Statistical Papers 40, 143–157 (1999)

Statistical Papers

Rounding with multiplier methods: An efficient algorithm and applications in statistics

Gregor Dorfleitner and Thomas Klein*

Received: August 22, 1997; revised version: December 11, 1997

Ordinary rounding does not always satisfy a summation restriction on the rounding results. This can be resolved by applying multiplier methods, for which we present an easy-to-implement algorithm complemented by remarks on special families of multiplier methods, the arithmetic-mean and power-mean method, and a previously unaddressed family, the geometric-mean methods. Finally, several applications in statistics are pointed out, i.e. rounding percentages in descriptive statistics, rounding optimal designs of experiments, and rounding optimal sample allocations.

Keywords: rounding with summation restrictions, apportionment methods, multiplier rounding methods, multiple solutions, efficient algorithms, rounding percentages, optimal design of experiments, optimal sampling allocation.

1 Introduction

Ordinary rounding cannot round a finite set of proportions or probabilities in such a way as to ensure the rounded numbers to sum up to one. For instance, using this procedure for rounding weights of 0.144, 0.534 and 0.322 to percents results in a sum of rounded weights equal to 99 instead of 100. If we modify the weights slightly to 0.146, 0.535 and 0.319, then the rounded weights total up to 101. Real data abound with examples suffering from this failure, see Balinski and Rachev (1993), or Happacher and Pukelsheim (1996a).

*The authors would like to thank F. Pukelsheim and M. Happacher for valuable discussion and the anonymous referee for thoughtful comments.

The issue can be resolved by applying alternative rounding methods that are tailored to the problem. These methods have been used for a long time in the political sciences in order to solve the *problem of apportionment*, that is, rounding electoral quotas to a given number of seats. Balinski and Young (1982) outline the historical background of this problem and the mathematical issues generated by it. Remarkably enough, many methods applied for rounding electoral quotas suffer from severe paradoxes, such as the so called population, Alabama, and new states paradoxes, see Balinski and Young (1982).

Table 1 illustrates one of the fatal deficiencies of Hamilton's method, that is used for the German Bundestag elections: When applied to the data of the 1880 US census, the state of Alabama is alloted 8 seats in the House of Representatives out of a total of 299, but is alloted 7 when the total increases to 300 (Balinski and Young (1982, page 39)).

	Alabama	Texas	Illinois	others	total
weights	0.02557	0.03224	0.06234	0.87985	1
seats	8	10	18	263	299
seats	7	10	19	264	300

Table 1. *Hamilton's method suffers from a lack of monotonicity, also known as the Alabama paradox.*

The three paradoxes mentioned above are avoided by the so called *multiplier rounding methods*. One specific multiplier method is the classical *Jefferson method*, also known as *d'Hondt's method*. Table 2 shows how this method avoids the Alabama paradox when applied to the same data as in Table 1:

	Alabama	Texas	Illinois	others	total
weights	0.02557	0.03224	0.06234	0.87985	1
seats	7	9	18	265	299
seats	7	9	18	266	300

Table 2. *Jefferson's method avoids the Alabama paradox.*

Balinski and Young (1982, page 70) showed that multiplier methods are the only rounding methods not suffering from the mentioned paradoxes, which makes them the most recommendable rounding methods. Their application, however, is not straightforward due to the summation restriction on the rounding results.

Apart from the electoral context, rounding problems with summation restrictions occur in many other areas of statistical interest, e.g. design of experiments, sampling from stratified populations, and descriptive statistics. The objective of using rounding methods that avoid the paradoxes mentioned above also applies to these contexts.

The rest of this paper is organized as follows: In Section 2, we give a brief introduction to rounding methods, especially multiplier methods. Section 3 contains the main feature of the article, a detailed description of a general algorithm for solving rounding problems with multiplier methods. For three special families of multiplier methods, the arithmetic-mean, power-mean and geometric-mean methods, specific adaptations of this algorithm are discussed in Section 4. In Section 5, several applications of this algorithm in different statistical areas are illustrated by examples, and the paper is wound up with conclusions in Section 6 and some additional remarks on an existing implementation of the algorithm presented.

