THE DISTRIBUTION OF SUMS OF ROUNDED PERCENTAGES

FREDERICK MOSTELLER, CLEO YOUTZ, AND DOUGLAS ZAHN*

RESUMEN

Cuando se calcula porcentajes para recuentos en diversas categorías o para varias medidas positivas, tomando cada una como una fracción de su suma, a menudo, los porcentajes redondeados no suman 100 porciento. Investigamos la frecuencia con que ocurre este error y cuales son las distribuciones de las sumas de los porcentajes redondeados, para (1) un conjunto de datos empíricos; (2) la distribución polinomial en muestras pequeñas; (3) espaciamientos entre puntos ubicados en un intervalo, el modelo de la barra quebrada; y (4) para la simulación de varias categorías. Los diversos métodos producen distribuciones similares.

Hallamos que en promedio, la probabilidad de que la suma de los porcentajes redondeados alcance exactamente a 100 porciento, es evidente para dos categorías; es cerca de tres cuartos para tres categorías; cerca de dos tercios para cuatro categorías; y cerca de $\sqrt{6/c\pi}$ *para un mayor número de categorías c, cuando las categorias no son improbables.*

SUMMARY

When percentages are computed for counts in several categories or for several positive measurements each taken as a fraction of their sum, the rounded percentages often fail to add to 100 percent. We investigate how frequently this failure occurs and what the distributions of sums of rounded percentages are for (1) an empirical set of data, (2) the multinomial distribution in small samples, (3) spacings between points dropped on an interval—the broken-stick model—; and (4) for simulation for several categories. The several methods produce similar distributions.

We find that the probability that the sum of rounded percentages adds to exactly 100 percent is certain for two categories, about three-fourths for three categories, about two-thirds for four categories, and about $\sqrt{6/c\pi}$ *for larger numbers of categories, c, on the average when categories are not improbable.*

SUMS OF ROUNDED PERCENTAGES

In tabulating percentages, even for exhaustive categories, everyone finds that the sums of the rounded percentages often fail to add to 100. For example, the accompanying four-category tabulation of counts was percentaged and rounded off to the nearest 10 percent.

	Category				
	1	2	3	4	Total (percent)
Counts.....	1	3	2	3	9
Rounded percent..	10	30	20	30	90

This failure to add to 100 percent occurs so frequently that, if very many sums of percentages do add to 100 percent in a set of reported tables, one begins to suspect the reporter of fudging. How often should the percentages fail to add up correctly? How wrong are they likely to be?

We investigate these questions (1) by illustrating the distribution of the sums of rounded percentages for a large collection of tables; (2) by examining the distribution of sums for the multinomial distribution, for small samples; (3) by turning to the continuous analogue of the multinomial for large samples—the broken-stick model—and getting exact distributions, means and variances, and asymptotic approximations; and (4) by simulating the broken-stick model for larger numbers of categories.

We find that the probability of the sum of rounded percentages adding to exactly 100 percent is certain for two categories, about 3/4 for three categories, 2/3 for four categories, and $\sqrt{6/c\pi}$ for larger numbers

* Harvard University. This work was facilitated by grants from the National Science Foundation.

of categories—c, on the average, when categories are not improbable.

The details of the probability distribution of sums would, of course, depend upon such things as the number of categories and the fineness of the rounding. Our numerical work is mainly with the standard decimal roundings: to the nearest 10 percent, the nearest 1 percent, or the nearest 0.1 percent. We convert these quantities to proportions and call them the grid spacings Δ, 0.1, 0.01, and 0.001. Once in a while, people round to the nearest 5 percent; $\Delta = 0.05$.

A grid width Δ that divides the interval from 0 to 1 into an integral number of parts yields $n = 1/\Delta$ parts. The $n + 1$ possible values that could occur as a result of rounding to the nearest Δ are 0 percent, 100Δ percent, 200Δ percent, . . . , $100n\Delta$ percent. For example, in a sample of 9 rounded to the nearest 10 percent ($\Delta = 0.1$), a count of 0 would yield 0 percent; a count of 1, 10 percent; of 2, 20 percent; 3, 30 percent; 4, 40 percent; 5, 60 percent; 6, 70 percent; 7, 80 percent; 8, 90 percent; and of 9, 100 percent. Note that although 50 percent was available, it was not used.

