Prüfer Codes

Drew Armstrong
armstron@math.umn.edu

March 24, 2009

In 1889, Arthur Cayley showed that the number of labeled trees with n vertices is n^{n-2}, a result today known as Cayley's Tree Formula. We present here the most famous proof of Cayley's formula, due to Heinz Prüfer (1918).

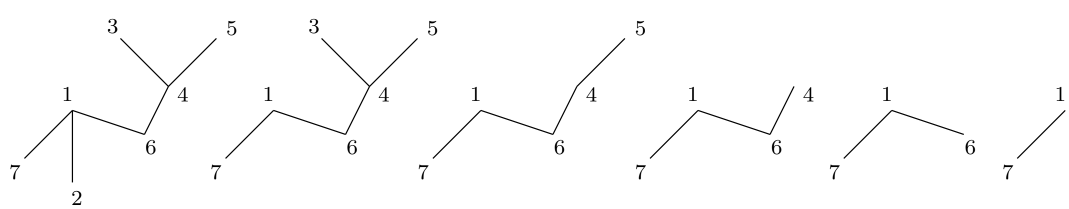
Prüfer's proof answers the following question: How can we efficiently generate all labeled trees? He answered this problem by assigning to each labeled tree a simple code, from which the tree can be easily recovered. Here is the encoding process.

The Prüfer Code of a Tree. Given a labeled tree T with n vertices, repeat the following step

- delete the leaf with the smallest label and record the label of its parent

until only a single edge remains. The resulting sequence of $n-2$ labels is called the Prüfer code of the tree.

For example, the following figure shows a tree with Prüfer code 14461.



Note the important fact that the number of occurrences of the label i in the Prüfer code is one less than the degree of the vertex i in the original tree. Our strategy for counting trees will be to establish a bijection between the set of trees and a set of codes and then to count the codes. Note that the number of words of length $n-2$ from the alphabet $\{1, 2, \ldots, n\}$ is equal to n^{n-2} since there are n choices for each letter. If we can show that each of these words is the Prüfer code of a unique tree, we will have proved Cayley's Tree Formula.

In order to show this, we define an auxiliary structure.

The Extended Prüfer Code. Define a $2 \times (n-1)$ matrix

$$\begin{pmatrix} x_1 & x_2 & \cdots & x_{n-1} \\ y_1 & y_2 & \cdots & y_{n-1} \end{pmatrix}$$

in which (x_i, y_i) is the ith edge deleted in the process of constructing the Prüfer code, and x_i is the leaf that we deleted. The final column (x_{n-1}, y_{n-1}), with $x_{n-1} < y_{n-1}$ will represent the final edge remaining. This matrix is called the extended Prüfer code of the tree.

For example, referring to the above figure we find the extended Prüfer code

$$\begin{pmatrix} 2 & 3 & 5 & 4 & 6 & 1 \\ 1 & 4 & 4 & 6 & 1 & 7 \end{pmatrix}.$$

Note the following:

- The entries $(y_1, y_2, \ldots, y_{n-2})$ are the usual Prüfer code.
- The entry y_{n-1} is always equal to n, since at each step we will always have at least two leaves to choose from, of which we will delete the smaller one. Hence we will never delete the vertex labeled n.
- Since no vertex can be deleted twice and each vertex but n is deleted, the entries of the first row are a permutation of the set $\{1, 2, \ldots, n-1\}$.

Now we wish to show that there is a bijection between labeled trees and Prüfer codes. To do this, we must show that the map that sends a tree to its code is 1:1 (different trees go to different codes) and onto (every code corresponds to some tree). This can be done by constructing the inverse map which takes a code back to its tree, and noting that a function is bijective if and only if it is invertible.

Note that the original tree can be easily recovered from the extended Prüfer code, since its edges are given by the columns of the matrix. We will show how to recover the "extended Prüfer code" from an arbitrary word of length $n-2$ from the alphabet $\{1, 2, \ldots, n\}$ (a "Prüfer code"). Then we will show that the graph obtained from this extended code is a tree and that it is indeed a tree whose Prüfer code is the original word.

