Sainte-Laguë’s chi-square divergence for the rounding of probabilities and its convergence to a stable law

Lothar Heinrich, Friedrich Pukelsheim, Udo Schwingenschlögl

Received: September 18, 2003; Accepted: March 3, 2004

Summary: For rounding arbitrary probabilities on finitely many categories to rational proportions, the multiplier method with standard rounding stands out. Sainte-Laguë showed in 1910 that the method minimizes a goodness-of-fit criterion that nowadays classifies as a chi-square divergence. Assuming the given probabilities to be uniformly distributed, we derive the limiting law of the Sainte-Laguë divergence, first when the rounding accuracy increases, and then when the number of categories grows large. The latter limit turns out to be a Lévy-stable distribution.

1 Introduction

Let \( W = (W_1, \ldots, W_c) \) be some “arbitrary” probability vector, for a fixed number of categories \( c \). We model “arbitrariness” by taking \( W \) to follow a uniform distribution on the probability simplex

\[
S_c = \left\{ (w_1, \ldots, w_c) \in [0, 1]^c : \sum_{j \leq c} w_j = 1 \right\}.
\]

When quoting the weights \( W_j \) in print, or when processing them otherwise, it is common practice to round them to rational proportions of the form \( N_j/n \), for some prescribed integer accuracy \( n \). For instance, in contingency table analysis the weights, often relative frequencies of some raw data, are usually rounded to multiples of a percent \( n = 100 \), or to multiples of a tenth of a percent \( n = 1000 \). Of course, in order that the proportions \( N_j/n \) again form a valid probability vector, the numerators \( N_1, \ldots, N_c \) must sum to \( n \).

The individual rounding of each of the scaled weights \( nW_j \) in a standard fashion to integers \( nW_j \) does not, however, guarantee that these integers achieve the correct total \( n \).

AMS 1991 subject classifications: Primary 60F05, 62E20; Secondary 11K38
Key words and phrases: apportionment methods, divisor method with standard rounding, Lévy-stable distribution, multiplier method with standard rounding, rounding error analysis, rounding of probability vectors, uniform distribution, Webster method
Instead, there may remain a nonzero discrepancy

\[ D_{c,n} = \left( \sum_{j \leq c} (nW_j) \right) - n, \]

see Happacher (2001) and the references given there. For this reason many statistical publications include a salvatory clause that “percentages may not sum to 100 due to rounding errors” or the like.

There are other spheres of life, though, that do not tolerate such a liberal attitude towards rounding errors. Most noticeably this concerns apportionment methods for proportional representation in electoral systems. There, the \( c \) categories signify the political bodies participating in the apportionment process, and the accuracy \( n \) is the number of seats to be apportioned among them. For instance, the \( n = 435 \) seats of the US House of Representatives are apportioned to the \( c = 50 \) States, proportionally to their population. Or the \( n = 598 \) seats in the German Bundestag are apportioned among \( c = 5 \) eligible parties, proportionally to their electoral votes. In the political arena it is plainly not acceptable that an apportionment procedure would terminate with a nonzero discrepancy, leaving some seats unaccounted for or creating surplus seats, “due to rounding errors”. In fact, the field of politics abounds with apportionment methods properly partitioning the total \( n \) into integers \( N_1, \ldots, N_c \). The seminal monograph of Balinski and Young (1982) is an excellent source for the political history of proportional representation, as well as for the mathematical theory of apportionment methods that flows from the historical experience.

One of the early contributors to the subject was the French mathematician André Sainte-Laguë [sɛ̃ lagu] (1882–1950), see the bibliographical note in Marshall, Olkin and Pukelsheim (2002, p. 888). Sainte-Laguë (1910a, b) set out to minimize a Gaussian-type squared-error goodness-of-fit criterion, determining the values \( N_1, \ldots, N_c \) so as to satisfy

\[
\sum_{j \leq c} \left( \frac{N_j - nW_j}{W_j} \right)^2 = \min \left\{ \sum_{j \leq c} \left( \frac{n_j - nW_j}{W_j} \right)^2 : n_1, \ldots, n_c \in \{0, 1, \ldots, n\}, \sum_{j \leq c} n_j = n \right\}.
\]

This approach gives rise to a well-defined apportionment method, the multiplier method with standard rounding. We will call

\[
S_{c,n} = \sum_{j \leq c} \left( \frac{N_j - nW_j}{W_j} \right)^2
\]

the Sainte-Laguë divergence, for \( c \) categories and accuracy \( n \).

Sainte-Laguë (1910a, b) motivated his criterion from the viewpoint of electoral applications. If party \( j \) gains weight \( W_j = V_j/V \), calculated from their vote count \( V_j \) and the vote total \( V = \sum_{i \leq c} V_i \), then

\[
S_{c,n} = \frac{n^2}{V} \sum_{j \leq c} V_j \left( \frac{N_j/n}{V_j/V} - 1 \right)^2
\]
Sainc-Lagué’s chi-square divergence

is the sum, for each of the \( V_j \) voters, of the squared difference between the voter’s realized success values \( \frac{N_j}{V_j} \) and the common ideal success value 1. These numerical quantities directly relate to the qualitative requirements that the German Federal Constitutional Court demands of electoral equality, see Pukelsheim (2000). By mapping qualitative, constitutional principles into a quantitative, operational formula, the Sainte-Lagué divergence \( S_{c,n} \) provides a measure for assessing the goodness of an apportionment method that is particularly persuasive from the viewpoint of constitutional law and political sciences.

Because of the striking similarity of the Sainte-Lagué divergence \( S_{c,n} \) with the \( \chi^2 \)-statistics, the multiplier method with standard rounding may well be the apportionment method that is most appropriate also for contingency table analysis, see Wainer (1998), Pukelsheim (1998). However, the statistical assumptions underlying the \( \chi^2 \)-analysis do not apply. This raises the question as to the distributional properties of \( S_{c,n} \), which is the topic of the present paper.