What sums are possible? For either one or two categories it is not possible to have a correct sum failing to add to 100 percent exactly. (If a number ends in a 5 that must be rounded, our rule is to round to the nearest even digit, but a high-speed computer often has views of its own about this, since it may produce 2.5 as 2.4999999, depending on just how it carried out the calculation. Consequently, we set this rounding 5's problem aside for the moment and regard it as a rare event not

Table 1.—Empirical Distribution of Sums of Rounded Percentages for Rounding to the Nearest 10 Percent, 1 Percent, and 0.1 Percent, for Each Number of Categories from 3 to 9

Sum of rounded Percentages	Number of categories						
	3	4	5	6	7	8	9
Δ = 10 percent							
Total[a]	100	100	100	100	100	100	100
120	...	...	0.29	...	...	...	...
110	8.97	17.91	18.93	15.90	12.41	15.59	17.65
100	68.79	63.39	58.63	50.87	59.31	49.81	26.47
90	22.24	18.70	21.66	31.21	28.28	31.94	41.18
80	...	...	0.49	2.02	...	2.66	14.71
Δ = 1 percent							
Total	100	100	100	100	100	100	100
102	...	0.16	0.10	0.87	0.00	1.52	0.00
101	14.48	17.27	20.49	17.34	24.83	28.14	35.29
100	72.93	66.40	58.73	58.38	46.90	39.54	42.65
99	12.59	16.01	20.49	21.68	27.59	27.76	22.06
98	...	0.16	0.20	1.73	0.69	3.04	0.00
Δ = 0.1 percent							
Total	100	100	100	100	100	100	100
100.2	...	...	0.29	1.16	2.76	1.90	1.47
100.1	12.59	16.96	20.49	21.10	22.76	21.67	19.12
100.0	74.48	66.40	58.44	53.18	53.10	52.85	48.53
99.9	12.93	16.64	20.59	23.70	20.00	20.15	23.53
99.8	...	...	0.20	0.87	1.38	3.42	7.35
Number of lines.	580	631	1025	346	145	263	68

[a] Column sums may fail to add to 100 percent because of rounding.

worth special attention in this study of average results.)

Sometimes a category practically never occurs. In such cases we suppose that the distribution of the sums of the rounded percentages behaves much as if this category were not present. We have not investigated this point.

One could use rules for rounding other than rounding to the nearest grid point, but we do not investigate these rules here. If a rule is used that forces the totals to add to 100 percent, it is probably well to state it.

EMPIRICAL DISTRIBUTIONS FROM THE NATIONAL HALOTHANE STUDY

To illustrate the behavior of distributions of the sums of the rounded percentages, we computed percentages for a considerable collection of tables from the National Halothane Study[1]—a study of death rates associated with surgery and anesthesia (sponsored by the National Academy of Sciences–National Research Council). These tables included large numbers of lines of counts associated with varying numbers of categories. Some typical variables are reports of numbers dying categorized by class intervals of age, by physical status (a 7-category scale), and by anesthetic administered (5 categories). Tables for surgical patients were constructed by using the same variables. The total for any given line is likely to be in the hundreds or the thousands. Thus the totals are fairly large, a matter whose relevance will be clarified below.

Table 1 shows, at the bottom, the numbers of lines that we used for each number of categories from 3 to 9. The same line was *not* reused, as it could have been—a five-category line could make ten three-category lines. But the same patient may be represented in several lines. We do not think this overlap affects the conclusions.

We converted each line of data to decimal fractions, rounded to the nearest 10 percent, 1 percent, and 0.1 percent, and, for each of these rounding grids, formed the distribution of the sums. Table 1 gives these distributions.