2 Multiplier rounding methods

In order to establish a general mathematical model for rounding problems we assume *weights* $w_1, \ldots, w_c > 0$ (with $c \geq 2$) and an integral *number of units* $n \geq 1$ to be given. Our objective is to round $w_1, \ldots, w_c$ to nonnegative integers $n_1, \ldots, n_c$ satisfying

$$n_i \approx \frac{n w_i}{\sum_j w_j} \quad (i = 1, \ldots, c), \quad \sum_{i=1}^{c} n_i = n \, .$$

In the political setting introduced in the previous section, the weights $w_1, \ldots, w_c$ correspond to electoral quotas of c competing parties, the number of units n stands for the number of seats to be alloted, and the rounded weights $n_1, \ldots, n_c$ are the numbers of seats alloted to each party.

Since there are rounding problems refusing a unique solution, e.g. the problem of rounding c equal weights to a number of units $n = c+1$, any rounding method meeting the summation restriction must give multiple solutions. Therefore, a rounding method (for fixed n) is a set-valued function mapping the weights vector $(w_1, \ldots, w_c)$ to a set of solutions of the corresponding rounding problem. Moreover, most rounding methods can not be explicitly represented in terms of a closed form expression, but are rather given as algorithms.

Multiplier methods are essentially based on one algorithm, that has not been published in a general, easy-to-implement form so far: After computing the normalized weights $\tilde{w}_i = w_i / \sum w_j$, $i = 1, \ldots, c$, a multiplier $\nu \in (0, \infty)$ is used for computing the *pseudo-quotas* $\nu\tilde{w}_1, \ldots, \nu\tilde{w}_c$. Since the pseudo-quotas are not necessarily integers, they are rounded according to a *signpost sequence*. Note that the resulting first "guess" for a solution of the rounding problem does not necessarily satisfy the summation restriction $\sum n_i = n$, and so this initial guess has to be adjusted in several iterative steps. This is done by incrementing or decrementing components according to criteria determined by the signpost sequence.

A few remarks on some of the notions mentioned above should be added:

The multiplier

The algorithm presented below works with *any* multiplier $\nu \in (0, \infty)$. Still it is instructive to think of $\nu = n$, which is in fact the most common choice. The value of ν only affects the number of iterative steps that have to be taken. An answer to the problem of finding a multiplier so as to minimize the number of iterative steps is known only in special cases, see Section 3.

The signpost sequence

Balinski and Young (1982, page 99) define a signpost sequence as a strictly increasing sequence $(s_k)_k$ with $s_k \in [k, k+1]$ for $k \in \mathbb{N}_0$. The signpost s_k marks the boundary between the set of numbers that are rounded down to k and those that are rounded up to $k+1$. With an additional signpost $s_{-1} = 0$ this induces a *rounding function* $R : [0, \infty) \to \mathbb{N}_0$ by

$$R(x) = k \quad \text{for} \quad x \in [s_{k-1}, s_k), \; k \in \mathbb{N}_0 .$$

Note that R is increasing.

The signpost sequence $(s_k)_k$ also induces two functions, the *incrementation criterion*

$$I : \mathbb{N}_0 \times (0, \infty) \to [0, \infty), \quad (k, w) \mapsto \frac{s_k}{w} ,$$

and the *decrementation criterion*

$$D : \mathbb{N}_0 \times (0, \infty) \to [0, \infty), \quad (k, w) \mapsto \frac{s_{k-1}}{w} .$$

These functions are used in the iterative steps of the algorithm in order to determine the rounded weights that have to be incremented or decremented, see Section 3.

Section 4 gives an overview of three families of multiplier methods and their specific implications for the algorithm presented below.

Handling multiple solutions

Although rounding problems may have multiple solutions, the set of solutions produced by a rounding method can be represented by essentially one: Given a solution of a rounding problem, any other solution can be obtained by incrementing or decrementing certain components. The algorithm presented in the next section uses *signs* $z_i = -1, 0, 1, i = 1, \ldots, c$, in order to indicate components that may be decremented by 1, must be left unchanged, or may be incremented by 1, respectively. Of course, incrementing and decrementing weights must conserve the summation restriction. Generally, if the numbers of signs equal to -1 and 1 are a and b, respectively, then there are $\binom{a+b}{a}$ solutions to the rounding problem, see Happacher and Pukelsheim (1996a, Theorem 1).

