## THE DOUBLE DIXIE CUP PROBLEM

DONALD J. NEWMAN, Brown University, AND LAWRENCE SHEPP, Princeton University

The familiar childhood occupation of obtaining a complete set of pictures of baseball players, movie stars, etc., which appear on the covers of dixie cups raises some interesting questions. One, which has already been answered, is the "single dixie cup problem," that of determining the expected number, $E(n)$, of dixie cups which must be purchased before a complete set of $n$ pictures is obtained: $E(n)=n(1+1/2+\cdots+1/n)$ ([1] p. 213).

Some time ago W. Weissblum asked how long, on the average, it would take to obtain *two* complete sets of $n$ pictures. This corresponds to the situation observed when two tots collect cooperatively, *i.e.*, "trading" takes place.

This "double dixie cup" problem cannot be handled by the same device used for the problem of the single set and in this paper we find a new method which allows us to write down the solution, $E_m(n)$, (as an easily evaluated definite integral) for the problem of collecting $m$ sets.

For $m$ fixed and $n$ large the expected number of dixie cups turns out to be $n(\log n+(m-1)\log\log n+o(1))$. Thus, although the first set "costs" $n \log n$, all further sets cost $n \log\log n$.

Suppose $m$ sets are desired. Let $p_i$ be the probability of failure of obtaining $m$ sets up to and including the purchase of the $i$th dixie cup. Then the expected number of dixie cups $E_m(n)=\sum_{i=0}^{\infty} p_i$, by a well-known argument ([1] p. 211). Now $p_i=N_i/n^i$ where $N_i$ is the number of ways that the purchase of $i$ dixie cups can fail to yield $m$ copies of each of the $n$ pictures in the set. If we represent the pictures by $x_1, \cdots, x_n$, then $N_i$ is simply $(x_1+\cdots+x_n)^i$ expanded and evaluated at $(1, \cdots, 1)$ after all the terms have been removed which have each exponent for each variable larger than $m-1$.

Now consider $m$ fixed and introduce the following notation. If $P(x_1, \cdots, x_n)$ is a polynomial or power series we define $\{P(x_1, \cdots, x_n)\}$ to be the polynomial, or series, resulting when all terms having all exponents $\geqq m$ have been removed. In terms of this notation $p_i$ is $\{(x_1+\cdots+x_n)^i\}/n^i$ evaluated at $x_1=\cdots=x_n=1$.

If we now make the definition

$$S_m(t) = \sum_{k<m} \frac{t^k}{k!} \tag{1}$$

and consider the expression

$$F = \exp(x_1+\cdots+x_n) - (e^{x_1}-S_m(x_1))\cdots(e^{x_n}-S_m(x_n)), \tag{2}$$

it is easily seen that $F$ has no terms with all exponents $\geqq m$; but $F$ does not have all terms of $\exp(x_1+\cdots+x_n)$ with at least one exponent $<m$. We conclude that

(3) $$F = \{\exp(x_1 + \cdots + x_n)\} = \sum \{(x_1 + \cdots + x_n)^i\}/i!.$$

By contrast, we have seen that

(4) $$E_m(n) = \sum p_i = \sum \{(x_1 + \cdots + x_n)^i\}/n^i$$

at $x_1 = \cdots = x_n = 1$, and so all we need now is a method for replacing $1/i!$ by $1/n^i$. This is afforded by the identity

$$n\int_0^\infty \frac{t^i}{i!} e^{-nt} dt = \frac{1}{n^i}$$

and the result is

(5) $$n\int_0^\infty [\exp(x_1 + \cdots + x_n)^t - (e^{x_1 t} - S_m(tx_1)) \cdots (e^{x_n t} - S_m(tx_n)]e^{-nt} dt.$$

Setting $x_1 = \cdots = x_n = 1$, finally, gives, by (4) and (5),

THEOREM 1. $n\int_0^\infty [1-(1-S_m(t)e^{-t})^n]dt = E_m(n)$.

This is the solution and it is readily integrable for small $m$ and $n$. It is perhaps worthwhile to mention that if only a particular $k$ of the pictures are desired the expected number is:

$$n\int_0^\infty 1 - \left(1 - e^{-t}\left(1 + t \cdots + \frac{t^{m-1}}{(m-1)!}\right)\right)^k dt.$$

This may also be easily generalized to the case where $m_k$ copies of the $k$th pictures are desired.

For large $m$, by the law of large numbers, the number required is asymptotic to $mn$. It remains to obtain the asymptotic form for large $n$.

We now prove

THEOREM 2 .$E_m(n) = n[\log n + (m-1) \log\log n + C_m + o(1)]$ *for* $m$ *fixed and* $n \to \infty$.

It suffices to prove that

$$\frac{E_m(n+1)}{n+1} - \frac{E_m(n)}{n} = \frac{1}{n+1} + \frac{m-1}{n \log n} + \lambda_n,$$

where $\sum |\lambda_n| < \infty$.

Now, by Theorem 1,

$$\frac{E_m(n+1)}{n+1} - \frac{E_m(n)}{n} = \int_0^\infty e^{-t}S_m(t)[1 - e^{-t}S_m(t)]^n dt$$

and, changing variables by

(6) $$1 - e^{-t}S_m(t) = x,$$

we have $dx=e^{-t}[(t^{m-1})/(m-1)!]dt$ and so the above is equal to

$$(7)\quad \int_0^1 x^n S_m(t)\frac{(m-1)!}{t^{m-1}}\,dx=\int_0^1 x^n\left[1+\frac{m-1}{t}+\frac{(m-1)(m-2)}{t^2}+\cdots\right]dx.$$

[$t=t(x)$ is of course defined by (6)].