The 100 percent line of each panel of Table 1 shows that, as we expect, increasing numbers of categories generally lead to decreasing percentages of sums that add to exactly 100 percent. The middle line of the bottom panel suggests that when the rounding is very fine for three categories the fraction totaling exactly 100 percent is about 3/4; for four categories, about 2/3; and for five categories, about 7/12. Below, we relate these results to theoretical formulations. The behavior of the empirical rounding error distributions seems fairly regular as a function of the grid width Δ and of the number of categories $k + 1$.

We note that the distributions for rounding to the nearest 1 percent or 0.1 percent are approximately symmetrical. The 10 percent rounding led to distributions that had more lines producing less than 100 percent than lines producing more than 100 percent. We shall return to this asymmetry below.

DISTRIBUTION OF THE SUMS OF ROUNDED PERCENTAGES FOR SMALL SAMPLES FROM THE MULTINOMIAL DISTRIBUTION HAVING EQUALLY LIKELY CATEGORIES

One way to view tables of counts is as arising from observed samples drawn from multinomial distributions. Here we consider the distributions of sums of rounded percentages arising from the multinomial distribution. Recall that if there are equally likely categories and a sample of size N drawn from this distribution, the probability of observing a count of x_i in category i, where $i = 1, 2, \ldots, c$, is given by

$$P(x_1,x_2, \ldots, x_c) = \frac{N!(1/c)^N}{x_1!x_2! \ldots x_c!}, \quad (1)$$

[1] J. P. Bunker, W. H. Forrest, Jr., F. Mosteller, and L. D. Vandam (eds.), *The National Halothane Study. A Study of the Possible Association Between Halothane Anesthesia and Postoperative Hepatic Necrosis* (Report of the Subcommittee on the National Halothane Study of the Committee on Anesthesia, National Academy of Sciences–National Research Council [in press].)

where $\Sigma x_i = N$. To get the final probability of a particular kind of partition, we must further multiply by the number of arrangements A of the partition itself.

For a given number of categories, we can write out all the possible samples of N for such a multinomial distribution, compute the Δ-rounded percentages for each category, and add. Then, associated with each possible sum, we will have a probability which is the sum of the probabilities associated with the samples that gave the sum. Thus we get the exact distribution of the sum.

Table 2 shows the details of the procedure for a particular case, $N = 7$ with four categories. Panel A shows the prototypic partitions of 7 into four categories and their associated probabilities. For example, $P(7,0,0,0) = 7!(1/4)^7/(7!0!0!0!) = 1/4^7$ and $A = 4!/(3!1!) = 4$; therefore, $P \times A = 1/4^6 = 1/4098 \approx 0.00024$. Panel B shows the percentages computed for each partition to two decimals. Panel C shows the percentages rounded to the nearest 1 percent, and their totals. Panel D gives the final distribution of sums. For example, 101 occurs on lines 8 and 11, and their probabilities in panel A add to 0.30762 or 30.76 percent.

Table 3 summarizes the percentages of samples that add to exactly 100 percent for each sample size from $N = 1$–20 and each number of categories from 3 to 6 for our three standard rounding grids. The pattern of the numbers is informative. For

Table 2.—ILLUSTRATION OF MULTINOMIAL CALCULATIONS FOR SAMPLES OF 7 IN 4 CATEGORIES

Panel A						Panel B			
Partitions of N = 7 into c = 4 categories						Percentages to be rounded			
Partition number	Category				$P(x_1,x_2,x_3,x_4)A$	Category			
	1	2	3	4		1	2	3	4
1.....	7	0	0	0	.00024	100.00	0.00	0.00	0.00
2.....	6	1	0	0	.00513	85.71	14.29	0.00	0.00
3.....	5	2	0	0	.01538	71.43	28.57	0.00	0.00
4.....	4	3	0	0	.02563	57.14	42.86	0.00	0.00
5.....	5	1	1	0	.03076	71.43	14.29	14.29	0.00
6.....	4	2	1	0	.15381	57.14	28.57	14.29	0.00
7.....	3	3	1	0	.10254	42.86	42.86	14.29	0.00
8.....	3	2	2	0	.15381	42.86	28.57	28.57	0.00
9.....	4	1	1	1	.05127	57.14	14.29	14.29	14.29
10.....	3	2	1	1	.30762	42.86	28.57	14.29	14.29
11.....	2	2	2	1	.15381	28.57	28.57	28.57	14.29
					1.00000				

