Approximating the Number of Bases for Almost All Matroids *

Brian Cloteaux
National Institute of Standards and Technology
Gaithersburg, Maryland, USA
brian.cloteaux@nist.gov

Abstract

We define a class of matroids $\mathscr{A}$ for which a fully polynomial randomized approximation scheme (fpras) exists for counting the number of bases of the matroids. We then show that as the number of elements in a matroid increases, the probability that a matroid belongs to $\mathscr{A}$ goes to 1. We thus provide a fpras for counting the number of bases that applies to almost all matroids.

The general problem of counting the bases of matroids represented by an independence oracle is known to be #*P*-complete and so it strongly suggests that no polynomial time algorithm for this problem exists. Further, it has been shown that under this model, even finding a deterministic approximation for the number of bases is #*P*-complete [1]. Consequently, a majority of research on this problem has been in creating randomized approximation schemes for estimating the number of bases (for example [3, 2]).

While several randomized schemes have been produced, these schemes tend to be complicated and are able to count only specialized classes of matroids. In contrast, Chavez-Lomeli and Welsh [2] gave a very simple randomized algorithm for counting the bases of a matroid provided that the matroid belongs to a class with a property that they termed *frequent*. We extend the result of Chavez-Lomeli and Welsh to show that the bases of almost all matroids can be approximated using their algorithm. Our result does this by proving a conjecture of Chavez-Lomeli and Welsh [7] that there exists a frequent class containing a majority of all matroids.

To describe the Chavez-Lomeli and Welsh result, we first define the frequent property. We use the notation that if M is a matroid on a set E, then we denote $b(M)$ and r as the number of bases and the rank of M respectively. The class of

all matroids is denoted as $\mathscr{G}$ and the set of all matroids with exactly n elements is $\mathscr{G}_n$. A class $\mathscr{C}$ of matroids is *frequent* if there exists a polynomial p such that if $M \in \mathscr{C}$ on n elements has r rank, then

$$b(M) \cdot p(n) \geq \binom{n}{r}$$

Chavez-Lomeli and Welsh's main result is that for any frequent class there exists a *fully polynomial randomized approximation scheme* (fpras) for estimating the number of bases for the matroids in that class.

Our contribution is to show that if we pick any number $\varphi > 5/2$ then for the bound $p(n) = n^{\varphi}$ almost all matroids satisfy this condition. We define the following classes of matroids

$$\mathscr{A}_n = \left\{ M \mid M \in \mathscr{G}_n \text{ and } b(M) \geq \binom{n}{r} / n^{\varphi} \right\}$$

and

$$\mathscr{A} = \bigcup_{i \geq 0} \mathscr{A}_i$$

From a result in the Chavez-Lomeli and Welsh paper, it already follows that all paving matroids must be in $\mathscr{A}$.

We use the class $\mathscr{A}$ to create Algorithm 1. This algorithm is a slight modification of the original Chavez-Lomeli and Welsh algorithm and the proof that it is a fpras for all matroids in $\mathscr{A}$ follows directly from their paper. What we show is that almost all matroids are in $\mathscr{A}$ and so the Algorithm 1 must almost always work. More formally, we give the following theorem.

Theorem 1

$$\lim_{n \to \infty} \frac{|\mathscr{A}_n|}{|\mathscr{G}_n|} = 1$$

We start by defining a function f between the class of all matroids and a subset of the strings over the alphabet $\{0,1\}$. In other words, $f : \mathscr{G} \to S$ where $S \subseteq \{0,1\}^*$. A matroid M is uniquely encoded by f with the following procedure. We enumerate all the ways to select r base elements from the n element ground set. Each base in M is specified by its order in the enumeration of the selection of its rank elements. To make the string unique, we list all the bases in their enumeration order. Specifying all the bases of M requires $b(M) \cdot \lceil \log_2 \binom{n}{r} \rceil$ bits. In addition, we need to record the sizes of n and r. They require $\lceil \log_2 n \rceil$ and $\lceil \log_2 r \rceil$ bits respectively and we add them as prefixes to the string encoding the bases. In order to be able to distinguish the sizes of n and r from the encoding of the bases, we use the standard technique of repeating each bit in the n and r fields and using the

Require: a matroid $M \in \mathscr{A}$
Ensure: $\Pr\left[\left|\frac{Z}{b(M)} - 1\right| > \varepsilon\right] \leq \frac{1}{4}$
  Determine rank r of M
  $n \leftarrow |E|$
  $t \leftarrow \lceil 4 \cdot n^{\varphi} \cdot \varepsilon^{-2} \rceil$
  for $i \leftarrow 1$ to t **do**
    Select uniformly at random an r-subset A of E
    if A is base of M **then**
      $Z_i \leftarrow 1$
    else
      $Z_i \leftarrow 0$
    end if
  end for
  $\hat{Z} \leftarrow \frac{Z_1 + Z_2 + \ldots + Z_t}{t}$
  $Z \leftarrow \hat{Z} \cdot \binom{n}{r}$
  return Z