In Section 2 we review the multiplier method with standard rounding, and set it up for the ensuing asymptotic analysis. Section 3 singles out the case of two categories, because it is particularly transparent. The main step is to show that the rounding residuals \( U_{1,n} = nW_1 - \langle nW_1 \rangle \) are, asymptotically, uniformly distributed and independent of \( W_1 \). Section 4 derives the asymptotic distribution of the Sainte-Lagué divergence \( S_{c,n} \) for a fixed number of categories \( c \geq 3 \), as the accuracy \( n \) tends to infinity:

\[
S_{c,n} \xrightarrow{D} S_c.
\]

Though not of any known type, the distribution of the limiting random variable \( S_c \) can be simulated by computer, as shown in Figure 3.1. Section 5 tackles the asymptotics when the number of categories, \( c \), grows large, assuming that the weight vector \( W \) follows a uniform distribution on the probability simplex \( S_c \). With appropriate scale constants \( a_c \) and shift constants \( b_c \) we obtain convergence,

\[
a_cS_c - b_c \xrightarrow{D} S,
\]

where \( S \) has a Lévy-stable distribution. Figure 5.1 shows that the (simulated) densities of \( a_cS_c - b_c \) converge to the density of \( S \), though rather slowly. In Section 6 we indicate some possible generalizations, such as assuming a Dirichlet distribution \( D(\alpha) \) for \( W \), with form parameter \( \alpha \in (0, 2] \).

2 The multiplier method with standard rounding

The idea underlying Sainte-Lagué’s apportionment method is simple: Given a set, \( w_1, \ldots, w_c \), each weight \( w_j \) is first scaled by some common multiplier \( \mu > 0 \) and then rounded in a standard fashion, \( N_j = \langle \mu w_j \rangle \), where the multiplier \( \mu \) is adjusted so as to achieve the correct total, \( \sum_{j=1}^{c} N_j = n \). Despite its simplicity we need to take a closer look at the method, in order to set it up for our asymptotic analysis.
As a prerequisite we consider the rounding of individual numbers. Standard rounding maps a real number \( x \in \mathbb{R} \) into its nearest integer \( \langle x \rangle \in \mathbb{Z} \) (Abramowitz and Stegun 1970, p. 223; Garfunkel 2003, p. 523). More precisely, if \( x \in [z, z + 1) \) lies in an interval delimited by two neighboring integers, \( z \) and \( z + 1 \), then
\[
\langle x \rangle = \begin{cases} 
    z + 1 & \text{for } x > z + 1/2, \\
    z & \text{for } x \leq z + 1/2.
\end{cases}
\]
Standard rounding enjoys the important stationarity property that \( \langle x + z \rangle = \langle x \rangle + z \), for all \( x \in \mathbb{R} \) and \( z \in \mathbb{Z} \).

There is no general agreement in the literature on where to round the dividing points \( z + 1/2 \). We have chosen to always round down, \( \langle z + 1/2 \rangle = z \). In order to evade a downward trend, one could alternatively round \( z + 1/2 \) to the nearest even integer, see Wallis and Roberts (1956, p. 175), Bronstein and Semendjajew (1991, Sect. 2.1.1.2). Balinski and Young (1982, p. 99) resolve a tie \( x = z + 1/2 \) by rounding \( x \) either up or down, thus generating multiple apportionment solutions. For our subsequent stochastic analysis, the distributional assumptions make ties disappear in Lebesgue nullsets (Happacher and Pukelsheim 1996, p. 378), and hence are of no further concern.

Now we turn to apportionment methods, that is, the rounding of sets of numbers subject to achieving a prespecified total \( n \). Depending on the weights \( w_1, \ldots, w_c \) there exists a multiplier \( \mu > 0 \) such that the individual roundings \( N_j = \langle \mu w_j \rangle \) satisfy \( \sum_{j \leq c} N_j = n \). The multiplier \( \mu \) is thus used as a degree of freedom for the individual roundings to correctly sum to the desired total. In this way any procedure for individual roundings gives rise to a multiplier method of apportionment. The method of Sainte-Laguë is the specific method built on standard rounding.

The definition of standard rounding entails \( N_j - 1/2 \leq \mu w_j < N_j + 1/2 \) (neglecting ties). Therefore the relation between the final apportionment \( N_1, \ldots, N_c \) and an appropriate multiplier \( \mu \) is captured by the max-min inequality
\[
\max_{j \leq c} \frac{N_j - 1/2}{w_j} < \mu < \min_{j \leq c} \frac{N_j + 1/2}{w_j},
\]
see Balinski and Young (1982, p. 100). However, initially neither the numbers \( N_j \) nor a multiplier \( \mu \) are given. Hence all apportionment methods make use of an algorithmic approach of some sort or other.

The following Adjustment Algorithm is tailored to our asymptotic needs. In essence it dates back to the computational scheme proposed by Hagenbach-Bischoff (1905, p. 9), and it is akin to the rank-index methods of Balinski and Young (1982, p. 142), or to the two-step algorithm of Happacher and Pukelsheim (1996, p. 378) and Dorfleitner and Klein (1999, p. 147). Since \( n \) is a reasonable choice to initialize the multiplier (Happacher and Pukelsheim 1996, 2000), this initialization is used to first jump to \( (nw_j) \) and then adjust these numbers if need be. Let \( \text{sgn}(d) = -1, 0, 1 \) according as \( d \) is negative, zero, or positive.

