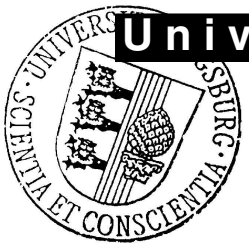

Universität Augsburg

Institut für Mathematik

Lothar Heinrich, Friedrich Pukelsheim, Vitali Wachtel

The Variance of the Discrepancy Distribution of Rounding Procedures, and Sums of Uniform Random Variables

Preprint Nr. 02/2016 — 08. Juni 2016

Institut für Mathematik, Universitätsstraße, D-86135 Augsburg http://www.math.uni-augsburg.de/

Impressum:

Herausgeber:

Institut für Mathematik
Universität Augsburg
86135 Augsburg
`http://www.math.uni-augsburg.de/de/forschung/preprints.html`

ViSdP:

Lothar Heinrich
Institut für Mathematik
Universität Augsburg
86135 Augsburg

The Variance of the Discrepancy Distribution of Rounding Procedures, and Sums of Uniform Random Variables

Lothar Heinrich, Friedrich Pukelsheim and Vitali Wachtel [1]

Abstract

When ℓ probabilities are rounded to integer multiples of a given accuracy n, the sum of the numerators may deviate from n by a nonzero discrepancy. It is proved that, for large accuracies $n \to \infty$, the limiting discrepancy distribution has variance $\ell/12$. The relation to the uniform distribution over the interval $[-1/2, 1/2]$, whose variance is $1/12$, is explored in detail.

Keywords : rounding residual, Euler-Maclaurin formula, Euler-Frobenius polynomial, Fourier-analytic approach

AMS 2010 MSC : Primary: 62E15, 62P25; Secondary: 65B15, 60E10

1 Introduction

Suppose we are given ℓ probabilities, $p_1, \ldots, p_\ell$. In printed publications the probabilities are rounded usually to percentages, or to multiples of tenths of a percent. That is, they are converted into integer multiples of $n = 100$, or of $n = 1000$. More generally each probability p_j is rounded into a fraction n_j/n, with some integer numerator n_j relative to a given "accuracy" n. The fractions n_j/n provide a valid distribution only if the sum of the numerators is equal to the denominator, $n_1 + \cdots + n_\ell = n$. It is well-known that individual rounding of the probabilities p_j may fail to satisfy this equation. Rather, a discrepancy $Z = n_1 + \cdots + n_\ell - n$ is observed which may be nonzero. Happacher [4] calculates the distribution of Z for finite accuracy n when the probability vector $\mathbf{p}^{(\ell)} = (p_1, \ldots, p_\ell)$ is uniformly distributed over the probability simplex. He also shows that these unwieldy distributions converge for large accuracies to the elegant

[1] Institute of Mathematics, University of Augsburg, 86135 Augsburg, Germany
E-mail: heinrich@math.uni-augsburg.de, pukelsheim@math.uni-augsburg.de, vitali.wachtel@math.uni-augsburg.de

distribution in display (2.1) below, the "discrepancy distribution". Because of symmetry its expectation is zero. Its variance is zero when $\ell = 1$ or $\ell = 2$. Using Maple, Mathematica or similar software it is easily verified computationally that the variance is equal to $\ell/12$, for all $\ell \geq 3$ that computers can handle.

In Section 2 the variance formula is proved in a rigorous way. Our proof is based on the Euler-Maclaurin formula and complements an alternative approach using characteristic functions (short: CF) that is due to Gawronski and Neuschel [3] and Janson [7]. Section 3 discusses an invariance principle according to which the discrepancy distribution is the limiting distribution whenever the distribution of the probability vector $\mathbf{p}^{(\ell)} = (p_1, \ldots, p_\ell)$ is absolutely continuous. The invariance principles explains the universal applicability of the discrepancy distribution. Section 4 reviews briefly the approach of Gawronski and Neuschel [3] and Janson [7]. An alternative Fourier-analytic derivation of the CF function of the discrepancy distribution is given in Section 5.

2 Discrepancy Variance

When rounding two or more proportions $p_1, \ldots, p_\ell$ to integer percentages $n_1, \ldots, n_\ell$, the resulting percentages do not necessarily sum exactly to 100 percentage points, but possibly leave a positive or negative discrepancy $z = (n_1 + \cdots + n_\ell) - 100$. Considering many and varied sets of $\ell \geq 2$ proportions, we may view the discrepancy to be a random variable Z taking its values in $\mathbb{Z}$, the set of all integers. Happacher [4] shows that, for all practical purposes, the distribution of the discrepancy Z is modelled well by $\mathbf{P}(Z = z) = g_\ell(z)$, where the probabilities are given by

$$g_\ell(z) = \sum_{k=0}^{\ell} \frac{(-1)^k}{(\ell-1)!} \binom{\ell}{k} \left(\frac{\ell}{2} + z - k\right)_+^{\ell-1}, \qquad z \in \mathbb{Z}. \tag{2.1}$$

The notation y_+^n is short for $(y_+)^n$, where $y_+ = y$ in case $y > 0$ and $y_+ = 0$ otherwise.

It is not hard to see (as reviewed below) that the probabilities are symmetric around zero, $g_\ell(z) = g_\ell(-z)$. Hence all odd moments of Z vanish. In particular the expectation of the discrepancy is zero, that is, the instances when the sum of the rounded percentages is larger than 100 outweigh the instances when the sum is smaller.