The theoretical background

The algorithm's theoretical background has been explored in Happacher and Pukelsheim (1996a) and Happacher (1996). Happacher and Pukelsheim (1996a) show that there always exists a compact interval of multipliers ν^* for which the summation restriction on the rounding results is satisfied without any iterative steps. Yet there is no direct way of computing this interval, which may be very small or, even worse, degenerates to a singleton in the case of multiple solutions. This makes clear that there is no reliable way of finding an element of this interval. Happacher and Pukelsheim have therefore given an alternative approach to the rounding problem, which is reflected by the iterative steps of the algorithm below.

3 The algorithm

Given the assumptions and notations of the previous section, rounding with multiplier methods can be carried out according to the following algorithm:

MULTIPLIER STEP

1. Compute the normalized weights $\tilde{w}_i = w_i / \sum_{j=1}^{c} w_j$, $i = 1, \ldots, c$.
2. Compute the rounded pseudo-quotas $n_i = R(\nu \tilde{w}_i)$, $i = 1, \ldots, c$.

DISCREPANCY LOOP

3. Compute the discrepancy $d = (\sum_{i=1}^{c} n_i) - n$.
4. If $d = 0$, jump to step 7.
5. If $d < 0$, choose $i \in \{1, \ldots, c\}$ with $I(n_i, w_i) = \min_{j=1,\ldots,c} I(n_j, w_j)$, replace n_i with $n_i + 1$ and go back to step 3. *(incrementation)*
6. If $d > 0$, choose $i \in \{1, \ldots, c\}$ with $D(n_i, w_i) = \max_{j=1,\ldots,c} D(n_j, w_j)$, replace n_i with $n_i - 1$ and go back to step 3. *(decrementation)*

MULTIPLE SOLUTIONS STEP

7. Compute $s = \min_{j=1,\ldots,c} I(n_j, w_j)$ and $t = \max_{j=1,\ldots,c} D(n_j, w_j)$.
8. If $t < s$, set $z_i = 0$, $i = 1, \ldots, c$. If $t = s$, set

$$z_i = \begin{cases} 1 & \text{for } I(n_i, w_i) = s \\ -1 & \text{for } D(n_i, w_i) = t \\ 0 & \text{otherwise} \end{cases}, \quad i = 1, \ldots, c\,.$$

9. Return the rounded weights $n_1, \ldots, n_c$ and the signs $z_1, \ldots, z_c$.

In step 8, only the cases $t < s$ and $t = s$ are distinguished. This is due to a result of Happacher and Pukelsheim (1996a), who showed that

$$\max_{i=1,\ldots,c} D(n_i, w_i) \leq \min_{i=1,\ldots,c} I(n_i, w_i) .$$

This inequality also shows the aforementioned fact that the set of multipliers ν leading to a *first step discrepancy* $d = 0$ (where the first step discrepancy is the discrepancy obtained without running the iterative steps 3–6) forms a compact interval, see Happacher and Pukelsheim (1996a).

Clearly, the above algorithm terminates, as every iterative step decrements the absolute value of the discrepancy by one. Note that this algorithm can be implemented in the general form given above. Implementations of specific multiplier methods are then easily obtained by supplying the signpost sequence $(s_k)_k$. Such specializations are pointed out in the following section.

Apart from its intuitive appeal, the common choice $\nu = n$ allows the following worst case analysis:

Proposition 1. *Using the multiplier $\nu = n$, the discrepancy loop of the above algorithm terminates after at most c iterations.*

Proof. Denoting the usual floor function by $\lfloor . \rfloor$, the inequality $\lfloor n\tilde{w}_i \rfloor \leq R(n\tilde{w}_i) \leq \lfloor n\tilde{w}_i \rfloor + 1$ readily follows from the definition of the rounding function R. Summation yields

$$\sum_{i=1}^{c} \lfloor n\tilde{w}_i \rfloor - n \leq \sum_{i=1}^{c} R(n\tilde{w}_i) - n \leq \sum_{i=1}^{c} \lfloor n\tilde{w}_i \rfloor + c - n .$$