**Lemma 2.1 (Adjustment Algorithm)** For Lebesgue almost all weight vectors \( (w_1, \ldots, w_c) \in S_c \) and for all prespecified totals \( n \), the multiplier method with standard round-
Sainte-Laguë’s chi-square divergence yields a unique apportionment \( (N_1, \ldots, N_c) \), which may be obtained as follows. With initial multiplier \( \mu_0 = n \), first calculate the “discrepancy” \( d = (\sum_{j \leq c} \langle nw_j \rangle) - n \), which is an integer satisfying \(|d| \leq \langle (c-1)/2 \rangle \). Then, for \( j = 1, \ldots, c \), adjust the roundings \( \langle nw_j \rangle \) to obtain

\[
N_j = \langle nw_j \rangle - \text{sgn}(d)m_{j,n}(d),
\]

where \( m_{j,n}(d) \) is the count how often index \( j \) appears among the \(|d|\)-smallest quotients

\[
\frac{k - \text{sgn}(d)\langle nw_i \rangle + 1/2}{w_i}, \quad i = 1, \ldots, c, \quad k = 0, \ldots, |d|.
\]

**Proof:** The event that one of the weights is zero, \( w_j = 0 \), or that, for a given multiplier \( \mu > 0 \), a scaled component \( \mu w_j \) hits a dividing point \( z + 1/2 \), is a Lebesgue nullset, see Happacher and Pukelsheim (1996, p. 378). In all other cases, the Sainte-Laguë apportionment \( N_1, \ldots, N_c \) is unique, and the max-min inequality is strict,

\[
\max_{j \leq c} \frac{N_j - 1/2}{w_j} < \mu < \min_{j \leq c} \frac{N_j + 1/2}{w_j}.
\]

The discrepancy \( d \) is an integer, with values in the interval \([-c/2, c/2]\); when \( c \) is even, the event that \( d \) hits a boundary point \( \pm c/2 \) is a Lebesgue nullset (Happacher 2001, p. 173). Hence the discrepancy \( d \) equals one of the whole numbers \(-\langle (c-1)/2 \rangle, \ldots, \langle (c-1)/2 \rangle\).

In case of a negative discrepancy, \( d < 0 \), the initial multiplier \( \mu_0 = n \) is too small and additional assignments are called for, \( m_{j,n}(d) := N_j - \langle nw_j \rangle \geq 0 \). Upon defining the quotients

\[
Q(k, z, w) = \frac{k+z+1/2}{w},
\]

the max-min inequality gives \( \mu_0 < \min_{i \leq c} Q(0, \langle nw_i \rangle, w_i) = Q(0, \langle nw_i \rangle, w_{i_1}) = \mu_1 \), say. As \( \mu \) grows past \( \mu_1 \), a first additional unit is apportioned to \( i_1 \). Another, second unit is assigned when \( \mu \) next passes the smallest among the quotients \( Q(0, \langle nw_i \rangle, w_i) \) for \( i \neq i_1 \) and \( Q(1, \langle nw_{i_1} \rangle, w_{i_1}) \), etc. In other words, \( m_{j,n}(d) \) is the count how often index \( j \) appears among the \(|d|\)-smallest values, of the quotients \( Q(k, \langle nw_i \rangle, w_i) \) for \( i = 1, \ldots, c \) and \( k = 0, \ldots, |d| \). We obtain \( N_j = \langle nw_j \rangle + m_{j,n}(d) = \langle nw_j \rangle - \text{sgn}(d)m_{j,n}(d) \).

In case of a vanishing discrepancy, \( d = 0 \), no adjustment is needed and \( N_j = \langle nw_j \rangle = \langle nw_j \rangle - \text{sgn}(0)m_{j,n}(0) \).

In case of a positive discrepancy, \( d > 0 \), the multiplier \( \mu_0 = n \) is too large, and we need to remove \( m_{j,n}(d) := \langle nw_j \rangle - N_j \geq 0 \) units. The first unit is removed from \( i_1 \) as soon as \( \mu \) falls below \( \max_{i \leq c} ((\langle nw_i \rangle - 1/2)/w_i = ((\langle nw_i \rangle - 1/2)/w_{i_1}) \), say. It transpires that now \( m_{j,n}(d) \) is the count of how often index \( j \) ranks among the \( d \)-largest values of the quotients \( ((\langle nw_i \rangle - k - 1/2)/w_i) \). Reversing the sign of the quotients, we may equivalently count occurrences among the \( d \)-smallest quotients \( Q(k, -\langle nw_i \rangle, w_i) \). Now we obtain \( N_j = \langle nw_j \rangle - m_{j,n}(d) = \langle nw_j \rangle - \text{sgn}(d)m_{j,n}(d) \). \( \square \)
3 Two categories

We now turn to the stochastic behavior and hence assume that we no longer have a deterministic weight vector \( \mathbf{w} \), but a random weight vector \( \mathbf{W} \). For \( c = 2 \) categories, the multiplier \( \mu_0 = n \) works for almost all weight vectors \( \mathbf{W} = (\mathbf{W}_1, \mathbf{W}_2) \) (Happacher and Pukelsheim, 1996, p. 379). Theorem 3.1 shows that the rounding residuals \( nW_1 - \langle nW_1 \rangle \) are, asymptotically, stochastically independent of \( W_1 \) and uniformly distributed, for a wide class of distributions of \( W_1 \).

**Theorem 3.1** If \( W_1 \) has a Riemann integrable Lebesgue density on \((0, 1)\), then

\[
\left( nW_1 - \langle nW_1 \rangle \right) \xrightarrow{D} n \to \infty \left( U_1 \right),
\]

where the random variable \( U_1 \) is uniformly distributed on \((-1/2, 1/2)\) and independent of \( W_1 \). Moreover, we have

\[
S_{2,n} \xrightarrow{D} n \to \infty S_2 := \frac{U_1^2}{W_1} + \frac{U_1^2}{1 - W_1}.
\]

