B.I.P.M.
F-92310 SEVRES

Rapport BIPM-74/3

A simple derivation of the Sheppard correction

Jörg W. Müller

1. Introduction

For practical or other reasons, the measured results of a continuous variate x are often grouped into a set of classes of finite widths d_i. Thus for all data x which fall within the range

$$y_i - \varepsilon_i \leq x < y_i + \delta_i , \quad \text{with } \varepsilon_i + \delta_i = d_i , \tag{1}$$

the same value y_i is retained for convenience and the resulting data y then form a grouped sample. Common choices for ε_i are 0 or $d_i/2$, which correspond to truncation or rounding, and in most cases equal widths $d_i = d$ will be preferred. Besides, y is often an integer (e.g. channel number of a kicksorter).

For a reference point at the center of the interval (thus for $\varepsilon = \delta = d/2$), the original variate, which is supposed to have a probability density f(x), is transformed by this grouping into the discrete random variable y, the probability of which is

$$p(y) = \int_{y-d/2}^{y+d/2} f(x)\, dx . \tag{2}$$

The frequency classes p(y) for a given experiment can be represented graphically by a histogram, and such a simplified plot may be very convenient for many applications. Nevertheless, besides being an incomplete and somewhat "deformed" representation of the measurements, it involves also two arbitrarily chosen parameters, namely the class width d and the "origin" y_o, since all grouped quantities y are now of the form

$$y_i = y_o + i \cdot d , \tag{3}$$

where i is an integer and y_o some conveniently chosen origin. Changing either y_o or d may considerably alter the shape of a histogram for a given set of original measurements x. Therefore a statistical **decision** should never be based on the aspect of a histogram alone.

In many cases, the analysis of a frequency distribution of the measurements involves at some stage the computation of the moments. For a continuous variate x which has been grouped artificially into classes (centered around y_i), one actually determines the "grouped" moments (of order n)

$$M_n(y) \equiv \sum_{i=-\infty}^{\infty} y_i^n \cdot p(y_i) \,. \tag{4}$$

However, what we would rather like to know are the corresponding moments of the original, continuous variate x, namely

$$m_n(x) \equiv \int_{-\infty}^{\infty} x^n \cdot f(x)\, dx \,. \tag{5}$$

In order to pass from the calculated empirical moments $M_n(y)$ to unbiased estimates of $m_n(x)$, an adjustment has to be made by means of formulae which are known as Sheppard's corrections [1,2].

2. The traditional solution

The usual derivation of these correction formulae (compare e.g. [3], [4] or [5]) is not quite simple. In mathematics, the relation between a sum and an integral, which are supposed to be of the types represented by (4) and (5) respectively, is described by the Euler - MacLaurin summation formula. There, a sum of equidistant ordinates is expressed as an integral over the corresponding range to which corrective terms have to be added which involve derivatives of odd order taken at the two boundaries. Assuming that the width d is sufficiently small and that the density f(x) has a contact of higher order with the x-axis at the borders, the following general relation can be obtained where the grouped moments are given as linear combinations of the original moments:

$$M_n(y) = \frac{1}{n+1} \sum_{k=0}^{N} \binom{n+1}{2k+1} \cdot \left(\frac{d}{2}\right)^{2k} \cdot m_{n-2k}(x) \,, \tag{6}$$

where $N \equiv \left[\left[\frac{n+1}{2}\right]\right]$ is the largest integer below $\frac{n+1}{2}$.

By a simple but lengthy elimination process, the original moments $m_n(x)$ can be obtained successively. Complete induction then allows to verify that the following general relation holds

$$m_n(x) = \sum_{k=0}^{N} \binom{n}{2k} \cdot (2^{1-2k} - 1) \cdot B_{2k} \cdot d^{2k} \cdot M_{n-2k}(y) , \tag{7}$$

where the coefficients B are the Bernoulli numbers as defined and tabulated in [6]. The first values are numerically $B_o = 1$, $B_2 = 1/6$, $B_4 = -1/30$, $B_6 = 1/42$, etc.

3. The new approach

In view of the important part that grouped measurements play today in the various fields of data handling - as regards nuclear physics it may be sufficient to mention the lots of results obtained by means of a multichannel pulse height analyzer -, and the increasing need for optimum analysis of measurements which have been performed with metrological care, it seemed worthwhile to sketch here briefly a very elementary new attempt to derive these corrections, although by somewhat heuristic arguments. It avoids the more troublesome mathematical arguments inherent in the usual approach and permits to arrive directly at the final result (7).