What about the discrepancy variance? It is not hard to see (and reviewed below) that the probability $g_\ell(z)$ is positive only for $|z| \leq L$, where here and throughout we put $L := \lfloor (\ell-1)/2 \rfloor$. With $\ell = 2$ the discrepancy attains the value $z = 0$ with probability one, whence the variance is zero. For three or more proportions this section establishes the following.

Theorem 2.1. *For $\ell \geq 3$ the discrepancy variance is $\ell/12$.*

Theorem 2.1 focusses on variances because of its use in Theorem 3.1. However, the proof of Theorem 2.1 shows that the statement extends to all even moments of order less than ℓ, as pointed out in the Remark at the end of this section.

Before turning to the proof of Theorem 2.1 we take the time for some preliminary remarks. The theorem states that a sum which appears to be rather elementary admits a simple evaluation,

$$\sum_{z\in\mathbb{Z}} z^2 g_\ell(z) = \frac{\ell}{12}. \tag{2.2}$$

Identity (2.2) looks like an innocuous exercise for an introductory probability course. The appearance is deceiving, we do not know of an easy proof of (2.2). Our proof builds on the Euler-Maclaurin formula, see Abramowitz and Stegun [1] (p. 806, formula 23.1.30). Another proof, due to Gawronski and Neuschel [3] and Janson [7] and to be reviewed in Section 4, identifies (2.1) as the Euler-Frobenius distribution $\mathcal{E}_{\ell-1,\ell/2}$.

The Euler-Maclaurin formula establishes a relation between a sum—the left hand side of (2.2)—and an integral. The right hand side of (2.2) becomes an integral by noting that it is the variance of $V_1+\cdots+V_\ell$, where $V_1,\dots,V_\ell$ are independent and identically distributed random variables whose common distribution is uniform over the interval $[-1/2,1/2]$.

The (well-known) Lebesgue density of $V_1+\cdots+V_\ell$ is obtained by way of the convolution lemma,

$$g_\ell(x) = \int g_{\ell-1}(y)g_1(x-y)\,dy = \int_{x-1/2}^{x+1/2} g_{\ell-1}(y)\,dy.$$

Starting from the indicator function $g_1(x)=\mathbb{1}_{[-1/2,1/2]}(x)$ for $\ell=1$, the density for $\ell\geq 2$ is found to be

$$g_\ell(x) = \sum_{k=0}^{\ell} \frac{(-1)^k}{(\ell-1)!}\binom{\ell}{k}\left(\frac{\ell}{2}+x-k\right)_+^{\ell-1}, \qquad x\in\mathbb{R}. \tag{2.3}$$

This is the same function as in (2.1), except that the domain of definition is extended from $\mathbb{Z}$ in (2.1), to $\mathbb{R}$ in (2.3).

The interrelation between (2.1) and (2.3) entails three useful implications. Firstly, the probabilities in (2.1) add to unity as they should,

$$\sum_{z\in\mathbb{Z}} g_\ell(z) = \sum_{z\in\mathbb{Z}} \int_{z-1/2}^{z+1/2} g_{\ell-1}(y)\,dy = \int_{\mathbb{R}} g_{\ell-1}(y)\,dy = 1.$$

Secondly, the function g_ℓ is symmetric. This is obvious for g_1. Assuming $g_{\ell-1}$ is symmetric, $g_{\ell-1}(y)=g_{\ell-1}(-y)$, so is g_ℓ, as seen by

$$g_\ell(x) = \int_{x-1/2}^{x+1/2} g_{\ell-1}(y)\,dy = \int_{x-1/2}^{x+1/2} g_{\ell-1}(-y)\,dy = \int_{-x-1/2}^{-x+1/2} g_{\ell-1}(y)\,dy = g_\ell(-x).$$

Thirdly, the Lebesgue density in (2.3) takes positive values on the open interval $(-\ell/2, \ell/2)$. Indeed, for $x \le -\ell/2$ we have $\ell/2 + x - k \le 0$. With all positive parts in (2.3) vanishing we get $g_\ell(x) = 0$. For $x \ge \ell/2$ symmetry entails $g_\ell(x) = g_\ell(-x) = 0$. For $x \in (-\ell/2, \ell/2)$ some positive parts are nonzero, whence $g_\ell(x) > 0$. From $g_\ell(z) = \int_{z-1/2}^{z+1/2} g_{\ell-1}(y)\,dy$ it follows that the probabilities in (2.1) are positive only for the integers $z \in \{0, \pm 1, \ldots, \pm L\}$. Upon introducing $f_\ell(x) = x^2 g_\ell(x)$ we may write identity (2.2) as

$$\sum_{z=-L}^{L} f_\ell(z) = \int_{-\ell/2}^{\ell/2} f_\ell(x)\,dx. \tag{2.4}$$

The Euler-Maclaurin formula relates sum and integral whenever f_ℓ is a smooth function. However, $g_\ell(x)$ in (2.3) is smooth only piecewise, on the integer intervals $[z, z+1]$ when ℓ is even, and on the shifted intervals $[z - 1/2, z + 1/2]$ when ℓ is odd. Therefore the proof of Theorem 2.1 treats even ℓ and odd ℓ separately.

