How I wasted too long finding a concentration inequality for sums of geometric variables

Daniel G. Brown

David R. Cheriton School of Computer Science
University of Waterloo
Waterloo ON N2L 3G1 Canada
browndg@uwaterloo.ca

Abstract. I wanted a concentration inequality for sums of iid geometric random variables. This took way too long.

1 What I wanted, and how I didn't get it

I needed a concentration inequality for sums of iid geometric random variables. I need this because I am interested in the height of a certain type of random tree; in what follows, you can pretend it's a random binary search tree, but that doesn't really matter.

First, some reminders. A *geometrically distributed* random variable $G(p)$ with expectation p has probability distribution function $\Pr[G(p) = i] = (1/p)(1 - 1/p)^{i-1}$. One way to think about it is that it's the index of the first coin flip that comes up "heads" if you repeatedly toss a coin that's "heads" with probability $1/p$ and "tails" with probability $1 - 1/p$.

Let the variable $Y(n, p)$ be the sum of n iid $G(p)$ random variables. (The name of such a random variable is "negative binomial", but I'd forgotten that.)

Because expectations add, $E[Y(n, p)] = np$. But how strongly concentrated is that? In my case, each of the $G(p)$ variables talks about the number of additions I have to make to the path in a tree from the root to a given leaf before the subtree at the current node is dropped by a factor of at least $6/5$ (starting from T, the number of leaves in the tree). In our application, p turns out to be a constant. (Perhaps 21, but that doesn't matter much.) The length of the path from a root to a leaf is a random variable bounded above by a variable with distribution $Y(\log_{6/5} T, 21)$; since $\ln 1.2 \approx .18$, this is approximately $Y(5.5 \ln T, 21)$. Again, the expectation of this random variable is $5.5 \cdot 21 \ln T$, or $115 \ln T$.

But how strongly concentrated is it? For example, what's the probability that it's more than twice its mean?

The standard way for computer scientists to address this is with concentration inequalities, like Chernoff bounds or Azuma's inequality or Bernstein's inequality. But all of the versions of this that I know assume that the individual summands (the geometric variables) are bounded above by some known amount. That's not true here; geometric variables have unbounded value (since you could have a huge number of coin flips before it finally comes up heads; see [5]). To be more proper, these summands aren't Lipschitz.

You might imagine, "what if I bound it by some really big bound beyond which the probability is really tiny?" But that doesn't seem to work well: geometric distributions are pretty pernicious, and all of the methods I know for that don't work in this domain. One recent book gives relatively little time over to sums of unbounded variables; when it does cover them, it's in the context of graph properties much more complicated than the ones I cared about [3]. (It also gives an interlude on this called "The Infamous Upper Tail," which seemed really quite ominous.)

But it has to be possible to give this sort of bound, right? I mean, this is the sum of n iid variables with mean p and standard deviation p; everyone knows that converges to a normally distributed variable with mean np and standard deviation $p\sqrt{n}$. What the heck?

Well, I'm dumb, so I kept looking. I remembered, finally, that these were "negative binomial" random variables, so I did the obligatory Google search for things like "negative binomial Chernoff" and "negative binomial tail inequality" with no real luck. Well, that's not even true. There's a paper [4] that might prove something almost identical to what I wanted, but I can't read past the second paragraph, despite many attempts. I don't even see anything like an exponentially decaying tail. I don't know what their D operator does. It mostly reminded me how bad of a mathematician I am.

I vaguely think that Gord Willmot (from here at Waterloo, even!) may have proved something like what I want [6], but we don't have that journal here.

And good grief, this should be easy.

2 The way to get it

Then I looked at "negative binomial distribution" in Wikipedia [2], and it of course has a link to the cumulative distribution function for negative binomial variables: it's the *regularized incomplete β function* [1]. Oh, dear. That's scary. It has calculus in it. I hate calculus.

And then the answer is right there. Consider $\Pr[Y(n,p) > knp]$. This is the probability that it takes more than knp trials before we get to the n-th head, right? Well, that's the probability that in knp trials, there are fewer than n heads. (And yes, this is pointed out in the Wikipedia article [2].)

Let $\{X_i(1/p)\}$ for $i > 0$ be a collection of iid Bernoulli trials, each of which has probability $1/p$ of having value 1 and probability $1 - 1/p$ of having value 0. Then $\Pr[Y(n,p) > knp] = \Pr[\sum_{i=1\dots knp} X_i(1/p) < n]$. But that latter sum is a binomial random variable, $B(knp, 1/p) = \sum_{i=1\dots knp} X_i(1/p)$, and I know how to study them. I just want $\Pr[B(knp, 1/p) < n]$.

OK, now what? Well, the mean of $B(knp, 1/p)$ is kn, so being less than n means being less than $1/k$ times the mean. A standard theorem ([3], p. 6) says that for a variable of the sort of $B(knp, 1/p)$, $\Pr[X < (1-\varepsilon)E[X]] \leq \exp(-\frac{\varepsilon^2}{2}E[X])$. Here, $\varepsilon = 1 - 1/k$ and $E[X] = kn$, so this is just $\Pr[B(knp, 1/p) <$

$n] \leq \exp(\frac{-kn(1-1/k)^2}{2})$, which, indeed, is the kind of strong concentration I wanted.

3 A proper statement

Theorem 1. *Let $Y(n,p)$ be a negative binomially distributed random variable that arises as the sum of n iid geometrically distributed random variables with expectation p.*

Then $E[Y(n,p)] = np$, and for $k > 1$, $\Pr[Y(n,p) > knp] \leq \exp(\frac{-kn(1-1/k)^2}{2})$.

4 The upshot

In my specific case, p is 21 and n is $5.5 \ln T$, so we wind up with $\Pr[Y > 2 \cdot E[Y]] \leq T^{-1.35}$, which means that the path in the tree to any leaf isn't likely to be more than twice our upper bound on its expected length. And indeed, since the tree has T leaves, the probability that *any* leaf has path length more than twice our upper bound on the expected length is at most $T \cdot T^{-1.35} = T^{-0.35} \approx \frac{1}{\sqrt[3]{T}}$, which is $o(1)$.

And now I feel really stupid over how long this took.

References

1. Beta function. Wikipedia article (September 2010)
2. Negative binomial distribution. Wikipedia article (September 2010)
3. Dubhashi, D., Panconesi, A.: Concentration of measure for the analysis of randomized algorithms. Cambridge University Press (2009)
4. Prokhorov, Y.V., Viskov, O.V., Khokhlov, V.I.: Analogues of the chernoff inequality for negative binomial ditribution. Teor. Veroyatnost. i Primenen. 50(2), 379382 (2005), in Russian
5. Stoppard, T.: Rosencrantz and Guildenstern are dead. Faber and Faber (1980)
6. Willmot, G., Lin, X.: Upper bounds for the tail of the compound negative binomial distribution. Scandinavian Actuarial Journal pp. 138–148 (1997)