Distributions with Fixed Marginals and Related Topics
IMS Lecture Notes – Monograph Series Vol. 28, 1996

How Should Data Be Rounded?

By Michel L. Balinski
C.N.R.S. and
Laboratoire d'Econométrie de l'Ecole Polytechnique

This paper considers the problem of how to round numerical results (such as health statistics, population censuses, or ...) that are to be reported, and describes two distinct approaches: rules and methods of rounding. Rounded percentages often fail to be "justified" – to add to 100% – and it is reasonable to address the question of how "best" to round. A *rule* of rounding is an independent rounding of each datum. The conventional rule of rounding – round to the closest integer – is "best" only in limited circumstances: the choice of rule should depend upon the distribution of the raw data. A *method* of rounding depends upon all of the data and guarantees justified results.

Introduction. Every day each and every one of us is confronted by numbers: election returns, income distributions, health statistics, laboratory results, population censuses, $\cdots$ For the most part these numbers are **rounded** in some way or other, but usually we never know exactly how. Frequently the numbers are reported in terms of percentages – presumably because in this light they are more telling – but often these percentages do not add up to precisely 100%. When the data is tabular, with row and column sums having significance of their own, this failure to have rounded data that is "justifed" in rows and in columns as well as in total, is even more prevalent. What **should** be done is the question I address. Whenever justified answers are a must, the roundings may be viewed as distributions with fixed marginals that depend upon the distributions of the original numbers and how the roundings are obtained.

The headline of *Le Monde* of September 22, 1992 announced that the Treaty of Maastricht had been approved by 51.04% of French voters and disapproved by 48.95%. Nothing was said about the other .01% of the voters - some 2580 unaccounted for persons. In fact the margin was 51.0461% for and 48.9539% against (rounded to the nearest .0001%), so one might reasonably

AMS 1991 Subject Classification: Primary 60E99; Secondary 60F99, 41A29

Key words and phrases: Rounding vectors, rounding tables, rounding proportions, probability metrics, apportionment.

have expected the headline to read 51.05% for and 48.95% against. Perusal of more detailed data given in accompanying tables confirmed the fact that the "rule" of rounding used by that august journalistic institution was to drop the last digits. This is not the best rule to use if one wishes the sum to be 100%, but the fact is that no "rule" of rounding can guarantee that the sum of the roundings is equal to the rounding of the sum.

It seems that relatively little attention has been given to the problem of how to round. In 1967 Mosteller, Youtz and Zahn computed the probability that the sum of percentages, each rounded to the nearest significant digit, equals 100%, under several different probabilistic models for the underlying data. In 1979 Diaconis and Freedman extended the analysis by computing the same probability in the limit as the rounding becomes more and more accurate and when the underlying data is uniform. Otherwise, the literature on rounding is primarily concerned with the propagation of error in computing with limited accuracy (von Neumann and Goldstine (1947), Turing (1948), Wilkinson (1963)).

This expository paper considers the problem of how to round when the goal is simply that of rounding numerical results that are to be reported, for example in terms of percentages. It shows through examples why the problem is of interest, describes two distinct approaches and summarizes several of the principal results. The details and proofs may be found in joint papers with S. T. Rachev (1993a and b).

The first approach was originally motivated by the idea of finding a "rule" of rounding – meaning an independent rounding of each datum – that maximizes the probability that the result is justified, that is, that the sum of the roundings equals the rounding of the sum. Raw data can be imagined as coming from some underlying probability distribution. A rule of rounding generates new data that has its own distribution, and different rules of rounding will of course engender different distributions. The essence of our results to date is that the original goal is best met by choosing a rule so that the distribution of the roundings is "as close as may be" to the distribution of the raw data (Balinski and Rachev (1993b)). The analysis draws heavily on the methods of probability metrics, more particularly on ideal metrics (Rachev (1991)). A noteworthy result is that while for uniform distributions of raw data the nearest significant digit rule is "best", the same is decidedly not true for other distributions.

