Sharp and Simple Bounds for the raw Moments of the Binomial and Poisson Distributions

Thomas D. Ahle* thomas@ahle.dk
University of Copenhagen, BARC, Facebook

29th March 2021

Abstract

We prove the inequality $\mathrm{E}[(X/\mu)^k] \le (\frac{k/\mu}{\log(k/\mu+1)})^k \le \exp(k^2/(2\mu))$ for sub-Poissonian random variables, such as Binomially or Poisson distributed random variables with mean μ. This improves over previous uniform bounds for the raw moments of those distributions by a factor exponential in k.

1 Introduction

Suppose we are given an urn of n balls, each colored red with probability p and otherwise blue. What is the probability that a sample of k balls, with replacement, from this urn has only red balls? Such questions are important in sample efficient statistics and the derandomization of algorithms.

If $X \sim \mathrm{Binomial}(n,p)$ denotes the number of red balls in the urn, the answer is of course given by $P = \mathrm{E}(X/n)^k$. A simple application of Jensen's inequality shows $P \ge p^k$. In the range $k \le \sqrt{np}$, this is nearly tight, as previous bounds (see (1) below) were able to show $P \le C^k p^k$, for some universal constant $C > 1$. In this note we improve the upper bounds on Binomial moments, getting rid of the k in the exponent of C. Specifically we get $P \le p^k e^{k^2/(2pn)}$, which is asymptotically tight with the lower bound as $k = o(\sqrt{np})$.

A direct approach to computing the Binomial moments expands them using the Stirling numbers of the second kind [9]: $\mathrm{E}\,X^k = \sum_{i=0}^k \left\{ {k \atop i} \right\} n^{\underline{i}} p^i$. This can be derived [5, 10], for example, by combining the much easier to compute factorial moments $\mathrm{E}\,X^{\underline{k}} = \binom{n}{k} p^k$. Taking the first two terms, one finds that $\mathrm{E}\,X^k = (np)^k \left(1 + \frac{k(k-1)(1-p)}{2np} + O(1/n^2)\right)$. However, it is not clear what happens when k can not be considered constant with respect to n. Similarly, for the Poisson distribution the moments can be written as the so-called Touchard polynomials [12] in μ: $\mathrm{E}\,X^k = \sum_{i=0}^k \left\{ {k \atop i} \right\} \mu^i$, but this again doesn't yield a simple, uniform bound.

*The author would like to thank Robert E. Gaunt for encouragement and helpful suggestions.

Specializing a beautiful result on moments of independent random variables by Latała [6], we get for Binomial and Poisson random variables (also Pinelis [8]):

$$\left(c\,\frac{k/\mu}{\log(1+k/\mu)}\right)^k \le \mathrm{E}(X/\mu)^k \le \left(C\,\frac{k/\mu}{\log(1+k/\mu)}\right)^k \tag{1}$$

for some universal constants $c < 1 < C$. Latała's methods are general and allow for bounds of the central moments of similar strength. However, as $k/\mu \to 0$, the upper bound (1) is dominated by the factor C^k, which we manage to remove in this note.

Another option is the use a Rosenthal bound, such as the following by Berend and Tassa [1] (see also Johnson et al. [4])

$$\mathrm{E}\,X^k \le B_k \max\{\mu, \mu^k\}.$$

Here B_k is the kth Bell number. Berend and Tassa [1] show $B_k < \left(\frac{0.792k}{\log(k+1)}\right)^k$ and it is known [3] that $B_k^{1/k} = \frac{k}{e\log k}(1+o(1))$ as $k \to \infty$. Indeed, recently Ostrovsky [7] gave a bound the Bell numbers using a technique very similar to that of this paper. These results show that for $k > \mu$ it is possible to improve upon the constant in our result; however for k/μ small such bounds are less useful.

