Chapter 7
Cubes, Carries, and an Amazing Matrix (Supplemental)

7.1 Slicing a cube

In this supplemental chapter we will find the Eulerian numbers cropping up in some surprising places.

First, consider cutting up the n-dimensional cube $[0,1]^n$ according to the braid arrangement. For example, Figure 7.1 shows this in three dimensions.

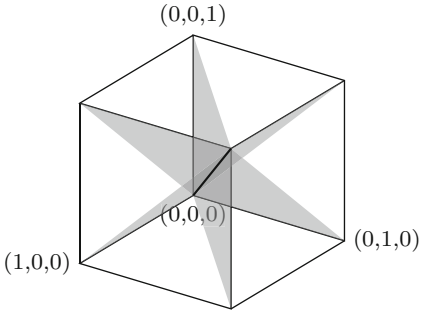

Fig. 7.1 Slicing a cube with the braid arrangement, looking down the line $x = y = z$.

Ignoring overlaps on the boundaries, each region here is a simplex of the form

$$\mathcal{S}_w = \{\mathbf{x} \in \mathbb{R}^n : 0 \leq x_{w(1)} \leq x_{w(2)} \leq \cdots \leq x_{w(n)} \leq 1\},$$

where $w \in S_n$. By symmetry, each of these regions has the same volume, and since their union has volume 1, we get

$$\mathrm{vol}(\mathcal{S}_w) = \frac{1}{n!}.$$

T.K. Petersen, *Eulerian Numbers*, Birkhäuser Advanced Texts Basler Lehrbücher, DOI 10.1007/978-1-4939-3091-3_7

Now consider slicing the cube by level sets. For fixed n, and any $k = 0, 1, \ldots, n-1$, let

$$\mathcal{R}_k = \{\mathbf{y} \in [0,1]^n : k \le y_1 + y_2 + \cdots + y_n \le k+1\}.$$

For three dimensions, we have illustrated these slices in Figure 7.2. The following proposition suggests how to compute the volume of these slices.

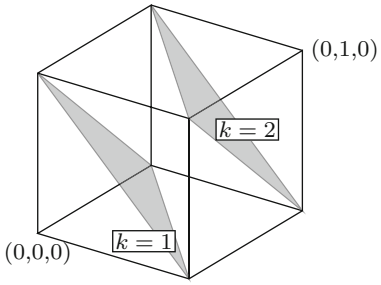

Fig. 7.2 Slicing a cube with level sets.

Proposition 7.1. *The volume of the kth slice of the n-cube is given by:*

$$\mathrm{vol}(\mathcal{R}_k) = \frac{\left\langle {n \atop k} \right\rangle}{n!},$$

where $\left\langle {n \atop k} \right\rangle$ *is the number of permutations of n with k descents.*

This result is mentioned in Dominique Foata's 1977 paper [67], in which he asks for a combinatorial proof. Richard Stanley provided a beautifully simple proof in a note at the end of Foata's paper, which we describe here. (This is Problem 51 in Stanley's textbook [154].)

Let

$$\mathcal{S}_k = \bigcup_{\mathrm{des}(w^{-1})=k} \mathcal{S}_w,$$

denote the union of points in the cones corresponding to permutations with k descents. We will define a map $\phi : \mathcal{S}_k \to \mathcal{R}_k$ that is "generically" a bijection, in that it is bijective for all points such that no two coordinates are equal. (Such points have measure zero and are irrelevant for the volume calculation.)

The map is given explicitly by $\phi(x_1, \ldots, x_n) = (y_1, \ldots, y_n)$ with

$$y_i = \begin{cases} x_{i+1} - x_i & \text{if } x_i < x_{i+1}, \\ 1 + x_{i+1} - x_i & \text{if } x_i > x_{i+1}, \end{cases}$$

where $x_{n+1} = 1$. If $x_i = x_{i+1}$ for some i, ϕ is undefined.

Suppose $\mathbf{x} = (x_1, \ldots, x_n)$ is a generic point in $\mathcal{S}_w$. To say that $x_i > x_{i+1}$ is to say that $i+1$ appears to the left of i in w, i.e., $w^{-1}(i+1) < w^{-1}(i)$. In other words i is a descent of w^{-1}. Notice that if $\mathrm{des}(w^{-1}) = k$, then $\sum y_i = k + 1 - x_1$. Thus ϕ maps points from $\mathcal{S}_k$ to $\mathcal{R}_k$.