**Proof:** Let \( f \) be a Riemann integrable density of \( W_1 \). In the joint Fourier transform of \( U_{1,n} = nW_1 - \langle nW_1 \rangle \) and \( W_1 \) we substitute \( nw = u \) to obtain, for \( s, t \in \mathbb{R} \),

\[
Ee^{isU_{1,n} + itW_1} = \int_0^1 \exp \left( is \left( n\langle w \rangle - \langle n\langle w \rangle \rangle \right) + itw \right) f(w) \, dw
\]

\[
= \int_0^n \exp \left( is \left( u - \langle u \rangle \right) + it \frac{u}{n} \right) f \left( \frac{u}{n} \right) \frac{du}{n}
\]

\[
= \int_0^1 \frac{1}{n} \sum_{k=0}^{n-1} \exp \left( is \left( u - \langle u \rangle \right) + it \frac{u + k}{n} \right) f \left( \frac{u + k}{n} \right) \frac{du}{n},
\]

where the last step uses the stationarity property \( u + k - \langle u + k \rangle = u - \langle u \rangle \). Dominated convergence and Riemann integrability entail

\[
\lim_{n \to \infty} Ee^{isU_{1,n} + itW_1} = \int_0^1 \int_0^1 \exp \left( is \left( u - \langle u \rangle \right) + itw \right) f(w) \, dw \, du
\]

\[
= \int_{-1/2}^{1/2} e^{isu} du \cdot \int_0^1 e^{itw} f(w) \, dw
\]

\[
= Ee^{isU_1} \cdot Ee^{itW_1},
\]

since any integrable function \( h \) fulfills

\[
\int_0^1 h \left( u - \langle u \rangle \right) \, du = \int_0^{1/2} h(u) \, du + \int_{1/2}^1 h(u - 1) \, du = \int_{-1/2}^{1/2} h(u) \, du.
\]

Moreover, in \( S_{2,n} \) we substitute \( W_2 = 1 - W_1 \), and \( nW_2 - \langle nW_2 \rangle = \langle nW_1 \rangle - nW_1 = -U_{1,n} \). Convergence of \( S_{2,n} \) to \( S_2 \) now follows from the Continuous Mapping Theorem (Pollard 2002, p. 175). \( \square \)
When \( W_1 \) is uniformly distributed on \((0, 1)\), the density of \( S_2 \) is given by the following, where \( I_A(s) \) is the usual indicator function of the set \( A \).

**Corollary 3.2** If \( W = (W_1, W_2) \) is uniformly distributed on the probability simplex \( S_2 \), then the Sainte-Laguë divergences \( S_{2,n} \) converge to a random variable \( S_2 \) with Lebesgue density \( g_2 \) given by, for \( s > 0 \),

\[
g_2(s) = \frac{\pi}{8\sqrt{s}} - \frac{1}{4\sqrt{s}} \left( \frac{\sqrt{s-1}}{s} + \arctan \frac{\sqrt{s-1}}{s} \right) I_{(1,\infty)}(s).
\]

**Proof:** If \( W \) is uniformly distributed on \( S_2 \), then \( W_1 \) is uniform on \((0, 1)\). We have \( S_2 = U_1^2 / \{(W_1(1 - W_1)) \}. \) The variables \( X = U_1^2 \) and \( Y = W_1(1 - W_1) \) have densities \( f_X(x) = 1/\sqrt{x} \) and \( f_Y(y) = 2/\sqrt{1 - 4y} \) on \((0, 1/4)\). The quotient \( S_2 = X/Y \) then has density \( g_2(s) = \int f_X(sy) f_Y(y) |y| \, dy \) which, after some calculations, takes the form given above.

The density \( g_2 \) is continuous, and is strictly decreasing from infinity at \( s = 0 \) to zero at \( s = \infty \). For large \( s \) it behaves like \( 1/(6s^2) \), that is, \( \lim_{s \to \infty} g(s)/(6s^2) = 1 \). Hence \( S_2 \) has all moments of order less than one finite, while its expectation is infinite. See Figure 3.1.

![Figure 3.1 Limiting density \( g_c \) for \( c = 2, 3, 4 \) categories, of the Sainte-Laguë divergences \( S_{c,n} \) as \( n \to \infty \). The density \( g_2 \) is given in Corollary 3.2, while \( g_3 \) and \( g_4 \) are obtained by simulating one million replicates according to Theorem 4.1. The starting points are \( g_2(0) = \infty, g_3(0) = 0.376 \), and \( g_4(0) = 0 \) for \( c \geq 4 \). The tails for \( s \to \infty \) are of order \( s^{-2} \), whence the moment of order \( p \) is finite if and only if \( p < 1 \).](image)
4 Three or more categories

For \( c \geq 3 \) categories the rounding residuals \( nW_j - \langle nW_j \rangle \) again become independent of \( W_j \) and uniform as \( n \) grows large. Convergence of the Sainte-Laguë divergences \( S_{c,n} \) builds on the Adjustment Algorithm from Lemma 2.1.

**Theorem 4.1** If the weight vector \( W_{(c-1)} := (W_1, \ldots, W_{c-1}) \) has a Riemann integrable Lebesgue density on \( \{(w_1, \ldots, w_{c-1}) \in [0, 1]^{c-1} : \sum_{j \leq c} w_j < 1\} \), then

\[
\left( \begin{array}{c} nW_1 - \langle nW_1 \rangle \\ \vdots \\ nW_{c-1} - \langle nW_{c-1} \rangle \\ W_{(c-1)} \end{array} \right) \xrightarrow{D, n \to \infty} \left( \begin{array}{c} U_1 \\ \vdots \\ U_{c-1} \\ W_{(c-1)} \end{array} \right),
\]

where the random variables \( U_1, \ldots, U_{c-1} \) are uniformly distributed on \((-1/2, 1/2)\), and independent of each other as well as independent of \( W_{(c-1)} \). Moreover, we have

\[
S_{c,n} \xrightarrow{n \to \infty} S_c := \sum_{d = -((c-1)/2)}^{((c-1)/2)} I \left\{ \sum_{j < c} U_j = -d \right\} \frac{\sum_{j < c} (U_j + \text{sgn}(d) m_j(d))^2}{W_j},
\]

where \( U_c := \langle \sum_{j < c} U_j \rangle - \sum_{j < c} U_j \), and \( m_j(d) \) is the count how often index \( j \) appears among the \(|d|\)-smallest quotients

\[
\frac{k - \text{sgn}(d) U_i + 1/2}{w_i}, \quad i = 1, \ldots, c, \quad k = 0, \ldots, |d|.
\]

**Proof:** The first convergence result follows from a multivariate extension of the Fourier transform argument used in the proof of Theorem 3.1.