Lemma 1. *The extended code is determined by the Prüfer code.*

Proof. We are given $(y_1, y_2, \ldots, y_{n-2})$. First, set y_{n-1}. Then for each $i = 1, 2 \ldots, n-1$, let x_i be the smallest label not in the set $\{x_1, \ldots, x_{i-1}\} \cup \{y_i, \ldots, y_{n-1}\}$. That is, the entry x_i is the smallest possible label not appearing to its left in the first row or its right in the second row. For example,

$$\begin{pmatrix} & & & & & \\ 1 & 4 & 4 & 6 & 1 & \end{pmatrix} \to \begin{pmatrix} 2 & 3 & 5 & 4 & 6 & 1 \\ 1 & 4 & 4 & 6 & 1 & 7 \end{pmatrix}.$$

If our original sequence is indeed the Prüfer code of some tree, then note that the numbers not in the set $\{x_1, \ldots, x_{i-1}\} \cup \{y_i, \ldots, y_{n-2}\}$ are precisely the leaves of the current tree after i

deletions. Since the Prüfer coding process will choose the smallest current leaf, we conclude that the matrix generated by the above process is indeed the extended Prüfer code of the original tree. □

Now, given a $2 \times (n-1)$ matrix with entries in $\{1, 2, \ldots, n\}$ we define the graph whose edges are the columns of the matrix. If we begin with an arbitrary word of length $n-2$ from the alphabet $\{1, 2, \ldots, n\}$ and convert it to an "extended code" by the above process, we claim that the resulting graph is a tree, and moreover, the Prüfer code of this tree is the sequence we began with.

Lemma 2. *Given a "code," the columns of the corresponding "extended code" are a tree whose Prüfer code is the original "code."*

Proof. We will build the graph by starting with the edge (x_{n-1}, y_{n-1}) and then successively adding $(x_{n-1-i}, y_{n-1-i}$ for $i = 1, 2, \ldots, n-2$. We will show by induction that the resulting graph is a tree. Indeed, we begin with a single edge, which is a tree. Now suppose that the graph H consisting of the edges $(x_{k+1}, y_{k+1}), \ldots, (x_{n-1}, y_{n-1})$. We claim that the new graph H' obtained from H by adding the edge (x_k, y_k) is also a tree. [This is homework.]

Finally, it is easy to check that if we compute the Prüfer code of the resulting tree, we obtain our original "code." □

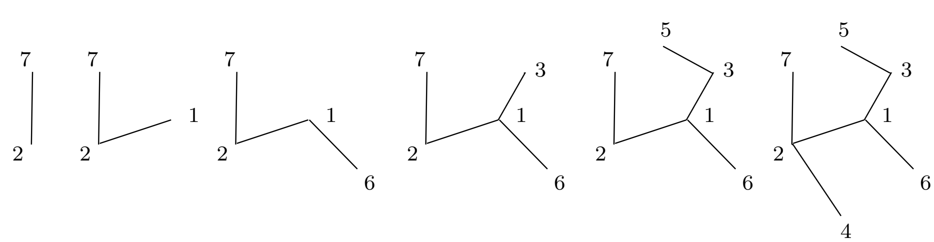
Combining these two lemmas, we have a procedure for converting a "code" into a tree:

$$\text{code} \;\rightarrow\; \text{extended code} \;\rightarrow\; \text{tree}$$

For example, if we begin with the code 23112 we obtain extended code

$$\begin{pmatrix} & & & & \\ 2 & 3 & 1 & 1 & 2 \end{pmatrix} \rightarrow \begin{pmatrix} 4 & 5 & 3 & 6 & 1 & 2 \\ 2 & 3 & 1 & 1 & 2 & 7 \end{pmatrix},$$

which in turn generates the following tree:



Corollary 3. *Let $d_1, d_2, \ldots, d_n$ be non-negative integers such that $\sum_i d_i = 2(n-1)$. Then the number of trees with vertices $\{1, 2, \ldots, n\}$ in which $\deg(\mathrm{i}) = \mathrm{d_i}$ is equal to the multinomial coefficient*

$$\binom{n-2}{d_1 - 1, d_2 - 1, \ldots, d_n - 1} = \frac{(n-2)!}{(d_1-1)!(d_2-1)!\cdots(d_n-1)!}.$$

Proof. We note that the number of trees of this type is equal to the number of Prüfer codes in which the symbol i appears $d_i - 1$ times. Given the collection of symbols in the word, there are $(n-2)!$ ways to put them in a sequence. However, we consider the $d_i - 1$ occurrences of the symbol i in the word to be indistinguishable. Hence we should divide by $(d_i - 1)!$ to cancel all of the possible permutations of these symbols. ☐

This brings us full circle, back to the original theorem of Cayley. From here, we can use the multinomial theorem (as Cayley did) to obtain the total of n^{n-2} trees.