The second approach concerns a different problem where one wishes rounded results that must *necessarily* be justified. The analysis of "methods" of rounding (Balinski and Rachev (1993a)) – meaning procedures for rounding data that depend upon all of the data and guarantee justified results

– is closely related to the problem of political apportionment (see Balinski and Young (1982), Balinski and Demange (1989)) and relies on a practical axiomatic formulation. Roughly speaking the idea is to view a method of rounding as a correspondence that assigns to each (vector or matrix) problem at least one justified rounding and to then postulate the properties that such a correspondence should reasonably enjoy. The specific methods that meet these properties are then derived, ... or shown not to exist.

1. Rules of Rounding. A *vector problem of rounding* is defined by any positive real t and vector $\boldsymbol{p} = (p_j), j \in S = \{1, \cdots, s\}$, where the p_j are real numbers. A *rule of* $(1/t)$*-rounding* is a mapping ρ_t

$$\rho_t : \boldsymbol{p} \rightarrow \{\boldsymbol{x} = (x_j) : x_j = k_j/t, k_j \text{ integer}, \ j \in S\}.$$

This will be written $\boldsymbol{x} = \rho_t(\boldsymbol{p})$.

The conventional rule of $(1/t)$-rounding – when x_j is taken to be p_j rounded to the nearest $1/t$ – and the truncation rule used by *Le Monde* are both particular instances of *divisor rules of* $(1/t)$*-rounding* based on the function d which is defined on the integers and satisfies $d(k) \in [k, k+1]$. Such a rule is a mapping ρ_t^d, defined by

$$x_j := [p_j]_t^d := k/t \text{ if } d(k-1) < tp_j \leq d(k), k \text{ integer}.$$

Thus $d(k)$ is simply a "threshold" for rounding: above it round up, at or below it round down. The *stationary rule of rounding based on* λ is a divisor rule with $d(k) = k + \lambda$ for all k, where $0 \leq \lambda \leq 1$. The conventional rule is the stationary rule based on $1/2$, with $d(k) = k + 1/2$ for all k; the truncation rule is the stationary rule based on 0, with $d(k) = k$ for all k. The K-*stationary rule based on* $(\lambda_0, \cdots, \lambda_{K-1}, \lambda), 0 \leq \lambda \leq 1, 0 \leq \lambda_j \leq 1$ for $j = 0, \cdots, K-1$, is the divisor rule with

$$d(k) = \begin{matrix} k + \lambda_k & \text{if } 0 \leq k \leq K-1, \\ k + \lambda & \text{otherwise.} \end{matrix}$$

In the sequel, for $S = \{1, \cdots, s\}$ and $\mathbf{y} = (y_j)$, define $y_S := \Sigma_S y_j$.

The following theorem strengthens a result of Diaconis and Freedman (1979).

THEOREM 1. *Suppose* $\boldsymbol{p}$ *is uniformly distributed (or absolutely continuously distributed) on the simplex* $\{\boldsymbol{p} \geq \mathbf{0} : p_S = 1\}$. *Then the maximum of the limiting probability* $\lim_{t\to\infty} Pr(\Sigma_S x_j = 1)$ *over the set of all K-stationary rules is attained with* $\lambda = 1/2$ *and* $\lambda_0, \cdots, \lambda_{K-1}$ *arbitrary.*

However, the rate of convergence can be very slow:

$$\lim_{t\to\infty} \Pr(\Sigma_S x_j = 1) \approx \{6/\pi(s-1)\}^{1/2} + o(s^{1/2}).$$

Suppose that $\boldsymbol{p} \geq \mathbf{0}$ consists of s i.i.d. random variables, that p_1 has a continuous distribution and that $\mathbf{x}$ is obtained from $\mathbf{p}$ by some $(1/t)$-rule of rounding. The question is: what is the deviation between p_S and x_S. The answer depends on the distribution of p_1, the measure μ of "deviation" and the rule of rounding.

A rule of rounding $\boldsymbol{x}^* = \rho_t^*(\boldsymbol{p})$ is *optimal* with respect to the metric μ over a class of rules $\mathfrak{R}$ if for any $\boldsymbol{p}$

(1) $\mu(p_S, x_S^*) = \min_{\rho_t}\{\mu(p_S, x_S) : \boldsymbol{x} = \rho_t(\boldsymbol{p}), \rho_t \in \mathfrak{R}\}$ and

(2) $\mu((1/s)p_S, (1/s)x_S^*) \to 0$ as $s \to \infty$.

