# On Sara's Dove Bar Habit

D. M. Einstein
C. C. Heckman
T. S. Norfolk

For those unfamiliar with it, the sci.math newsgroup hierarchy is a collection of thread-based bulletin boards dealing with all aspects of mathematics. Of that hierarchy, the group sci.math is devoted to the discipline in general. On any given day, topics posted there might be by high school or college students looking for help with homework, people suggesting solutions to classical conjectures, cranks disputing well-known mathematical results, teachers discussing methods and curricula, researchers looking for help, and others with mathematical problems of personal interest. One such was posted in 2007 (see [2]):

Sara and David share a box of mixed vanilla and chocolate Dove ice cream bars. Sara prefers vanilla, and David only eats chocolate, so they use the following chivalrous protocol, which is applied repeatedly until the box is empty:

1. If there is a single bar in the box, Sara eats it, regardless of flavour.
2. If there are at least two bars left, Sara selects one at random.
   (a) If it is vanilla, she eats it.
   (b) If it is chocolate, then David selects one at random. If he selects a chocolate one, then they both eat theirs. If he selects a vanilla one, then they switch, so that she eats the vanilla bar, and he eats the chocolate one.

The question posed was to compute the expected number of chocolate bars eaten by Sara, assuming $n$ vanilla and $n$ chocolate bars in a box. There are $\binom{2n}{n}$ ways of taking the bars in order out of the box, and Sara eats a chocolate bar exactly when both she and David select one, except for the terminal case when the box only contains a single chocolate bar.

Using standard counting techniques, we enumerate the number of ways that Sara can eat exactly $k$ chocolates as follows:

1. If we consider the possible orderings where there is not a final chocolate bar, there must be $k$ pairs of the form $CC$, $n-2k$ pairs of the form $CV$, and $2k$ individual vanillas, all disjoint. These $n+k$ structures can then be ordered in $\binom{n+k}{k}\binom{n}{n-2k}$ ways.

2. For the orderings for which the process terminates in a single chocolate, we put that chocolate at the end of the ordering, form $k-1$ pairs of the form $CC$, $n+1-2k$ pairs of the form $CV$, and $2k-1$ individual vanilla bars, which can be ordered in $\binom{n+k-1}{k-1}\binom{n}{n+1-2k}$ ways.

This yields the expected number of chocolate bars eaten by Sara as

$$\begin{aligned}\mathcal{E}_S &= \frac{1}{\binom{2n}{n}}\left\{\sum_{k=0}^{\lfloor\frac{n}{2}\rfloor} k\binom{n+k}{k}\binom{n}{n-2k}\right. \\ &+ \left.\sum_{k=1}^{\lfloor\frac{n+1}{2}\rfloor} k\binom{n+k-1}{k-1}\binom{n}{n-2k+1}\right\} \qquad (1)\end{aligned}$$

where $\lfloor x \rfloor$ denotes the greatest integer (floor) function.

A heuristic analysis, based on large $n$, and neglecting the terminal condition, suggests that Sara should asymptotically eat $n/3$ chocolate bars, since the chocolate bars appear (approximately) equally in pairs of the form $CV$ and $CC$, and she only eats a chocolate bar in the latter case.

However, extensive numerical evaluation of the expectation in (1) indicates that

$$\mathcal{E}_S = \frac{n}{3} + \frac{4}{27} + o(1), \qquad (2)$$

which, of course, leads to the question of the origin of the term $4/27$.

Several respondents to the thread [2] attempted an explanation, including one who noted that this term could not be due only to the boundary condition, since allowing David to eat the last bar, if it is chocolate, yields an expectation

$$\mathcal{E}'_S = \frac{n}{3} - \frac{5}{27} + o(1). \qquad (3)$$

Another presented a Central Limit approach to asymptotically approximate the summation in (1), which is difficult to make rigorous.

The most promising approach presented in [2] suggested generalizing the problem, by writing the expected number $S_{c,v}$ of chocolate bars eaten by Sara, given $c$ chocolate and $v$ vanilla bars, in the recurrence relation

$$S_{c,v} = \frac{v}{c+v} \cdot S_{c,v-1} + \frac{c}{c+v} \cdot \frac{v}{c+v-1} \cdot S_{c-1,v-1} + \frac{c}{c+v} \cdot \frac{c-1}{c+v-1} \cdot \left(1 + S_{c-2,v}\right), \tag{4}$$

with initial conditions

$$S_{c,0} = \left\lfloor \frac{c+1}{2} \right\rfloor, \ S_{0,v} = 0 \text{ and } S_{1,v} = \frac{1}{v+1}. \tag{5}$$

Our approach will be to take the recurrence relation (4,5) above and modify it to compute the total number of chocolate bars eaten by Sara, solving the counting problem exactly by a two-variable generating function. For a good introduction to the theory of generating functions, see [3].

