Everett’s Formula for Bivariate Interpolation and
Throwback of Fourth Differences

It is well known \[1\], \[2\], \[6\], \[7\], \[8\], \[9\], that it is possible in the case of
univariate tables for use with Everett’s formula, to eliminate columns of higher
order differences with practically no loss of accuracy by modification of one or
more lower order differences through a process known as throwback. That the
same thing is possible with bivariate tables (and, presumably, with other multi-
variate tables) seems not to have been recorded in print.

Everett’s formula for bivariate interpolation, as far as fourth order differences,
can be written as follows, using symbolism similar to that of \[2\] (see also \[3\],
p. 8–9 and \[4\], p. viii–xiv).

\[
(1) \quad u_{p, p'} = u(x_0 + ph, y_0 + p'k)
\]

\[
= qq'u_0,0 + qp'u_{0,1} + pq'u_{1,0} + pp'u_{1,1}
- E_2(q')\{q'\delta_2^2u_{0,0} + p'\delta_2^2u_{0,1}\} - E_2(p)\{q'\delta_2^2u_{1,0} + p'\delta_2^2u_{1,1}\}
- E_2(q')\{q'\delta_2^2u_{0,0} + p'\delta_2^2u_{0,1}\} - E_2(p)\{q'\delta_2^2u_{1,0} + p'\delta_2^2u_{1,1}\}
+ E_4(q')\{q'\delta_2^3u_{0,0} + p'\delta_2^3u_{0,1}\} + E_4(p)\{q'\delta_2^3u_{1,0} + p'\delta_2^3u_{1,1}\}
+ E_4(q')\{q'\delta_2^3u_{0,0} + p'\delta_2^3u_{0,1}\} + E_4(p)\{q'\delta_2^3u_{1,0} + p'\delta_2^3u_{1,1}\}
+ E_4(q')\{q'\delta_2^3u_{0,0} + p'\delta_2^3u_{0,1}\} + E_4(p)\{q'\delta_2^3u_{1,0} + p'\delta_2^3u_{1,1}\}
+ E_4(q')\{q'\delta_2^3u_{0,0} + p'\delta_2^3u_{0,1}\} + E_4(p)\{q'\delta_2^3u_{1,0} + p'\delta_2^3u_{1,1}\}
\]

where

\[ p + q = 1 \quad \text{and} \quad p' + q' = 1 \]

and where

\[
E_2(x) = -\left(x + 1\right)
\]

and

\[
E_4(x) = \left(x + 2\right).
\]

We write the above as follows:

\[
(2) \quad u_{p, p'} = qq'u_{0,0} + qp'u_{0,1} + pq'u_{1,0} + pp'u_{1,1}
- E_2(q)\{q'\delta_2^2u_{0,0} - c\delta_2^4u_{0,0} - d\delta_2^2\delta_2^2u_{0,0}\}
- E_2(q')\{q'\delta_2^2u_{0,0} - c'\delta_2^4u_{0,0} - d'\delta_2^2\delta_2^2u_{0,0}\}
- E_2(q)\{q'\delta_2^2u_{0,1} - c\delta_2^4u_{0,1} - d\delta_2^2\delta_2^2u_{0,1}\}
- E_2(q')\{q'\delta_2^2u_{0,1} - c'\delta_2^4u_{0,1} - d'\delta_2^2\delta_2^2u_{0,1}\}
- E_2(q)\{q'\delta_2^2u_{1,0} - c\delta_2^4u_{1,0} - d\delta_2^2\delta_2^2u_{1,0}\}
- E_2(q')\{q'\delta_2^2u_{1,0} - c'\delta_2^4u_{1,0} - d'\delta_2^2\delta_2^2u_{1,0}\}
- E_2(q)\{q'\delta_2^2u_{1,1} - c\delta_2^4u_{1,1} - d\delta_2^2\delta_2^2u_{1,1}\}
- E_2(q')\{q'\delta_2^2u_{1,1} - c'\delta_2^4u_{1,1} - d'\delta_2^2\delta_2^2u_{1,1}\}
+ R,
\]
where \( c, c', d, \) and \( d' \) are constants to be selected so that \(|R|\) is “small” for all \( p, p' \) on ranges \( 0 \leq p \leq 1, 0 \leq p' \leq 1 \), and where

\[
R = \left[ E_4(q) - cE_2(q) \right] q\delta_x u_{0,0} + \left[ E_4(q') - c'E_2(q') \right] q'\delta_y u_{0,0} + E_2(q)\left[ E_2(q') - dq' - d'q \right] \delta_x^2 \delta_y^2 u_{0,0} + E_2(q')\left[ E_2(q) - dp' - d'p \right] \delta_x^2 \delta_y^2 u_{0,0} + \left[ E_4(p) - cE_2(p) \right] p\delta_x u_{1,0} + \left[ E_4(p') - c'E_2(p') \right] p\delta_x u_{1,0} + E_2(p)\left[ E_2(p') - dp' - d'p \right] \delta_x^2 \delta_y^2 u_{1,0} + E_2(p')\left[ E_2(p) - dp' - d'p \right] \delta_x^2 \delta_y^2 u_{1,0}.
\]