2 Bounds

Theorem 1. *Let X be a non-negative random variable with mean $\mu > 0$ and moment-generating function $\mathrm{E}[\exp(tX)]$ bounded by $\exp(\mu(e^t - 1))$ for all $t > 0$. Then for all $k > 0$:*

$$\mathrm{E}[X^k] \le \left(\frac{k}{\log(1+k/\mu)}\right)^k.$$

A standard logarithmic bound, $\frac{x}{\log(1+x)} \le 1 + x/2$ (see e.g. [11] eq. 6), implies the corollary

$$\mathrm{E}\left[(X/\mu)^k\right] \le \left(1 + k/(2\mu)\right)^k \le \exp\left(k^2/(2\mu)\right).$$

Random variables satisfying the requirement $\mathrm{E}[\exp(tX)] \le \exp(\mu(e^t - 1))$ are known as sub-Poissonian and include many simple distributions, such as the Poisson or Binomial distribution. We give more examples in Section 3.

Technically our bound is shown using the moment-generating function. This in turn involves some particularly sharp bounds on the Lambert-W function, which is defined by $W(x)e^{W(x)} = x$.[1] We will use the following lemma:

Lemma 1 (Hoorfar and Hassani [2])**.** *For all $y > 1/e$ and $x > -1/e$,*

$$e^{W(x)} \le \frac{x+y}{1+\log y}. \tag{2}$$

[1] The Lambert-W function has multiple branches. We are interested in the main one, corresponding to $x > 0$.

Proof. Starting from $1+t \le e^t$ substitute $\log(y)-t$ for t to get $1+\log y - t \le ye^{-t}$. Multiplying by e^t we get $e^t(1+\log y) \le te^t + y$. Let $t = W(x)$ s.t. $te^t = x$. Rearranging we get (2). □

Taking $y = e^{W(x)}$ in (2) makes the two sides equal. Hoorfar and Hassani make various substitutions, resulting in different bounds. We will be particularly interested in $y = x+1$.

Proof of Theorem 1. Let $m(t) = \mathrm{E}[\exp(tX)]$ be the moment-generating function. We will bound the moments of X by

$$\mathrm{E}[X^k] \le m(t)(k/(et))^k, \tag{3}$$

which holds for all $k \ge 0$ and $t > 0$. This follows from the basic inequality $1+x \le e^x$, where we substitute $tx/k - 1$ for x to get $tx/k \le e^{tx/k-1} \implies x^k \le e^{tx}(k/(et))^k$. Taking expectations we get (3).

We now define $B = k/\mu$ and take t such that $te^t = B$. Using the Lambert-W function, this means $t = W(B)$. We note that $t > 0$ whenever $B > 0$.

We can thus bound

$$\begin{aligned}
\mathrm{E}[(X/\mu)^k] &\le m(t)\mu^{-k}(k/(et))^k \\
&\le \exp(\mu(e^t-1))\left(\frac{k}{e\mu t}\right)^k \\
&= \exp(\mu(B/t-1))\left(\frac{e^t}{e}\right)^k \\
&= \exp((k/B)(B/t-1) + k(t-1)) \\
&= \exp(kf(B)),
\end{aligned}$$

where $f(B) = 1/t + t - 1 - 1/B$.

It remains to show $\exp(f(B)) \le \frac{B}{\log(B+1)}$. Taking logarithms, this means showing the bound

$$1/W(x) + W(x) - 1 - 1/x \le \log(x) - \log\log(1+x) \tag{4}$$

for all $x > 0$, where $W(x)$ is the Lambert-W function.

We define $g(x) = \mathrm{LSH} - \mathrm{RHS} = 1/W(x) + W(x) - 1 - 1/x - \log(x) + \log\log(1+x)$. For $x \to 0$ we have $1/W(x) = 1/x + 1 + O(x)$ and $\log\log(1+x) = \log(x) + O(x)$. Thus $g(x) \to 0$ as $x \to 0$. We proceed to show $g(x)$ is non-increasing for $x > 0$, from which (4) follows.