Panel C					Panel D	
Percentages rounded to the nearest 1 percent					Distribution of sums	
Partition number	Category				Total	Sum	Percentage of samples occurring
	1	2	3	4			
1.....	100	0	0	0	100	101	30.76
2.....	86	14	0	0	100	100	61.04
3.....	71	29	0	0	100	99	8.20
4.....	57	43	0	0	100		
5.....	71	14	14	0	99		
6.....	57	29	14	0	100		
7.....	43	43	14	0	100		
8.....	43	29	29	0	101		
9.....	57	14	14	14	99		
10.....	43	29	14	14	100		
11.....	29	29	29	14	101		

example, $N = 9$ generates a good deal of rounding error for the standard grids. Note also that some otherwise nice numbers like $N = 4$ start off poorly for the wider grids before producing sums of exactly 100 percent for all finer standard decimal grids.

Beyond the pattern of results displayed in Table 3, we can look at the average behavior over the first twenty values of N as shown in Table 4. Turning at once to the finest of the standard grids, we see that the average percentages of exactly 100 percent sums is again nearly 3/4 for 3 categories, 2/3 for 4, and 7/12 for 5, as it was for the empirical data of the National Halothane Study. It is interesting that such small N's as 20 and less give averages similar to those obtained for very large N's. As before, we note that increasing the number of categories reduces the average percentage adding to exactly 100 percent.

Space precludes showing all the distributions of sums of rounded percentages, but Table 5 gives the rounded results for $c = 3$ and 5 for those N's for which all partitions do not add to 100 percent, for $\Delta = 0.001$. The bottom line of this table shows the average for the first twenty N's, not just the tabled ones, and we see that for three categories the distribution is asymmetric around 100 percent with about $\frac{1}{6}$ falling at 100.1 and $\frac{1}{12}$ falling at 99.9. We do not expect this asymmetry for large values of N. Furthermore, we note that this asymmetry goes in a direction opposite to that for the halothane data where the sample sizes are very large.

Table 3.—PERCENTAGES OF MULTINOMIAL SAMPLES HAVING SUMS OF ROUNDED PERCENTAGES TOTALING EXACTLY 100 PERCENT, FOR SAMPLE SIZES 1 TO 20; 3 TO 6 CATEGORIES AND THREE ROUNDING GRIDS

N \ c	Δ = .1				Δ = .01				Δ = .001			
	3	4	5	6	3	4	5	6	3	4	5	6
1.....	100	100	100	100	100	100	100	100	100	100	100	100
2.....	100	100	100	100	100	100	100	100	100	100	100	100
3.....	78	62	52	44	78	62	52	44	78	62	52	44
4.....	26	16	10	7	100	100	100	100	100	100	100	100
5.....	100	100	100	100	100	100	100	100	100	100	100	100
6.....	75	71	62	52	75	71	62	52	75	71	62	52
7.....	75	51	34	23	65	61	63	62	52	28	17	11
8.....	53	45	47	52	25	13	7	4	100	100	100	100
9.....	43	20	10	5	43	20	10	5	43	20	10	5
10.....	100	100	100	100	100	100	100	100	100	100	100	100
11.....	37	14	6	3	37	14	6	3	37	14	6	3
12.....	69	41	19	9	78	70	62	57	78	70	62	57
13.....	77	61	48	32	70	59	55	53	31	10	4	1
14.....	74	70	56	49	64	67	47	27	75	73	62	47
15.....	78	70	63	58	78	70	63	58	78	70	63	58
16.....	72	51	43	39	63	48	34	28	25	12	6	3
17.....	71	66	60	55	66	59	29	13	81	51	45	48
18.....	75	74	62	47	75	74	62	47	75	74	62	47
19.....	75	50	31	19	75	63	51	39	75	62	49	43
20.....	61	54	46	34	100	100	100	100	100	100	100	100