Proof. of Theorem 2.1 Parts I and II deal with ℓ even. Part I restricts attention to integer intervals $[z, z+1]$ to apply the Euler-Maclaurin formula. Part II aggregates the interval results in order to establish the desired identity (2.4). Part III handles odd ℓ.

I. We fix an integer $z \in \mathbb{Z}$. For $x \le z + 1$ the term $(\ell/2 + x - k)_+^{\ell-1}$ in (2.3) is positive only for $0 < \ell/2 + x - k \le \ell/2 + z + 1 - k$, that is, $k < \ell/2 + z + 1$ and $k \le \ell/2 + z$. If additionally $x \ge z$ then $\ell/2 + x - k \ge \ell/2 + z - k \ge 0$. The passage to positive parts becomes superfluous, $(\ell/2 + x - k)_+^{\ell-1} = (\ell/2 + x - k)^{\ell-1}$. This shows that on the interval $[z, z+1]$ the function $g_\ell(x)$ is a polynomial of degree $\ell - 1$, namely,

$$g_\ell(x) = \sum_{k=0}^{\ell/2+z} \frac{(-1)^k}{(\ell-1)!} \binom{\ell}{k} \left(\frac{\ell}{2} + x - k\right)^{\ell-1} \quad \text{for} \quad z \le x \le z+1\,. \tag{2.5}$$

Hence on $[z, z+1]$ the function $f_\ell(x) = x^2 g_\ell(x)$ is a polynomial of degree $\ell + 1$.

Let $g_\ell^{(q)}(z+)$ and $f^{(q)}(z+)$ denote the derivatives of order q at the left endpoint of $[z, z+1]$, and $g_\ell^{(q)}(z+1-)$ and $f^{(q)}(z+1-)$ those at the right endpoint of $[z, z+1]$. The Euler-Maclaurin formula invokes the Bernoulli numbers B_{2k} and the odd derivatives $f^{(2k-1)}$ which we include up to order $\ell - 1$. The formula finishes with a balancing term depending on the $(\ell+2)$nd derivative of f_ℓ which is zero. Therefore the balancing term disappears, in our application. Thus the Euler-Maclaurin formula yields an identity,

$$\begin{aligned} \frac{1}{2} f_\ell(z) + \frac{1}{2} f_\ell(z+1) &= \int_z^{z+1} f_\ell(x)\,dx \\ &\quad + \sum_{k=1}^{\ell/2} \frac{B_{2k}}{(2k)!} \left(f^{(2k-1)}(z+1-) - f^{(2k-1)}(z+) \right). \end{aligned} \tag{2.6}$$

The derivatives of f_ℓ are obtained from the Leibniz rule,

$$\begin{aligned} f^{(q)}(x) &= \sum_{j=0}^{q} \binom{q}{j} \frac{d^j}{dx^j} x^2 \frac{d^{q-j}}{dx^{q-j}} g_\ell(x) \\ &= x^2\, g_\ell^{(q)}(x) + 2q\, x\, g_\ell^{(q-1)}(x) + q\,(q-1)\, g_\ell^{(q-2)}(x). \end{aligned} \tag{2.7}$$

The derivatives of g_ℓ in (2.5) are

$$g_\ell^{(q)}(x) = \begin{cases} \displaystyle\sum_{k=0}^{\ell/2+z} \frac{(-1)^k}{(\ell-1-q)!} \binom{\ell}{k} \left(\frac{\ell}{2} + x - k\right)^{\ell-1-q} & \text{if} \quad q < \ell - 1\,, \\ \displaystyle\sum_{k=0}^{\ell/2+z} (-1)^k \binom{\ell}{k} & \text{if} \quad q = \ell - 1\,, \\ 0 & \text{if} \quad q > \ell - 1\,. \end{cases} \tag{2.8}$$

For orders $q < \ell - 1$ the qth derivative $g_\ell^{(q)}(z+)$ at the left endpoint of $[z, z+1]$ coincides with the qth derivative $g_\ell^{(q)}(z-)$ at the right endpoint of the preceding interval $[z-1, z]$,

$$\begin{aligned} g_\ell^{(q)}(z+) &= \sum_{k=0}^{\ell/2+z} \frac{(-1)^k}{(\ell-1-q)!} \binom{\ell}{k} \left(\frac{\ell}{2} + z - k\right)^{\ell-1-q} \\ &= \sum_{k=0}^{\ell/2+z-1} \frac{(-1)^k}{(\ell-1-q)!} \binom{\ell}{k} \left(\frac{\ell}{2} + z - k\right)^{\ell-1-q} = g_\ell^{(q)}(z-). \end{aligned}$$

Therefore g_ℓ and f_ℓ are q times differentiable also at the knots $z \in \mathbb{Z}$, with derivatives $g_\ell^{(q)}(z) = g_\ell^{(q)}(z+) = g_\ell^{(q)}(z-)$ and $f^{(q)}(z) = f^{(q)}(z+) = f^{(q)}(z-)$, as long as $q < \ell - 1$.