We begin by making the transformation

$$T_{c,v} = \binom{v+c}{c} S_{c,v}, \tag{6}$$

and slightly changing the boundary condition by having Sara eat a proportion $\alpha$ of the last bar in the box, if it is chocolate. This leads to the modified recurrence relation

$$T_{c,v} = T_{c,v-1} + T_{c-1,v-1} + T_{c-2,v} + \binom{v+c-2}{c-2}, \tag{7}$$

valid for $c \geq 2$ and $v \geq 1$, with initial conditions

$$T_{0,v} = 0, \ T_{1,v} = \alpha, \ T_{2k,0} = k, \text{ and } T_{2k+1,0} = k + \alpha. \tag{8}$$

We write $F(x,y) = \sum_{c=0}^{\infty} \sum_{v=0}^{\infty} T_{c,v} x^c y^v$, multiply the equation (7) by $x^c y^v$ and sum for $c \geq 2$ and $v \geq 1$. Considering each sum separately, using the initial conditions (8), and making extensive use of the extended binomial expansion

$$\frac{1}{(1-r)^m} = \sum_{k=0}^{\infty} \binom{m-1+k}{k} r^k \tag{9}$$

yields

$$\begin{aligned}
\sum_{v=1}^{\infty}\sum_{c=2}^{\infty} T_{c,v}x^c y^v &= F(x,y) - \sum_{c=0}^{\infty} T_{c,0}x^c - \sum_{v=1}^{\infty} T_{0,v}y^v - x\sum_{v=1}^{\infty} T_{1,v}y^v \\
&= F(x,y) - \sum_{k=0}^{\infty} kx^{2k} - \sum_{k=0}^{\infty}(k+\alpha)x^{2k+1} - \alpha x\sum_{v=1}^{\infty} y^v \\
&= F(x,y) - \frac{x^2+x^3}{(1-x^2)^2} - \frac{\alpha x}{1-x^2} - \frac{\alpha xy}{1-y}\ . \qquad (10)
\end{aligned}$$

Similarly,

$$\begin{aligned}
\sum_{v=1}^{\infty}\sum_{c=2}^{\infty} T_{c,v-1}x^c y^v &= yF(x,y) - \frac{\alpha xy}{1-y}, \qquad (11) \\
\sum_{v=1}^{\infty}\sum_{c=2}^{\infty} T_{c-1,v-1}x^c y^v &= xyF(x,y), \qquad (12) \\
\sum_{v=1}^{\infty}\sum_{c=2}^{\infty} T_{c-2,v}x^c y^v &= x^2F(x,y) - \frac{x^4+x^5}{(1-x^2)^2} - \frac{\alpha x^3}{1-x^2}, \qquad (13) \\
\sum_{v=1}^{\infty}\sum_{c=2}^{\infty} \binom{v+c-2}{c-2} x^c y^v &= x^2 \sum_{v=1}^{\infty}\sum_{k=0}^{\infty} \binom{v+k}{k} x^k y^v \\
&= \frac{x^2}{1-x}\sum_{v=1}^{\infty}\left(\frac{y}{1-x}\right)^v \\
&= \frac{x^2 y}{(1-x)(1-x-y)}. \qquad (14)
\end{aligned}$$

Rearranging and combining terms in (7) and (10–14) yields the compact form

$$F(x,y) = \frac{x^2}{(1+x)(1-x-y)^2} + \frac{\alpha x}{(1+x)(1-x-y)}. \qquad (15)$$

In order to obtain the necessary asymptotics, we re-expand $F(x,y)$ as follows:

$$\begin{aligned}
\frac{x^2}{(1+x)(1-x-y)^2} &= \frac{x^2}{(1+x)(1-x)^2}\left(1-\frac{y}{1-x}\right)^{-2} \\
&= \sum_{v=0}^{\infty}\sum_{c=2}^{\infty}\sum_{j=1}^{c-1}(v+1)\binom{v+c-j}{c-1-j}(-1)^{j-1}x^c y^v \qquad (16)
\end{aligned}$$

and

$$\frac{\alpha x}{(1+x)(1-x-y)} = \alpha\sum_{v=0}^{\infty}\sum_{c=1}^{\infty}\sum_{j=1}^{c}\binom{v+c-j}{c-j}(-1)^{j-1}x^c y^v. \qquad (17)$$

Hence, we obtain the modified version of Sara's chocolate expectation (where she eats the proportion $\alpha$ of the last chocolate bar) as

$$\begin{aligned} S_{c,v,\alpha} &= \frac{T_{c,v}}{\binom{v+c}{c}} \\ &= \frac{1}{\binom{v+c}{c}} \sum_{j=1}^{c-1} (v+1) \binom{v+c-j}{c-1-j} (-1)^{j-1} \\ &+ \frac{\alpha}{\binom{v+c}{c}} \sum_{j=1}^{c} \binom{v+c-j}{c-j} (-1)^{j-1}. \end{aligned} \tag{18}$$

