

CORRECTION FOR THE MOMENTS OF A FREQUENCY DISTRIBUTION IN TWO VARIABLES¹

By

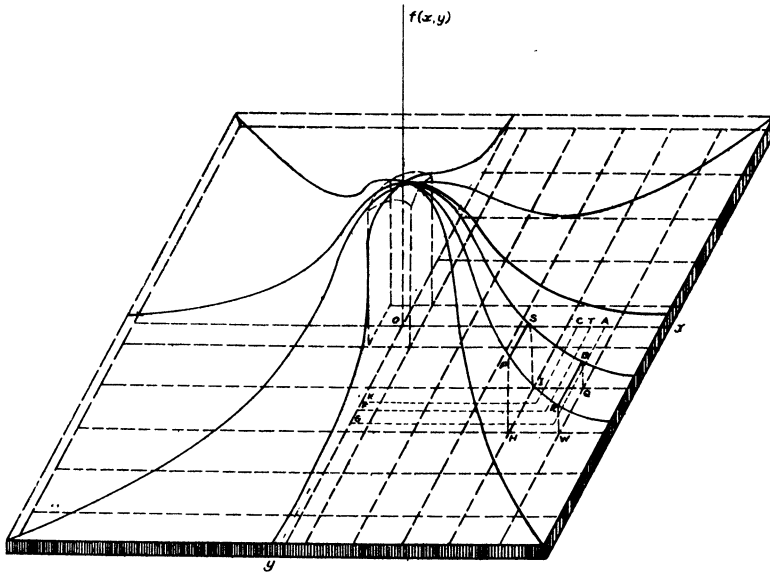
WILLIAM DOWELL BATEN
University of Michigan

In certain statistical problems it is beneficial to divide the given data into classes or groups and investigate the distribution in this form. The moments determined for the distribution divided into classes differ from the moments determined from the original data. It is the object of this article to show how to modify the former to secure the latter for a frequency distribution in two variables.

After the data, given for a frequency distribution of one variable, have been divided into classes the class mark is then the representative of the items in a class. This is assuming that the mean of the items falling in a class is equal to the class mark. For a large number of items in a class, distributed throughout the entire class, the class mark differs very little from the average of the items in the class. But the average of the items raised to a power is not equal to the class mark raised to the same power. Hence corrections should be made to the moments determined from a distribution which is divided into classes.

For a distribution of two variables x and y the data are divided into xy -classes, where the class mark of an x -class

¹Presented to the American Mathematical Society, Sept. 12, 1930.



is considered to be the representative of the items falling in this class, while the class mark of a y -class is the representative of all items in this particular class. The coordinates of the point in the xy -plane, whose abscissa is the class mark of the x -class and whose ordinate is the class mark of the y -class, may be considered to be the class mark of the double class or the xy class.

Let the frequencies of the distribution be represented by the volumes of the volume-compartments as shown in the figure. The sum of all such compartments is the total of the frequencies and should be equal to the number of items in the distribution. The little solid $HWQI-SPRD$ is the frequency of the items falling in the 5th x -class and in the 3rd y -class. OT and OF are the class marks of this x -class and this y -class. $(OT)^n(OF)^m$ multiplied by the frequency of the items falling in this double xy -class may differ considerably from the sum $(OC)^n(OK)^m + (OA)^n(OG)^m + \dots$, hence corrections must be made to the moments obtained from the distribution divided into classes where the double class marks are the representatives of the items in the class. If the class units are made smaller and are allowed to become very near to zero the errors are not so large, for it must be remembered that our results are only approximations.

By definition the n 'th, m 'th moment of the distribution which is divided into classes is

$$V'_{n,m} = \frac{\sum \sum x_i^n y_j^m}{N} \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} f(x_i + h, y_j + k) dh dk,$$

where (x_i, y_j) is considered to be the class mark of the i, j -class, and the double summation extends over all the classes. It is further assumed that $f(x_i + h, y_j + k)$ is such a function which can be expanded into a Taylor series. The above becomes