For example, generic points in the region $\mathcal{S}_{631425}$ satisfy

$$0 < x_6 < x_3 < x_1 < x_4 < x_2 < x_5 < 1,$$

and these get mapped to

$$(y_1, y_2, y_3, y_4, y_5, y_6) = (x_2 - x_1, 1 + x_3 - x_2, x_4 - x_3, x_5 - x_4, 1 + x_6 - x_5, 1 - x_6).$$

The sum of the coordinates under this map is $\sum y_i = 3 - x_1$, so $2 < \sum y_i < 3$, as expected since $\mathrm{des}(w^{-1}) = 2$. Notice that on $\mathcal{S}_w$, the map ϕ is an affine transformation, given here by:

$$\mathbf{y} = \phi(\mathbf{x}) = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix} \mathbf{x}.$$

The determinant of the linear part of this transformation has absolute value 1, so it is volume-preserving.

It remains to show that ϕ is invertible.

To reverse the map ϕ, we work from right to left, exploiting the observation that $x_i = x_{i+1} - y_i$ or $x_i = 1 + x_{i+1} - y_i$. Since $0 < x_i < 1$, only one of these expressions can be correct. By convention $y_n = 1 - x_n$, so we get started with $x_n = 1 - y_n$. Otherwise, once we have calculated x_{i+1} we get:

$$x_i = \begin{cases} x_{i+1} - y_i & \text{if } x_{i+1} > y_i, \\ 1 + x_{i+1} - y_i & \text{if } x_{i+1} < y_i. \end{cases}$$

To take an example, suppose

$$\mathbf{y} = (.3, .14, .1592, .6, .53, .58, .97).$$

Working through the coordinates one at a time we conclude that

$$\begin{aligned} x_7 &= 1 - y_7 = .03, \\ x_6 &= 1 + x_7 - y_6 = .45, \\ x_5 &= 1 + x_6 - y_5 = .92, \\ x_4 &= x_5 - y_4 = .32, \\ x_3 &= x_4 - y_3 = .1608, \end{aligned}$$

$$x_2 = x_3 - y_2 = .0208,$$
$$x_1 = 1 + x_2 - y_1 = .7208\,.$$

One can check that these coordinates define a point in the region corresponding to $w = 2734651$, and applying ϕ will take $\mathbf{x}$ back to $\mathbf{y}$.

A more succinct way to express the inverse transformation is to collect partial sums from right to left, taking only the fractional part of the partial sum as we go:

$$x_i = 1 - ((y_i + \cdots + y_n) \mod 1)\,.$$

Since the x_i must be generic, we leave this inverse map undefined whenever any subset of the y_j sums to an integer. But if the y_j are generic, this will never happen, so for the volume calculation this set has measure zero.

We have shown that ϕ is generically bijective and volume-preserving. Thus Proposition 7.1 follows.

7.2 Carries in addition

The volume calculation we just carried out turns out to have a surprising application in the problem of the distribution of "carries" in addition.

Consider adding two numbers in base ten with the usual addition algorithm. As we move from right to left we "carry" a 1 to the next column if the sum in the previous column (plus the previous carried digit) adds up to ten or more. How many carries will we expect to have?

Here is the sum of two thirty digit numbers:

carries: 000001 01011 11010 00101 00110 1001

```
    27182 81828 45904 52353 60287 47135
+   31415 92653 58979 32384 62643 38328
-----------------------------------------
    58598 74482 04883 84738 22930 85463
```

We carried a one in thirteen of the thirty columns, or about forty-three percent of the time. Intuition tells us that we will carry a one about half the time, and this is indeed what will bear out.

But now consider adding three numbers. Here we can carry 0, 1, or 2. For example, here is the sum of three thirty digit numbers:

carries: 121011 11121 12111 11102 00001 0121

```
    57721 56649 01532 86060 65120 90082
    69314 71805 59945 30941 72321 21458
+   16449 34066 84822 64364 72415 16665
-----------------------------------------
   143485 62521 46300 81367 09857 28205
```

Of the thirty columns, seven carried zero, five carried two, and eighteen carried a one. It certainly doesn't seem that each carry is equally likely. Symmetry should suggest that carrying a zero has the same probability as carrying a two. The fact that we carry a one much more frequently is suggested by the fact that there are many more ways to obtain a number between 10 and 19 as a sum of three digits than there are ways to write a single digit number as a sum of three digits. But what exactly is the probability of getting a carry of two?

This is the problem considered by John Holte in [91]. (The title of this chapter is a nod to his fine paper.) To quote Holte's motivating question,

> What is the long-run frequency of each possible carry value when we add any number of long numbers represented in any base?