Using the inequality $n\tilde{w}_i - 1 < \lfloor n\tilde{w}_i \rfloor \leq n\tilde{w}_i$ we have

$$-c < \sum_{i=1}^{c} \lfloor n\tilde{w}_i \rfloor - n \leq 0 \qquad \text{and} \qquad 0 \leq \sum_{i=1}^{c} \lfloor n\tilde{w}_i \rfloor + c - n \leq c$$

and hence

$$\left| \sum_{i=1}^{c} R(n\tilde{w}_i) - n \right| \leq c .$$

Obviously, the left hand side of this inequality is the number of runs through the discrepancy loop. □

The last inequality of the proof is sharp for $n \geq c$, as we can choose the rounding function $R(.) = \lfloor . \rfloor + 1$ and (non-normalized) weights $w_1, \ldots, w_c \in \mathbb{N}$ with $\sum_{i=1}^{c} w_i = n$.

Since each run through the discrepancy loop takes c computations of incrementation or decrementation criteria plus c comparisons, and assuming that the number of operations needed for computing a signpost is $O(1)$, the above proposition implies a worst case overall complexity of $O(c^2)$.

4 Special multiplier methods

The classical Jefferson rounding method mentioned in the introduction and other popular rounding methods can be embedded into several families of rounding methods, which are based on the idea of defining signposts as means. This previously neglected idea encloses the arithmetic mean and power mean signpost sequences explored by Happacher and Pukelsheim (1996a) and naturally leads to the new family of geometric mean signpost sequences.

Arithmetic-mean rounding methods

The family of *arithmetic-mean rounding methods* is defined by the family of signpost sequences $\{(s_k^{(q)})_k \,|\, q \in [0,1]\}$ with

$$s_k^{(q)} = (1-q)\,k + q\,(k+1) = k+q$$

for $k \in \mathbb{N}_0$ and $q \in [0,1]$. Happacher and Pukelsheim (1996a) introduce this family as q-stationary signpost sequences, but we prefer to call these sequences arithmetic-mean signpost sequences, because this name displays the principle of defining signposts as means, which is common to all families of signpost sequences presented in this section. Since $s_{k+1}^{(q)} - s_k^{(q)} = 1$ holds for any k and q, these signpost sequences are stationary in the sense defined by Balinski and Rachev (1993).

The family of arithmetic-mean methods includes the three popular methods of Adams, Webster and Jefferson, for $q = 0$, $q = 0.5$, and $q = 1$, respectively, where Jefferson's method is also known as the d'Hondt method.

Happacher and Pukelsheim (1996a) showed that the multiplier

$$\nu = n + c\left(q - \frac{1}{2}\right)$$

results in an expected first step discrepancy d that vanishes asymptotically for $n \to \infty$ when the normalized weights are uniformly distributed on the probability simplex. Happacher and Pukelsheim (1996b) added the results that under the same assumption this multiplier asymptotically maximizes the probability of a vanishing first step discrepancy and minimizes the expected absolute first step discrepancy.

Even without any probabilistic assumptions on the weights, this multiplier can be motivated by the fact that the worst case bound for the number of iterations given in Proposition 1 can be improved:

Proposition 2. *Using the multiplier* $\nu = n + c\left(q - \frac{1}{2}\right)$ *and an arithmetic-mean signpost sequence, the discrepancy loop of the above algorithm terminates after at most* $\frac{c}{2}$ *iterations.*

Proof. Let $R_q(.)$ denote the rounding function induced by the arithmetic-mean signpost sequence $(s_k^{(q)})_k$. Then, for $x \in [0, \infty)$, we have

$$R_q(x) = R_1(x + 1 - q) = \lfloor x + 1 - q \rfloor \quad .$$

By the same arguments as in the proof of Proposition 1 we get the inequality

$$\sum_{i=1}^{c} (\nu \tilde{w}_i - q) - n < \sum_{i=1}^{c} R_q(\nu \tilde{w}_i) - n \leq \sum_{i=1}^{c} (\nu \tilde{w}_i + 1 - q) - n \quad .$$

Substituting ν by $n + c\left(q - \frac{1}{2}\right)$ establishes the claim. □

Again, the worst case overall complexity of the algorithm is $O(c^2)$.