$$\begin{aligned}
& \frac{\Sigma \Sigma x_i^n y_j^m}{N} \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} f(x_i+h, y_j+k) dh dk = \frac{\Sigma \Sigma x_i^n y_j^m}{N} \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} \left\{ f(x_i, y_j) \right. \\
& + h \frac{\partial f(x_i, y_j)}{\partial x_i} + k \frac{\partial f(x_i, y_j)}{\partial y_j} + \frac{1}{2!} \left[\frac{h^2 \partial^2 f(x_i, y_j)}{\partial x_i^2} + \frac{2hk \partial^2 f(x_i, y_j)}{\partial x_i \partial y_j} \right. \\
& + \left. \left. \frac{k^2 \partial^2 f(x_i, y_j)}{\partial y_j^2} \right] + \frac{1}{3!} \left[\frac{h^3 \partial^3 f(x_i, y_j)}{\partial x_i^3} + \frac{3h^2 k \partial^3 f(x_i, y_j)}{\partial x_i^2 \partial y_j} + \frac{3hk^2 \partial^3 f(x_i, y_j)}{\partial x_i \partial y_j^2} \right. \\
& + \left. \left. \frac{k^3 \partial^3 f(x_i, y_j)}{\partial y_j^3} \right] + \dots \right\} dh dk = \frac{\Sigma \Sigma x_i^n y_j^m}{N} \left\{ f(x_i, y_j) \right. \\
& + \left[\frac{\partial^2 f(x_i, y_j)}{2^2 3! \partial x_i^2} + \frac{\partial^2 f(x_i, y_j)}{2^2 3! \partial y_j^2} \right] + \left[\frac{\partial^4 f(x_i, y_j)}{2^4 5! \partial x_i^4} + \frac{2 \partial^4 f(x_i, y_j)}{2^4 3 \cdot 5! \partial x_i^2 \partial y_j^2} \right. \\
& + \left. \frac{\partial^4 f(x_i, y_j)}{2^4 5! \partial y_j^4} \right] + \frac{1}{6!} \left[\frac{\partial^6 f(x_i, y_j)}{2^6 \cdot 7 \partial x_i^6} + \frac{\partial^6 f(x_i, y_j)}{2^6 \partial x_i^4 \partial y_j^2} \right. \\
& + \left. \frac{\partial^6 f(x_i, y_j)}{2^6 \partial x_i^2 \partial y_j^4} + \frac{\partial^6 f(x_i, y_j)}{2^6 \cdot 7 \partial y_j^6} \right] + \dots \left. \right\}.
\end{aligned}$$

Now use the Euler-Maclaurin Summation* formula for two variables for finding the value of this double summation. This formula is

*This formula is developed on pages 317-319.

$$\int_c^d \int_a^b U(x,y) = \int_c^{d+1} \int_a^{b+1} U(x,y) dx dy - \frac{1}{2} \int_c^{d+1} U(x,y) dy \Big|_a^{b+1} - \frac{1}{2} \int_a^{b+1} U(x,y) dx \Big|_c^{d+1}$$

$$+ \frac{1}{12} \frac{\partial}{\partial y} \int_a^{b+1} U(x,y) dx \Big|_c^{d+1} + \frac{1}{12} \frac{\partial}{\partial x} \int_c^{d+1} U(x,y) dy \Big|_a^{b+1} + \frac{U(x,y)}{4} \Big|_c^{d+1} \Big|_a^{b+1} - \frac{\partial U(x,y)}{24 \partial x} \Big|_c^{d+1} \Big|_a^{b+1}$$

$$- \frac{\partial U(x,y)}{24 \partial y} \Big|_c^{d+1} \Big|_a^{b+1} - \frac{\partial^3}{720 \partial x^3} \int_c^{d+1} U(x,y) dy \Big|_a^{b+1} - \frac{\partial^3}{720 \partial y^3} \int_a^{b+1} U(x,y) dx \Big|_c^{d+1}$$

$$+ \frac{\partial^2 U(x,y)}{144 \partial x \partial y} \Big|_c^{d+1} \Big|_a^{b+1} + \frac{\partial^3 U(x,y)}{1440 \partial x^3} \Big|_c^{d+1} \Big|_a^{b+1} + \frac{\partial^3 U(x,y)}{1440 \partial y^3} \Big|_c^{d+1} \Big|_a^{b+1} + \dots$$

which is the double summation of the function $U(x,y)$ from a to b on the x -axis and from c to d along the y -axis. Applying this formula to the double summation above