The only type of metric that seems able to meet these requirements is an *ideal metric* of order $r > 0$ (see Rachev (1991)) which satisfies

$$\mu(c\Sigma_1^s X_j, c\Sigma_1^s Y_j) \leq c^r \Sigma_1^s \mu(X_j, Y_j) \quad \text{for any } c > 0,$$

where $\boldsymbol{X} = (X_j)$ and $\boldsymbol{Y} = (Y_j)$ are random vectors each with i.i.d. components. An example of an ideal metric of order $r = 1 + 1/\alpha, \alpha \geq 1$, is

$$\theta_r(X, Y) = \sup[|E(f(X) - f(Y))| : f \in \mathcal{F}_r],$$

where $\mathcal{F}_r$ is the set of all functions f whose second derivative has a bounded β-norm: $\|f''\|_\beta = [\int |f''|^\beta]^{1/\beta}$, with $1/\alpha + 1/\beta = 1$. It is easy to check that

(3) $\theta_r((1/s)p_S, (1/s)x_S) \leq O(s^{-1/\alpha})$ as $s \to \infty$ whenever $\theta_r(p_1, x_1) < \infty$.

By the definition of $\theta_r, \theta_r(X, Y) < \infty$ implies $E(X - Y) = 0$. In fact, $\theta_r(X, Y) \geq \sup[|E(aX - aY)| : a > 0] = +\infty$ if $E(X - Y) \neq 0$. So a *necessary* condition for $\boldsymbol{x}^* = \rho_t^*(\boldsymbol{p})$ to be an optimal stationary rule with respect to θ_r is the equality of the first moments of p_1 and x_1^*. This condition becomes sufficient under the mild condition that $Ep_1^r < \infty$. Thus,

THEOREM 2. *Suppose* $\boldsymbol{p}$ *consists of* s *i.i.d. random variables and* Ep_1^r *is finite for some* $r \in (1, 2)$. *Then* $\theta_r(p_S, x_S) = \infty$ *and* $\theta_r((1/s)p_S, (1/s)x_S) = \infty$ *for any rule of* $(1/t)$*-rounding with* $Ep_1 \neq Ex_1^*$. *However, if* $Ep_1 = Ex_1^*$ *for some stationary rule* $\boldsymbol{x}^* = \rho_t^*(\boldsymbol{p})$ *then* ρ_t^* *is optimal with respect to* θ_r *over the class of all stationary rules, and (3) holds.*

Thus, to determine an optimal stationary rule with respect to Q_r it suffices to choose λ so that

(4) $tEp_1 = \Sigma_0^\infty \Pr(tp_1 > k + \lambda)$,

an equation which has a unique solution λ for any $t > 0$ provided that the distribution function $Fp_1(x)$ is strictly increasing.

EXAMPLE 1. Suppose p_1 is uniform over the interval $(0, 1)$. Then for any $t \in N = \{1, 2, \cdots\}$, (4) is satisfied by $\lambda = 1/2$. On the other hand, if $t \in N + 1/2$, then (4) is satisfied by $\lambda = (t - 1/4)(2t + 1) < 1/2$.

EXAMPLE 2. Suppose p_1 is distributed according to the "first digit law": $Fp_1(x) = \log_{10}(1 + x)$, $0 \leq x \leq 9$. Then for $t = 10$, (4) is satisfied by $\lambda = 0.4984 \cdots$, again close to $1/2$ but not exactly on the mark!

The class of stationary rules, while natural enough, is very restricted, and the convergence of (2) can be slow. Given several optimal rules with respect to some μ the *preferred* rule is the one which has the fastest rate of convergence of (2). Accordingly, ρ_t^* is *optimal of order* δ with respect to μ over the class of rules $\Re$ if it is the preferred optimal rule and $\mu((1/s)p_S, (1/s)x_S^*) \to O(s^{-\delta})$ as $s \to \infty$. Thus Theorem 2 asserts that there exists an optimal stationary rule of order $r - 1$ with respect to θ_r if the r-th moment of p_1 is finite. One immediately asks: would a different ideal metric μ yield a different result? The answer is "no" provided that μ has the following "law of large numbers property": $\mu((1/s)\Sigma_1^s X_j, EX_1) \to 0$ as $s \to \infty$ for any nonnegative i.i.d. X_j with finite EX_1.

