinh k\vartheta/2} \cdot \frac{n \sinh l\omega/2n}{\sinh l\omega/2} = e^{1/2!\,(a_2\vartheta^2+b_2\omega^2) + 1/4!\,(a_4\vartheta^4+b_4\omega^4) + \cdots}$$

$$= 1 + 1/2!\,(\alpha_{20}\vartheta^2 + \alpha_{02}\omega^2) + 1/4!\,(\alpha_{20}\vartheta^2 + \alpha_{02}\omega^2)^{(2)} + \cdots,$$

with now,

$$(\alpha_{20} + \alpha_{02})^{(h)} = \sum \frac{(2h)!\,(a_2 + b_2)^r\,(a_4 + b_4)^s \cdots}{(2!)^r\,(4!)^s \cdots r!\,s! \cdots},$$

the summation to be over all positive integral values of $r, s, \cdots$ for which

$$r + 2s + \cdots = h$$

and in which the parameters ϑ and ω may be omitted without ambiguity.

The formula for the corrected moments can now be written

$$(14) \qquad (\mu'_{10} + \mu'_{01})^{(p)} = \sum_{q=0}^{[p/2]} \binom{p}{2q} (\alpha_{20} + \alpha_{02})^{(q)}\,(\nu'_{10} + \nu'_{01})^{(p-2q)}.$$

This gives

$$\begin{aligned}
&\mu'_{10} + \mu'_{01} = \nu'_{10} + \nu'_{01}\\
&(\mu'_{10} + \mu'_{01})^{(2)} = (\nu'_{10} + \nu'_{01})^{(2)} + (\alpha_{20} + \alpha_{02})\\
(15)\quad &(\mu'_{10} + \mu'_{01})^{(3)} = (\nu'_{10} + \nu'_{01})^{(3)} + 3(\nu'_{10} + \nu'_{01})\,(\alpha_{20} + \alpha_{02})\\
&(\mu'_{10} + \mu'_{01})^{(4)} = (\nu'_{10} + \nu'_{01})^{(4)} + 6(\nu'_{10} + \nu'_{01})^{(2)}\,(\alpha_{20} + \alpha_{02}) + (\alpha_{20} + \alpha_{02})^{(2)}\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots
\end{aligned}$$

Noting that,

$$(\alpha_{20} + \alpha_{02})^{(2)} = a_4 + b_4 + 3(a_2 + b_2)^2,$$

we get the following formulas for the correction of the product moments about an arbitrary origin:

$$\begin{aligned}
\mu'_{11} &= \nu'_{11}\\
\mu'_{21} &= \nu'_{21} - (1 - 1/m^2)\,\nu'_{01}\,k^2/12\\
\mu'_{12} &= \nu'_{12} - (1 - 1/n^2)\,\nu'_{10}\,l^2/12\\
\mu'_{31} &= \nu'_{31} - (1 - 1/m^2)\,\nu'_{11}\,k^2/4\\
\mu'_{22} &= \nu'_{22} - (1 - 1/m^2)\,\nu'_{20}\,l^2/12 - (1 - 1/n^2)\,\nu'_{02}\,k^2/12\\
&\qquad - (1 - 1/m^2)(1 - 1/n^2)\,k^2l^2/144\\
\mu'_{13} &= \nu'_{13} - (1 - 1/n^2)\,\nu'_{11}\,l^2/4.
\end{aligned}$$

The above results give the corrections for moments about the mean, merely by dropping the primes and setting $\nu_{10} = \nu_{01} = 0$. In practice the corrections needed are for moments about the mean, and though there would be no difficulty in computing additional results for an arbitrary origin, I shall give here only the additional results for moments about the mean through the sixth order, omitting those obtained merely by permutation of subscripts and interchange of k and m with l and n respectively.

First, the necessary extension of (15) is

$$(\mu_{10} + \mu_{01})^{(5)} = (\nu_{10} + \nu_{01})^{(5)} + 10\,(\nu_{10} + \nu_{01})^{(3)}\,(\alpha_{20} + \alpha_{02})$$

(15) $$(\mu_{10} + \mu_{01})^{(6)} = (\nu_{10} + \nu_{01})^{(6)} + 15\,(\nu_{10} + \nu_{01})^{(4)}\,(\alpha_{20} + \alpha_{02}) + 15\,(\nu_{10} + \nu_{01})^{(2)}\,(\alpha_{20} + \alpha_{02})^{(2)} + (\alpha_{20} + \alpha_{02})^{(3)}\,.$$

We need the additional relation:

$$(\alpha_{20} + \alpha_{02})^{(3)} = a_6 + b_6 + 15(a_4 + b_4)\,(a_2 + b_2) + 15\,(a_2 + b_2)^3\,.$$

The additional formulas for product moments about the mean follow:

$$\mu_{41} = \nu_{41} - (1 - 1/m^2)\,\nu_{21}\,k^2/12$$

$$\mu_{32} = \nu_{32} - (1 - 1/n^2)\,\nu_{30}\,l^2/2 - (1 - 1/m^2)\,\nu_{12}\,k^2/4$$

$$\mu_{51} = \nu_{51} - (1 - 1/m^2)\,5\nu_{31}\,k^2/6 + (1 - 1/m^2)\,(7 - 3/m^2)\,\nu_{11}\,k^4/48$$

$$\mu_{42} = \nu_{42} - (1 - 1/n^2)\,\nu_{40}\,l^2/12 - (1 - 1/m^2)\,\nu_{22}\,k^2/2 + (1 - 1/m^2)\,(1 - 1/n^2)\,\nu_{20}\,k^2l^2/24$$
$$+ (1 - 1/m^2)\,(7 - 3/m^2)\,\nu_{02}\,k^4/240 - (1 - 1/m^2)\,(7 - 3/m^2)\,(1 - 1/n^2)\,k^4l^2/2880$$

$$\mu_{33} = \nu_{33} - (1 - 1/m^2)\,\nu_{13}\,k^2/4 - (1 - 1/n^2)\,\nu_{31}\,l^2/4 + (1 - 1/m^2)\,(1 - 1/n^2)\,\nu_{11}\,k^2l^2/16\,.$$

For m and n infinite these results give the formulas for two continuous variables already found by Baten[5] and Wold.[6]

The reader will note that this development does not impose the "high contact" condition, except in so far as it assumes the existence of the moments that occur in the formulas. And it exhibits in the clearest fashion that Sheppard's corrections are corrections on the average.

UNIVERSITY OF MICHIGAN.

---

[5] W. D. Baten: Corrections for the Moments of a Frequency Distribution in Two Variables; Annals of Mathematical Statistics, vol. II (1931), pp. 309–319.

[6] Loc. cit., p. 253.