$$V_{n:m}^i = \iint x^n y^m \left\{ f(x,y) + \frac{1}{2!2^2} \left[\frac{\partial^2 f(x,y)}{3 \partial x^2} + \frac{\partial^2 f(x,y)}{3 \partial y^2} \right] \right.$$

$$\left. + \frac{1}{4!2^4} \left[\frac{\partial^4 f(x,y)}{5 \partial x^4} + \frac{\binom{4}{2}}{3 \cdot 3} \frac{\partial^4 f(x,y)}{\partial x^2 \partial y^2} + \frac{\partial^4 f(x,y)}{5 \partial y^4} \right] \right\}$$

$$\begin{aligned}
 & + \frac{1}{6!2^6} \left[\frac{\partial^6 f(x,y)}{7 \partial x^6} + \frac{\binom{6}{2} \partial^6 f(x,y)}{5 \cdot 3 \partial x^4 \partial y^2} + \frac{\binom{6}{4} \partial^6 f(x,y)}{3 \cdot 5 \partial x^2 \partial y^4} + \frac{\partial^6 f(x,y)}{7 \partial y^6} \right] \\
 & + \dots \\
 & + \frac{1}{s!2^s} \left[\frac{\partial^s f(x,y)}{(s+1) \partial x^s} + \frac{\binom{s}{2} \partial^s f(x,y)}{(s-2+1)(2+1) \partial x^{s-2} \partial y^2} + \frac{\binom{s}{4} \partial^s f(x,y)}{(s-4+1)(4+1) \partial x^{s-4} \partial y^4} \right. \\
 & + \frac{\binom{s}{6} \partial^s f(x,y)}{(s-6+1)(6+1) \partial x^{s-6} \partial y^6} + \dots + \frac{\binom{s}{t} \partial^s f(x,y)}{(s-t+1)(t+1) \partial x^{s-t} \partial y^t} \\
 & \left. + \dots + \frac{\partial^s f(x,y)}{(s+1) \partial y^s} \right] \Big\} dx dy + 0 + 0 + \dots;
 \end{aligned}$$

t is an even number. In obtaining this result it was assumed that $f(x,y), f'(x,y), x^k y^w f(x,y), x^k y^w f'(x,y)$ vanish or become negligible at the limits on the x and y axes, k and w are positive integers.

Therefore

$$\begin{aligned}
 \nu_{n:m} = & \mu'_{n:m} + \frac{2!}{2^2 3!} \binom{n}{2} \mu'_{n-2:m} + \frac{2!}{2^2 3!} \binom{m}{2} \mu'_{n:m-2} \\
 & + \frac{1}{4!2^4} \left[\frac{4!}{5} \binom{n}{4} \mu'_{n-4:m} + \frac{(2!)^2}{3 \cdot 3} \binom{4}{2} \binom{n}{2} \binom{m}{2} \mu'_{n-2:m-2} + \frac{4!}{5} \binom{m}{4} \mu'_{n:m-4} \right. \\
 & + \frac{1}{6!2^6} \left[\frac{6!}{7} \binom{n}{6} \mu'_{n-6:m} + \frac{4!2!}{5 \cdot 3} \binom{6}{4} \binom{n}{2} \binom{m}{2} \mu'_{n-4:m-2} + \frac{2!4!}{3 \cdot 5} \binom{6}{4} \binom{n}{2} \binom{m}{4} \mu'_{n-2:m-4} \right. \\
 & \left. + \frac{6!}{7} \binom{m}{6} \mu'_{n:m-6} \right] + \dots \\
 & + \frac{1}{s!2^s} \left[\frac{s!}{(s+1)} \binom{n}{s} \mu'_{n-s:m} + \frac{(s-2)!2!}{(s-2+1)(2+1)} \binom{s}{2} \binom{n}{s-2} \binom{m}{2} \mu'_{n-s-2:m-2} \right.
 \end{aligned}$$

$$\begin{aligned}
 & + \frac{(s-4)!4!}{(s-4+1)(5)} \binom{s}{4} \binom{n}{s-4} \binom{m}{4} \mu'_{n-s-4:m-4} \\
 & + \frac{(s-6)!6!}{(s-6+1)(7)} \binom{s}{6} \binom{n}{s-6} \binom{m}{6} \mu'_{n-s-6:m-6} + \dots \\
 & \dots + \frac{(s-t)!t!}{(s-t+1)(t+1)} \binom{s}{t} \binom{n}{s-t} \binom{m}{t} \mu'_{n-s-t:m-t} + \dots \\
 & + \frac{s!}{(s-1)} \binom{m}{s} \mu'_{n:m-s} \Big] + \dots
 \end{aligned}$$

If $m=0$ the formula becomes the formula for obtaining the moments about a fixed origin for one variable. This has been done by Sheppard and Carver.