A K-*stationary rule* $\boldsymbol{x}^* = \rho_t^*(\boldsymbol{p})$ *of order* $\delta = r - 1$ is *optimal* with respect to the metric θ_r over the class of K-stationary rules if for any $\boldsymbol{p}$

$$\theta_r(p_S, x_S^*) = \min_{\rho^t}\{\theta_r(p_S, x_S) : \boldsymbol{x} = \rho_t\ (\boldsymbol{p}), \rho_t \text{ K-stationary}\}$$

and

$$\theta_r((1/s)p_S, (1/s)x_S^*) = O(s^{1-r}) \quad \text{as } s \to \infty.$$

THEOREM 3. *Suppose $Ep_1^r < \infty$ with $r = K + 1 + 1/p$. Then $\boldsymbol{x}^* = \rho_t^*(\boldsymbol{p})$ is an optimal K-stationary rule of order $\delta = r - 1$ with respect to the metric θ_r if and only if the thresholds $\lambda_0, \cdots, \lambda_{K-1}, \lambda$ are chosen so that*

$$E(p_1^j - x^{*j}) = 0 \text{ for } j = 1, \cdots, K + 1.$$

Thus, to determine an optimal K-stationary rule the thresholds must be chosen to satisfy

(5) $\quad Ep_1^j = \Sigma_0^\infty (k/t)^j \Pr(k - 1 + \lambda_{k-1} < tp_1 < k + \lambda_k),$

for $j = 1, \cdots, K + 1$, where $\lambda_k = \lambda$ for $k \geq K$.

EXAMPLE 3. Suppose p_1 is uniform over the interval $(0, 1)$, t is an integer and $K = 1$. Then (5) determines $\lambda_0 = 1/3, \lambda = (3t - 2)/(6t - 6)$.

If $t = 2$, then p_1 must be rounded to either $0, 1/2$ or 1. The conventional rule with rounding x_1^c has as its thresholds $1/4$ and $3/4$. The first moments agree, $Ep_1 = Ex_1^c = 1/2$, but not the second moments, $1/12 = \text{Var } p_1 < \text{Var } x_1^c = 1/8$.

The optimal 1-stationary rule with rounding x_1^* has as its thresholds $1/6$ and $5/6$, and of course the first and second moments agree. Therefore the central limit theorem applies, and for the Kolmogorov metric $K(p_S, x_S^*) = \sup\{|\Pr(p_S \leq y) - \Pr(x_S^* \leq y)| : y \text{ real}\}$,

$$K(p_S, x_S^*) \approx O(s^{-1/2}) \text{ as } s \to \infty,$$

whereas the conventional rule yields

$$K(p_S, x_S^c) \to K(N_{(0,1)}, N_{(0,\sqrt{(2/3)})}) > 0 \text{ as } s \to \infty,$$

where $N_{(m,\sigma)}$ is the normal distribution with mean m and standard deviation σ.

Another evaluation of the deviations of x_S^* and x_S^c from p_S is as follows. Let the deviation of the sum of the roundings x_S^c from p_S be

$$\Delta_s^c = \sup\{|\Pr(a < p_S \leq b) - \Pr(a < x_S^c \leq b)| : a < b\}$$

and let Δ_s^* be the corresponding deviation of x_S^* from p_S. Then

$$\lim_{s\to\infty} \Delta_s^c \geq 0.049 \quad \text{whereas} \quad \lim_{s\to\infty} \Delta_s^* = 0.$$

EXAMPLE 4. Suppose p_1 is uniform over the interval $(0, 1)$ and $t = 3$, so that p_1 must be rounded to 0, 1/3, 2/3 or 1. The conventional rule rounds at the thresholds 1/6, 1/2, and 5/6 and only the first moments agree. The optimal 2-stationary rule has as its thresholds 1/8, 1/2 and 7/8 and the first three moments agree.

To attack the problem of rounding vectors with n independent nonidentically distributed random variables consider the matrix $\boldsymbol{p} = (p_{ij}) \geq 0, i \in N, j \in S$, with each row $\boldsymbol{p}_{i\cdot} = (p_{i1}, \cdots, p_{is})$ consisting of s observations of the same variable.