In equation (2) above, it is implied that the entire term in (1) involving \( \delta_x^4 u_{i,j} \) is to be thrown back onto \( \delta_x^2 u_{i,j} \), and that the entire term in (1) involving \( \delta_y^4 u_{i,j} \) is to be thrown back onto \( \delta_y^2 u_{i,j} \), while the single term in (1) involving \( \delta_x^2 \delta_y^2 u_{i,j} \) may be thrown back onto both \( \delta_x^2 u_{i,j} \) and \( \delta_y^2 u_{i,j} \). Let us define \( R_x \) as that part of \( R \) involving \( \delta_x^4 u_{i,j} \), \( R_y \) as that part of \( R \) involving \( \delta_y^4 u_{i,j} \), and \( R_{xy} \) as that part involving \( \delta_x^2 \delta_y^2 u_{i,j} \). Thus

\[
R = R_x + R_y + R_{xy}.
\]

Since \( \delta_x^4 u_{i,j}, \delta_y^4 u_{i,j} \) and \( \delta_x^2 \delta_y^2 u_{i,j} \) are essentially independent of each other, it seems natural to select \( c \) and \( c' \) so that \(|R_x|\) and \(|R_y|\) are small, and take care of \( R_{xy} \) later. This has the further advantage that it will take care of the special case where \( \delta_x^2 \delta_y^2 u \) is zero or nearly so. We do this by the method of [1] and [2]. In [2] it is shown (in effect) that if \( c = c_0 = (3 + \sqrt{2})/24 \approx 0.1839256 \), then the absolute value of

\[
R' = \left[ E_4(p) - cE_2(p) \right] \delta_x u_1 + \left[ E_4(q) - cE_2(q) \right] \delta_y u_0
\]

is less than \( L/2000 \) provided \(|\delta_x u| < L\) and \(|\delta_y u| < 0.07L\). In equation (3) we have four groups of terms, two each in \( R_x \) and \( R_y \), which resemble \( R' \). For example, \( R_x \) can be written as

\[
R_x = q'\left[ \left[ E_4(q) - cE_2(q) \right] \delta_x^2 u_{0,0} + \left[ E_4(p) - cE_2(p) \right] \delta_x^2 u_{0,1} \right] + p'\left[ \left[ E_4(q) - cE_2(q) \right] \delta_x^2 u_{1,0} + \left[ E_4(p) - cE_2(p) \right] \delta_x^2 u_{1,1} \right].
\]

The terms in braces are precisely analogous to the univariate expression \( R' \).

Thus, if we select \( c = c_0 \), it follows immediately (since \( q' + p' = 1 \)) that

\[
|R_x| < L/2000 \quad \text{provided} \quad |\delta_x^4 u| < L \quad \text{and} \quad |\delta_y^4 u| < 0.07L.
\]

If we also select \( c' = c_0 \), we get, in a similar fashion

\[
|R_y| < M/2000 \quad \text{provided} \quad |\delta_y^4 u| < M \quad \text{and} \quad |\delta_y^4 u| < 0.07M.
\]

Now \( R_{xy} \) consists of four terms, of which a typical one is

\[
\phi(q, q') \delta_x^2 \delta_y^2 u_{0,0},
\]
where
\[ \phi(q, q') = E_2(q)[E_2(q') - dq' - d'q]. \]

The natural method of attack on the problem of minimizing the maximum absolute contribution of \( R_{xy} \) is to proceed in a manner similar to that of [2], viz., to write \( \delta_x^2 \delta_y^2 u_{i,1} = \delta_x^2 \delta_y^2 u_{i,0} + \delta_x^2 \delta_y^2 u_{i,1}, \quad i = 0, 1. \) After these substitutions we can write \( R_{xy} \) as follows:

\[
R_{xy} = G \delta_x^2 \delta_y^2 u_{0,0} + H \delta_x^2 \delta_y^2 u_{1,0} + A \delta_x^2 \delta_y^2 u_{0,1} + B \delta_x^2 \delta_y^2 u_{1,1},
\]

where
\[
A = E_2(q)[E_2(p') - dp' - d'q],
B = E_2(p)[E_2(q') - E_2(p) - d - 2d'q],
G = E_2(q)[E_2(q') + E_2(p) - d - 2d'q],
H = E_2(p)[E_2(q') + E_2(p) - d - 2d'q].
\]