II. Now we sum (2.6) over $z \in \mathbb{Z}$. Since (2.1) has support points $-L, \ldots, L$ and (2.3) has support $(-\ell/2, \ell/2)$ aggregation of the Euler-Maclaurin formulas yields

$$\begin{aligned} \sum_{z=-L}^{L} f_\ell(z) &= \int_{-\ell/2}^{\ell/2} f_\ell(x)\, dx \\ &\quad + \sum_{k=1}^{\ell/2} \frac{B_{2k}}{(2k)!} \sum_{z=-L}^{L} \Big(f^{(2k-1)}(z+1-) - f^{(2k-1)}(z+) \Big)\,. \end{aligned} \tag{2.9}$$

Thus it suffices to show that

$$\sum_{z=-L}^{L} \Big(f^{(2k-1)}(z+1-) - f^{(2k-1)}(z+) \Big) = 0, \qquad k = 1, \ldots, \ell/2. \tag{2.10}$$

For $k < \ell/2$ the order is $q = 2k-1 < \ell-1$ for which f_ℓ is q times differentiable. Thus (2.10) is a telescope sum and simplifies to a plain difference,

$$\sum_{z=-\ell}^{\ell} \Big(f^{(q)}(z+1) - f^{(q)}(z) \Big) = f^{(q)}(\ell+1) - f^{(q)}(-\ell) = 0.$$

For $k = \ell/2$ with ensuing order $q = \ell - 1$ we must evaluate the sum

$$S = \sum_{z=-L}^{L} \Big(f^{(\ell-1)}(z+1-) - f^{(\ell-1)}(z+) \Big) .$$

The Leibniz rule (2.7) includes lower order derivatives $g_\ell^{(\ell-2)}$ and $g_\ell^{(\ell-3)}$. They, too, lead to telescope sums that vanish and hence contribute nothing to S. As for the $(\ell-1)$st derivative, (2.8) gives $g_\ell^{(\ell-1)}(z+) = g_\ell^{(\ell-1)}(z+1-) = \sum_{k=0}^{\ell/2+z} (-1)^k \binom{\ell}{k}$. We obtain

$$S = \sum_{z=-L}^{L} \Big((z+1)^2 - z^2 \Big) \sum_{k=0}^{\ell/2+z} (-1)^k \binom{\ell}{k} = \sum_{k=0}^{\ell} (-1)^k \binom{\ell}{k} \sum_{z=k-\ell/2}^{\ell/2} \Big((z+1)^2 - z^2 \Big)$$

with

$$\sum_{z=k-\ell/2}^{\ell/2} \Big((z+1)^2 - z^2 \Big) = \Big(\frac{\ell}{2} + 1 \Big)^2 - \Big(k - \frac{\ell}{2} \Big)^2 = (\ell+1) + (\ell-1)k - k(k-1).$$

The binomial theorem and multiple uses of the identity $\binom{\ell}{k} k = \ell \binom{\ell-1}{k-1}$ result in

$$S = (\ell+1)(1-1)^\ell - \ell(\ell-1)(1-1)^{\ell-1} - \ell(\ell-1)(1-1)^{\ell-2}. \tag{2.11}$$

This gives $S = 0$ whenever $\ell \geq 3$, and establishes (2.10). Now (2.9) reduces to (2.4). The proof of Theorem 2.1 for even ℓ is complete.

III. In case ℓ is odd we again start out with a fixed integer $z \in \mathbb{Z}$. Since ℓ is odd the polynomial structure of the function g_ℓ holds true on the shifted interval,

$$g_\ell(x) = \sum_{k=0}^{(\ell-1)/2+z} \frac{(-1)^k}{(\ell-1)!} \binom{\ell}{k} \Big(\frac{\ell}{2} + x - k \Big)^{\ell-1} \quad \text{for} \quad z - \frac{1}{2} \leq x \leq z + \frac{1}{2} .$$

The derivatives of g_ℓ are the same as in (2.8), except that the upper summation limit now reads $(\ell-1)/2$. The function $f_\ell(x) = x^2\, g_\ell(x)$ continues to be a polynomial of degree $\ell+1$.

As in (2.6) the Euler-Maclaurin formula yields the identity

$$\frac{1}{2} f\Big(z - \frac{1}{2} \Big) + \frac{1}{2} f\Big(z + \frac{1}{2} \Big) = \int_{z-1/2}^{z+1/2} f_\ell(x)\, dx + \sum_{k=1}^{(\ell+1)/2} \frac{B_{2k}}{(2k)!} \Delta(k),$$

where we have set $\Delta(k) := f^{(2k-1)}(z+1/2-) - f^{(2k-1)}(z-1/2+)$ for short. The formula permits a refinement by dividing the underlying interval into two equal parts in order to pick up the value of f_ℓ at the interval midpoint z,

$$\frac{1}{2}f\Big(z-\frac{1}{2}\Big)+f_\ell(z)+\frac{1}{2}f\Big(z+\frac{1}{2}\Big)=2\int_{z-1/2}^{z+1/2}f_\ell(x)\,dx+\sum_{k=1}^{(\ell+1)/2}\frac{1}{2^{2k-1}}\frac{B_{2k}}{(2k)!}\Delta(k).$$

Subtraction of the first equation from the second yields the version to be pursued further,

$$f_\ell(z)=\int_{z-1/2}^{z+1/2}f_\ell(x)\,dx-\sum_{k=1}^{(\ell+1)/2}\left(1-\frac{1}{2^{2k-1}}\right)\frac{B_{2k}}{(2k)!}\Delta(k).$$