Convergence of \( S_{c,n} \) is established by re-expressing it in terms of \( W_j \) and \( U_{j,n} := nW_j - \langle nW_j \rangle \) for \( j < c \). We set \( W_c = 1 - \sum_{j < c} W_j \), as usual, and define

\[
U_{c,n} := \left( \sum_{j < c} U_{j,n} \right) - \sum_{j < c} U_{j,n} = nW_c - \langle nW_c \rangle.
\]

The last equality is a straightforward consequence of the stationarity property of standard rounding. Thus \( U_{c,n} \), itself is a rounding residual, of the negative cumulative sum of the rounding residuals \( U_{j,n} \) for \( j < c \), or of \( nW_c \).

In case \( D_{c,n} = d \), the Adjustment Algorithm from Lemma 2.1 yields

\[
nW_j - N_j = U_{j,n} + \text{sgn}(d) m_{j,n}(d),
\]

for all \( j \leq c \), where \( m_{j,n}(d) \) counts the occurrences of index \( j \) among the \(|d|\)-smallest quotients \( Q(k, -\text{sgn}(d) \langle nW_i \rangle, W_i) \), with \( Q(k, z, w) \) defined as in the proof of Lemma 2.1.
The latter equal \( Q(k, \text{sgn}(d) U_{i,n}, W_i) - \text{sgn}(d)n \), from which the common shift \( \text{sgn}(d)n \) may be dropped since it does not affect the ranking. This establishes the representation

\[
S_{c,n} = \sum_{d=-((c-1)/2)}^{((c-1)/2)} I \{ D_{c,n} = d \} \sum_{j \leq c} \frac{(U_{j,n} + \text{sgn}(d)m_{j,n}(d))^2}{W_j}.
\]

By definition, we have \( \langle \sum_{j \leq c} U_{j,n} \rangle = \sum_{j \leq c} U_{j,n} = -D_{c,n} \); see also Diaconis and Freedman (1979, p. 361), Häppacher and Pükelsheim (2000, p. 155). Now an appeal to the Continuous Mapping Theorem completes the proof.

For three categories and uniform weights \( W = (W_1, W_2, W_3) \) on the probability simplex \( S_3 \), the density \( g_3(s) \) of \( S_3 \) may be shown to satisfy

\[
g_3(0) = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \text{Pr} \left( \frac{U_1^2}{W_1} + \frac{U_2^2}{W_2} + \frac{(U_1 + U_2)^2}{1 - W_1 - W_2} \leq \epsilon \right) = \frac{4\pi^2}{105} = 0.376.
\]

From \( c \geq 4 \) categories on the densities start out from \( g_c(0) = 0 \). See Figure 3.1.

In Theorem 4.1, \( U_c = \langle \sum_{j \leq c} U_j \rangle - \sum_{j \leq c} U_j \) is evidently not independent of \( U_1, \ldots, U_{c-1} \), but it shares their distribution by also being uniformly distributed on the interval \((-1/2, 1/2)\). This follows from Theorem 3.1, or by a direct Fourier transform argument using the uniform convolution density formula \( f_{c-1}(v) = \int_{v-1/2}^{v+1/2} f_{c-2}(x) \, dx \):

\[
E e^{it \left( \sum_{j \leq c} U_j \right) - \sum_{j \leq c} U_j} = \sum_{d=-((c-1)/2)}^{((c-1)/2)} \int_{d-1/2}^{d+1/2} e^{it(d-x)} f_{c-1}(x) \, dx = \int_{-1/2}^{1/2} e^{itu} C \, du,
\]

with

\[
C := \sum_{d=-((c-1)/2)}^{((c-1)/2)} f_{c-1}(d-u) = \sum_{d=-((c-1)/2)}^{((c-1)/2)} \int_{d-u-1/2}^{d-u+1/2} f_{c-2}(x) \, dx = \int_{\mathbb{R}} f_{c-2}(x) \, dx = 1.
\]

Section 5 will show that the asymptotic behavior of \( S_c \) is determined by the sum over the leading terms \( U_j^2/W_j \). To this end we provide the following bounds.

**Corollary 4.2** With the assumptions and notations of Theorem 4.1, we have

\[
0 \leq S_c - \sum_{j \leq c} \frac{U_j^2}{W_j} \leq c \left| \sum_{j \leq c} U_j \right|.
\]
Proof: Expanding the squares \( (U_j + \text{sgn}(d)m_j(d))^2 \) we get

\[
S_c - \sum_{j \leq c} \frac{U_j^2}{W_j} = \sum_{d=-((c-1)/2)}^{((c-1)/2)} I \left\{ \left( \sum_{j \leq c} U_j \right) = -d \right\} \sum_{j \leq c} \frac{2\text{sgn}(d)U_j + m_j(d)}{W_j} m_j(d).
\]

All terms in the last sum are nonnegative, which is evident for \( m_j(d) = 0 \), and which for \( m_j(d) \geq 1 \) follows from \( 2\text{sgn}(d)U_j + m_j(d) \geq 0 \). This verifies the lower bound zero.