Contrary to this worst case approach, Happacher (1996) analyzes the asymptotically expected first step discrepancy assuming a uniform distribution of the normalized weights. Happacher's results state that the expected absolute value of the first step discrepancy is not greater than $\sqrt{c/12}$ for $n \to \infty$ and $c \geq 12$. This implies an asymptotically expected overall complexity of $O(c^{3/2})$.

Power-mean rounding methods

Another important family of methods is the family of power-mean methods. Here, the signpost sequences are given by

$$t_k^{(p)} = \left(\frac{k^p + (k+1)^p}{2} \right)^{\frac{1}{p}}$$

for $k \in \mathbb{N}_0$ and $p \in [-\infty, \infty]$. Note that for $p \in \{-\infty, 0, \infty\}$ this is interpreted by taking limits, i.e.

$$t_k^{(-\infty)} = k \, , \quad t_k^{(0)} = \sqrt{k(k+1)} \, , \quad t_k^{(\infty)} = k + 1$$

for $k \in \mathbb{N}_0$. This family is given by Happacher and Pukelsheim (1996a), who call it the family of p-mean rounding methods.

On the one hand, power-mean methods are important due to the fact that they include the five classical multiplier methods of Adams ($p = -\infty$), Dean ($p = -1$), Hill ($p = 0$), Webster ($p = 1$), and Jefferson ($p = \infty$). For an overview of classical rounding methods see Balinski and Young (1982). On the other hand, the asymptotic behavior for $n \to \infty$ of power-mean methods with $|p| < \infty$ is that of Webster's method, see Happacher (1996). Due to this fact, Happacher (1996) gives asymptotic analysis only for arithmetic-mean methods, since they asymptotically offer a greater range of possibilities, see Pukelsheim and Rieder (1992).

An asymptotic optimality theory on the choice of the multiplier ν comparable to that of Happacher (1996) for arithmetic-mean methods is not available for power-mean methods. Similarly, the arguments used in the proof of Proposition 2 do not apply to this situation. However, $\nu = n$ is a reasonable choice due to Proposition 1.

Geometric-mean rounding methods

Apart from the two above families, which are already mentioned in the literature, we suggest the family of geometric-mean rounding methods based on the signposts

$$u_k^{(r)} = k^{1-r}(k+1)^r$$

for $k \in \mathbb{N}_0$ and $r \in [0,1]$. For any $r \in [0,1]$ we have

$$\lim_{k\to\infty} u_k^{(r)} - k = r\,,$$

and hence these signposts sequences are asymptotically stationary just like the power-mean signpost sequences, and approximate the arithmetic-mean signpost sequences. As a consequence, the asymptotic behavior of geometric-mean rounding methods is as rich as that of arithmetic-mean methods, while power-mean methods asymptotically degenerate to only three cases as seen above. Special cases included in the family of geometric-mean methods are Adams' method ($r = 0$), Hill's method ($r = 0.5$) and Jefferson's method ($r = 1$).

A noteworthy property of geometric-mean signpost sequences is that $u_0^{(r)} = 0$ for any $r < 1$. Therefore, except for the pathological case $n < c$, weights are not rounded to zero—a property that may be desirable in some contexts. In the family of arithmetic-mean methods, only Adams' method shows such behavior.

Again there is neither an asymptotic optimality theory nor a special worst case result for the choice of the multiplier ν, and so, owing to Proposition 1, we suggest $\nu = n$.

5 Applications in Statistics

In the following we present applications of multiplier rounding methods in various statistical fields, namely descriptive statistics, design of experiments and sampling from stratified populations.

Rounding percentages in descriptive statistics

Tables of rounded percentages failing to sum up to 100% can be encountered in many types of publications. Balinski and Rachev (1993) give examples of this phenomenon drawn from newspapers and scientific publications, Diaconis and Freedman (1979) compute the probability of a table with ordinarily rounded entries summing up to 100% under some distribution assumptions for the entries.