Summation over $z\in\mathbb{Z}$ leads to

$$\sum_{z=-L}^{L}f_\ell(z)=\int_{-\ell}^{\ell}f_\ell(x)\,dx-\sum_{k=1}^{(\ell+1)/2}\left(1-\frac{1}{2^{2k-1}}\right)\frac{B_{2k}}{(2k)!}\sum_{z=-L}^{L}\Delta(k). \tag{2.12}$$

We aim to verify that the last sum is zero, that is,

$$\sum_{z=-L}^{L}\Big(f^{(2k-1)}\Big(z+\frac{1}{2}-\Big)-f^{(2k-1)}\Big(z-\frac{1}{2}+\Big)\Big)=0,\qquad k=1,\dots,\frac{\ell+1}{2}. \tag{2.13}$$

For $k<(\ell+1)/2$ the order of the derivative is $q=2k-1<\ell$. Since q and ℓ are odd this forces $q<\ell-1$, whence the sum in (2.13) is a telescope sum that vanishes.

For $k=(\ell+1)/2$ the order is $q=2k-1=\ell$. The sum in (2.13) becomes

$$S=\sum_{z=-L}^{L}\Big(f^{(\ell)}\Big(z+\frac{1}{2}-\Big)-f^{(\ell)}\Big(z-\frac{1}{2}+\Big)\Big).$$

Applying the Leibniz rule (2.7) to $f^{(\ell)}$ we find that the first term depends on $g_\ell^{(\ell)}$ which is zero throughout. The third term, involving $g_\ell^{(\ell-2)}$, leads to another telescope sum that vanishes. Thus S is determined by the second term,

$$S=2\ell\sum_{z=-L}^{L}\sum_{k=0}^{(\ell-1)/2+z}(-1)^k\binom{\ell}{k}=2\ell\sum_{k=0}^{\ell}(-1)^k\binom{\ell}{k}\sum_{z=k-(\ell-1)/2}^{(\ell+1)/2}1.$$

Since the last sum counts $(\ell+1)-k$ ones, the result is

$$S=2(\ell+1)\ell(1-1)^\ell+2\ell^2(1-1)^{\ell-1}.$$

Now $S=0$ verifies (2.13), whence (2.12) turns into (2.4). Thus, Theorem 2.1 is completely proved. □

Remark *The proof generalizes to functions of the type* $f_\ell(x) = x^p\, g_\ell(x)$*, with* $p < \ell$ *even. Hence the discrepancy* Z *and the sum* $V_1 + \cdots + V_\ell$ *share not only the same variance* $\ell/12$*, but also all even moments up to order* $p < \ell$*.*

The aggregated Euler-Maclaurin identity (2.9) remains valid also when $\ell = 2$. The discrepancy variance on the left is 0, as mentioned in the paragraph prior to Theorem 2.1. On the other hand (2.11) results in $S = -2$. Inserting $B_2 = 1/6$ turns the right hand side of (2.9) into $2/12 + (1/6)(-2)/2 = 0$, too. In the same way identity (2.12) maintains its validity when $\ell = 1$.

Theorem 2.1 has repercussions on the correlation structure of the rounding residuals which transpire to jointly constitute the discrepancy. Rounding residuals, though dependent, are uncorrelated, as discussed in the next section.

3 Discrepancy Representation as a Sum of Uniform Random Variables

Happacher [4] derives the discrepancy distribution (2.1) under the assumption that the vector of proportions $\mathbf{p}^{(\ell)} = (p_1, \ldots, p_\ell)$ follows a uniform distribution on the probability simplex $\Omega_\ell = \{(p_1, \ldots, p_\ell) \in (0,1)^\ell \,|\, p_1 + \cdots + p_\ell = 1\}$. This assumption is too specific to justify the universal applicability of the discrepancy distribution (2.1). Rather, the justification originates from an invariance principle that allows to replace the uniform distribution on Ω_ℓ by an arbitrary absolutely continuous distribution on Ω_ℓ.

The general task is to round proportions p_j to integer multiples n_j of a preordained accuracy n. The accuracy is $n = 100$ for percentages, $n = 1000$ for tenths of a percent, etc. An obvious approach is to multiply a proportion by n, and to round the scaled quantity np_j to an integer n_j. We designate the standard rounding function by angle brackets $\langle x \rangle$, as do Abramowitz and Stegun [1] (p. 223). That is, if the fractional part of $x > 0$ is smaller than one half then x is rounded downwards, $\langle x \rangle = \lfloor x \rfloor$. If the fractional part is larger than or equal to one half then x is rounded upwards, $\langle x \rangle = \lceil x \rceil$.

The rounding procedure gives rise to the rounding residuals $U_j(n) = \langle np_j \rangle - np_j$. As $\langle np_j \rangle = n_j$ the discrepancy Z coincides with the sum of the rounding residuals,

$$Z = (n_1 + \cdots + n_\ell) - n = U_1(n) + \cdots + U_\ell(n).$$

The representation has dramatic consequences. The distribution of Z depends an the distributional assumption for the proportions $(p_1, \ldots, p_\ell)$ on Ω_ℓ only through the induced distribution of the rounding residuals $U_1(n), \ldots, U_\ell(n)$. While every specific distributional assumption for

the proportions $(p_1, \ldots, p_\ell)$ sparks objections as to its universal validity, this is not so for rounding residuals.

