ANALYSIS OF THE BEHAVIOR OF S_t/t

MICHAEL KANE

As part of the research being done on modeling stochastic process, specifically with respect to (Achlioptas and Molloy 1997), this paper investigates the ratio of the sum of binomial random variables with a time dependent p parameter to time.

1. INTRODUCTION

Consider an urn which starts out with n balls. At the first time step, W_1 balls are taken where

$$W_1 \sim \text{Bin}(n, c/n).$$

That is, a binomial number of balls is taken. More generally, this can be written as

$$W_{t+1}|W_t \sim \text{Bin}(n - W_t, c/n)$$

We would like to get the distribution function for

$$S_t = \sum_{i=1}^{t} W_t$$

At each time there is a c/m change that a ball in the urn will be picked to be removed. The probability that a ball is not chosen at time 1 is $1 - c/m$. The probability that a ball is not chosen at time 2 is the probability of the intersection of the events that it is not chosen at time 1 and time 2, i.e. $(1 - c/m)^2$. Inductively it can be seen that the probability that a ball is not chosen at time t is $(1 - c/m)^t$. Then,

$$\sum_{i=1}^{t} b_t = \text{Bin}(n, 1 - (1 - c/m)^t).$$

Now, consider the case where it is guaranteed that at least 1 ball is take at each time. This can similarly be modeled as a sum of binomials plus 1 at each time step. In other words, for

$$b_t = \text{Bin}(n - 1, 1 - (1 - c/(n-1)^t)$$

Date: May 24, 2007.

we get

$$S_t = t + \sum_{i=1}^{t} b_t.$$

2. THE EXPECTED BEHAVIOR OF S_t/t

Let $\eta = \mathbb{P}_t b_{t+1}$. Using the approximation $b_t \sim \text{Bin}(n, 1-(1-c/n)^t)$ we can write

$$\begin{aligned} \frac{S_t}{t} &= \frac{S_0 + b_1 + 1 + b_2 + 1 + ... + b_t + 1}{t} \\ &= \frac{b_1 - \eta_1 + b_2 - \eta_2 + ... + b_t - \eta_t}{t} + \frac{\eta_1 + 1 + \eta_2 + 1 + ... + \eta_t + 1}{t}. \end{aligned}$$

Neglecting S_0, and taking the expectation we see that to understand the behavior of S_t/t we need to understand the expression

$$\begin{aligned} \mathbb{P}\left[\frac{\eta_1 + 1 + \eta_2 + 1 + ... + \eta_t + 1}{t}\right] &= \frac{n}{t}\sum_{i=1}^{t}\left(1 - \left(1 - \frac{c}{n}\right)^i\right) + 1 \\ &= \frac{n}{t}\left(t - \sum_{i=1}^{t}\left(1 - \frac{c}{n}\right)^i\right) + 1 \end{aligned} \tag{2.1}$$

Remembering

$$\sum_{i=0}^{\infty} a^i = \frac{1}{1 + a^{-1}} \text{ for } |a| < 1,$$

Equation (2.1) can be rewritten as

$$\begin{aligned} \frac{n}{t}\left(t - \frac{n}{c} + 1 + \sum_{i=t+1}^{\infty}\left(1 - \frac{c}{n}\right)^i\right) &= \frac{n}{t}\left(t - \frac{n}{c} + 1\right) + 1 + o\left(n\left(1 - \frac{c}{n}\right)^t\right) \\ &\approx \frac{n}{t}\left(t - \frac{n}{c} + 1\right) + 1. \end{aligned} \tag{2.2}$$

When looking at this approximation it should be noted that $o(n(1-c/n)^t)$ is asymptotic in t. Since $t \leq n$, this means that for this approximation to be close, n must be sufficiently large and t must be of a significant proportion to n. It should also be noted that this term is strictly positive for $c \leq n$. This means that the approximation also provides an upper bound.