Table 4.—AVERAGE PERCENTAGE SUMMING TO EXACTLY 100 PERCENT

Average over N =	Categories			
	3	4	5	6
1(1)10				
Δ = 0.1.....	75	66	62	58
0.01.....	79	73	69	67
0.001.....	85	78	74	71
1(1)20				
Δ = 0.1.....	72	61	52	46
0.01.....	75	68	60	55
0.001.....	75	66	60	56

Table 5.—DISTRIBUTION OF SUMS OF ROUNDED FREQUENCIES FOR $\Delta = 0.001$ FOR N's NOT ALWAYS YIELDING SUMS OF 100 PERCENT, THROUGH N = 20, FOR THE MULTINOMIAL WITH 3 AND 5 CATEGORIES

N	3 categories			5 categories				
	99.9	100	100.1	99.8	99.9	100.0	100.2	100.2
3	22	78			48	52		
6	12	75	12		6	62	33	
7		52	48			17	83	
9	57	43			90	10		
11		37	63			6	94	
12	11	78	11		19	62	18	
13		31	69			4	96	
14	7	75	18	0[a]	32	62	6	
15	11	78	11		19	63	18	
16		25	75			6	63	31
17	14	81	5	5	43	45	7	
18	20	75	5		6	62	31	0
19	12	75	13		5	49	43	3
Average for 20	8.3	75.2	16.7	0.2	13.4	60.0	24.6	1.7

[a] Blanks mean nothing occurred, 0 means the entry rounded to zero.

THE BROKEN-STICK MODEL AND ITS RELATION TO THE MULTINOMIAL

Insofar as we use the exact multinomial model for the rounding error of percentages, we are limited by what we can do on the computer, and of course, $N!$ soon outruns the capabilities of even the largest computer. To pursue larger values of N, we have to turn to other theory. In large samples there is an intimate relation between the distribution of the spacings between the order statistics in a sample from a unit uniform distribution and the distribution of the observed proportions in the equally-likely-category multinomial.[2] Let us describe the relation.

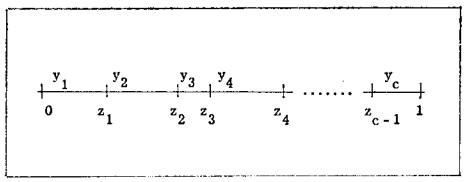

FIG. 1.—Illustrating the relation of the order statistics z_i to the interval lengths y_i.

When a sample of $c - 1$ points is dropped at random on the unit interval, these points, together with the two endpoints, can be used to form c adjacent intervals. If we label the dropped points, after their values have been ordered as $z_1, z_2, \ldots, z_{c-1}$, then the intervals will have lengths $y_1, y_2, \ldots, y_c$, as shown in Figure 1. Note that $y_1 = z_1$, $y_i = z_i - z_{i-1}$, $i = 2, \ldots, c - 1$, and $y_c = 1 - z_{c-1}$. The z's are called the order statistics of the sample. The y's must add to 1, just as the unrounded percentages in the multinomial problem must always add to 100 percent. Then y_i plays the role of a proportion in a category in the *equally likely* multinomial problem.

If x_1/N is the proportion of observations in the first category of a c-category multinomial, for example, then for large N its distribution is approximately that of the length of the leftmost interval created when a sample of $c - 1$ observations is drawn from the unit uniform. Further, x_i/N is distributed like that of the length of the ith interval, and the joint distribution of all the x_i/N is approximately that of the joint distribution of the lengths of the intervals $y_1, y_2, \ldots, y_c$. When N grows large, this distribution theory for the y's matches that for the x_i/N's very closely. However, some specific N's such as 10,000, will be commensurate with standard decimal Δ's, and our rounding error theory will be inaccurate for such values of N. We neglect this in what follows. We are discussing large N's "on the average."