We first note that the displacement imposed on an original result x, when this is to be identified with the central value y, can be taken as a random shift which may be described by a rectangular probability density of total width d about an origin in the center, since the classes are supposed to be formed according to (2). The effect on an individual point x, ignoring the position of the origin y_o, is thus as if a convolution with a rectangular density had been performed. However, this reasoning may not be applied offhand to the original distribution as a whole since the displacements of individual points (for fixed parameters y_o and d) may be strongly correlated. As a matter of fact, the result does not look at all like a "smeared-out" version of the original density f(x), but can be described by a series of equidistant delta functions (i.e. a "Dirac comb"), namely

$$\sum_{j=-\infty}^{\infty} p(j) \cdot \delta(y - j) , \tag{8}$$

where the "weights" p(j) are given by (2), with d = 1. Nevertheless, for the application we have in mind, it is permitted all the same to do as if the original density as a whole would indeed undergo a convolution, i.e. just as if the grouped data y followed a density

$$F(y) = f(y) * r(y) , \tag{9}$$

$$\text{where} \quad r(y) = \begin{cases} 1/d & \text{for } |y| < d/2 \\ 0 & \text{otherwise} . \end{cases} \tag{10}$$

The reason for this is that (8) and (9), although they have little in common at first sight, are equivalent for deriving such quantities which involve an integration over the whole range of the variate, as their mutual replacement then amounts substantially to reversing the order of two summations.

With the help of (9), the determination of the moments is now straightforward. For the sake of comparison, we shall first derive (6). Afterwards, a method will be sketched to arrive directly at (7). The moments of the rectangular density (10) are easily shown to be given by

$$\mu_k = \begin{cases} \dfrac{d^k}{2^k\,(k+1)} & \text{for } k \text{ even} \\[2ex] 0 & \text{" } k \text{ odd.} \end{cases} \tag{11}$$

By applying the general rule which holds for the moments of a convolution product [7], we are led to

$$\begin{aligned} M_n &= \sum_{j=0}^{N} \binom{n}{2j} \cdot m_{n-2j} \cdot \mu_{2j} \\ &= \frac{1}{n+1} \sum_{j=0}^{N} \binom{n+1}{2j+1} \cdot \left(\frac{d}{2}\right)^{2j} \cdot m_{n-2j} \,, \end{aligned}$$

which is identical with (6).

When numerical values are inserted, we find for the first few moments:

$$\begin{aligned} M_o &= m_o = 1, \\ M_1 &= m_1 \,, \\ M_2 &= m_2 + \frac{1}{12}\, d^2 \,, \\ M_3 &= m_3 + \frac{1}{4}\, d^2\, m_1 \,, \\ M_4 &= m_4 + \frac{1}{2}\, d^2\, m_2 + \frac{1}{80}\, d^4 \,, \\ M_5 &= m_5 + \frac{5}{6}\, d^2\, m_3 + \frac{1}{16}\, d^4\, m_1 \,, \\ M_6 &= m_6 + \frac{5}{4}\, d^2\, m_4 + \frac{3}{16}\, d^4\, m_2 + \frac{1}{448}\, d^6 \,. \end{aligned} \tag{12}$$

From these equations, the original moments $m_n(x)$ could be determined successively. However, this procedure would not easily yield a general formula. We therefore prefer to use a more elegant approach where the explicit evaluation of the grouped moments (12) is avoided.

By applying integral transformations, e.g. in the form of the two-sided Laplace transform, we get for (10)

$$\mathcal{L}\left\{r(x),\ p\right\} \equiv \tilde{r}(p) = \frac{e^{dp/2} - e^{-dp/2}}{dp} = \frac{\sinh(dp/2)}{dp/2}\ , \tag{13}$$

and (9) therefore corresponds to

$$\tilde{f}(p) = \tilde{F}(p) \cdot \frac{dp/2}{\sinh(dp/2)} = \tilde{F}(p) \cdot \frac{dp}{2} \cdot \operatorname{csch}(dp/2)\ . \tag{14}$$

By means of the series development (see e.g. [6], p. 85)

$$\operatorname{csch} x = -2 \sum_{k=0}^{\infty} \frac{2^{2k-1} - 1}{(2k)!} \cdot B_{2k} \cdot x^{2k-1}\ , \tag{15}$$

we obtain from (14)

$$\tilde{f}(p) = \left(\sum_{j} \alpha_j \cdot p^j\right) \cdot \left(\sum_{k} \beta_{2k} \cdot p^{2k}\right) = \sum_{n=0}^{\infty} \gamma_n \cdot p^n\ , \tag{16}$$

where $\alpha_j = \dfrac{(-1)^j \cdot M_j}{j\,!}$ and $\beta_{2k} = \dfrac{2\,(1 - 2^{2k-1})}{(2k)\,!} \cdot B_{2k} \cdot \left(\dfrac{d}{2}\right)^{2k}$.