Figure 1: Fully polynomial random approximation scheme for counting the number of bases for any matroid in the class $\mathscr{A}$

sequence 01 as a separator between these fields. Thus the number of bits needed to uniquely encode any matroid M using f is

$$|f(M)| = b(M) \cdot \lceil \log_2 \binom{n}{r} \rceil + 2 \cdot \lceil \log_2 n \rceil + 2 \cdot \lceil \log_2 r \rceil + 4$$

Using this function we now define a class of matroids which uses the Kolmogorov complexity $K(s)$ of its string mapping s as its membership criteria

$$\mathscr{F}_n^c = \{M | M \in \mathscr{G}_n \text{ and } K(f(M)) \geq \lfloor \log_2 |\mathscr{G}_n| \rfloor - c\}$$

for the natural number $c \geq 0$. In other words, $\mathscr{F}_n^c$ is the set of all matroids with n elements and are c-complex. From a simple counting argument (see theorem 2.2.1 in [5]), it follows that

$$\frac{|\mathscr{F}_n^c|}{|\mathscr{G}_n|} > 1 - 2^{-c}$$

We want to show that for all $c \geq 0$ there exists a constant $p_c \in \mathbb{N}$ such that for all $n \geq p_c$ if $M \in \mathscr{F}_n^c$ then $b(M) \cdot n^{\varphi} \geq \binom{n}{r}$. That implies that for all $n \geq p_c$

$$\mathscr{F}_n^c \subseteq \mathscr{A}_n$$

and so our result would be proved.

To see this, choose an arbitrary $M \in \mathscr{F}_n^c$. From the definition of $\mathscr{F}_n^c$, we know that

$$\lfloor \log_2 |\mathscr{G}_n| \rfloor - c \leq K(f(M)) \leq b(M) \cdot \log_2 \binom{n}{r} + 2 \cdot \lceil \log_2 n \rceil + 2 \cdot \lceil \log_2 r \rceil + 4 + a$$

where a is a constant that represents the additional number of bits needed to encode a Turing machine which has the string $f(M)$ encoded in it and simply prints the string out. Using a result of Knuth[4], we put a lower bound on the size of the class $\mathscr{G}_n$.

$$\log_2 \log_2 |\mathscr{G}_n| \geq n - \frac{3}{2} \log_2 n + O(\log \log n)$$

Combining these two results, we get

$$n - \frac{3}{2} \log_2 n + O(\log \log n) \leq \log_2 K(f(M)) + c \leq$$
$$\log_2 b(M) + \log_2 \log_2 \binom{n}{r} + O(\log \log n)$$

If we assume that $M \notin \mathscr{A}$, i.e. $b(M) < \binom{n}{r} / n^{\varphi}$ then

$$n - \frac{3}{2} \log_2 n + O(\log \log n) < \log_2 \binom{n}{r} - \varphi \log_2 n + \log_2 \log_2 \binom{n}{r} + O(\log \log n)$$

Since $n > \log_2 \binom{n}{r}$ then

$$n - \frac{3}{2} \log_2 n + O(\log \log n) < n - \varphi \log_2 n + \log_2 n + O(\log \log n)$$

or equivalently

$$\left(\varphi - \frac{5}{2} \right) \cdot \log_2 n < O(\log \log n)$$

Since $\varphi > \frac{5}{2}$ then for n large enough this statement is always false. Thus, for any c there can be at most a finite number of sets $\mathscr{F}_n^c$ which contain a matroid M where $b(M) \cdot n^{\varphi} < \binom{n}{r}$. $\square$

There are two notes concerning the practicality of this algorithm we should mention. The first is that even though we do not know if an arbitrary matroid is in the class $\mathscr{A}$, we can use the Chernoff bound while our algorithm is performing its sampling to verify the condition.

A second point is that although we have shown that the algorithm can be applied with a probability approaching 1 to large matroids there is a question about how well it applies for smaller matroids. While not a complete answer, it can be quickly verified using Mayhew and Royle's database of small matroids [6] that all matroids with at most nine elements are in the class $\mathscr{A}$ for any $\varphi > 5/2$. It seems that finding a matroid not in $\mathscr{A}$, even for small matroids, is the rare exception.

References

[1] Y. AZAR, A. Z. BRODER, AND A. M. FRIEZE, *On the problem of approximating the number of bases of a matroid*, Information Processing Letters, 50 (1994), pp. 9–11.

[2] L. CHAVEZ-LOMELI AND D. WELSH, *Randomised approximation of the number of bases*, Contemporary Mathematics, 197 (1996), pp. 371–376.

[3] T. FEDER AND M. MIHAIL, *Balanced matroids*, in STOC '92: Proceedings of the twenty-fourth annual ACM symposium on Theory of computing, New York, NY, USA, 1992, ACM, pp. 26–38.

[4] D. E. KNUTH, *The asymptotic number of geometries*, Journal of Combinatorial Theory (A), 16 (1974), pp. 398–400.

[5] M. LI AND P. VITÁNYI, *An Introduction to Kolmogorov Complexity and its Applications*, Springer-Verlag New York, Inc., second ed., 1997.

[6] D. MAYHEW AND G. F. ROYLE, *Matroids with nine elements*, Journal of Combinatorial Theory (B), 98 (2008), pp. 415–431.

[7] D. WELSH, *Some problems on approximate counting in graphs and matroids*, in Research Trends in Combinatorial Optimization, 2009, pp. 523–544.