If n and m take on integral values

$$\begin{aligned}
 V'_{1:1} &= \mu'_{1:1}, \\
 V'_{2:1} &= \mu'_{2:1} + \frac{1}{12} \mu'_{0:1}, \quad V'_{1:2} = \mu'_{1:2} + \frac{1}{12} \mu'_{1:0}, \\
 V'_{2:2} &= \mu'_{2:2} + \frac{1}{12} \mu'_{0:2} + \frac{1}{12} \mu'_{2:0} + \frac{1}{144}, \\
 V'_{3:1} &= \mu'_{3:1} + \frac{1}{4} \mu'_{1:1}, \quad V'_{1:3} = \mu'_{1:3} + \frac{1}{4} \mu'_{1:1}, \\
 V'_{3:2} &= \mu'_{3:2} + \frac{1}{4} \mu'_{1:2} + \frac{1}{12} \mu'_{3:0} + \frac{1}{48} \mu'_{1:0}, \\
 V'_{2:3} &= \mu'_{2:3} + \frac{1}{12} \mu'_{0:3} + \frac{1}{4} \mu'_{2:1} + \frac{1}{48} \mu'_{0:1}, \\
 V'_{3:3} &= \mu'_{3:3} + \frac{1}{4} \mu'_{1:3} + \frac{1}{4} \mu'_{3:1} + \frac{1}{16} \mu'_{1:1}, \\
 V'_{4:1} &= \mu'_{4:1} + \frac{1}{2} \mu'_{2:1} + \frac{1}{60} \mu'_{0:1}, \quad V'_{1:4} = \mu'_{1:4} + \frac{1}{2} \mu'_{1:2} + \frac{1}{60} \mu'_{1:0}, \\
 V'_{4:2} &= \mu'_{4:2} + \frac{1}{2} \mu'_{2:2} + \frac{1}{12} \mu'_{4:0} + \frac{1}{60} \mu'_{0:2} + \frac{1}{24} \mu'_{2:0} + \frac{1}{960},
 \end{aligned}$$

$$V'_{2:4} = \mu'_{2:4} + \frac{1}{12} \mu'_{0:4} + \frac{1}{2} \mu'_{2:2} + \frac{1}{80} \mu'_{2:0} + \frac{1}{24} \mu'_{0:2} + \frac{1}{960},$$

$$V'_{4:3} = \mu'_{4:3} + \frac{1}{2} \mu'_{2:3} + \frac{1}{4} \mu'_{4:1} + \frac{1}{80} \mu'_{0:3} + \frac{1}{8} \mu'_{2:1} + \frac{1}{260} \mu'_{0:1},$$

$$V'_{3:4} = \mu'_{3:4} + \frac{1}{2} \mu'_{3:2} + \frac{1}{4} \mu'_{1:4} + \frac{1}{80} \mu'_{3:0} + \frac{1}{8} \mu'_{1:2} + \frac{1}{260} \mu'_{1:0},$$

$$V'_{4:4} = \mu'_{4:4} + \frac{1}{2} \mu'_{2:4} + \frac{1}{2} \mu'_{4:2} + \frac{1}{80} \mu'_{0:4} + \frac{1}{4} \mu'_{2:2} + \frac{1}{80} \mu'_{4:0}$$

$$+ \frac{1}{160} \mu'_{2:0} + \frac{1}{160} \mu'_{0:2} + \frac{1}{6400},$$

.....