THEOREM 4. *Let $\mathbf{p} = (p_{ij}) \geq 0, i \in N,\ j \in S$ and suppose that $\{p_{ij} : j \in S\}$ are i.i.d. copies of p_i and that Ep_i^r exists for each i. For each $i \in N$ let $\boldsymbol{x}_{i\cdot} = \rho_t^{(i)}(\boldsymbol{p}_{i\cdot})$ be the K-stationary rule of rounding uniquely determined by*

$$Ex_i^k = Ep_i^k \quad \text{for } k = 1, \cdots, K + 1.$$

Then the matrix $\boldsymbol{x} = (x_{ij})$ *consists of roundings that are optimal K-stationary of order* $r-1$ *with respect to the metric* θ_r

(i) over each row $i \in N$,

(ii) over each column $j \in S$, *and*

(iii) over the entire matrix.

The rules that optimally round the rows (Theorem 3) are also optimal for each column and for the entire matrix.

As a concrete illustration of how these results might be "used" – despite the fact that they are asymptotic in nature – consider two tables concerning the American presidential election printed in the *International Herald Tribune* on November 5, 1992. In the first table, 51 rows gave the vote totals received by each of the three major candidates – Bush, Clinton and Perot – in each of the 50 states and the District of Columbia. In the second one, also 51 by 3, integer percentages of the share of votes received by each candidate in each of the states and D.C. were given. The *I.H.T.* used the conventional rule to round raw percentages into integers. 34 of the 51 row sums turned out to equal 100%, 11 gave 101% and 6 gave 99%. Was this the "best" rule to use? Ignoring the obvious dependence among the three percentages, the above results suggest that a "best" stationary rule would be to round "by columns": for each column chose the stationary rule that equates "as near as may be" the averages of the raw and the rounded percentages. In this way the "best" rounding of the rows should also be obtained. In fact, by this rule 35 of the 51 sums were equal to 100%, 8 gave 101% and 8 gave 99%.

2. Methods of Rounding. A *vector problem of justified rounding* is defined by any positive real t and pair $(\boldsymbol{p}, h)$, where $\boldsymbol{p} = (p_j) \geq \mathbf{0}, j \in S$, is a nonzero vector of reals and $h > 0$ is a real number. The *set of justified* $(1/t)$*-roundings* for the problem $(\boldsymbol{p}, h)$ is

$$R_t(h) = \{\boldsymbol{x} = (x_j), j \in S : x_j = k_j/t, \ \ k_j \ \text{ integer}, \ x_S = h\}.$$

A problem is *feasible* whenever $R_t(h)$ is nonempty, and this is the case if and only if $h = k/t$ for some integer k. A *method of* $(1/t)$*-rounding* φ_t is a point to set mapping that assigns at least one justified $(1/t)$-rounding to every feasible problem $\varphi_t(\boldsymbol{p}, h) \subset R_t(h)$. The possibility of multiple solutions is unavoidable: the problem of 1-rounding ((2.5, 4.5), 7) cannot a *priori* exclude either of the roundings (3,4) and (2,5).

The question is: what methods φ_t **should** be used? Our approach is to impose properties on the qualitative behavior of φ_t that seem eminently reasonable in the context of the problem, and to deduce from them what should be done.

To begin it is intuitively clear that every $(1/t)$-rounding should be equivalent to an integer (or 1-) rounding via a change of scale: $\varphi_t(\mathbf{p}, h) = \varphi_1(\boldsymbol{p}, th)$. Accordingly, almost all mention of t is herewith dropped, $R_1(h) = R(h) = \{\boldsymbol{x} = (x_j), j \in S : x_j \text{ integer}, x_S = h\}, R(h) \neq \emptyset$ if $h > 0$ is integer, and $\varphi_1(\boldsymbol{p}, h) = \varphi(\boldsymbol{p}, h) \subset R(h)$.