The Lambert function is simple to differentiate using the identity $W'(x) = \frac{W(X)}{x(1+W(x))}$. Taking derivatives and simplifying we get $g'(x) = \frac{1-e^{W(x)}}{x^2} + \frac{1}{(1+x)\log(1+x)}$. Rearranging, it suffices to show the following simple bound on the Lambert function:

$$e^{W(x)} \le \frac{x^2}{(1+x)\log(1+x)} + 1.$$

Using Lemma 1 with $y = x+1$ yields the bound

$$e^{W(x)} \le \frac{2x+1}{1+\log(1+x)}.$$

We then need to show $\frac{2x+1}{1+z} \leq \frac{x^2}{(1+x)z} + 1$, where $z = \log(1+x)$. Solving the quadratic in z, we see that the inequality holds when $x > 0$ and $\frac{x}{1+x} \leq z \leq x$, but we have exactly that inequality for $\log(1+x)$, which finishes the proof. $\square$

3 Sub-Poissonian Random Variables

We call a non-negative random variable X sub-Poissonian if $\mathrm{E}\, X = \mu$ and the moment-generating function, mgf., $\mathrm{E}\exp(tX) \leq \exp(\mu(e^t - 1))$ for all $t > 0$. If $X_1, \dots, X_n$ are sub-Poissonian with mgf. $m_1(t), \dots, m_n(t)$, then $\sum_i X_i$ is sub-Poissonian as well, since

$$\mathrm{E}\exp\big(t\sum_i X_i\big) = \prod_i m_i(t) \leq \prod_i \exp\big(\mu_i(e^t - 1)\big) = \exp\big((\sum_i \mu_i)(e^t - 1)\big).$$

A random variable bounded in $[0, 1]$ with mean μ has mgf.

$$\mathrm{E}\exp(tX) = 1 + \sum_{k=1}^{\infty} \frac{t^k E X^k}{k!} \leq 1 + \mu \sum_{k=1}^{\infty} \frac{t^k E[1^{k-1}]}{k!} = 1 + \mu(e^t - 1) \leq \exp(\mu(e^t - 1)).$$

Hence if $X = X_1 + \cdots + X_n$ where each $X_i \in [0, 1]$ we have $\mu = EX = \sum_i EX_i$ and by Theorem 1 that $E[(X\mu)^k] \leq \frac{k/\mu}{\log(k/\mu+1)}$. In particular this captures sum of Bernoulli variables with distinct probabilities.

An example of a non-sub-Poissonian distribution is the exponential distribution with mean $1/p = \mu$. The moment generation is $m(t) = \frac{1}{1-\mu(e^t-1)}$, which is larger than $\exp(\mu(e^t - 1))$ for all $t > 0$. It is possible however that similar methods to those in the proof of Theorem 1 will still be applicable.

References

[1] Daniel Berend and Tamir Tassa. Improved bounds on bell numbers and on moments of sums of random variables. *Probability and Mathematical Statistics*, 30(2):185–205, 2010.

[2] Abdolhossein Hoorfar and Mehdi Hassani. Inequalities on the lambert w function and hyperpower function. *J. Inequal. Pure and Appl. Math*, 9(2):5–9, 2008.

[3] Rustam Ibragimov and Sh Sharakhmetov. On an exact constant for the rosenthal inequality. *Theory of Probability & Its Applications*, 42(2):294–302, 1998.

[4] William B Johnson, Gideon Schechtman, and Joel Zinn. Best constants in moment inequalities for linear combinations of independent and exchangeable random variables. *The Annals of Probability*, pages 234–253, 1985.

[5] Andreas Knoblauch. Closed-form expressions for the moments of the binomial probability distribution. *SIAM Journal on Applied Mathematics*, 69(1):197–204, 2008.

[6] Rafał Latała et al. Estimation of moments of sums of independent real random variables. *The Annals of Probability*, 25(3):1502–1513, 1997.

[7] E Ostrovsky and L Sirota. Non-asymptotic estimation for bell function, with probabilistic applications. *arXiv preprint arXiv:1712.08804*, 2017.

[8] Iosif Pinelis. Optimum bounds on moments of sums of independent random vectors. *Siberian Adv. Math*, 5(3):141–150, 1995.

[9] John Riordan. Moment recurrence relations for binomial, poisson and hypergeometric frequency distributions. *The Annals of Mathematical Statistics*, 8(2):103–111, 1937.

[10] Maciej Skorski. Handy formulas for binomial moments. *arXiv preprint arXiv:2012.06270*, 2020.

[11] Flemming Topsøe. Some bounds for the logarithmic function. *Inequality theory and applications*, 4(01), 2007.

[12] Jacques Touchard. Sur les cycles des substitutions. *Acta Mathematica*, 70(1):243–297, 1939.