Since $j + 2k = n$, the coefficients of p^n are

$$\gamma_n = \sum_{k=0}^{N} \alpha_{n-2k} \cdot \beta_{2k} = \sum_{k} \frac{(-1)^{n-2k}\, M_{n-2k}}{(n-2k)\,!} \cdot \frac{2\,(1 - 2^{2k-1})}{(2k)\,!} \cdot B_{2k} \cdot \left(\frac{d}{2}\right)^{2k}. \tag{17}$$

However, by a series expansion, any transformed probability density can be written in the form

$$\tilde{f}(p) = 1 - m_1 \cdot p + \frac{m_2}{2} \cdot p^2 - \frac{m_3}{6} \cdot p^3 \pm \ldots$$

$$= \sum_{n=0}^{\infty} \frac{(-1)^n}{n\,!} \cdot m_n \cdot p^n\ , \tag{18}$$

where m_n are the moments of $f(x)$.

A comparison with (16) shows that

$$m_n = (-1)^n \cdot n\,! \cdot \gamma_n\ , \tag{19}$$

and by means of (17) we therefore arrive at the required relation

$$m_n = \sum_{k=0}^{N} \frac{n\,!}{(n-2k)\,!\;(2k)\,!} \cdot (1 - 2^{2k-1}) \cdot 2 \cdot B_{2k} \cdot \left(\frac{d}{2}\right)^{2k} \cdot M_{n-2k}$$

$$= \sum_{k=0}^{N} \binom{n}{2k} \cdot \left(\frac{1 - 2^{2k-1}}{2^{2k-1}}\right) \cdot B_{2k} \cdot M_{n-2k} \cdot d^{2k} \; .$$

This is the final formula for determining the original moments m which the measurements had before they were ranged into classes, when only the grouped moments M are available. It obviously agrees with the expression (7) given before for the Sheppard corrections*.

When numerical values are inserted, we get the following expressions for the corrected moments of lowest order:

$$
\begin{aligned}
m_o &= M_o = 1 \, , \\
m_1 &= M_1 \, , \\
m_2 &= M_2 - \frac{1}{12}\, d^2 \, , \\
m_3 &= M_3 - \frac{1}{4}\, d^2\, M_1 \, , \\
m_4 &= M_4 - \frac{1}{2}\, d^2\, M_2 + \frac{7}{240}\, d^4 \, , \\
m_5 &= M_5 - \frac{5}{6}\, d^2\, M_3 + \frac{7}{48}\, d^4\, M_1 \, , \\
m_6 &= M_6 - \frac{5}{4}\, d^2\, M_4 + \frac{7}{16}\, d^4\, M_2 - \frac{31}{1344}\, d^6 \, .
\end{aligned}
\qquad (20)
$$

We note in particular that the variance has to be reduced by $d^2/12$, whereas the third moment remains unchanged if central moments are taken, i.e. for $M_1 = 0$.

As regards the exact conditions of validity for the Sheppard corrections, see e.g. [4]. Essentially, the class width has to be sufficiently small, but demanding $d^2 \ll m_2 - m_1^2$ is often more restrictive than would be needed.

---

* We note that (7) is indeed equivalent to Wold's form of the equation

$$m_n = \sum_{k=0}^{n} \binom{n}{k} \cdot (2^{1-k} - 1) \cdot B_k \cdot d^k \cdot M_{n-k} \; , \qquad (7')$$

which is usually given in the textbooks, as in (7') all the terms with odd k vanish (for $k = 1$ because $2^o - 1 = 0$, and for the others since $B_{2j+1} = 0$ for $j \geqslant 1$).

In particular, the apparent difference is therefore not due to the Bernoulli numbers, as one might first suspect, for which an alternative set B_k' is often applied which is related to the definition used above by $B_k' = (-1)^{k+1} \cdot B_{2k}$, for $k \geqslant 1$.

References

[1] W.F. Sheppard: "On the calculation of the most probable values of frequency-constants for data arranged according to equidistant divisions of a scale", Proc. London Math. Soc. 29, 353 (1898)

[2] W.F. Sheppard: "The calculation of moments of a frequency-distribution", Biometrika 5, 450 (1907)

[3] H. Wold: "Sulla correzione di Sheppard", Giorn. Ist. Italiano d. Attuari 5, 304 (1934)

[4] H. Cramér: "Mathematical Methods of Statistics" (University Press, Princeton, 1946), p. 359 ff.

[5] M.G. Kendall, A. Stuart: "The Advanced Theory of Statistics, Vol. 1" (Griffin, London, 1969[3]), p. 75 ff.

[6] "Handbook of Mathematical Functions", ed. by M. Abramowitz and I.A. Stegun, NBS, AMS 55 (Washington, 1965)

[7] J.W. Müller: "Traitement statistique des résultats de mesure", Rapport BIPM-108 (1969 ff.), p. 54

Papers [1] and [2] have not been available to us.

(April 1974)