From the above the μ' 's can be obtained.

$$\mu'_{1:1} = V'_{1:1},$$

$$\mu'_{2:1} = V'_{2:1} - \frac{1}{12} M_y, \quad \mu'_{1:2} = V'_{1:2} - \frac{1}{12} M_x,$$

$$\mu'_{3:1} = V'_{3:1} - \frac{1}{4} V'_{1:1}, \quad \mu'_{1:3} = V'_{1:3} - \frac{1}{4} V'_{1:1},$$

$$\mu'_{3:2} = V'_{3:2} - \frac{1}{4} V'_{2:1} - \frac{1}{12} V'_{3:0}, \quad \mu'_{2:3} = V'_{2:3} - \frac{1}{4} V'_{1:2} - \frac{1}{12} V'_{0:3},$$

$$\mu'_{3:3} = V'_{3:3} - \frac{1}{4} V'_{3:1} - \frac{1}{4} V'_{1:3} + \frac{1}{16} V'_{1:1}, \quad \text{etc.}$$

By translating the origin to (M_x, M_y)

$$\begin{aligned} \mu_{1:1} &= V_{1:1} \\ \mu_{2:1} &= V_{2:1} \quad , \quad \mu_{1:2} = V_{1:2} \\ \mu_{2:2} &= V_{2:2} - \frac{1}{12} (V_{0:2} + V_{2:0}) + \frac{1}{144} \\ \mu_{3:1} &= V_{3:1} - V_{1:1} \quad , \quad \mu_{1:3} = V_{1:3} - V_{1:1} \\ \mu_{3:2} &= V_{3:2} - \frac{1}{2} V_{2:1} - \frac{1}{12} V_{3:0} \\ \mu_{2:3} &= V_{2:3} - \frac{1}{2} V_{1:2} - \frac{1}{12} V_{0:3} \\ \mu_{3:3} &= V_{3:3} - \frac{1}{4} V_{1:3} - \frac{1}{4} V_{3:1} + \frac{1}{16} V_{1:1} \end{aligned}$$

etc.

In making corrections for the double moments it must be remembered to correct the single moments of the x 's and the y 's.*

EULER-MACLAURIN SUMMATION FOR TWO VARIABLES

Suppose it is possible to find a function $g(x, y)$ such that $g(x+1, y+1) - g(x+1, y) - g(x, y+1) + g(x, y) = f(x, y)$, or $\Delta_x \Delta_y g(x, y) = f(x, y)$ or $\Delta_x^{-1} \Delta_y^{-1} f(x, y) = g(x, y)$, where Δ represents finite difference and Δ^{-1} represents finite integration. If $g(x, y)$ is such a function, then

$$\begin{aligned} g(a+1, c+1) - g(a+1, c) - g(a, c+1) + g(a, c) &= f(a, c), \\ g(a+2, c+1) - g(a+2, c) - g(a+1, c+1) + g(a+1, c) &= f(a+1, c), \\ g(a+1, c+2) - g(a+1, c+1) - g(a, c+2) + g(a, c+1) &= f(a, c+1), \\ g(a+2, c+2) - g(a+2, c+1) - g(a+1, c+2) + g(a+1, c+1) &= f(a+1, c+1), \end{aligned}$$

*See Frequency Curves by H. C. Carver in Handbook of Math. Statistics.

.....

$$g(b, d) - g(b, d-1) - g(b-1, d) + g(b-1, d-1) = f(b-1, d-1),$$

$$g(b+1, d+1) - g(b+1, d) - g(b, d+1) + g(b, d) = f(b, d).$$

Add: $g(b+1, d+1) - g(b+1, c) - g(a, d+1) + g(a, c) = \sum_c^d \sum_a^b f(x, y).$

Or
$$\sum_c^d \sum_a^b f(x, y) = g(x, y) \Big]_{c}^{d+1} \Big]_{a}^{b+1}$$

If it is possible to find the function $g(x, y)$ then the double sum $\sum_c^d \sum_a^b f(x, y)$ can be found. Expand $g(x+1, y+1)$ in a Taylor series.

$$g(x+1, y+1) = g(x, y) + \frac{\partial g}{\partial x} + \frac{\partial g}{\partial y} + \frac{1}{2!} \left[\frac{\partial^2 g}{\partial x^2} + \frac{2 \partial^2 g}{\partial x \partial y} + \frac{\partial^2 g}{\partial y^2} \right]$$

$$+ \frac{1}{3!} \left[\frac{\partial^3 g}{\partial x^3} + \frac{3 \partial^3 g}{\partial x^2 \partial y} + \frac{3 \partial^3 g}{\partial x \partial y^2} + \frac{\partial^3 g}{\partial y^3} \right] + \dots$$

$$= (e^{\frac{\partial}{\partial x} + \frac{\partial}{\partial y}} g(x, y) = (e^{D_x + D_y}) g(x, y),$$

where $D_x, D_y^h, D_x^r D_y^s$ represent respectively

$$\frac{\partial}{\partial x} g(x, y), \frac{\partial^h}{\partial x^h} g(x, y), \frac{\partial^{r+s}}{\partial x^r \partial y^s} g(x, y).$$

