Asymptopia of $m(n)$

Recall the theorem proven in previous notes.

**Theorem 0.1 (Erdős 1964)** If there exists $v$ with

$$2^v \left(1 - \frac{2(v/2)^2}{(v)}\right)^m < 1,$$

then $m(n) \leq m$, i.e., there is a family $A_1, A_2, \ldots, A_m \subset \{1, \ldots, v\}$ with no 2-coloring.

We wish to solve for $m$ and then choose $v$ to maximize $m$. We begin by using the inequality $1 - p \leq e^{-p}$ to find

$$m \geq \left\lceil \frac{v \ln 2}{2^{(v/2)^2}/(v)} \right\rceil$$

Define

$$p = \frac{2^{(v/2)^2}}{(v)}$$

We need to explore the asymptotics of $p$. To gain some intuition about how tight a bound we need, recall what the quantity $p$ represents. In our original proof, $p$ was the probability that a randomly chosen set from a two-colored $v$-sized universe is monochromatic. This is equivalent to the probability in the following experiment: Fill a bag with $\frac{v}{2}$ red balls and $\frac{v}{2}$ blue balls. Draw $n$ balls at random without replacement. Then $p$ is the probability that all $n$ balls have the same color. One might be tempted to bound $p$ using the approximation $\binom{n}{k} \approx \frac{n^k}{k!}$. This approximation yields $p = 2^{1-n}$, which is the same as the probability of drawing $n$ monochromatic balls with replacement. Intuitively it is clear that for small $v$, drawing with replacement and drawing without replacement yield different results. As we are trying to minimize $m$ (and thus will be dealing with small $v$), we look for a better approximation.

First we rewrite $p$.

$$p = \frac{2^{(v/2)^2}}{(v)}$$
\[
\begin{align*}
\prod_{i=0}^{n-1} \frac{v/2 - i}{v - i} &= 2^{1-n} \prod_{i=0}^{n-1} \left(1 - \frac{i}{v - i}\right) \\
&= 2^{1-n} \exp\left(\sum_{i=0}^{n-1} \ln\left(1 - \frac{i}{v - i}\right)\right)
\end{align*}
\]

By the Taylor expansion,
\[
\ln(1 - \epsilon) \sim -\epsilon - \frac{\epsilon^2}{2} - \ldots
\]

Therefore, for (see later for what happens when this doesn’t hold) \(v \gg n^{1.5}\),
\[
\ln\left(1 - \frac{i}{v - i}\right) \sim -\frac{i}{v}
\]

Note we can safely ignore the second order term of the Taylor expansion because \(\sum_{i=0}^{n} i^2 \sim n^3\) and, by assumption, \(v^2 \gg n^3\), so the \((\frac{i}{v-i})^2\) term will tend to zero after the summation.

This gives our final approximation for \(p\),
\[
p \sim 2^{1-n} \exp\left(-\frac{n^2}{2v}\right)
\]

from where
\[
m(n) \leq \left\lceil 2^{n-1} (\ln 2) ve^{n^2/2v} \right\rceil
\]

for \(v \gg n^{1.5}\). Let \(z = ve^{n^2/2v}\). We want to find \(v\) that minimizes \(z\). Using standard calculus techniques,
\[
y \equiv \ln z = \ln v + \frac{n^2}{2v}
\]
\[
y' = \frac{1}{v} - \frac{n^2}{2v^2} = 0
\]
giving
\[
v = \frac{n^2}{2}
\]

for a final result
\[
m(n) \leq \left\lceil e \ln 2 \frac{2n^2}{4} \right\rceil
\]

While this is the correct answer we haven’t completed our rigorous argument. We assumed \(v \gg n^{1.5}\) in making our approximations. We now need to check that if \(v\) is not that large then the value is bigger than \(cn^22^n\). This is a somewhat typical (and annoying) part because here the value will be very much bigger than \(n^22^n\) – so we can use crude tools. (Often in asymptotics the calculations are easier when the results are tighter.) Let’s take, leaving room, \(v \leq n^{1.51}\). The value of \(p\) is (remember that \(n\) is fixed) an increasing function of \(v\) (check the representation as a product) and so it suffices to look at \(v = n^{1.51}\). But there the approximation \(p \sim 2^{1-n} \exp(-n^2/2v)\) does hold and this gives a \(p\) which is exponentially small and so a bound on \(m\) which is exponentially large, so it is way off from the real value.
Asymptopia of $\binom{n}{k}$

In many of the problems in this field, we are faced with an $\binom{n}{k}$ that we need to approximate. There is no single approximation for $\binom{n}{k}$ which always works. Rather, the approximation one should use depends greatly on the relationship between $n$ and $k$. Here we discuss several approximations of $\binom{n}{k}$ for various relationships between $n$ and $k$.

First, if both $n$ and $k$ are fixed, $\binom{n}{k}$ has a definite value. Use it. Also, if just $k$ is fixed, clearly $\binom{n}{k} \sim \frac{n^k}{k!}$ is a good approximation. When both $n$ and $k$ grow, things get more complicated. First, we consider the case where $k = o\left(\frac{n}{1/2}\right)$. Notice $\binom{n}{k} = \frac{n^k}{k!} \prod_{i=0}^{k-1} \left(1 - \frac{i}{n}\right)$

Using Stirling’s formula, we can approximate $k! \sim k^k e^{-k} (2\pi k)^{1/2}$, and $n^k$ is fine as is. What we have left to deal with is the $\ln(1 - \frac{i}{n})$ term. Here we can use the Taylor expansion

$$\ln(1 - \epsilon) = -\epsilon - \frac{\epsilon^2}{2} - O(\epsilon^3)$$

from where

$$\sum_{i=0}^{k-1} \ln(1 - \frac{i}{n}) = -\frac{k^2}{2n} - \frac{k^3}{6n^2} - O\left(\frac{k^4}{n^3}\right)$$

Of course, we can retain even more terms of the Taylor expansion if we need to. For each extra term we get yet another estimate that generalizes the previous estimate.