Or, when adding n random numbers in base b, what is the probability of having a carry of k? Remarkably, we will see the answer depends only on n and k, but not the base b. Let us denote the probability by $p_{n,k}$.

Theorem 7.1. *When adding n numbers in base b, the probability of having a carry of k is*

$$p_{n,k} = \frac{\left\langle {n \atop k} \right\rangle}{n!},$$

where $k = 0, 1, \ldots, n-1$.

The form this answer takes suggests that we make a connection between Holte's question and Foata's question. That is, we will show that the volume calculation in Proposition 7.1 implies Theorem 7.1.

To see the connection, suppose we are adding n numbers in base b, and that in a particular column we add digits $d_1, d_2, \ldots, d_n$, with $0 \le d_i \le b-1$. If we carried a j from the previous column, then to say that we carry k into the next column means

$$bk \le j + d_1 + d_2 + \cdots + d_n < b(k+1). \tag{7.1}$$

Now split j into n equal pieces so to write

$$j + d_1 + d_2 + \cdots + d_n = (d_1 + j/n) + (d_2 + j/n) + \cdots + (d_n + j/n).$$

Since $0 \le j \le n-1$, we have $0 \le j/n < 1$ and so $0 \le (d_i + j/n) < b$. Thus, dividing (7.1) by b, we obtain

$$k \le x_1 + x_2 + \cdots + x_n < k+1, \tag{7.2}$$

where

$$0 \le x_i = \frac{d_i + j/n}{b} < 1.$$

Let ψ denote the map from integer n-tuples to the cube $[0,1]^n$ given by $\psi(d_i) = (d_i + j/n)/b$, depending on the prior carry of j in $\{0, 1, \ldots, n-1\}$.

Thus having a carry of k corresponds to a point in the kth slice of the n-cube as discussed in Section 7.1. For fixed n and b, there are only finitely many points $(j, d_1, \ldots, d_n)$ in $[0, n-1] \times [0, b-1]^n$. Thus, the image of these points under ψ is finite as well. We want to argue that despite the discrete nature of this problem, we can use the volume calculation to obtain the result here. This can certainly be done if our points x_i are geometrically uniform in the n-cube.

If the digits d_i are uniformly random in $\{0, 1, \ldots, b-1\}$, intuition tells the points x_i are distributed roughly uniformly in the interval $[0, 1)$. While perfect uniformity won't always occur, we get something close enough to uniform. For fixed j, $d_i + j/n$ is just a slight shift away from uniform, and taking all j together splits $[0, 1)$ into n subintervals on which the x_i are identically distributed:

$$[0, 1/n) \cup [1/n, 2/n) \cup \cdots \cup [1 - 1/n, 1).$$

So whatever the probability of having a carry of j come in, this distribution is repeated in n intervals of equal size in $[0, 1)$, and this is good enough to conclude that probability is proportional to volume.

Hence we can conclude Theorem 7.1 from the geometric result: choosing n random digits in base b that results in a carry of k is equal to the probability of choosing a random point in the kth slice of the unit cube. However while [154] mentions this geometric argument, it was not the technique used by Holte. We present his argument next.

7.3 The amazing matrix

Holte's approach to the carries problem is to view the "carries process" as a Markov chain. This is natural, since carrying a k depends only on the digits in the column being added and the number j that was carried into that column.

Thus for fixed b and n, let $\pi(j, k)$ denote the probability of "carrying out" k given that we "carry in" j to a particular column. Then by (7.1),

$$\pi(j, k) = \frac{(\text{number of solutions } (d_1, \ldots, d_n) \text{ to } (7.1))}{b^n}.$$

To count the integer solutions to (7.1) is to ask for the number of integer solutions to

$$c + d_1 + d_2 + \cdots + d_n = b(k+1) - 1 - j, \tag{7.3}$$

where $0 \le c, d_1, d_2, \ldots, d_n \le b - 1$. If we let $r = b(k+1) - 1 - j$, then the number of solutions to Equation (7.3) is the coefficient of z^r in

$$(1+z+z^2+\cdots+z^{b-1})^{n+1} = \frac{(1-z^b)^{n+1}}{(1-z)^{n+1}}.$$

Expanding both numerator and denominator as series in z, we find

$$\begin{aligned}\frac{(1-z^b)^{n+1}}{(1-z)^{n+1}} &= \sum_{l=0}^{n+1}(-1)^l\binom{n+1}{l}z^{bl}\sum_{m\geq 0}\binom{m+n}{n}z^m,\\ &= \sum_{m,l\geq 0}(-1)^l\binom{n+1}{l}\binom{n+m}{n}z^{bl+m},\\ &= \sum_{r\geq 0}\left(\sum_{l=0}^{n+1}(-1)^l\binom{n+1}{l}\binom{n+r-bl}{n}\right)z^r.\end{aligned}$$

Given that $\binom{n+r-bl}{n} = 0$ if $r < bl$, the coefficient of z^r only ranges over $l \leq r/b = k+1-(j+1)/b$. We therefore have the following explicit formula for $\pi(j,k)$.

