JOURNAL OF COMBINATORIAL THEORY 3, 279-282 (1967)

Extensions of the Birthday Surprise

M. S. KLAMKIN AND D. J. NEWMAN*

Scientific Laboratory, Ford Motor Company, Dearborn, Michigan 48121

Communicated by Mark Kac

ABSTRACT

The so-called "birthday surprise" is the fact that, on the average, one need only stop about 24 people at random to discover *two* who have the same birthday. Here we determine, asymptotically, the expected number of people in order for n of them to have the same birthday. In particular, for three birthdays, it is about 83 people.

The so-called "birthday suprise" is the fact that, on the average, one need only stop about 24 people to discover *two* who have the same birthday. The computations used to establish this fact are quite simple because the underlying probabilities are so easily obtained. But let us ask for the number of people we must stop, on the average, in order to discover *three* who have the same birthday.[1] Here the probabilities become very complicated and the computations unwieldy. Our purpose, in this note, is to supply an apparently new idea which permits this and analogous computations to be carried out easily.

In general, we will investigate the following problem: Given are n equally likely alternatives. We choose from among them repeatedly until we find that one of the alternatives has occurred k times. Our purpose is to find $E(n, k)$, the *expected* number of repetitions necessary for this success. (In particular, the birthday paradox number is $E(365, 2)$ which is around 24, and the question we asked above was to find $E(365, 3)$).

The method we utilize is an outgrowth of a method developed by Newman and Shepp [1] and is as follows:

If we identify the letters x_i with the n alternatives of the experiment, then the terms of the expansion of $(x_1 + x_2 + \cdots x_n)^N$ can be thought of

* And also Yeshiva University.

[1] The answer is 83 as we will see.

as the alternatives when the experiment is carried out N times. If, for example, $n = 3$, $N = 2$,

$$(x_1 + x_2 + x_3)^2 = x_1x_1 + x_1x_2 + x_1x_3 + x_2x_1 + x_2x_2 + x_2x_3 \\ + x_2x_1 + x_3x_2 + x_3x_3$$

and these nine terms represent the nine possibilities of two trials (e.g., x_2x_3 means alternative 2 followed by alternative 3, x_1x_1 means alternative 1 followed again by alternative 1 and so on).

Now introduce the "truncating" operation T_k. This operator when applied to a polynomial, or power series, has the effect of removing any term which contains any variable raised to a power which is $\geqslant k$, (e.g., T_1 simply gives the constant term, T_2 gives the constant, linear and xy terms, etc.). Then

$$T_k\{(x_1 + x_2 + \cdots x_n)^N\}$$

represents all the possible occurrences for N experiments such that no alternative has appeared k or more times. If, in this polynomial, the letters are then all replaced by the number $1/n$ we obtain, exactly, the *probability* that no alternative has appeared k or more times. Thus,

$$T_k\{(x_1 + x_2 + \cdots x_n)^N\}]_{\frac{1}{n},\frac{1}{n},\ldots\frac{1}{n}}$$

is equal to the *failure* probability after N trials.

Now recall that the expected number of trials is exactly equal to the sum of these failure probabilities. We thus derive a formula, so far useless, for our desired expectancy, namely,

$$E = \sum_{N=0}^{\infty} T_k\{(x_1 + x_2 + \cdots x_n)^N\}]_{\frac{1}{n},\frac{1}{n},\ldots,\frac{1}{n}} .$$

We can transform this formula into a useful one by the following device: Consider, instead of the above series, the series

$$F(t) = \sum_{N=0}^{\infty} T_k\{(x_1 + x_2 + \cdots x_n)^N\} \frac{t^N}{N!} .$$

