THE 4 × 4 CASE

With a deck of sixteen cards composed of 4 suits of 4 cards each, the number of possible runs is known to be 63,063,000. The corresponding frequency table (computed by the same direct elementary process) is given as Table B.

TABLE B

Score	Freq.	Rel. freq.
0	748,521	.011,869,42
1	3,662,976	.058,084,39
2	8,607,744	.136,494,36
3	12,880,512	.204,248,32
4	13,731,616	.217,744,41
5	11,042,688	.175,105,66
6	6,928,704	.109,869,56
7	3,458,432	.054,840,91
8	1,395,126	.022,122,73
9	453,888	.007,197,37
10	122,016	.001,934,83
11	25,344	.000,401,88
12	4,824	.000,076,50–
13	512	.000,008,12
14	96	.000,001,52
15	0	.0
16	1	.000,000,02
	63,063,000	1.000,000,00

The mean is 4, and the standard deviation is $\sqrt{(3.2)}$, or 1.788,854,38.

THE 5 × 5 CASE

The case of greatest current interest concerns a deck of twenty-five cards composed of 5 suits of 5 cards each. With such a deck the number of possible runs is 623,360,743,125,120. The labor involved in a direct computation of the exact frequencies in this case appears to be prohibitive. By the use of more advanced methods, however, a solution for the 5 × 5 case has been worked out by my colleague, Dr. T. E. Sterne, and is presented in the following note.

EDWARD V. HUNTINGTON
DEPARTMENT OF MATHEMATICS
HARVARD UNIVERSITY

THE SOLUTION OF A PROBLEM IN PROBABILITY

IF a deck of 25 playing-cards, composed of 5 suits of 5 cards each, is shuffled at random and then compared with any fixed sequence of a similar deck, then the number, r, of coincidences of suit can take any value, with the exception of 24, from 0 to 25 inclusive. The theoretical probability of r, which I call $p(r)$, is a quantity of some current interest. The simple Binomial Law is not strictly applicable, since the coincidences are not independent.

By considering the probabilities in a 5 × 5 contingency table, I have computed the exact values of $p(r)$ for a few large values of r. They are:

r	$p(r)25!/(5!)^5$
21	44125
22	2500
23	250
24	0
25	1

I see no way to extend the exact computations practically to much smaller values of r. It is possible, however, to compute the moments of the distribution $p(r)$ exactly. The mean of r I find to be 5, and the second, third and fourth moments about the mean to be 25/6, 125/46 and 3625/69, respectively. All these values are exact. The distribution is thus of Pearson's Type I, and I have fitted a Type I curve to the above mean and moments with the results given in the second column of Table 1.

TABLE 1

r	$p(r)$	Binomial
0	.0043	.0038
1	.0247	.0236
2	.0730	.0708
3	.1374	.1358
4	.1852	.1867
5	.1919	.1960
6	1597	.1634
7	.1096	.1108
8	.0633	.0624
9	.0311	.0294
10	.0131	.0118
11	.0047	.0040
12	.0014	.0012
13	.0004	.0003
14	.0001	.0001
15	.0000	.0000
..	....	....
25	.0000	.0000

The values given by the Binomial Law appear in the third column, and appear to be fairly good approximations to the true values. The Binomial Law has the correct mean, but its second, third and fourth moments, 4, 2.4 and 48.16, are all somewhat too small. The Binomial Law would be exact were the coincidences of suit independent.

The values given in the second column above are expected to be correct to within a few units in the last place. Pearson Type I curves were fitted to the moments of the distributions of r in the 3 × 3 and 4 × 4 cases, and then compared with the exact values $p(r)$ computed for those cases by Professor Huntington. The accuracy of the Pearson curve in the 4 × 4 case was much greater than in the 3 × 3 case, and is shown in the comparison given in Table 2.

TABLE 2

r	True $p(r)$	Type I $p(r)$
0	.0119	.0112
1	.0581	.0570
2	.1365	.1374
3	.2042	.2059
4	.2177	.2176
5	.1751	.1741
6	.1099	.1095
7	.0548	.0552
8	.0221	.0224
9	.0072	.0073
10	.0019	.0019
11	.0004	.0004
12	.0001	.0001
13	.0000	.0000
..	....	....
16	.0000	.0000

Since the accuracy of the Type I approximation increases rapidly with the size of the deck, the values given in the second column for the 5 × 5 case are expected to be correct to within a very few units in the last place. Even in the 4 × 4 case, the accuracy of the Pearson Type I approximation is already so great that some 170,000 trials (a trial being an observation of *r*) would be required before one could distinguish with moderate certainty, from the observations, between the set of true probabilities and the direct Pearson approximation. To distinguish observationally between the true values and the binomial values, in the 5 × 5 case, would require some 30,000 trials.