Rounding residuals $U_j(n)$ are "known" to be uniformly distributed over the interval $[-1/2, 1/2]$. This knowledge is scientific commonplace to an extent that every reference which makes use of it can be surpassed by a prior reference that has done so earlier. Examples abound, as Seal [10] demonstrates in his witty note "Spot the prior reference"; see Happacher [4] for a reprint of the note. While there exist numerous publications modeling a rounding residual by the uniform distribution, we know of just a handful of sources proposing a rigorous argument how the uniform distribution comes into being.

The work of Happacher [4] implies that for finite accuracy n the discrepancy fails the distribution (2.1), and that the rounding residuals cannot have a uniform distribution. Even assuming that each rounding residual is uniformly distributed over the interval $[-1/2, 1/2]$ they cannot be jointly independent, because their sum has the discrete distribution (2.1) and not the convolution distribution (2.3).

It seems natural to resort to an asymptotic approach. Janson [6] assumes that the accuracies n are uniformly distributed over a finite range $\{1, \ldots, N\}$. Then he lets the range tend to infinity, $N \to \infty$. In the present paper we assume an absolutely continuous distribution for the vector of proportions. Then we let the accuracy tend to infinity, $n \to \infty$. In this setting Heinrich et al. [**?**] prove that the limiting distribution of the rounding residual becomes uniform if the underlying distribution admits a Riemann integrable density. Janson [6] shows that the conclusion remains valid if the underlying distribution is absolutely continuous. Only recently did we spot the prior reference Tukey [12] who establishes the same result. The state of the art is summarized by the following invariance principle.

Theorem 3.1. *Assume that the proportions $\mathbf{p}^{(\ell)} = (p_1, \ldots, p_\ell)$ follow an absolutely continuous distribution on the probability simplex Ω_ℓ. Then the vector of rounding residuals $\mathbf{U}^{(\ell)}(n) = (U_1(n), \ldots, U_\ell(n))$ for accuracy n and the vector of proportions $\mathbf{p}^{(\ell)}$ jointly converge in distribution,*

$$(\mathbf{U}^{(\ell)}(n), \mathbf{p}^{(\ell)}) \xrightarrow[n\to\infty]{\text{in distribution}} (\mathbf{U}^{(\ell)}, \mathbf{p}^{(\ell)}),$$

where the components of the limit random vector $\mathbf{U}^{(\ell)} = (U_1, \ldots, U_\ell)$ are uniformly distributed over the interval $[-1/2, 1/2]$, exchangeable, uncorrelated, and independent of $\mathbf{p}^{(\ell)}$. Omitting an arbitrary component U_j, the remaining $\ell - 1$ variables U_k, $k \neq j$, are independent.

Proof. Theorem 3.1 coincides with Theorem 6.10 in Pukelsheim [8] where the assertions are proved, except for uncorrelatedness. As for the latter, we calculate the variance of the sum as the sum of the variances plus the sum of the covariances. Due to exchangeability all covariances are the same, whence $\mathbf{Var}(U_1 + \cdots + U_\ell) = \ell/12 + \ell(\ell-1)\,\mathbf{Cov}(U_1, U_2)$. But the sum of the

rounding residuals is equal to the discrepancy which has variance $\ell/12$, by Theorem 2.1. Now $\mathbf{Cov}(U_1, U_2) = 0$ establishes uncorrelatedness. □

Theorem 3.1 provides a solid justification for the commonplace assumption that rounding residuals follow a uniform distribution. Theorem 3.1 also justifies the universal applicability of the discrepancy distribution (2.1). Indeed, from $Z = U_1 + \cdots + U_\ell$ we see that the discrepancy Z attains a value $z \in \mathbb{Z}$ if and only if $U_1 + \cdots + U_{\ell-1} = z - U_\ell$. Evidently we have $z - U_\ell \in [z - 1/2, z + 1/2]$. This yields

$$\Big\{Z = z\Big\} = \Big\{U_1 + \cdots + U_{\ell-1} \in \Big[z - \frac{1}{2}, z + \frac{1}{2}\Big]\Big\}. \tag{3.1}$$

When for $n \to \infty$ the limit distributions take over, the probability of the event on the right hand side in (3.1) becomes $\int_{z-1/2}^{z+1/2} g_{\ell-1}(y)\, dy = g_\ell(z)$ as stipulated by (2.1). This shows that for large accuracies the distribution of the discrepancy Z is given by (2.1).

Standard rounding permits yet another representation. Since the right hand side in (3.1) may be expressed as $\{\langle U_1 + \cdots + U_{\ell-1}\rangle = z\}$, the discrepancy satisfies

$$Z = \langle U_1 + \cdots + U_{\ell-1}\rangle\,.$$

That is, the discrepancy Z behaves as if standard rounding is applied to the sum of $\ell - 1$ copies of uniform random variables.

4 Euler-Frobenius Distributions

The shifted discrepancy $Z + L$ is a discrete random variable that is nonnegative. Hence its probability generating function is a polynomial. Gawronski and Neuschel [3] identify this polynomial to be an Euler-Frobenius polynomial and study the induced distributions. Janson [7] calls them Euler-Frobenius distributions, and provides many additional results The distribution of the discrepancy Z is the Euler-Frobenius distribution $\mathcal{E}_{\ell-1,\ell/2}$.