We now show that

$$\text{(A)}\qquad \text{For } 1<k<m,\qquad \sum_{n=1}^{\infty}\int_0^1 \frac{x^n}{t^k}\,dx<\infty;$$

$$\text{(B)}\qquad \int_0^1 \frac{x^n}{t}\,dx=\frac{1}{n\log n}+\alpha n,\quad \text{where}\quad \sum|\alpha n|<\infty;$$

and, by (7), these will suffice to prove our theorem.

*Proof of* (A). By (6) we have $x=1-e^{-t}S_m(t)\leqq 1-e^{-t}$ so that,

$$(8)\qquad t\geqq \log\frac{1}{1-x}\cdot$$

On the other hand

$$x=1-e^{-t}S_m(t)=\int_0^t e^{-u}\frac{u^{m-1}}{(m-1)!}\,du\leqq\int_0^t u^{m-1}du\leqq t^m,$$

so that

$$(9)\qquad t\geqq x^{1/m}.$$

Now, the infinite series given in (A) is equal to

$$\int_0^1\frac{1}{1-x}\frac{dx}{t^k}=\int_0^{1/2}\frac{1}{1-x}\frac{dx}{t^k}+\int_{1/2}^1\frac{1}{1-x}\frac{dx}{t^k}\cdot$$

The first of the integrals on the right is finite by virtue of (9) and the fact that $k<m$, while the second integral is finite by (8) and the fact that $k>1$. This proves (A).

*Proof of* (B). By (8) we obtain

$$t\geqq x+\frac{x^2}{2}+\cdots\frac{x^r}{r}\geqq x^r\left[1+\frac{1}{2}+\cdots\frac{1}{r}\right]\geqq x^r\log r,$$

and so

$$(10)\qquad \int_0^1\frac{x^n}{t}\,dx\leqq\int_0^1\frac{x^{n-r}}{\log r}\,dx=\frac{1}{(n-r+1)\log r}\cdot$$

Now let $u\geqq 1$ be a parameter and set

$$(11)\qquad a=1-S_m(u)e^{-u}.$$

We have

$$\int_0^1 \frac{x^n}{t}\,dx \geqq \int_0^a \frac{x^n}{t}\,dx \geqq \frac{1}{u}\int_0^a x^n dx$$

$$\geqq \frac{1}{u}\left[\int_0^1 x^n dx - \int_a^1 dx\right] = \frac{1}{(n+1)u} - \frac{S_m(u)e^{-u}}{u} \qquad \text{by (11).}$$

Now note that by (1) and the fact that $u \geqq 1$, we obtain $S_m(u) \leqq eu^{m-1}$. Combining this with the previous inequality gives

$$\int_0^1 \frac{x^n}{t}\,dx \geqq \frac{1}{(n+1)u} - u^{m-2}e^{1-u}. \tag{12}$$

If we now set $r = [n/\log n]$ in (10) and $u = \log n + m \operatorname{loglog} n$ in (12) we obtain

$$\frac{1}{n\log n} - \frac{C \operatorname{loglog} n}{n \log^2 n} \leqq \int_0^1 \frac{x^n}{t}\,dx \leqq \frac{1}{n \log n} + \frac{C \operatorname{loglog} n}{n\log^2 n},$$

and this completes the proof since $\sum(\operatorname{loglog} n)/(n \log^2 n) < \infty$.


### Reference

1. W. Feller, Introduction to Probability Theory, vol. I, New York, 1950.


## THE ANTICENTER OF A GROUP

Norman Levine, University of Pittsburgh

In contrast to the center of a group $G$ we will define the rim of $G$, denoted by $R(G)$, to be the set of elements of $G$ which permute with no elements of $G$ except in trivial cases. We will define the anticenter of $G$, denoted by $AC(G)$, as the set of products of elements in $R(G)$. More precisely:

Definition 1. $R(G) \equiv \{a \mid ab = ba$ *implies* $\exists\, c \in G$ *such that* $a = c^i,\ b = c^j$ *for some integers* $i$ *and* $j.\}$

Definition 2. $AC(G) \equiv \{a_1 \cdots a_n \mid a_i \in R(G).\}$

Lemma 1. *The identity* $e$ *of* $G$ *belongs to* $R(G)$.

*Proof.* $eb = be$ implies $e = b^0$, $b = b^1$.

Lemma 2. *If* $a \in R(G)$ *then* $a^{-1} \in R(G)$.

*Proof.* Suppose $a^{-1}b = ba^{-1}$. Then $ab = ba$ and hence $a = c^i$ and $b = c^j$ or $b = c^j$ and $a^{-1} = c^{-i}$.

Lemma 3. *If* $a \in R(G)$ *then* $b^{-1}ab \in R(G)$ *for all* $b \in G$.

*Proof.* Let $b^{-1}abx = xb^{-1}ab$ for some $x$ in $G$. Then $abxb^{-1} = bxb^{-1}a$. Hence $a = c^i$ and $bxb^{-1} = c^j$ or $b^{-1}ab = b^{-1}c^ib = (b^{-1}cb)^i$ and $x = b^{-1}c^jb = (b^{-1}cb)^j$. This proves $b^{-1}ab \in R(G)$.