To establish the upper bound we consider the case \( \sum_{j \leq c} U_j = -d \), and characterize the adjustment counts \( m_j(d) \) using a multiplier \( \rho \), see Balinski and Young (1982, p. 100). That is, one has \( m_j(d) = \rho W_j - \text{sgn}(d)U_j \) for all \( j \leq c \) if and only if

\[
\max_{j \leq c} \frac{\text{sgn}(d)U_j + m_j(d) - 1/2}{W_j} < \rho < \min_{j \leq c} \frac{\text{sgn}(d)U_j + m_j(d) + 1/2}{W_j}.
\]

By construction we have \( \sum_{j \leq c} U_j = -d \) and \( \sum_{j \leq c} m_j(d) = |d| \), whence the right inequality yields

\[
\rho = \sum_{j \leq c} \rho W_j < -\text{sgn}(d)d + |d| + \frac{c}{2} = \frac{c}{2}.
\]

The left inequality is invoked only when \( m_j(d) \geq 1 \), giving

\[
\frac{2\text{sgn}(d)U_j + m_j(d)}{W_j} \leq 2\frac{\text{sgn}(d)U_j + m_j(d) - 1/2}{W_j} < 2\rho < c.
\]

Now \( c \sum_{d=-((c-1)/2)}^{((c-1)/2)} I \left\{ \left( \sum_{j \leq c} U_j \right) = -d \right\} |d| = c|\left( \sum_{j \leq c} U_j \right)| \) completes the proof. \( \square \)

5 Infinitely many categories

The behavior of \( S_c \) for a large number of categories \( c \) is studied under the assumption that, for \( c \) fixed, the weight vector \( W = (W_1, \ldots, W_c) \) is uniformly distributed on the probability simplex \( S_c \). By casting out some negligible terms we approximate the distribution of \( S_c \) by an average over stochastically independent and identically distributed random variables. This does away with the dependencies, of \( W_c \) on \( W_1, \ldots, W_{c-1} \), and of \( U_c \) on \( U_1, \ldots, U_{c-1} \).

Lemma 5.1 Suppose the distribution of \( W \) on \( S_c \) is uniform, and let the random variables \( V_1, E_1, V_2, E_2, \ldots \) be independent such that every \( V_j \) is uniformly distributed on \((-1/2, 1/2)\) and every \( E_j \) is exponentially distributed with mean one. Then the differences

\[
\frac{1}{c^2} S_c - \frac{1}{c} \sum_{j \leq c} \frac{V_j^2}{E_j}
\]

converge in probability to zero, as \( c \) tends to infinity.
Proof: Firstly we get, from Corollary 4.2 and using \( \sum_{j \leq c} U_j \leq (\sum_{j \leq c} U_j) + 1/2 \),
\[
0 \leq \frac{1}{c^2} S_c - \frac{1}{c^2} \sum_{j \leq c} \frac{U_j^2}{W_j} \leq \left| \frac{1}{c} \sum_{j \leq c} U_j \right| + \frac{1}{2c}.
\]
The latter converges in probability to zero, by the law of large numbers. Hence we may approximate \( c^{-2} S_c \) by \( c^{-2} \sum_{j \leq c} U_j^2 / W_j \).

Secondly we replace the dependent variable \( U_j \) by the independent variable \( V_j \). Happacher and Pukelsheim (2000, p. 153) show that \( P(W_c \leq t) = 1 - (1 - t)^{c-1} \). From \(|U_c^2 - V_c^2| \leq 1/4\) we get, for \( \epsilon > 0 \),
\[
\lim_{c \to \infty} P\left( \frac{1}{c^2} \left| \frac{U_c^2}{W_c} - \frac{V_c^2}{W_c} \right| \geq \epsilon \right) \leq \lim_{c \to \infty} P\left( \frac{1}{4c^2 W_c} \geq \epsilon \right) = 1 - \lim_{c \to \infty} \left( 1 - \frac{1}{4c^2} \right)^{c-1} = 0.
\]
Hence we may approximate \( c^{-2} \sum_{j \leq c} U_j^2 / W_j \) by \( c^{-2} \sum_{j \leq c} V_j^2 / W_j \).

Thirdly we represent the uniform distribution on \( S_c \) through \( W_j = E_j / \sum_{i \leq c} E_i \), see Aitchison (1986, p. 59). Thus the random variables
\[
\frac{1}{c^2} \sum_{j \leq c} \frac{V_j^2}{W_j} \quad \text{and} \quad \frac{\sum_{i \leq c} E_i}{c^2} \sum_{j \leq c} \frac{V_j^2}{E_j}
\]
have the same distribution. The average \( c^{-1} \sum_{i \leq c} E_i \) converges in probability to one, by the law of large numbers. Hence we may approximate \( c^{-2} \sum_{j \leq c} V_j^2 / W_j \) by \( c^{-1} \sum_{j \leq c} V_j^2 / E_j \). The proof is complete. \( \square \)

The random variables \( Z_j := V_j^2 / E_j \) live on \([0, \infty)\), and their common density \( f \) satisfies
\[
f(z) = \frac{1}{\sqrt{2}} \int_0^{1/(4z)} e^{-x \sqrt{x}} \, dx = \frac{1}{12 z^2} - \frac{1}{80 z^3} + \frac{1}{2688 z^4} + \cdots
\]
Clearly \( Z_j \) has infinite expectation, and the averages \( c^{-1} \sum_{j \leq c} Z_j \) do not obey the law of large numbers. Instead, \( Z_j \) lies in the domain of attraction of a stable distribution, see Ibragimov and Linnik (1971, p. 76). Let \( S \) be a random variable with Fourier transform given by
\[
E e^{itS} = \exp \left( -|t| - \frac{2it}{\pi} \log |t| \right),
\]
for \( t \in \mathbb{R} \). That is, \( S \) has a Lévy-stable distribution with characteristic exponent \( \alpha = 1 \) and skewness parameter \( \beta = 1 \), see Zolotarev (1986, p. 9). With Euler’s constant \( \gamma = \lim_{k \to \infty} (1 + 1/2 + \cdots + 1/k - \log k) = 0.5772 \cdots \), the limiting distribution of the suitably scaled and shifted statistics \( S_c \) is now given by the following.
Theorem 5.2  Suppose the distribution of $W$ on $\mathcal{S}$ is uniform. Then we have

$$a_c S_c - b_c \xrightarrow{\text{D}} S$$

with scale constants $a_c = \frac{24}{\pi c^2}$ and shift constants $b_c = \frac{2}{\pi} \left( \frac{5}{3} - 2\gamma + \log \frac{\pi c}{6} \right)$.