The Statistical Yearbook 1996 for the Federal Republic of Germany (Statistisches Bundesamt (1996)) also contains such examples, e.g. a table of origins of foreigners living in Germany on December 31st, 1995, distinguished by continents. The percentages of foreigners coming from Europe, Africa, America, Asia, Australia and Oceania, stateless persons, and persons of unknown nationality sum up to

100.1%, while the last line of the table claims a sum of 100%. Table 3 displays the original data and percentages given by the Statistical Yearbook and additionally shows rounded percentages obtained by Webster's method for $n = 1000$. Obviously, the number of units n can be viewed as the rounding accuracy, and so the choice $n = 1000$ corresponds to rounding to tenths of percents.

	persons (in thousands)	proportions (%) ordinary rounding	proportions (%) Webster or Hill
Europe	5 920.3	82.5	82.5
Africa	291.2	4.1	4.1
America	183.0	2.6	2.5
Asia	702.9	9.8	9.8
Australia/Oceania	9.2	0.1	0.1
stateless	19.3	0.3	0.3
unknown origin	48.0	0.7	0.7
total	7 173.9	100.1	100.0

Table 3. *Foreigners living in Germany, distinguished by continents of origin. The percentages in the third column are taken from Statistisches Bundesamt (1996) and fail to sum up to* 100%. *Webster's and Hill's methods cope with this summation restriction by rounding the proportion of Americans to* 2.5%, *as can be seen in the fourth column.*

Generally, we suggest Webster's or Hill's method for rounding percentages in descriptive statistics: Both of them are multiplier methods and thus avoid the paradoxes described by Balinski and Young (1982), see Section 1. Furthermore, Webster's method is a generalisation of the ordinary rounding procedure, i.e. ordinary rounding and rounding with Webster's method coincide in all cases where ordinary rounding happens to meet the summation restriction. Finally, Webster's and Hill's method each minimize certain measures of the average rounding error; more explicitly, given weights $w_1, \dots, w_c$ and a number of units n, Webster's method minimizes

$$\sum_{i=1}^{c} \frac{1}{w_i} \left(n_i - \frac{n w_i}{\sum_j w_j} \right)^2$$

among all multiplier methods, while Hill's method minimizes

$$\sum_{i=1}^{c} \frac{1}{n_i} \left(n_i - \frac{n w_i}{\sum_j w_j} \right)^2$$

in the class of all rounding methods meeting the summation restriction, see Balinski and Young (1982, pages 103–105). Note that the signposts defining Webster's and Hill's methods are asymptotically equal, and therefore these rounding methods coincide for $n \to \infty$.

Other classical multiplier methods minimize measures of maximal rounding errors and tend to favor small weights (Adams' or Dean's method) or large weights (Jefferson's method), which does not seem to be suitable in this context.

Rounding optimal experimental designs

Within the setting of polynomial regression models the theory of design of experiments supplies results on how to choose optimal designs, see Pukelsheim (1993). Generally, a design consists of regression vectors $x_1, \ldots, x_c$ and weights $w_1, \ldots, w_c > 0$ summing to 1. This specifies to take a proportion of w_i out of all observations under the experimental conditions that correspond to the regression vector x_i. Given the total number n of observations to be taken and an optimal design, one faces the problem of rounding the proportions given by the design to integral numbers n_i of observations summing up to n.

Pukelsheim and Rieder (1992) showed that rounding optimal designs with Adams' method maximizes a lower efficiency bound within the class of all rounding methods meeting the summation restriction $\sum n_i = n$. Therefore Pukelsheim (1993) calls Adams' method the *efficient design apportionment* and recommends it for rounding optimal designs. Note that, however, Adams' method maximizes a lower efficiency bound, not efficiency itself.

Table 4 reworks Exhibit 12.2 of Pukelsheim (1993, page 310). The weights $w_1 = \frac{1}{6}$, $w_2 = \frac{1}{3}$, $w_3 = \frac{1}{2}$ are given by an optimal design. With different total numbers n of observations, Adams' rounding method yields the following results for n_i, $i = 1, 2, 3$:

The total numbers of observations 1, 2, 7, 8, 13, 14, ... admit three solutions each, which is indicated by trailing '+' or '−' signs. For instance, the notation $(n_1, n_2, n_3) = (3-, 5-, 6+)$ (see Table 4, $n = 14$) is shorthand for the set of solutions $\{(2,5,7), (3,4,7), (3,5,6)\}$.