If we replace \( \delta_x^2 \delta_y^2 u_{0,0} \) by \( \delta_x^2 \delta_y^2 u_{0,0} + \delta_x^2 \delta_y^2 u_{1,0} \) we can write above equation as

\[
R_{xy} = G^* \delta_x^2 \delta_y^2 u_{0,0} + H \delta_x^2 \delta_y^2 u_{1,0} + A \delta_x^2 \delta_y^2 u_{0,1} + B \delta_x^2 \delta_y^2 u_{1,1},
\]

where
\[
G^* = G + H = [E_2(q') + E_2(p) - d][E_2(q) + E_2(p)] - 2d'[E_2(q) + pE_2(p)].
\]

We assume, for the moment, that the three terms in (11) involving fifth order mixed differences are “small” in comparison with the fourth order term; we examine the problem of selecting \( d \) and \( d' \) so as to minimize the maximum absolute deviation from zero of the fourth order term in the region \( D: 0 \leq p \leq 1, \ 0 \leq p' \leq 1. \)

Note that \( G^*(p, p') \) is symmetric with respect to \( p = 1/2 \) and with respect to \( p' = 1/2. \) Therefore, we may restrict attention to the region \( D^*: 0 \leq p \leq 1/2, \ 0 \leq p' \leq 1/2. \)

Evidently we may assume both \( d > 0 \) and \( d' > 0, \) since any positive extremum of \( G^* \) for \( d \leq 0 \) could be made smaller by taking \( d > 0; \) the same is true of any positive extremum for \( d' \leq 0. \)

We have
\[
\frac{\partial G^*}{\partial p} = (1 - 2p)[E_2(p') + E_2(q') - d]/2 - 2d'[2p^2 - 2p + 1]/3
\]
\[
= -4d'(1 - 2p)(p - p_1)(p - p_2)/3,
\]
where
\[
p_1 = 1/2 - \sqrt{g - 1/4}, \quad p_2 = 1/2 + \sqrt{g - 1/4},
g = \frac{3}{8d'2}[E_2(q') + E_2(p') - d].
\]

Also,
\[
\frac{\partial G^*}{\partial p'} = (1 - 2p')[E_2(p) + E_2(q)]/2.
\]
The extrema of $G^*$ either lie on the boundary of $D^*$ or satisfy the conditions 
\[ \frac{\partial G^*}{\partial p} = \frac{\partial G^*}{\partial p'} = 0. \]
The latter conditions imply that

(a) $p = 1/2, \ p_1$ or $p_2$

and

(b) $p' = 1/2$ or $p = 0.$

Since $G^*(0, p') \equiv 0,$ the only non-zero extrema of $G^*$ occur on one of the other three boundaries of $D^*.$

Along $p' = 0,$ $g = -3d/8d' < 0; \text{ hence both } p_1 \text{ and } p_2 \text{ are imaginary here, and therefore } \frac{\partial G^*}{\partial p} < 0 \text{ on } 0 \leq p < 1/2, = 0 \text{ at } p = 1/2.$ Thus, $G^*(1/2, 0) = -(d + d')/8 < 0$ is the minimum value of $G^*$ on boundary $p' = 0$ of $D^*.$

Obviously, $\frac{\partial G^*}{\partial p'} > 0$ on the interior of $D^*$ and along $p = 1/2,$ and $\frac{\partial G^*}{\partial p'} = 0$ in $D^*$ only along $p = 0$ and $p' = 1/2.$ Hence, we can infer that $G^*(1/2, 0)$ is the absolute minimum of $G^*$ in $D^*$ and the only possible maximum is along $p' = 1/2.$ We therefore next investigate $G^*$ along $p' = 1/2.$

Along $p' = 1/2,$ $g = 3(1/8 - d)/8d'$ and $g - 1/4 = (3 - 24d - 16d')/64d'.$ It can be shown, by examination of $\frac{\partial G^*}{\partial p},$ that

a. if $3 - 24d - 16d' \leq 0,$ $G^*(1/2, 1/2)$ is a (negative) minimum and the absolute maximum of $G^*$ on $D^*$ is therefore 0.