Using the results from Approximation (2.2),

$$\mathbb{P}\frac{S_t}{t} \approx \frac{S_0}{t} + \frac{n}{t}\left(t - \frac{n}{c} + 1\right) + 1. \tag{2.3}$$

Again, neglecting S_0 (or setting $S_0 = 0$), as t approaches n

$$\mathbb{P}\frac{S_t}{t} \approx n(1 - 1/c) + 2.$$

Recalling that $1 \leq c \leq n$ it can be seen that when c is small, as is the case of (Achlioptas and Molloy 1997), the slope goes to 2. This implies that we would expect S_t to reach n in about $n/2$ steps. In the more exotic case where c goes to n the slope goes to $n + 1$. However, in practice we know that the slope can not be more than n which implies that the $S_1 = n$. In summary, we can say that for large enough n and t,

$$2 \leq \mathbb{P}\frac{S_t}{t} \leq n. \tag{2.4}$$

Now, we would like to get some idea of how much variation we can expect around the slope.

3. The Tail Bound for S_t/t

Based on the expected behavior of S_t it is clear that S_t can reach n in 1 to $n/2$ time steps. It would be useful to get the tail probability of S_t/t. This will be done in a manner similar to Pollard (2001, Chapter 4), where the maximal inequality is found for the first time a sum of independent passes a threshold.

We would like to find β, a positive constant, such that

$$\mathbb{P}\{S_n - S_i \leq \varepsilon\} \geq 1/\beta.$$

Since $S_n = n$ and $S_i \geq i$,

$$S_n - S_i \leq n - i$$

Let ε be a positive constant such that $\varepsilon < n$, then

$$\mathbb{P}\{S_n - S_i \leq \varepsilon\} \geq \varepsilon\mathbb{P}[(S_n - S_i)^{-1}] \geq \frac{\varepsilon}{n - i} = 1/\beta.$$

Now, define

$$\tau := \{\text{first } i \text{ for which } S_i = n\}.$$

$$
\begin{aligned}
&\mathbb{P}\{\tau = i \text{ and } S_i \geq n \text{ for some } i\} \\
&\qquad = \sum_{i=1}^{n} \mathbb{P}\{\tau = i, S_i \geq n\} \\
&\qquad \leq \sum_{i=1}^{n} \mathbb{P}\{\tau = i, S_i \geq n\}\beta\mathbb{P}\{S_n - S_i \leq \varepsilon\} \\
&\qquad \leq \beta \sum_{i=1}^{n} \mathbb{P}\{\tau = i, S_i \geq n, S_n - S_i \leq \varepsilon\} \\
&\qquad \leq \beta \sum_{i=1}^{n} \mathbb{P}\{\tau = i, S_n \geq n - \varepsilon\}.
\end{aligned}
$$

Substituting for β and using the Markov Inequality,

$$
\begin{aligned}
&\mathbb{P}\{\tau = i \text{ and } S_i \geq n \text{ for some } i\} \\
&\qquad = \beta\mathbb{P}\{S_n \geq n - \varepsilon\} \\
&\qquad \leq \frac{(n-i)}{\varepsilon(n-\varepsilon)} n \left(1 - (1 - c/n)^n\right) \\
&\qquad \leq \frac{n(n-i)}{\varepsilon(n-\varepsilon)}.
\end{aligned}
$$

The denominator of the last inequality is maximized for $\varepsilon = n/2$ and we get

$$
\mathbb{P}\{\tau = i \text{ and } S_i \geq n \text{ for some } i\} \leq 4(1 - i/n). \tag{3.1}
$$

It should be noted that for this inequality to be useful, we need $i/n \geq 3/4$.

REFERENCES

Achlioptas, D. and M. Molloy (1997). The analysis of a list-coloring algorithm on a random graph (extended abstract). *Proceedings of the 38th IEEE Symposium on Foundations of Computer Science*, 204–212.

Pollard, D. (2001). *A User's Guide to Measure Theoretic Probability*. Cambridge University Press.