The divisor rules of rounding are now modified to obtain methods of rounding, but in this case the thresholds must be handled with greater care. Specifically, define a *divisor function* d to be an arbitrary monotone real function defined over the nonnegative integers where for any integer $a \geq 0$, $a \leq d(a) \leq a + 1$ and, moreover, there exists no pair of integers $a \geq 0$ and $b \geq 1$ with $d(a) = a + 1$ and $d(b) = b$. A *d-rounding* of a real number $z \geq 0$ is

$$[0]_d = 0, \qquad [z]_d = a \text{ if } d(a-1) \leq z \leq d(a) \text{ for } z \neq 0,$$

so $[d(a)]_d = a$ or $a+1$: at the threshold one can either round up or down. The *divisor method of rounding* φ^d *based on* d is

$$\varphi^d(\boldsymbol{p}, h) = \{\boldsymbol{x} = (x_j), j \in S : x_j = [\lambda p_j]_d, \ \lambda > 0 \text{ chosen so that } x_S = h\}.$$

These methods first arose in the study of the apportionment problem (see Balinski and Young (1982)): how to apportion h seats in a legislature among the states or provinces (or the political parties) S having populations (or vote totals) $\boldsymbol{p}$. Popular methods include: $d(k) = k$ (first proposed by John Quincy Adams), $d(k) = k + 1/2$ (first proposed by Daniel Webster), and $d(k) = k + 1$ (first proposed by Thomas Jefferson). The method used in the United States to apportion Congress since 1940 is based on $d(k) = \{k(k+1)\}^{1/2}$.

Methods admit multiple solutions, rules do not. But the essential difference between a divisor *rule* based on d and a divisor *method* based on d is that a method rounds a scaling of the raw data by a common factor $\lambda_d > 0$ (in general not unique) chosen so that the result is justified, rather than rounding the raw data itself. Thus λ_d may be viewed as a "distortion", the greater the deviation from 1 the greater the distortion. In this light Theorem 1 may be interpreted as follows. Suppose again that $\boldsymbol{p}$ is absolutely continuously distributed on the simplex, and for each K-stationary method φ let Λ_φ be the set of distortion factors λ_φ that can arise when φ is used. Then

$$\max\{\lim_{t\to\infty} \Pr(\lambda_\varphi = 1 : \lambda_\varphi \in \Lambda_\varphi) : \varphi \text{ a K-stationary method}\}$$

is achieved by any K-stationary method having $\lambda_k = 1/2$ for $k \geq K$. Asymptotically the conventional method of rounding (based on $d(k) = k + 1/2$) is "best" if distortion is to be avoided.

Another method has often been proposed (first by Alexander Hamilton, also in the context of "controlled rounding", Cox and Ernst (1982)). Equivalent descriptions are as follows. Let $q_j = hp_j/p_S$, and $[z]^-$ and $[z]^+$ be z rounded down and up, respectively. The "controlled rounding" method is

$$\Psi(\mathbf{p}), h) = \{\boldsymbol{x} = (x_j), j \in S : [q_j]^- \leq x_j \leq [q_j]^+, x_j \text{ integer}, x_S = h,$$
$$\Sigma_S|x_j - q_j| = \text{ minimum}\}.$$

More prosaically, $\Psi(\mathbf{p}, h)$ is obtained by: (i) let $x_j = [q_j]^-$; (ii) if $\Sigma_S x_k < h$ and $k \in S$ satisfies $q_k - x_k \geq q_j - x_j$ for all $j \in S$, augment x_k by 1 (and repeat if necessary).

Indeed, there are many methods! Minimize your favorite concept of the distance of $\mathbf{x} \in R(h)$ from $\mathbf{p}$ (or $\mathbf{q}$). Four properties that a method should enjoy immediately suggest themselves. First, a method of rounding can depend only on the magnitudes of the data $(\mathbf{p}, h)$ and not on the order in which $\mathbf{p}$ is presented. If $\mathbf{z}$ is a vector, let $\mathbf{z}^\sigma$ be $\mathbf{z}$ reordered by a permutation σ of its indices. A method φ is *anonymous* if $\mathbf{x} \in \varphi(\mathbf{p}, h)$ implies $\mathbf{x}^\sigma \in \varphi(\mathbf{p}^\sigma, h)$ for any permutation σ. Second, a method should be independent of the scale in which the data $\mathbf{p}$ is presented: φ is *homogeneous* if $\mathbf{x} \in \varphi(\mathbf{p}, h)$ implies $\mathbf{x} \in \varphi(\lambda\mathbf{p}, h)$ for every $\lambda > 0$. Third, if the data $\mathbf{p}$ is itself integer valued and justified it must constitute the unique solution: φ is *exact* if $\mathbf{p}$ integer valued and $p_S = h$ implies $\varphi(\mathbf{p}, h) = \{\mathbf{p}\}$. Fourth, a method of rounding should preserve the ordering of the magnitudes: φ is *weakly monotonic* if $\mathbf{x} \in (\mathbf{p}, h)$ and $p_i < p_j$ implies $x_i \leq x_j$.