b. if $3 - 24d - 16d' > 0,$
and (1) $3 - 24d - 32d' < 0, \ G^*(1/2, 1/2)$ is a maximum, absolute if positive,
and (2) $3 - 24d - 32d' \geq 0, \ G^*(1/2, 1/2) > 0$ and the absolute maximum of $G^*$ on $D^*.$

In all cases, $G^*(1/2, 1/2) = 1/64 - (d + d')/8.$

Since we wish to minimize the maximum absolute deviation of $G^*$ from 0, since $G^*(1/2, 1/2) - G^*(1/2, 0) = 1/64,$ and since both $G^*(1/2, 1/2)$ and $G^*(1/2, 0)$ are monotonic increasing functions of $(d + d'),$ we can evidently solve this minimax problem by making $G^*(1/2, 1/2) \geq 0$ and $= |G^*(1/2, 0)|.$ From this it follows that $d + d' = 1/16$ and maximum absolute deviation of $G^*$ from 0 is 1/128. Note that the condition $d + d' = 1/16$ is consistent with case (b) (2) above.

Having determined that we should select $d + d' = 1/16,$ whence $|G^*| \leq 1/128,$ we now consider the selection of $d$ (and of $d' = 1/16 - d$) and its effect on fifth order terms in $R_{xy}.$

An investigation of the coefficients $A, B,$ and $H$ of the fifth order terms in $R_{xy}$ by methods similar to the investigation of $G^*$ shows that (when $d' = 1/16 - d$)

(1) Maximum $|H|$ is a monotonic decreasing function of $d, 0 < d < 1/16.$

(2) Maximum $|A|$ and maximum $|B|$ are monotonic increasing functions of $d$ on above range, and have same values, although attained for different values of $(p, p').$

Thus, if $|\delta_x^2\delta_y^2u| \leq N,$ we can say that

(12) $|R_{xy}| < N/128 + \max |H| \max |\delta_x^2\delta_y^2u| + 2 \max |A| \max |\delta_x^2\delta_y^2u|.$
Unless we make some assumption about the relative magnitudes of \( \delta_2^2 \delta_y^2 u \) and \( \delta_2^3 \delta_y^3 u \), there is no rational basis for selection of the value of \( d \) other than a selection based on the fact that \( \max |H| \) is monotonic increasing and \( \max |A| \) is monotonic decreasing. For example, if we assume that \( \max |\delta_2^3 \delta_y^3 u| = 2 \max |\delta_2^2 \delta_y^2 u| = P \), then we would be confronted with a problem of minimizing \( P[\max |H| + \max |A|] \).

If this were the case, it turns out we should choose \( d \) close to 1/32 (midpoint of interval 0 < \( d < 1/16 \)) to minimize \([\max |H| + \max |A|]\). Since this choice at least has the virtue of symmetry (if \( d = 1/32 \), then \( d' = 1/32 \)), we make it.

With the above choice, we find that

\[
\max |H| < 0.00451
\]

and

\[
\max |A| < 0.00323.
\]

However, maxima above are attained for different \((p, p')\) from the pairs for which \( |G^*| \) has its maximum, viz., \((1/2, 1/2)\) and \((1/2, 0)\). The maximum values of \( |H| \) and \( |A| \) for these pairs of arguments are respectively \( 1/256 \approx 0.00391 \) and \( 1/512 \approx 0.00195 \).

In order to determine a reasonable set of restrictions on \( |\delta_2 \delta_y^2 u| \) and \( |\delta_2^2 \delta_y^3 u| \), we remember that we already have assumed the fifth order terms "small" in comparison with the fourth order term whose maximum absolute value is \( N/128 \). If, e.g., \( N \leq 60 \) so that \( N/128 \leq 0.46875 \), and this term constitutes most of \( |R_{xy}| \), then in (12) we may allow \( \max |\delta_2 \delta_y^2 u| = \max |\delta_2^2 \delta_y^3 u| \leq 4 \) and have \( |R_{xy}| < 1/2 \).

Thus, if \( |\delta_2 \delta_y^2 u| \leq N \) and \( |\delta_2^2 \delta_y^3 u| \), \( |\delta_2^3 \delta_y^3 u| \leq N/15 \), then \( |R_{xy}| < N/120 \).