With this relationship between the order statistics and the multinomial established we can push a bit further.

From the geometric theory of the order statistics, we have worked out the exact distribution of the sums of the rounded percentages (rounded y_i's expressed as percentages) for two, three, and four categories.

For two categories the theory is easy, because under our rule all sums add to 100 percent exactly. For three categories, matters are a bit harder, but the analysis only requires breaking up an equilateral triangle into many smaller ones and counting them. The result is simple. Let $n = 1/\Delta$, as before, then we have the accompanying tabulation.

Sum	Probability
$100 + \Delta$	$\frac{n-1}{8n}$
100	$\frac{6n}{8n} = \frac{3}{4}$
$100 - \Delta$	$\frac{n+1}{8n}$

For four categories, the geometry becomes three-dimensional and involved. The distribution of the sums is shown in the accompanying tabulation.

[2] J. W. Tukey, "Non-parametric Estimation II. Statistically Equivalent Blocks and Tolerance Regions—Continuous Case," *The Annals of Mathematical Statistics*, XVIII (1947), 529–39.

Sum	Probability
$100 + \Delta$.....	$\frac{2n(2n-1)(2n-2)}{48n^3} = \frac{(2n-1)(2n-2)}{24n^2}$
100.....	$\frac{4(2n+1)(2n)(2n-1)}{48n^3} = \frac{4(4n^2-1)}{24n^2}$
$100 - \Delta$.....	$\frac{(2n+2)(2n+1)(2n)}{48n^3} = \frac{(2n+2)(2n+1)}{24n^2}$

We have not carried the exact large-sample theory beyond four categories at this time.

Note, first, that sums larger than 100 percent are less likely than sums smaller than 100 percent, agreeing with the halothane data rather than with the average for the small N multinomials. Second, note that for large n (small Δ) the probability of a sum being exactly 100 percent is 1 for two categories, 3/4 for three categories, and tends to 2/3 for four categories. These results are gratifyingly close to those given by the halothane data and by the small-sample multinomial averages for a fine grid.

This theory also models the percentaging and rounding of continuous measurements, such as proportions of protein, fats, and carbohydrates in a diet.

MEAN AND STANDARD DEVIATION

Although we found this large-sample theory difficult to press beyond $c = 4$, it is tedious but fairly straightforward to obtain the mean and variance of these distributions. Let μ_{sum} be the mean of the sum of the rounded intervals and μ_I be the mean rounded percentage for a single interval, then

$$\mu_{\text{sum}} = c\mu_I = c\sum_{i=1}^{n}[1 - (i - \tfrac{1}{2})\Delta]^{c-1}, \quad (2)$$

and for $c \geq 3$, the variance of the sum is

$$\sigma^2_{\text{sum}} = 2c\sum_{i=1}^{n} i[1 - (i - \tfrac{1}{2})\Delta]^{c-1} + (c-1)c\sum_{i=1}^{n-1} i(1 - i\Delta)^{c-1} - \mu_{\text{sum}} - \mu^2_{\text{sum}}. \quad (3)$$

For $c = 1$ or 2 the variance is zero.

Limiting case. For fixed c, as the grid gets fine ($\Delta \to 0$), it can be shown that the mean of the sum of the rounded values tends to n in units of Δ, or to 1 in $n\Delta$. The variance tends to $c/12$ in units of Δ^2. This result is fairly reasonable if we assume that the rounding errors are uncorrelated and uniformly distributed over an interval of length Δ for each of the c pieces. Using the normal approximation for small Δ gives

$$P(\text{sum} = 1) \approx 1/\sqrt{2\pi c/12} \approx 1.38/\sqrt{c}. \quad (4)$$

For three categories, this gives 0.798 instead of 0.75; for four categories, 0.69 instead of 0.67, and the approximation improves with increasing values of c.