We then have

$$\begin{aligned} F(t) &= T_k \left\{ \sum_{N=0}^{\infty} (x_1 + x_2 + \cdots x_n)^N \frac{t^N}{N!} \right\} \\ &= T_k\{e^{(x_1+x_2+\cdots x_n)t}\} \\ &= T_k\{e^{x_1t}\}\, T_k\{e^{x_2t}\} \cdots T_k\{e^{x_nt}\} \\ &= S_k(x_1t)\, S_k(x_2t) \cdots S_k(x_nt), \end{aligned}$$

where $S_k(x)$ is the k-th partial sum of e^x, i.e.,

$$S_k(x) = \sum_{j<k} \frac{x^j}{j!}.$$

This gives an explicit formula for $F(t)$. But the formula

$$1 = \int_0^\infty \frac{t^N}{N!} e^{-t}\, dt$$

allows us to get rid of the $t^N/N!$. The result is

$$\sum_{N=0}^{\infty} T_k\{(x_1 + \cdots x_n)^N\} = \int_0^\infty F(t)\, e^{-t}\, dt = \int_0^\infty S_k(x_1 t) \cdots S_k(x_n t)\, e^{-t}\, dt$$

and finally setting all the $x_i = 1/n$, our formula gives us

THEOREM 1.

$$E(n, k) = \int_0^\infty \left[S_k\left(\frac{t}{n}\right)\right]^n e^{-t} dt.$$

This is our "handy" formula and we shall now use it to obtain good asymptotic estimates for E. We will prove

THEOREM 2. *For each fixed k, we have*

$$E(n, k) \sim \sqrt[k]{k!}\, \Gamma\left(1 + \frac{1}{k}\right) n^{1-1/k} \qquad \text{as} \quad n \to \infty.$$

This asymptotic formula gives for the original birthday problem

$$E(365,2) \approx \sqrt{2}\, \Gamma(\tfrac{3}{2}) \sqrt{365} \approx 23.94.$$

For the new birthday problem $E(365, 3)$, we obtain

$$\sqrt[3]{6}\, \Gamma\left(\frac{4}{3}\right) 365^{2/3}$$

which is roughly 82.87.

To obtain the asymptotic formula in Theorem 2, set $t = n^{1-1/k}u$ in the integral formula, Theorem 1. The result is

$$E = n^{1-1/k} \int_0^\infty \{S_k(un^{-1/k})\, e^{-un^{-1/k}}\}^n\, du.$$

Now

$$S_k(x)\,e^{-x} = \frac{1 + x + \dfrac{x^2}{2!} + \cdots + \dfrac{x^{k-1}}{(k-1)!}}{1 + x + \cdots + \dfrac{x^{k-1}}{(k-1)!} + \dfrac{x^k}{k!} + \cdots}$$

$$= \frac{1}{1 + \dfrac{x^k}{k!} + \text{higher terms}}$$

$$= 1 - \frac{x^k}{k!} + \text{higher terms}$$

so that

$$\{S_k(un^{-1/k})\,e^{-un^{-1/k}}\}^n \to e^{-u^k/k!} \qquad \text{as} \quad n \to \infty.$$

Furthermore,

$$S_k(x) \cdot \left(1 + \frac{x^k}{k!}\right) = 1 + x + \cdots + \frac{x^k}{k!} + \frac{x^{k+1}}{k!\,1!} + \cdots \leqslant e^x,$$

so that

$$S_k(un^{-1/k})\,e^{-un^{-1/k}} \leqslant \frac{1}{1 + u^k/nk!}\,.$$

Thus we obtain

$$\{S_k(un^{-1/k})\,e^{-un^{-1/k}}\}^n \leqslant \frac{1}{1 + u^k/k!}\,,$$

and this tells us that our integral is dominatedly convergent. We may thus take the limit inside the integral sign and obtain

$$\lim_{n\to\infty} \frac{E(n,k)}{n^{1-1/k}} = \int_0^\infty e^{-u^k/k!}\,du.$$

Finally, changing $u^k/k!$ into v, the above right-hand side becomes the usual Γ-function integral and we have

$$\lim_{n\to\infty} \frac{E(n,k)}{n^{1-1/k}} = \sqrt[k]{k!}\,\Gamma(1 + 1/k).$$

as required.

Reference

1. D. J. Newman and L. Shepp, The Double Dixie Cup Problem, *Amer. Math. Monthly* **67** (1960), 58–61.