Rounding optimal sample allocations

When drawing n observations from populations with c strata and known strata variances σ_i^2, $i = 1, \ldots, c$, the theory of stratified sampling supplies results on how to choose allocations, i.e. numbers n_i^* of observations to be drawn from stratum i with $\sum n_i^* = n$, which are optimal with respect to variance minimization of the stratified population mean estimator, see Cochran (1977) for details. More explicitly, if p_i denotes the relative size of stratum i within the total population, the optimal allocation $(n_1^*, \ldots, n_c^*)$ is given by

$$n_i^* = \frac{p_i \sigma_i n}{\sum_j p_j \sigma_j}, \quad i = 1, \ldots, c.$$

In general, the numbers n_i^* fail to be integral and therefore need to be rounded. Neglecting constant factors, this leads to a rounding problem with weights $w_i = p_i \sigma_i$ and a number of units n.

n	n_1		n_2		n_3	
1	1	−	0	+	0	+
2	1	−	1	−	0	+
3	1		1		1	
4	1		1		2	
5	1		2		2	
6	1		2		3	
7	2	−	2	+	3	+
8	2	−	3	−	3	+
9	2		3		4	
10	2		3		4	
11	2		4		5	
12	2		4		6	
13	3	−	4	+	6	+
14	3	−	5	−	6	+
15	3		5		7	
16	3		5		8	
17	3		6		8	
18	3		6		9	

Table 4. *Rounding an optimal design with Adams' method: For better readability, the signs z_i indicating multiple solutions have been replaced by trailing '+' or '−' signs; rounded weights marked with '+' or '−' may be incremented or decremented, respectively, in order to obtain alternative solutions to the rounding problem.*

Pukelsheim (1997) showed that rounding optimal allocations with Jefferson's method maximizes a lower bound for the variance efficiency within the class of all rounding methods respecting the summation restriction, and therefore suggested the name *efficient sample allocation*. Here, variance efficiency denotes the quotient of the optimum variance and the variance yielded by the regarded allocation. Jefferson's method, however, in general fails to maximize variance efficiency itself, as can be seen in Table 5. In this table we consider a population with $c = 3$ strata of sizes $p_1 = 0.35$, $p_2 = 0.45$, $p_3 = 0.2$, and strata variances $\sigma_1^2 = 36$, $\sigma_2^2 = 4$, and $\sigma_3^2 = 1$. The optimal allocation for sample size $n = 300$ is

$$(n_1^*, n_2^*, n_3^*) = (196.875, 84.375, 18.75),$$

which implies the minimum variance $0.0341\overline{3}$ of the stratified population mean estimator.

Apart from sub-optimality, Jefferson's method, and more general, all arithmetic-mean methods except for Adams' method, suffer from possibly rounding weights to zero. In the stratified sampling context this corresponds to neglecting strata and leads to biased population mean estimation. In order to illustrate this,

method	n_1	n_2	n_3	variance	efficiency
Adams	196	85	19	0.0341347	0.9999607
Webster	197	84	19	0.0341339	0.9999832
Jefferson	198	84	18	0.0341378	0.9998690

Table 5. *Rounding optimal sample allocations with different rounding methods yields different variances of the stratified population mean estimator. Jefferson's method is outperformed by Adams' and Webster's method in terms of variance efficiency. (Actually, Adams' method gives two solutions, one of which is identical to that given by Webster's method.)*

method	n_1	n_2	n_3	standard deviation	bias
Adams	1332	667	1	0.33569692	0
Webster	1333	667	0	0.33561145	4
Jefferson	1333	667	0	0.33561145	4
Hill	1333	666	1	0.33569691	0

Table 6. *Rounding optimal sample allocations with arithmetic-mean methods may lead to a biased stratified population mean estimator. The last two columns give the standard deviation and bias of this estimator. Note that the bias 4 is about twice as large as the 3σ range of the estimator.*

consider the following parameters, where μ_i denotes the mean of the observed variable in stratum i:

$$\begin{aligned} p_1 &= 0.6\,, & p_2 &= 0.2\,, & p_3 &= 0.2\,, \\ \sigma_1 &= 16.675\,, & \sigma_2 &= 25.02\,, & \sigma_3 &= 0.005\,, \\ \mu_1 &= 400\,, & \mu_2 &= 600\,, & \mu_3 &= 20\,. \end{aligned}$$

Rounding the optimal allocation

$$(n_1^*, n_2^*, n_3^*) = (1333.111259, 666.755496, 0.133245)$$

to a total of 2000 observations with different rounding methods gives the results displayed in Table 6.