It is rather interesting from the point of view of conjecture that the distribution of the sum of two *unit* uniforms has 3/4 of its probability within 1/2 a unit of its mean, thus corresponding to the case of $c = 3$. And the sum of three independent unit uniforms has, under its middle unit interval, the probability 2/3, thus corresponding to the case $c = 4$.

This raises the question whether the limiting probabilities could be given exactly by distributions computed from the sum of $c - 1$ independent uniforms. Z. W. Birnbaum[3] quotes Laplace as having discovered the areas to the left of z under the distribution of the sum of u uniform ran-

[3] Z. W. Birnbaum, "On Random Variables with Comparable Peakedness," *The Annals of Mathematical Statistics*, XIX (1948), 80.

dom variables, each on the interval -1 to 1 (not unit uniforms) to be given by

$$F(z) = \frac{1}{u!} \sum_{i \le (z+u)/2} (-1)^i \binom{u}{i} \times \left(\frac{z+u}{2} - i\right)^u. \quad (5)$$

For our problem $z = -1$, and the estimate is $1 - 2F(z)$. This formula breeds denominators that match those of the exact theory given earlier, $c = 2:8 = 2!2^2$, $c = 3:48 = 3!2^3$, and the next from the Laplace formula is $4!2^4$. In any case, as the number of categories grows the normal approximation based on c categories, and the Laplace numbers are bound to grow close to one another, since the sums of the uniforms are tending to the normal. As c grows large, the discrepancy 1/12 in variance makes less numerical difference. Thus the Laplace numbers, being exactly correct for small c (2, 3, 4) and correct asymptotically, cannot go far astray.

SIMULATION AND COLLATION

To check the results for the broken-stick model and to look into larger values of c, we ran 1,000 samples in which we drew randomly $c - 1$ observations from a uniform, thus generating c values of y; we Δ-rounded these and summed, and obtained the frequency distribution of totals. For this study we used $c = 4, 6, 8, 10, 25$. For $c = 25$, $\Delta = 0.001$, we observed a variance of $2.013\Delta^2$ whose coefficient is very close to the theoretical $25/12 = 2.0833$.

Table 6 gives the various estimates of the probability for the sum of the rounded percentages adding to exactly 100 percent. Both the Laplace result and the normal approximation are close to the simulated value for $c = 6, 8, 10, 25$, and for order of magnitude, either is adequate. All results give compatible probabilities. We could, if we wished, explore the distribution further, but the normal approximation should be adequate. The agreement would be improved still further if the normal approximation were moved one unit down in Table 6, except for the $c = 25$, which would be replaced by 0.276. The column would then agree closely with $\sqrt{6/c\pi}$. Apparently the latter formula handles the flatter shape of the true distribution for small c better than the normal we fitted does.

Table 6.—TRUE AND ESTIMATED PROBABILITIES OF THE SUM OF THE ROUNDED PERCENTAGES ADDING TO EXACTLY 100 PERCENT, FOR VARIOUS NUMBERS OF CATEGORIES FOR THE BROKEN-STICK MODEL

Number of categories	True	Normal approximations[a]	Laplace number[b]	Simulation	Halothane data[c]	Multinomial[d]
1.....	1	0.917	....	....	....	....
2.....	1	0.779	1	....	....	....
3.....	0.75	0.683	0.75	....	0.745	0.75
4.....	0.667	0.614	0.667	0.645	0.664	0.66
5.....	....	0.561	0.599	....	0.584	0.60
6.....	....	0.521	0.550	0.536	0.532	0.56
7.....	....	0.487	0.511	....	0.531	....
8.....	....	0.460	0.479	0.490	0.528	....
9.....	....	0.436	0.453	....	0.485	....
10.....	....	0.416	0.430	0.410	....	....
25.....	....	0.271	0.275	0.273	....	....

[a] Using integral with correction term.

[b] Based on c-1 categories.

[c] $\Delta = 0.001$.

[d] Average for first 20 N's, $\Delta = 0.001$.