Proof: In view of Lemma 5.1 and using Slutzky’s Theorem (Pollard 2002, p. 175) it suffices to show that $\frac{24}{\pi c} \left( \sum_{j \leq c} Z_j \right) - b_c$ converges in distribution to $S$. To this end we consider their Fourier transforms

$$\psi_c(t) := \mathbb{E} \exp \left( i \frac{24t}{\pi c} \left( \sum_{j \leq c} Z_j \right) - it b_c \right) = \left( \mathbb{E} e^{isZ_1 - \frac{it}{c} b_c} \right)^c,$$

with $t \in \mathbb{R}$ fixed, and upon setting $s := 24t/(\pi c)$. For a passage to the limit as $c$ tends to infinity, that is, as $s$ tends to zero, we shall establish the following approximation ($\ast$):

$$\mathbb{E} e^{isZ_1} = 1 - \frac{\pi |s|}{24} + \frac{i s}{12} \left( \frac{5}{3} - 2\gamma + \log 4 - \log |s| \right) + O\left( |s|^{3/2} \right) \quad \text{(*)}$$

We multiply by $e^{it\frac{b_c}{c}} = 1 - \frac{it}{c} b_c + o(1/c)$ and collect lower order terms to conclude

$$\lim_{c \to \infty} \psi_c(t) = \lim_{c \to \infty} \left( 1 - \frac{|t| + \frac{2it}{\pi} \log |t|}{c} + o\left( \frac{1}{c} \right) \right)^c = \mathbb{E} e^{is}.$$

In order to establish the approximation ($\ast$), we set $a := 1/\sqrt{s}$ and split the pertinent integral into two:

$$\mathbb{E} e^{isZ_1} - 1 = \int_0^a (e^{isz} - 1) f(z) dz + \int_a^\infty (e^{isz} - 1) f(z) dz =: I_1(a) + I_2(a).$$

In the integral $I_1(a)$, we use the inequality $|e^{isz} - 1 - isz| f(z) \leq s^2/24$, substitute $1/(4z) = y$, and integrate by parts, to obtain the first three of the following equalities:

$$I_1(a) = \int_0^a \sqrt{z} \int_0^{1/(4z)} e^{-x} \sqrt{x} dx \, dz + O(as^2)$$

$$= \frac{is}{8} \int_0^{1/(4a)} y^{-5/2} \int_0^y e^{-x} \sqrt{x} dx \, dy + O(as^2)$$

$$= \frac{is}{12} \left( (4a)^{3/2} \int_0^{1/(4a)} e^{-x} \sqrt{x} dx \right) + O(as^2)$$

$$= \frac{is}{12} \left( \frac{2}{3} - \gamma + \log 4 + \log \frac{|s|}{a} \right) + O\left( \frac{|s|}{a} \right) + O(as^2).$$
Sainte-Lagué’s chi-square divergence

The last equality follows from \(\int_0^{1/(4a)} e^{-x} \sqrt{x} \, dx = (2/3)(4a)^{-3/2} + O(1/a)\), while the exponential integral is evaluated from

\[
\int_{\epsilon}^{\infty} \frac{e^{-x}}{x} \, dx = -\gamma - \log \epsilon + \int_{\epsilon}^{1} \frac{1 - e^{-x}}{x} \, dx = -\gamma - \log \epsilon + O(\epsilon)
\]

for \(\epsilon \in (0, 1]\), see Abramowitz and Stegun (1970, Eq. 5.1.11).

In the integral \(I_2(a)\), the inequality \(|e^{isz} - 1| \leq |s| z\) implies that the remainder

\[
\int_a^{\infty} (e^{isz} - 1)(f(z) - 1/(12z^2)) \, dz
\]

is bounded of order \(|s|/a\). We substitute \(sz = x\) and \(e^{iz} = \cos x + i \sin x\), combine the two cases \(s > 0\) and \(s < 0\) into one, use

\[
\int_0^{\infty} (1 - \cos x)/x^2 \, dx = \pi/2,
\]

and integrate by parts, to obtain the first three of the following equalities:

\[
I_2(a) = \frac{1}{12} \int_a^{\infty} \frac{e^{isz} - 1}{z^2} \, dz + O \left( \frac{|s|}{a} \right)
\]

\[
= -\frac{|s|}{12} \int_{|a||s|}^{\infty} \frac{1 - \cos x}{x^2} \, dx + \frac{is}{12} \int_{|a||s|}^{\infty} \frac{\sin x}{x^2} \, dx + O \left( \frac{|s|}{a} \right)
\]

\[
= -\frac{|s|}{12} \left( \frac{\pi}{2} - \int_0^{a|s|} \frac{1 - \cos x}{x^2} \, dx \right)
\]

\[
+ \frac{is}{12} \left( \frac{\sin(a|s|)}{a|s|} + \int_{a|s|}^{\infty} \frac{\cos x}{x} \, dx \right) + O \left( \frac{|s|}{a} \right)
\]

\[
= -\frac{\pi |s|}{24} + \frac{is}{12} (1 - \gamma - \log a - \log |s|) + O(a^2s^2) + O \left( \frac{|s|}{a} \right).
\]

For the last equality, we use \(1 - \cos x \leq x^2/2\) to obtain \(\int_0^{a|s|} (1 - \cos x)/x^2 \, dx = O(a|s|)\).

From \((1 - (\sin x)/x \leq x^2/6\) for \(x > 0\) we get \((\sin a|s|)/(a|s|) = 1 + O(a^2s^2)\).