The first three terms listed above yield the following approximations:

- For $k = o(n^{1/2})$, $\sum_{i=0}^{k-1} \ln(1 - \frac{i}{n}) \sim 0$.
- For $k = o(n^{2/3})$, $\sum_{i=0}^{k-1} \ln(1 - \frac{i}{n}) \sim -\frac{k^2}{2n}$.
- For $k = o(n^{3/4})$, $\sum_{i=0}^{k-1} \ln(1 - \frac{i}{n}) \sim -\frac{k^2}{2n} + \frac{k^3}{6n^2}$.

If $k = o(n)$, we have a useful logarithmically asymptotic result. Since $k! \geq (\frac{k}{e})^k$ for all $k$,

$$\binom{n}{k} \leq \frac{n^k}{k!} \leq \left(\frac{ne}{k}\right)^k$$

Furthermore, for some computable constant $c$,

$$\binom{n}{k} \geq \frac{(n-k)^k}{k!} \geq \left(\frac{ne}{k}\right)^k ck^{-1/2} (1 - \frac{k}{n})^k$$
where \( c k^{-1/2}(1 - \frac{k}{n})^k = n^{o(k)} \). Combining these inequalities, we have the result

- For \( k = o(n) \), \( \binom{n}{k} = \left( \frac{ne}{k} \right)^k (1 + o(1)) \).

If \( k = \alpha n \) for some \( \alpha \in (0, 1) \), we can use Stirling’s formula to see

\[
\binom{n}{\alpha n} = \frac{n!}{(\alpha n)!!((1 - \alpha)n)!} \sim \frac{n^{n} e^{-n (2\pi n)}^{\frac{1}{2}}}{\left( \frac{\alpha n}{2\pi (1 - \alpha) n} \right)^{\frac{1}{2}}}
\]

For convenience, let \( c_\alpha \equiv (2\pi \alpha (1 - \alpha))^{-\frac{1}{2}} \). We use the entropy function \( H(\alpha) \) defined as

\[
H(\alpha) \equiv -\alpha \log \alpha - (1 - \alpha) \log (1 - \alpha)
\]

One sets \( H(0) = H(1) = 0 \) so that this is continuous on \([0, 1]\). \( H \) is symmetric about \( \alpha = \frac{1}{2} \) and assumes its maximum at \( \alpha = \frac{1}{2} \) with \( H(1/2) = 1 \).

The slope of \( H(\alpha) \) at 0 and 1 is \( \infty \) and \( -\infty \). We can estimate \( H(1/2 + \epsilon) \) as

\[
H(1/2) + H'(1/2) \epsilon + \frac{H''(1/2)}{2} \epsilon^2 + O(\epsilon^3) = 1 - \frac{1}{2} \epsilon^2 + O(\epsilon^3).
\]

Now we can succinctly state our result.

- For \( k = \alpha n \), \( \binom{n}{\alpha n} \sim c_\alpha n^{-\frac{1}{2}} 2^{n H(\alpha)} \).

### Improving bounds of \( m(n) \)

In 1963 Erdős gave a lower bound for \( m(n) \) of \( 2^{n - 1} \). In 1964, he gave an upper bound of \( \frac{e \ln 2}{4} n^{2n} \). These bounds are quite far apart, especially if you consider \( 2^n \) as 1. In 1978 Beck improved Erdős’s result giving a lower bound of \( 2^n n^{1/3} \). Then Radhakrishnan and Srinivasan improved the lower bound to \( 2^n \left( \frac{n}{\ln n} \right)^{1/2} \). There is, of course, still room for improvement, but these are the best known bounds to date.

Radhakrishnan and Srinivasan arrived at their result by proposing a randomized algorithm for two-coloring a family and then analyzing its failure probability. We present the algorithm they proposed \(^1\) The basic idea of the algorithm is to color all the points randomly and then, on a second pass, fix the points that cause sets to be monochromatic. Of course, the more clever the algorithm the harder it is to analyze. The art of the proof is to not let the algorithm be too clever so that you can’t analyze it but clever enough that it works. Not easy!

**Algorithm**

1. Color each point \( v \) red or blue randomly with probability \( \frac{1}{2} \).

\(^1\) we’ll indicate the analysis in the lectures, time permitting
2. Assign to each point \( v \) a uniformly identically distributed birthtime \( t_v \in [0, 1] \). Let time go from 0 to 1. When \( v \) is born, flip a coin with probability \( p \). Switch the color of \( v \) if

(a) coin is heads

(b) \( v \) is still dangerous

where \( v \) is called still dangerous if for some \( i \),

(i) \( v \in A_i \)

(ii) \( A_i \) was monochromatic after step 1

(iii) no \( w \in A_i \) has had its color changed (i.e. \( A_i \) is still monochromatic)

Indeed, many of the more recent (and more exciting) results involve the analysis of Random Processes or Randomized Algorithms. Let us even give this a name:

**Modern Erdős Magic**: If there is a randomized algorithm that creates an object with desired properties with positive probability that that object must exist.

For example, let \( BAD \) be the event that at the end of the Radhakrishnan Srinivasan randomized algorithm there is a set \( A_i \) which is monochromatic. Once one shows (not easy!) that \( \Pr[BAD] < 1 \) then a good two-coloring must exist.