As a consequence, we cannot recommend a specific multiplier method for rounding optimal sample allocations, but rather suggest geometric-mean methods with parameter $r < 1$ in order to generate "candidates" for rounded sample allocations with maximum efficiency. To the experience of the authors, this is a reasonable way for practitioners, since variance efficiencies can be easily computed and compared.

6 Conclusions

Due to the paradoxes many other rounding methods suffer from, only multiplier rounding methods should be used for rounding weights with a summation restriction on the rounding results. When applying multiplier methods, an easy-to-implement algorithm can be utilised in order to avoid the problem of finding a multiplier yielding rounding results that meet the summation restriction. Different fields of applications may require specific properties of the rounding methods used, which can be guaranteed by using special families of multiplier methods.

As a topic of future research, Pukelsheim's and Rieder's (1992) efficient design apportionment and Pukelsheim's (1997) efficient sample allocation could be reviewed. Furthermore, Happacher's (1996) theory on the distributional properties of arithmetic-mean rounding methods and the worst case analysis of Proposition 2 could be extended to other families of multiplier methods.

Implementation and availability

The algorithm presented in Section 3 and its specialisations for the families of arithmetic-mean, power-mean, and geometric-mean rounding methods have been implemented at the University of Augsburg. These implementations are integrated in the software package *Roundpro* offering a GNU Emacs based user interface.

Roundpro is available at URL `http://www1.Math.Uni-Augsburg.DE/sta/` via World Wide Web and requires the freely available GNU Emacs editor and a C compiler on a UNIX system. Dorfleitner, Happacher, Klein, and Pukelsheim (1996) may serve as a manual for this software. It can be requested at the address given in the header of the present article.

References

Balinski, M.L. and Rachev, S.T. (1993). Rounding proportions: Rules of rounding. *Numerical Functional Analysis and Optimisation* **14**, 475–501.

Balinski, M.L. and Young, H.P. (1982). *Fair Representation. Meeting the Ideal of One Man, One Vote.* Yale University Press, New Haven, CT.

Cochran, W.G. (1977). *Sampling Techniques, Third Edition.* Wiley, New York.

Diaconis, P. and Freedman, D. (1979). On Rounding Percentages. *Journal of the American Statistical Association* **74**, 359–364.

Dorfleitner, G., Happacher, M., Klein, T., and Pukelsheim, F. (1996). Roundpro: An Emacs Lisp Implementation Of Rounding Procedures. Report No. 347, Institut für Mathematik der Universität Augsburg.

Happacher, M. (1996). *Die Verteilung der Diskrepanz von stationären Multiplikatormethoden zur Rundung von Wahrscheinlichkeiten.* Wißner, Augsburg.

Happacher, M. and Pukelsheim, F. (1996a). Rounding Probabilities: Unbiased Multipliers. *Statistics & Decisions* **14**, 373–382.

Happacher, M. and Pukelsheim, F. (1996b). Rounding Probabilities: Maximum Probability and Minimum Complexity Multipliers. Report No. 361, Institut für Mathematik der Universität Augsburg.

Pukelsheim, F. (1993). *Optimal Design of Experiments.* Wiley, New York.

Pukelsheim, F. (1997). Efficient Rounding of Sampling Allocations. *Statistics and Probability Letters* **35**, 141–143.

Pukelsheim, F. and Rieder, S. (1992). Efficient rounding of approximate designs. *Biometrika* **79**, 763–770.

Statistisches Bundesamt (ed.) (1996). *Statistisches Jahrbuch 1996 für die Bundesrepublik Deutschland.* Metzler-Poeschel, Stuttgart.

Gregor Dorfleitner
Institut für Statistik
und Mathematische Wirtschaftstheorie
Universität Augsburg
D-86135 Augsburg, Germany

Thomas Klein
Institut für Mathematik
Universität Augsburg
D-86135 Augsburg, Germany