The next property is more subtle, although so natural as to seem to be innocuous. Suppose $\mathbf{x} \in \varphi(\mathbf{p}, h)$, that T is some subset of S and T' its complement in S. Denote the corresponding vectors $\mathbf{p}^T, \mathbf{x}^T$, etc., so that $\mathbf{p} = (\mathbf{p}^T, \mathbf{p}^{T'})$ and $\mathbf{x} = (\mathbf{x}^T, \mathbf{x}^{T'})$. What should be the φ-rounding of $(\mathbf{p}^T, x_T)$? For example, if $\varphi^{1/2}$ is the conventional method (based on $d(k) = k+1/2$), then $\varphi^{1/2}$ ((42.53, 35.89, 21.58), 100)= (42, 36, 22), and $\varphi^{1/2}$ ((42.53, 21.58),64) = (42, 22). For any reasonable method one expects that any subset of a rounding should be a rounding of the corresponding subproblem. And if it happened that some other rounding for the subproblem occured then it should be substitutable to obtain another rounding for the parent problem. For example, if (42, 36, 22) $\in \varphi$ ((42.53, 35.89, 21.58),100) for some φ and both (42, 22) and (43, 21) belonged to φ ((42.53, 21.58),64) then (43, 36, 21) should also belong to φ ((42.53, 35.89, 21.58),100). A method of rounding φ is *consistent* if $(\mathbf{x}^T, \mathbf{x}^{T'}) \in \varphi((\mathbf{p}^T, \mathbf{p}^{T'}), h)$ implies $\mathbf{x}^T \in \varphi(\mathbf{p}^T, x_T)$; moreover, if also $\mathbf{y}^T \in \varphi(\mathbf{p}^T, x_T)$ then $(\mathbf{y}^T, \mathbf{x}^{T'}) \in \varphi((\mathbf{p}^T, \mathbf{p}^{T'}), h)$.

A sixth property is implicitly suggested by Hamilton and the "controlled rounding" advocates: a real number z expressing a precise real percentage should be rounded to either $[z]^-$ or $[z]^+$ but surely not to $[z]^- - 1$ or less, or to $[z]^+ + 1$ or more! A method of rounding is *adjacent* if $\mathbf{x} \in \varphi(\mathbf{p}, h)$ with $p_S = h$ implies $[p_j]^- \leq x_j \leq [p_j]^+$ for all $j \in S$ (and is *lower-adjacent* if only the first inequalities are implied, *upper-adjacent* if only the second inequalities are implied).

THEOREM 5. *There exists no method of rounding that is anonymous, homogeneous, exact, weakly monotone, consistent and adjacent.*

Facts are stubborn. Some desirable property or properties cannot be met. It is relatively easy to verify that divisor methods satisfy the first five of these properties, and to find examples for particular divisor methods that do not satisfy adjacency. It is also easy to see that the controlled rounding method satisfies all of the properties except consistency, and to find examples that show consistency can be violated.

There are many good reasons to discard the demand of adjacency. The common underlying cause for this is that finding justified roundings of data, say in terms of percentages, is at bottom a question of proportionality. Adjacency is not a proportional idea: it imposes much more stringent limitations on data that is large in magnitude than on data that is small. A qualitative expression of this is in terms of another concept of monotonicity that compares the roundings of two different problems, $(\mathbf{p}, h)$ and $(\mathbf{p}', h)$, both with index set S, and that any method of rounding should satisfy. φ is *monotone* if $\mathbf{x} \in \varphi(\mathbf{p}, h)$ and $\mathbf{x}' \in \varphi(\mathbf{p}', h)$, both with index set S, and $p'_i/p'_j > p_i/p_j$ implies that either $x'_i \geq x_i$ or $x'_j \leq x_j$. That is, if the relative difference between p'_i and p'_j is greater than that between p_i and p_j, then their roundings should surely not satisfy $x'_i < x_i$ and $x'_j > x_j$. Every divisor method is monotone; and it is trivial to find examples showing that controlled rounding is not. Adjacency and monotonicity are incompatible; adjacency and consistency are also incompatible. What happens then when adjacency is dropped?