Gawronski and Neuschel [3] (p. 7) and Janson [7] (p. 10) also calculate the CF of Euler-Frobenius distributions. The discrepancy distribution turns out to have CF

$$\varphi_\ell(s) = \mathrm{i}^{-\ell} e^{-\mathrm{i}\ell s/2} \left(e^{\mathrm{i}s} - 1\right)^\ell \sum_{k=-\infty}^{\infty} \frac{e^{-\pi \mathrm{i} k \ell}}{(s + 2\pi k)^\ell}. \tag{4.1}$$

Thus, the variance is found by calculating the negative second derivative of the CF (4.1) at $s = 0$ being equal to $\ell/12$. This approach provides another proof of Theorem 2.1. Section 5

concludes with an alternative derivation of the CF (4.1) in addition to those in Gawronski and Neuschel [3] and Janson [7].

5 An alternative Fourier-analytic approach

To complete this paper we present a further way to obtain the CF (4.1) which seems to be more direct and different from the methods used in Gawronski and Neuschel [3] and Janson [7]. Among others our approach is related with so-called sinc-integrals which recently attracted much interest due to their unexpected properties, see e.g. Schmid [9], Almkvist and Gustavsson [2]. The sinc-function is defined as follows: $\mathrm{sinc(t)} := \sin(\mathrm{t})/\mathrm{t}$ for $t \in \mathbb{R} \setminus \{0\}$ and $\mathrm{sinc}(0) := 1$. It is easily verified that the CF $u_\ell(t) := \mathbf{E}\exp\{\mathrm{i}t(V_1 + \cdots + V_\ell)\}$ of independent and uniformly on $[-1/2, 1/2]$ distributed random variables $V_1, \ldots, V_\ell$, see Section 2, can be expressed by $u_\ell(t) = \big(\mathrm{sinc(t/2)}\big)^\ell$. Since $u_\ell(t)$ is absolutely integrable for $\ell \geq 2$, the Fourier inversion theorem, see e.g. Taylor [11] (p. 271) combined with $\mathrm{sinc}(-\mathrm{t}) = \mathrm{sinc(t)}$ allows to express the symmetric density of $V_1 + \cdots + V_\ell$ given in (2.3) as Fourier integral

$$\begin{aligned} g_\ell(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\mathrm{i}tx}\, u_\ell(t)\, dt = \frac{1}{\pi}\int_{-\infty}^{\infty} e^{2\mathrm{i}tx}\,\big(\mathrm{sinc(t)}\big)^\ell\,\mathrm{dt} \\ &= \frac{1}{\pi}\int_{-\infty}^{\infty} \cos(2\,t\,x)\,\big(\mathrm{sinc(t)}\big)^\ell\,\mathrm{dt} > 0 \quad \text{for} \quad |\mathrm{x}| < \frac{\ell}{2}\,. \end{aligned} \tag{5.1}$$

The right hand integral disappears for $|x| \geq \ell/2$ and for $x \in \{0, \pm 1, \ldots, \pm L\}$ we get the symmetric lattice distribution (2.1) of the discrepancy introduced in Section 1. Its CF $\psi_\ell(s)$ is defined by

$$\psi_\ell(s) := \sum_{z=-L}^{L} e^{\mathrm{i}sz} g_\ell(z) = \frac{1}{\pi}\int_{-\infty}^{\infty} \sum_{z=-L}^{L} e^{\mathrm{i}(s+2\,t)z}\,\big(\mathrm{sinc(t)}\big)^\ell\,\mathrm{dt}\,. \tag{5.2}$$

We may rewrite this CF as follows:

Theorem 5.1. *For all $s \in \mathbb{R}$ we have the identity*

$$\psi_\ell(s) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\sin((2L+1)(\frac{s}{2}+t))}{\sin(\frac{s}{2}+t)}\,\big(\mathrm{sinc(t)}\big)^\ell\,\mathrm{dt} = \sum_{\mathrm{k}\in\mathbb{Z}} \Big(\mathrm{sinc}(\frac{\mathrm{s}}{2} + \mathrm{k}\pi)\Big)^\ell = \varphi_\ell(\mathrm{s})\,.$$

Proof. The first equality of Theorem 5.1 is immediately seen by inserting the Dirichlet kernel $\sum_{z=-L}^{L} e^{2\mathrm{i}\,x\,z} = \sin((2\,L+1)x)/\sin(x)$, see e.g. Taylor [11] (p. 162) in the right hand integral of (5.2) for $x = s/2 + t$. The third equality can be checked simply by rewriting all members of the doubly infinite series (4.1) by inserting Eulers formula $e^{\mathrm{ix}} = \cos(x) + \mathrm{i}\,\sin(\mathrm{x})$. It remains to verify the middle equality between the improper integral on the left and the double-sided infinite series on the right for any real s. It is rapidly seen that both $\psi_\ell(s)$ as well as $\varphi(s)$ (as

uniformly convergent series for $\ell \geq 2$) are even and 2π-periodic functions being infinitely often differentiable and taking the value 1 at $s = 0$. Here, $\psi_\ell(0) = 1$ is obvious for a CF but it also follows in the special case $a = 2L + 1, k = \ell$ from the "sinc integral"

$$\int_{-\infty}^{\infty} \frac{\sin(at)}{\sin(t)} \, (\mathrm{sinc}(t))^k \, dt = \pi \quad \text{for any real} \quad a > 0 \quad \text{and} \quad k = 1, \ldots, \lfloor a \rfloor + 1,$$