And the formulas for the sine integral in Abramowitz and Stegun (1970, Eqs. 5.2.2+27) lead to

\[
\int_{\epsilon}^{\infty} \frac{\cos x}{x} \, dx = -\gamma - \log \epsilon + \int_{0}^{\epsilon} \frac{1 - \cos x}{x} \, dx = -\gamma - \log \epsilon + O(\epsilon^2)
\]

for \(\epsilon \in (0, 1]\), whence \(\int_{a|s|}^{\infty} (\cos x)/x \, dx = -\gamma - \log a - \log |s| + O(a^2s^2)\). In view of \(a = 1/\sqrt{|s|}\) all remainder terms are seen to be bounded of order \(|s|^{3/2}\). Now

\[
1 + I_1(a) + I_2(a)
\]

sums to (*)\), thus completing the proof.

The density of \(S\) can be calculated numerically, see Nolan (1997, 1998). It is known to be positive over the whole real line, smooth, and unimodal, with mode in the negative axis, at about \(-0.435\). See Figure 5.1.
Figure 5.1 Convergence of $a_c S_c - b_c$ to a Lévy-stable law when $c \to \infty$. Theorem 5.2, specifying the scale constants $a_c$ and the shift constants $b_c$, only establishes convergence in distribution, while the figure exhibits convergence of the associated densities. For $c = 5, 20, 100$, the density of $S_c$ is obtained by simulating one million replicates according to Theorem 4.1. The density of the limiting law ($c = \infty$) is calculated using the program of Nolan (1997).

6 Concluding remarks

Our approach suggests generalizations in various directions. Thus we may assume that the random variables $E_j$ from Lemma 5.1 possess a Gamma density with form parameter $\alpha \in (0, 2]$ and scale parameter $1$,

$$\frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x}, \quad x > 0.$$ 

That is, on the simplex $S_c$ the weight vector $W$, with exchangeable components $W_j = E_j/\sum_{i \leq c} E_i$ for $j = 1, \ldots, c$, follows a Dirichlet distribution $D(\alpha)$, see Aitchison (1986, p. 59). Specifically, $\alpha = 1$ is the uniform case discussed in Section 5.

Generalizing Lemma 5.1, we find that now, as $c$ tends to infinity,

$$\frac{1}{\alpha e^{1+1/\alpha}} S_c - \frac{1}{c^{1/\alpha}} \sum_{j \leq c} \frac{V_j^2}{E_j}$$

vanishes in probability when $\alpha \in (0, 2)$, while for $\alpha = 2$ convergence takes place for

$$\frac{1}{2c \sqrt{c \log c}} S_c - \frac{1}{\sqrt{c \log c}} \sum_{j \leq c} \frac{V_j^2}{E_j}.$$
The density $f$ of the random variables $Z_j = V_j^2 / E_j$ becomes

$$f(z) = \frac{1}{\Gamma(\alpha) \sqrt{z}} \int_0^{1/(4z)} e^{-x} x^{\alpha-1/2} dx = \frac{1}{(1 + 2\alpha)^{4\alpha} \Gamma(\alpha) z^{1+\alpha}} + r(z).$$

It remains bounded of order $z^{-1/2}$ as $z$ tends to zero, and the remainder $r(z)$ is bounded of order $z^{-(2+\alpha)}$ as $z$ tends to infinity. For $\alpha > 1$, we obtain $E[Z_1] = 1 / (12(\alpha - 1))$.

For $\alpha \in (0, 2)$, $Z_1$ thus belongs to the domain of normal attraction of a stable distribution with characteristic exponent $\alpha$. For $\alpha = 2$, $Z_1$ lies in the non-normal domain of attraction of the Gaussian law. See Ibragimov and Linnik (1971) for details. After some lengthy calculations, similar to those in the proof of Theorem 5.2, it transpires that the appropriate norming constants turn out to be

$$a_c = \begin{cases} \frac{\Gamma(1 + \alpha) (1 + 2\alpha) (1 - \alpha)}{\Gamma(2 - \alpha) \sin(\pi(1 - \alpha)/2)}^{1/\alpha} \frac{4}{\alpha c^{1+1/\alpha}} & \text{for } \alpha \in (0, 1) \cup (1, 2), \\ \frac{4\sqrt{5}}{c \sqrt{c \log c}} & \text{for } \alpha = 2; \end{cases}$$

$$b_c = \begin{cases} 0 & \text{for } \alpha \in (0, 1), \\ \frac{\alpha c^2 a_c}{12(\alpha - 1)} & \text{for } \alpha \in (1, 2). \end{cases}$$

Then $a_c S_c - b_c$ is found to converge in distribution to the stable law with Fourier transform $\varphi_\alpha$ given by, for $\alpha \neq 1$,

$$\varphi_\alpha(t) = \exp \left(-|t|^\alpha \left(1 - i \text{sgn}(t) \tan \frac{\pi \alpha}{2}\right)\right).$$

Other generalizations are possible. With $f(x) = x^2$ the Sainte-Laguë divergence takes the form

$$S_{n,c} = n^2 \sum_{j \leq c} W_j f \left(\frac{N_j/n}{W_j} - 1\right).$$

This is just one choice out of many for generating meaningful statistical distances, see Liese and Vajda (1987). Other choices would entail a similar analysis as the one presented here, though their bearing on the rounding problem would need to be explored.

Acknowledgements

We are grateful to Stefan Golling for ample simulation evidence in the beginning phase of this research, and to Thomas Klein and the referees for valuable comments.

References


Lothar Heinrich
Institut für Mathematik
Universität Augsburg
86135 Augsburg, Germany
heinrich@math.uni-augsburg.de

Friedrich Pukelsheim
Institut für Mathematik
Universität Augsburg
86135 Augsburg, Germany
pukelsheim@math.uni-augsburg.de

Udo Schwingenschlögl
Institut für Physik
Universität Augsburg
86135 Augsburg, Germany
schwingenschloegl@physik.uni-augsburg.de