T. E. STERNE
HARVARD COLLEGE OBSERVATORY

SCIENTIFIC APPARATUS AND LABORATORY METHODS

THE MEASUREMENT OF TURBIDITY IN COLORED MEDIA

THE measurement of turbidity photometrically is based on the assumption that there is a linear relationship between the amount of suspended matter and the intensity of the Tyndall beam. Although this holds moderately well within narrow limits when the sample is colorless and faintly turbid, the relationship between the amount of turbidity and the Tyndall beam intensity is much more complex at higher concentrations of suspended material as well as when coloring matter is present in solution.

One of the most disturbing factors is the concomitant weakening of the intensity of the Tyndall beam by absorption. When the suspending liquid is also colored, the absorption effect may be so great that it overbalances the scattering to an extent such that increasing turbidity may give a lower intensity of the Tyndall beam.

According to Sauer,[1] under these conditions the absolute turbidity is given by the following equation:

$$\text{Absolute Turbidity} = A \times f(k) \times D \times t \qquad (1)$$

where

A = Relative intensity of the Tyndall beam as measured with the Pulfrich photometer.
D = A factor varying with the thickness of the absorption cell.
t = The absolute turbidity of the glass standard of the instrument.

If d = Depth of layer of solution in centimeters, and k the coefficient of extinction

then
$$f(k) = \frac{kd(\sqrt{2}-1)\ \log_e 10}{10^{-kd}\{1-10^{-kd(\sqrt{2}-1)}\}} \qquad (1a)$$

In previous work the writers[2,3,4] have shown that the two empirical equations:

$$-\log T = C + N \qquad (2)$$

$$N = \frac{R\,b^C}{a} \qquad (3)$$

hold for turbidity and color determinations for commercial raw cane sugar solutions as well as for white sugars.

C = Coloring matter expressed as − log T.
N = Turbidity expressed as − log T.
R = Intensity of the Tyndall beam = 100 A of Sauer.
T = Per cent. transmittancy.
(− log T) = 2 − log T.
a and b are constants.

We know now that the term b^C in equation (3) is equal to f(k), and that Sauer's equation is valid if C is substituted in place of the kd term in (1a). Hence the turbidity of a colored solution can be calculated either in terms of the absolute turbidity of Sauer, or else in the system used by the writers. Naturally the two methods are interrelated.

The final relationships are given by the following equations:

$$C = -\log T - \frac{R\,f(k)}{a} \qquad (4)$$

$$f(k) = \frac{\left[-\log T - \left(\frac{Rf(k)}{a}\right)\right](\sqrt{2}-1)\ \log_e 10}{10^{-\left(-\log T - \left(\frac{Rf(k)}{a}\right)\right)}\left\{1-10^{-\left[-\log T-\left(\frac{Rf(k)}{a}\right)\right](\sqrt{2}-1)}\right\}} \qquad (5)$$

When C is zero, f(k) is equal to 1.

Equation (5) is solved for varying values of C, which is the $-\log T - \frac{R\,f(k)}{a}$ term, and a table of corresponding f(k) values is thus obtained. Substitution of these C and f(k) values and increments of R into equation (4) yields a chart from which C is readily found from the − log T and the R values. Practically, C is found from curves based on the chart. The application to cane sugar solutions will be published elsewhere in the near future.

F. W. ZERBAN
LOUIS SATTLER
NEW YORK SUGAR TRADE LABORATORY,
NEW YORK CITY

---

[1] Sauer, *Zeits. techn. Phys.*, 12: 149, 1931. See also Landt and Witte: *Zeits. Ver. deut. Zuckerind.*, 84: 462, 1934.

[2] Zerban, Sattler and Lorge, *Ind. Eng. Chem. Anal. Ed.*, 6: 178, 1934.

[3] Zerban, Sattler and Lorge, *Ibid.*, 7: 157, 1935.

[4] Zerban and Sattler, *Ibid.*, 9: 229, 1937.