see Schmid [9] (p. 15-17). Now, we prove that $\psi_\ell(s)$ and $\varphi_\ell(s)$ have the same Fourier expansions. For doing this we show that the corresponding coefficients at $\cos(js)$ are identical for all $j \in \mathbb{Z}$. The expansion (5.2) reveals that the jth Fourier coefficient of $\psi_\ell(s)$ coincides with $g_\ell(j)$ as expressed in (5.1) for $j \in \mathbb{Z}$, where $g_\ell(j) = 0$ for $|j| > L$. Assuming an expansion $\varphi_\ell(s) = \sum_{j \in \mathbb{Z}} c_\ell(j) \cos(js)$ we can calculate the coefficients $c_\ell(j)$ as follows:

$$\begin{aligned} c_\ell(j) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \cos(js) \varphi_\ell(s) ds = \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \cos(2js) \sum_{k \in \mathbb{Z}} \big(\mathrm{sinc}(\mathrm{s} + \mathrm{k}\pi)\big)^\ell \mathrm{ds} \\ &= \frac{1}{\pi} \sum_{k \in \mathbb{Z}} \int_{-\pi/2+k\pi}^{\pi/2+k\pi} \cos\big(2j(s - k\pi)\big) \big(\mathrm{sinc}(\mathrm{s})\big)^\ell \mathrm{ds} = \frac{1}{\pi} \int_{-\infty}^{\infty} \cos(2\mathrm{js}) \big(\mathrm{sinc}(\mathrm{s})\big)^\ell \mathrm{ds} \, . \end{aligned}$$

Together with (5.1) we have $c_\ell(j) = g_\ell(j)$ for all $j \in \mathbb{Z}$ which was to be proved. □

Theorem 5.1 and the identity $\mathrm{sinc}(\frac{\mathrm{s}}{2} + \mathrm{k}\pi) = \mathrm{sinc}(\frac{\mathrm{s}}{2}) \, (-1)^{\mathrm{k}} \, \mathrm{s}/(\mathrm{s} + 2\mathrm{k}\pi)$ for $k \in \mathbb{Z}$ imply that

$$\log \psi_\ell(s) = \ell \, \log \mathbf{E} \exp\{\mathrm{i}sV_1\} + r_\ell(s) \quad \text{with} \quad r_\ell(s) = \log \left(\sum_{k \in \mathbb{Z}} \Big(\frac{(-1)^k s}{s + 2k\pi} \Big)^\ell \right). \tag{5.3}$$

The rules of differentiation yield that the derivatives of $r_\ell(s)$ up to order $\ell - 1$ disappear at $s = 0$. Using this fact the relation (5.3) reveals that the mth cumulant of the discrepancy distribution (2.1) coincides with the ℓ-fold of the mth cumulant of V_1 for $m = 1, \ldots, \ell - 1$, see also Theorem 5.3 in Janson [7]. In the particular case $m = 2$ this confirms once more the assertion of Theorem 2.1. Moreover, (5.3) allows to determine the fourth moment of the discrepancy distribution (2.1) which takes the values $\ell \, (5\,\ell - 2)/240$ for $\ell \geq 5$, whereas $1/4$ for $\ell = 3$ and $1/3$ for $\ell = 4$ follow directly from (2.1).

References

[1] Abramowitz, M., Stegun, I. A. (1970). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York.

[2] Almquist, G., Gustavsson, J. (2014). More remarkable sinc integrals and sums. *arXiv: 1405.1265v2 [math.CA]*.

[3] Gawronski, W., Neuschel, T. (2013). Euler-Frobenius numbers. *Integral Transforms and Special Functions* **24**, 817–830.

[4] Happacher, M. (2001). The discrepancy distribution of stationary multiplier rules for rounding probabilities. *Metrika* **53**, 171-181.

[5] Heinrich, L., Pukelsheim, F., Schwingenschlögl, U. (2004). Sainte-Laguë's chi-square divergence for the rounding of probabilities and its convergence to a stable law. *Statistics & Decisions* **22**, 43–59.

[6] Janson, S. (2013a). Asymptotic bias of some election methods. *Annals of Operations Research* **215**, 89–136.

[7] Janson, S. (2013b). Euler-Frobenius numbers and rounding. *Online Journal of Analytic Combinatorics* **8**, 34 pages.

[8] Pukelsheim, F. (2014). *Proportional Representation – Apportionment Methods and Their Applications. With a Foreword by Andrew Duff MEP*. Springer International Publishing, Cham (CH).

[9] Schmid, H. (2014). Two curious integrals and a graphic proof. *Elemente der Mathematik* **69**, 11–17.

[10] Seal, H. L. (1950). Spot the prior reference. *Journal of the Institute of Actuaries Students' Society* **10**, 255–258.

[11] Taylor, C. J. (1996). *An Introduction to Measure and Probability*. Springer, New York.

[12] Tukey, J. W. (1938). On the distribution of the fractional part of a statistical variable. *Recueil Mathématique Nouvelle série* [Matematicheskii Sbornik Novaya Seriya] **4**, 561–562. Reprinted in: *The Collected Works of John W. Tukey*, Volume VI (More Mathematical: 1938–1984), edited by Colin L. Mallows, Wadsworth & Brooks/Cole, Pacific Grove CA, 237–238.