THEOREM 6. *A method of rounding is anonymous, homogeneous, exact, weakly monotone and consistent if and only if it is a divisor method.*

A specific well defined class of methods of rounding obtains ... but instead of the paucity of the empty set one is confronted by the riches of an infinity of choice.

Underlying the idea of adjacency is that each rounding should individually be close to its corresponding exact value. A slight weakening of this expression is this: φ is *pairwise adjacent* if $\mathbf{x} \in \varphi(\mathbf{p}, h)$ with $p_S = h$ implies there exists no pair $i, j \in S$ satisfying

$$p_i - (x_i - 1) < x_i - p_i \quad \text{and} \quad x_j + 1 - p_j < p_j - x_j,$$

that is, if it is impossible to increase the rounding of one number by 1 and decrease that of another by 1 and thereby bring *both* roundings closer to their respective precise values.

THEOREM 7. *The unique method of rounding that is anonymous, homogeneous, exact, weakly monotone, consistent and pairwise adjacent is the conventional divisor method* φ^d *based on* $d(k) = k + 1/2$.

Indeed, the conventional method rarely violates adjacency, and in certain circumstances may be shown to be the divisor method that is least likely to violate adjacency. In any case the following is also true:

THEOREM 8. *If* $\mathbf{x} \in \varphi^d(\mathbf{p}, h)$, *and* φ^d *is a divisor method of rounding, then* $\mathbf{x}$ *satisfies either lower- or upper-adjacency.*

There are many questions, properties and operational procedures yet to be answered, to be investigated and to be developed. The results described here are only intended to impart the flavor of the approach. In particular, this approach has also been applied to rounding matrices of real data. The marginal sums of rows and columns are fixed – they are the roundings of the corresponding sums of raw data – and the problem is to round the entries of the matrix so that each row and column of the result is justified. A set of properties having the same qualitative "feel" as those above result in the definition and characterization of a class of "divisor" methods of rounding matrices.

I am indebted to the editors and the referees for their very helpful comments.

References

BALINSKI, M. L., and DEMANGE, G. (1989). An axiomatic approach to proportionality between matrices. *Math. Oper. Res.* **14**, 700–719.

BALINSKI, M. L., and RACHEV, S. T. (1993a). Rounding proportions: methods of rounding. In revision.

BALINSKI, M. L., and RACHEV, S. T. (1993b). Rounding proportions: rules of rounding. *Numer. Funct. Anal. Optim.* **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.

COX, L. H., and ERNST, L. R. (1982). Controlled rounding. *INFOR* **20**, 423–432.

DIACONIS, P., and FREEDMAN, D. (1979). On rounding percentages. *J. Amer. Statist. Assn.* **72**, 359–364.

MOSTELLER, F., YOUTZ, C., and ZAHN, D. (1969). The distribution of sums of rounded percentages. *Demography* **4**, 1223–1230.

NEUMANN, J. VON, and GOLDSTINE, H. H. (1947). Numerical inverting of matrices of high order. *Bull. Amer. Math. Soc.* **53**, 1021–1099.

RACHEV, S. T. (1991). *Probability Metrics.* Wiley, Chichester: New York.

TURING, A. M. (1948). Rounding-off errors in matrix processes. *Quart. J. Mech.* **1**, 287–308.

WILKINSON, J. H. (1963). *Rounding Errors in Algebraic Processes*, Prentice–Hall, Englewood Cliffs, N.J.

LABORATOIRE D'ECONOMÉTRIE
ECOLE POLYTECHNIQUE
1 RUE DESCARTES
75005 PARIS, FRANCE
balinski@poly.polytechnique.fr