Hence

$$g(x+1, y+1) - g(x+1, y) - g(x, y+1) + g(x, y)$$

$$= (e^{D_x + D_y} - e^{D_x} - e^{D_y} + 1) g(x, y)$$

$$= \{ (e^{D_x} - 1)(e^{D_y} - 1) \} g(x, y).$$

where the D 's are operators operating on the function $g(x, y)$. Therefore

$$g(x, y) = \frac{1}{[(e^{D_x} - 1)(e^{D_y} - 1)]} f(x, y),$$

where the operators are now operating upon the function $f(x, y)$.

To develop $\frac{1}{(e^u - 1)(e^v - 1)}$ into a Taylor series it is necessary to develop $\frac{uv}{(e^u - 1)(e^v - 1)}$ into a Taylor series and then divide by uv . This becomes after \mathcal{D}, \mathcal{D} replace u and v respectively,

$$\left\{ \frac{1}{[(e^{\mathcal{D}-1})x(e^{\mathcal{D}-1})]} \right\} f(x,y) = \left\{ \frac{1}{\mathcal{D}\mathcal{D}} - \frac{1}{2\mathcal{D}} - \frac{1}{2\mathcal{D}} + \frac{1}{2!6\mathcal{D}} + \frac{1}{2} + \frac{\mathcal{D}}{6\mathcal{D}} \right. \\ \left. - \frac{\mathcal{D}}{24} - \frac{\mathcal{D}}{24} - \frac{\mathcal{D}^3}{720\mathcal{D}} + \frac{\mathcal{D}\mathcal{D}}{144} - \frac{\mathcal{D}^3}{720\mathcal{D}} + \frac{\mathcal{D}^3}{1440} + \frac{\mathcal{D}^3}{1440} \right. \\ \left. + \frac{\mathcal{D}^5}{6!42\mathcal{D}} - \frac{\mathcal{D}\mathcal{D}^3}{6!12} - \frac{\mathcal{D}^3\mathcal{D}}{6!12} + \frac{\mathcal{D}^5}{6!42\mathcal{D}} \dots \right\} f(x,y),$$

where $\frac{1}{\mathcal{D}}, \frac{1}{\mathcal{D}}$ represent integration.

Using these results $\sum_c^d \sum_a^b f(x,y) = g(x,y) \Big|_c^{d+1} \Big|_a^{b+1}$ or

$$\sum_c^d \sum_a^b f(x,y) = \int_c^{d+1} \int_a^{b+1} f(x,y) dx dy - \frac{1}{2} \int_c^{d+1} f(x,y) dy \left[\frac{b+1}{a} - \frac{1}{2} \int_a^{b+1} f(x,y) dx \right]_c^{d+1} \\ + \frac{\partial}{12\partial y} \int_a^{b+1} f(x,y) dx \Big|_c^{d+1} + \frac{1}{12} \cdot \frac{\partial}{\partial x} \int_c^{d+1} f(x,y) dy \Big|_a^{b+1} + \frac{f(x,y)}{4} \Big|_c^{d+1} \Big|_a^{b+1} \\ - \frac{\partial f(x,y)}{24 \partial x} \Big|_c^{d+1} \Big|_a^{b+1} - \frac{\partial f(x,y)}{24 \partial y} \Big|_c^{d+1} \Big|_a^{b+1} - \frac{\partial^3}{720 \partial x^3} \int_c^{d+1} f(x,y) dy \Big|_a^{b+1} \\ + \frac{\partial^2 f(x,y)}{144 \partial x \partial y} \Big|_c^{d+1} \Big|_a^{b+1} - \frac{\partial^3}{720 \partial y^3} \int_c^{d+1} f(x,y) dx \Big|_a^{b+1} \\ + \frac{\partial^3 f(x,y)}{1440 \partial x^3} \Big|_c^{d+1} \Big|_a^{b+1} + \frac{\partial^3 f(x,y)}{1440 \partial y^3} \Big|_c^{d+1} \Big|_a^{b+1} \dots$$

W. D. Baten