Proposition 7.2. *Suppose we are adding a list of n numbers in base b. The probability of carrying out a k from one column to the next, given that we carry in a j is*

$$\pi(j,k) = \frac{1}{b^n}\sum_{0\leq l\leq k+1-(j+1)/b}(-1)^l\binom{n+1}{l}\binom{n+b(k+1-l)-1-j}{n}.$$

The transition matrix $\Pi_n = (\pi(j,k))_{0\leq j,k\leq n-1}$ is what Holte calls the "Amazing matrix." Here are the first two matrices:

$$\Pi_2 = \frac{1}{2b}\begin{pmatrix} b+1 & b-1 \\ b-1 & b+1\end{pmatrix}, \Pi_3 = \frac{1}{6b^2}\begin{pmatrix} b^2+3b+2 & 4b^2-4 & b^2-3b+2 \\ b^2-1 & 4b^2+2 & b^2-1 \\ b^2-3b+2 & 4b^2-4 & b^2+3b+2\end{pmatrix}.$$

It turns out that the matrix Π is diagonalizable, and its eigenvalues are $1, 1/b, 1/b^2, \ldots, 1/b^{n-1}$, though the eigenvectors are independent of b.

Let $V = V_n$ denote the matrix such that $V\Pi V^{-1} = D$, with D the diagonal matrix with the indicated eigenvalues. For example, one can check

$$\begin{pmatrix}1 & 4 & 1\\ 1 & 0 & -1\\ 1 & -2 & 1\end{pmatrix}\cdot\frac{1}{6b^2}\begin{pmatrix} b^2+3b+2 & 4b^2-4 & b^2-3b+2 \\ b^2-1 & 4b^2+2 & b^2-1 \\ b^2-3b+2 & 4b^2-4 & b^2+3b+2\end{pmatrix} = \begin{pmatrix}1 & 0 & 0\\ 0 & 1/b & 0\\ 0 & 0 & 1/b^2\end{pmatrix}\begin{pmatrix}1 & 4 & 1\\ 1 & 0 & -1\\ 1 & -2 & 1\end{pmatrix},$$

so

$$V_3 = \begin{pmatrix} 1 & 4 & 1 \\ 1 & 0 & -1 \\ 1 & -2 & 1 \end{pmatrix}.$$

Let $V_n = (v(j,k))_{0 \le j,k \le n-1}$. It turns out that

$$v(j,k) = \sum_{l=0}^{k} (-1)^l \binom{n+1}{l} (k+1-l)^{n-j}.$$

The matrices V_4 and V_5 are shown here:

$$V_4 = \begin{pmatrix} 1 & 11 & 11 & 1 \\ 1 & 3 & -3 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -3 & 3 & -1 \end{pmatrix}, V_5 = \begin{pmatrix} 1 & 26 & 66 & 26 & 1 \\ 1 & 10 & 0 & -10 & -1 \\ 1 & 2 & -6 & 2 & 1 \\ 1 & -2 & 0 & 2 & -1 \\ 1 & -4 & 6 & -4 & 1 \end{pmatrix}.$$

Notice the Eulerian numbers appearing in the top row! This is because if $j = 0$,

$$v(0,k) = \sum_{l=0}^{k} (-1)^l \binom{n+1}{l} (k+1-l)^n,$$

which is the formula given in Equation (1.11) for the Eulerian number $\left\langle {n \atop k} \right\rangle$.

For fixed j, $v(j,k)$ is the coefficient of t^k in

$$\left(\sum_{l \ge 0} (-1)^l \binom{n+1}{l} t^l \right) \left(\sum_{m \ge 0} (m+1)^{n-j} t^m \right) = (1-t)^{n+1} \frac{S_{n-j}(t)}{(1-t)^{n+1-j}},$$

where the second sum is the Carlitz identity given in Equation (1.10). Thus we have a simpler way to describe the entries of V:

$$\sum_{k \ge 0} v(j,k) t^k = (1-t)^j S_{n-j}(t),$$

where $S_{n-j}(t)$ is the Eulerian polynomial.