Obviously, other sets of restrictions are possible; e.g., we might allow the two fifth order differences different maxima, or we might allow them to be larger fractions of total \( |R_{xy}| \). A similar result would be: if \( |\delta_2 \delta_y^2 u| \leq N \) and \( |\delta_2^2 \delta_y^3 u| \), \( |\delta_2^3 \delta_y^3 u| \leq .28N \), then \( |R_{xy}| < N/100 \).

However, it seems natural to select a restriction set corresponding roughly to that chosen for \( |\delta_2^4 u| \) and \( |\delta_5^5 u| \) in [2], viz., \( L \) and 0.07\( L \) (thus, \( 1/15 \approx 0.07 \)).

In summary, then, we can say that with the choice \( c = c' = c_0 \) and \( d = d' = 1/32 \) for the "throwback" constants,

\[
(13) \quad |R| < L/2000 + M/2000 + N/120
\]

provided

\[
|\delta_2^4 u| \leq L, \quad |\delta_2^5 u| \leq 0.07L, \\
|\delta_4^4 u| \leq M, \quad |\delta_4^5 u| \leq 0.07M, \\
|\delta_2^3 \delta_y^3 u| \leq N, \quad |\delta_2^5 \delta_y^3 u| \leq N/15,
\]

and

\[
|\delta_2^3 \delta_y^3 u| \leq N/15.
\]

In some instances, it may be desirable to have available the general result corresponding to (13), which does not necessarily assume that the fifth order differences are "small" with respect to the fourth order terms.

In order to see how we get such a result, consider, as in [2], that we have written equation (5) as follows:

\[
(5') \quad R' = G \delta^4 u_2 + [E_4(p) - cE_2(p)] \delta^4 u_4
\]
where
\[ G = p(p^2 - 1)(p - 2)/24 + c(p - 1)/2. \]

With the choice \( c = c_0 \) (which minimizes maximum \( |G| \)), we find that \( |G| < .000447 \) and \( |E_4(p) - cE_2(p)| < .000792 \). Consequently, if \( |\delta^4u_0| = 1000 \) and \( |\delta^4u_4| < 70 \), we would have \( |R'| < .503 \), which is approximately \( 1/2 \). Indeed, if we make use of the fact that \( \max |G| \) and \( \max |E_4(p) - cE_2(p)| \) occur for different values of \( p \), and that at \( p = .88265 \), \( |G| = \max \), whereas \( |E_4(p) - cE_2(p)| \approx .000744 \) instead of .000792, we find that, probably, \( |R'| < .499 \). From this arbitrary selection of \( \max |\delta^4u| = 1000 \) comes, then, the result stated in connection with equation (5). Evidently, the corresponding general (and conservative) result is that if \( |\delta^4u| \leq L \) and \( |\delta^4u| \leq KL \), then

\[ (5') \quad |R'| < L(.000447 + .000792K). \]

Obviously, this implies that

\[ (7') \quad |R_z| < L(.000447 + .000792K_i) \quad \text{if} \quad |\delta_z^4u| \leq L \quad \text{and} \quad |\delta_z^4u| \leq K_iL. \]

Similarly, we have

\[ (8') \quad |R_z| < M(.000447 + .000792K_2) \quad \text{if} \quad |\delta_z^4u| \leq M \quad \text{and} \quad |\delta_z^4u| \leq K_2M. \]

(In the above, the only restrictions on the \( K_i \) are that they be \( >0 \).)

From the discussion in connection with equation (12) above, we see that we have

\[ (12') \quad |R_{xy}| < N[1/128 + .00451K_{22} + .00646K_{23}] \]

if

\[ |\delta_x^2\delta_y^2u| \leq N, \quad |\delta_x^2\delta_y^2u| \leq K_{22}N, \]

and

\[ |\delta_x^2\delta_y^2u| \leq K_{22}N, \]

where, again, the only condition on the \( K_{ij} \) is that they be positive.

Consequently, we have the following general result to replace (13):

\[ (13') \quad |R| < L(.000447 + .000792K_i) + M(.000447 + .000792K_2) + N(1/128 + .00451K_{22} + .00646K_{23}) \]

under the hypotheses governing (7'), (8'), (12').

The advantages of the use of modified second differences as in equation (2) are obvious—a published table need contain no fourth order differences and formula (2) requires only 24 multiplications to evaluate 12 terms as opposed to the 48 multiplications required to evaluate 24 terms in formula (1).