The terms in each sum decrease geometrically as $j$ increases, so we may use Stirling's approximation for small $j$,

$$\ln m! = \frac{1}{2}\ln 2\pi + \left(m+\frac{1}{2}\right)\ln m - m + \frac{1}{12m} + O\left(\frac{1}{m^3}\right), \tag{19}$$

and, after a little algebra, with the aid of Maple,

$$\begin{aligned} \frac{\binom{v+c-j}{c-j}}{\binom{v+c}{c}} &= \frac{(v+c-j)!c!}{(c-j)!(v+c)!} \\ &= \exp\left(j\ln\left(\frac{c}{v+c}\right) - \frac{v}{2c(v+c)} j(j-1)\left\{1+O\left(\frac{1}{c}\right)\right\}\right) \\ &= \left(1 - \frac{v}{2c(v+c)} j(j-1)\left\{1+O\left(\frac{1}{c}\right)\right\}\right)\left(\frac{c}{v+c}\right)^j. \end{aligned} \tag{20}$$

With the observation that

$$(v+1)\binom{v+c-j}{c-1-j} = (c-j)\binom{v+c-j}{c-j}, \tag{21}$$

we obtain

$$\begin{aligned} S_{c,v,\alpha} &= \left(\frac{c}{v+c}\right) \sum_{j=1}^{c-1} \left\{c - \frac{c}{v+c}\binom{j}{j-1} - \frac{v}{v+c}\binom{j+1}{j-1}\right\}\left(-\frac{c}{v+c}\right)^{j-1} \\ &+ \alpha\left(\frac{c}{v+c}\right) \sum_{j=1}^{c} \left(-\frac{c}{v+c}\right)^{j-1} + O\left(\frac{v}{c(v+c)}\right), \end{aligned} \tag{22}$$

from which, again using the extended binomial expansion,

$$S_{c,v,\alpha} = \frac{c^2}{v+2c} - \frac{c^2}{(v+2c)^2} - \frac{vc(v+c)}{(v+2c)^3} + \frac{\alpha c}{v+2c} + O\left(\frac{v}{c(v+c)}\right). \qquad (23)$$

In the original problem, we have $v = c = n$ and $\alpha = 1$, whence

$$S_{n,n,1} = \frac{n}{3} - \frac{1}{9} - \frac{2}{27} + \frac{1}{3} + O\left(\frac{1}{n}\right) = \frac{n}{3} + \frac{4}{27} + O\left(\frac{1}{n}\right), \qquad (24)$$

as originally claimed. This derivation also clearly demonstrates that the extra term $4/27$ is due to $c$ being large, but not infinite.

The flavour of the bivariate expansion above is similar to those discussed more generally in [1]. The methods discussed there indeed give the same results as in our expansion, but require significantly more manipulation.

While we have chosen to solve this problem using generating functions, it is possible to give a classical combinatorial argument to show that the given expression for $S_{c,v,\alpha}$ is indeed the generalization of the original problem.

Using (18) and (21), we may write

$$\binom{v+c}{c} S_{c,v,\alpha} = \sum_{j=1}^{c-1} (c-j) \binom{v+c-j}{c-j} (-1)^{j-1} + \alpha \sum_{j=1}^{c} \binom{v+c-j}{c-j} (-1)^{j-1} .$$

For the first of these summations, consider the following selection procedure: For fixed $j$, order $v$ vanilla and $c-j$ chocolate bars, and insert the remaining $j$ chocolate bars immediately in front of one of the latter. This can be done in $(c-j)\binom{v+c-j}{j}$ ways. Clearly, each sequence of chocolate bars of length exactly $j+i$ will be generated $i$ times. Thus, in the first summation, which amounts to a use of the principle of inclusion and exclusion, a sequence of length $2n$ will be counted $(2n-1) - (2n-2) + \cdots + 1 = n$ times, and a sequence of length $2n+1$ will be counted $2n - (2n-1) + \cdots - 1 = n$ times. That is, the summation counts the number of chocolate bars eaten by Sara. For the second summation, the argument is similar, except that we only insert the remaining $j$ chocolate bars at the end of the sequence. Thus, for fixed $j$, the term counts the number of chocolate sequences of length at least $j$. A direct application of the principle of inclusion and exclusion means that the given sum is exactly the number of sequences which end in an odd number of chocolate bars.

## References

[1] M. Drmota, Asymptotic distributions and a multivariate Darboux method in enumeration problems, *J. Comb. Theory, Ser. A* **67** (1994) 169–184.

[2] D. M. Einstein, On Sara's method of eating Dove Bars (October 2007), available at http://mathforum.org/kb/forum.jspa?forumID=13.

[3] H. S. Wilf, *generatingfunctionology*, 2nd ed., A K Peters, Wellesley, MA, 2006.

*deinst@gmail.com*

*The School of Mathematical and Statistical Sciences, Arizona State University, Tempe, AZ 85287, checkman@math.asu.edu*

*Department of Theoretical and Applied Mathematics, The University of Akron, Akron, OH 44325, norfolk@uakron.edu*