Now let us verify that $V\Pi = DV$.

We want to show that

$$\sum_{k=0}^{n-1} v(j,k)\pi(k,l) = \frac{v(j,l)}{b^j},$$

for $0 \le j, l \le n-1$.

Using the formulas we've derived, we have

$$\begin{aligned}
&\sum_{k=0}^{n-1} v(j,k)\pi(k,l) \\
&\quad = \frac{1}{b^n}\sum_{k=0}^{n-1}\sum_{m=0}^{l+1-(k+1)/b}(-1)^m\binom{n+1}{m}\binom{n-1-k+(l+1-m)b}{n}v(j,k), \\
&\quad = \frac{1}{b^n}\sum_{m=0}^{l}(-1)^m\binom{n+1}{m}\sum_{k=0}^{(l+1-m)b-1}\binom{n-1-k+(l+1-m)b}{n}v(j,k).
\end{aligned} \tag{7.4}$$

If we let $M = (l+1-m)b-1$, we can rewrite the inner sum here as

$$\sum_{k=0}^{M}\binom{n+M-k}{n}v(j,k),$$

which we can recognize as the coefficient of t^M in

$$\left(\sum_{r\geq 0}\binom{n+r}{n}t^r\right)\left(\sum_{k\geq 0}v(j,k)t^k\right) = \frac{1}{(1-t)^{n+1}}(1-t)^j S_{n-j}(t).$$

Using the Carlitz identity once more, we find

$$\begin{aligned}
\frac{1}{(1-t)^{n+1}}(1-t)^j S_{n-j}(t) &= \frac{S_{n-j}(t)}{(1-t)^{n+1-j}}, \\
&= \sum_{M\geq 0}(M+1)^{n-j}t^M,
\end{aligned}$$

and therefore

$$\sum_{k=0}^{M}\binom{n+M-k}{n}v(j,k) = (M+1)^{n-j}.$$

Returning to Equation (7.4), we now obtain

$$\begin{aligned}
\sum_{k=0}^{n-1} v(j,k)\pi(k,l) &= \frac{1}{b^n}\sum_{m=0}^{l}(-1)^m\binom{n+1}{m}((l+1-m)b)^{n-j}, \\
&= \frac{1}{b^j}\sum_{m=0}^{l}(-1)^m\binom{n+1}{m}(l+1-m)^{n-j}, \\
&= \frac{v(j,l)}{b^j},
\end{aligned}$$

as desired.

Since the largest eigenvalue of Π is 1, the Perron-Frobenius theorem tells us the first row of V is proportional to the stable distribution for the carries process. Hence Theorem 7.1 follows.

We finish this chapter by remarking that the Amazing Matrix has reappeared in some surprising places. For instance Francesco Brenti and Volkmar Welker rediscovered the Amazing Matrix in commutative algebra [36], where Π is essentially the transformation of a Hilbert series of a graded ring to its bth "Veronese algebra." In terms of generating functions, this is the map

$$\frac{h(t)}{(1-t)^d} = \sum_{k\geq 0} a_k t^k \mapsto \sum_{k\geq 0} a_{bk} t^k = \frac{h^{\langle b\rangle}(t)}{(1-t)^d}.$$

The transformation matrix for $h \mapsto h^{\langle b\rangle}$ is (after deleting the first row and column) the Amazing Matrix.

Brenti and Welker analyze this transformation as they did for the barycentric subdivision transformation, which is discussed in Chapter 9. Since the stable distribution for the Amazing Matrix is the Eulerian distribution, they find that repeatedly applying the Veronese map takes any h-polynomial to the Eulerian polynomial in the limit. In particular, applying the map enough times yields a real-rooted h-polynomial.

In a different direction, the Amazing Matrix shows up in the analysis of card shuffling. Persi Diaconis and Jason Fulman have several papers on this topic [57–59]. A "b"-shuffle of a deck of cards is a generalization of the usual riffle shuffle, which is a b-shuffle for $b = 2$. In a b-shuffle we split the deck into b piles of sizes $c_1, \ldots, c_b$ with probability

$$\frac{\binom{n}{c_1,\ldots,c_b}}{b^n}.$$

Then we drop cards randomly from each of the piles, with probability proportional to the size of the pile. The connection between carries in addition and shuffling is most succinctly summarized by Theorem 1.1 of [58], which we quote directly here:

> The probability that the base-b carries chain goes from 0 to j in r steps is equal to the probability that the permutation in S_n obtained by performing r successive b-shuffles (started at the identity) has j descents.

The reader is encouraged to read [57] for a very friendly introduction to this story.