It should be noted that the fraction \( 1/120 \) is disappointingly large in comparison with \( 1/2000 \). However, there really seems to be nothing one can do about this situation. Even so, we have "gained" in some sense, since the condition for negligibility of terms involving fourth order (pure and mixed) differences is that

\[ (14) \quad L/42 + M/42 + N/64 < 1/2 \quad \text{(units of last decimal place tabulated)} \]
whereas the condition for "throwback ability" of the same differences is

\[(15) \quad \frac{L}{2000} + \frac{M}{2000} + \frac{N}{120} < \frac{1}{2},\]

in the presence of auxiliary conditions in (13).

As an example of the use of throwback, we may consider its application to the tables in [5]. Cursory examination of [5] indicates that in a large proportion of the table, inequality (15) (or its more general version corresponding to (13')) is satisfied, so that modified second differences could be made to work. As a specific illustration, we take the numerical example on p. xxiv–xxv, viz., computation of \(I(4.025, 7.05)\). We use formula (2) as modified by introduction of the throwback constants \(c, c', d, d'\) selected as above, viz.,

\[u_{p, p'} = q q' u_{0,0} + q p' u_{0,1} + p q' u_{1,0} + p p' u_{1,1} - E_2(q) q' M_x^2 u_{0,0} - E_2(q) p' M_x^2 u_{0,1} - E_2(p) q' M_x^2 u_{1,0} - E_2(p) p' M_x^2 u_{1,1} - E_2(q') q' M_y^2 u_{0,0} - E_2(q') p' M_y^2 u_{0,1} - E_2(p') q' M_y^2 u_{1,0} - E_2(p') p' M_y^2 u_{1,1}\]

where \(M_{x,y}^2\) is the modified second difference of \(u\), with respect to the variable \(z\), evaluated at "point" \(i, j\).

At the point in question, \(I(4.0, 7.0) \sim u_{0,0}, I(4.1, 7.0) \sim u_{1,0}, I(4.0, 7.2) \sim u_{0,1},\) and \(I(4.1, 7.2) \sim u_{1,1}\). Thus, the full array of differences for use in the unmodified formula would be:

\[
\begin{array}{cccc}
\delta_x^2 & \delta_y^2 & \delta_x^4 & \delta_y^4 \\
(0, 0) & -17520 & -1747 & 63 & 0 & 16 \\
(0, 1) & -17985 & -1745 & 80 & 0 & 17 \\
(1, 0) & -16319 & -1643 & 30 & 0 & 10 \\
(1, 1) & -16802 & -1649 & 43 & 0 & 13 \\
\end{array}
\]

This is replaced by the following set of modified differences:

\[
\begin{array}{cccc}
M_x^2 & M_y^2 \\
(0, 0) & -17532 & -1748 \\
(0, 1) & -18000 & -1746 \\
(1, 0) & -16325 & -1643 \\
(1, 1) & -16810 & 01649 \\
\end{array}
\]
Here we have $p = 0.25$, $p' = 0.25$. Using (2'), together with a table of $E_2(x)$, as in [2], we get $I(4.025, 7.05) \approx .878, 5410$, which agrees with result obtained by Pearson.

In principle, throwback of other kinds in the bivariate Everett formula is possible. There is, also, no a priori reason why it could not be accomplished in the trivariate Everett formula, which is stated in [4], page x.

Thomas H. Southard

University of California
Los Angeles, California

This paper was prepared under the sponsorship of the Office of Naval Research and the Office of Ordnance Research, U. S. Army.


TECHNICAL NOTES AND SHORT PAPERS

On the Treatment of Monte Carlo Methods in Text Books

Though Monte Carlo methods have not yet reached the text book stage, they are gradually being afforded a brief mention in new books on numerical analysis. This is probably stimulated by a desire for completeness but the brevity of treatment often results in a very narrow picture of Monte Carlo. For any given problem there may be several methods of solution by Monte Carlo and there are general principles which should guide one in choosing between these. A description of a single method, leaving the reader with the impression that this is the sole method, is therefore a mistake. Care must also be taken to see that, though a method may be illustrated on a simple example for which Monte Carlo would not normally be used, it is easily extensible to problems for which Monte Carlo would be needed and provides a reasonably practical method of solving these.

In two otherwise excellent books by Householder [2] and Kopal [4] a technique is given for evaluating integrals by Monte Carlo. It is the only example of Monte Carlo in either book and is rather an unfortunate one; for it is not easily applicable in practice to multiple integrals though it is largely for these that Monte Carlo can be useful; in addition, it is always less efficient, and often much less so, than another well-known simple technique (defined as crude Monte Carlo by Hammersley and Morton [1]). Their comparative efficiency is also an admirable illustration of a general precept to be used when choosing even the simplest Monte Carlo methods.