

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

9th July 2021

Abstract

We prove the inequality $E[(X/\mu)^k] \leq (\frac{k/\mu}{\log(1+k/\mu)})^k \leq \exp(k^2/(2\mu))$ for sub-Poissonian random variables X , such as Binomially or Poisson distributed variables, with mean μ . The asymptotic behaviour $E[(X/\mu)^k] = 1 + O(k^2/\mu)$ matches a lower bound of $1 + \Omega(k^2/\mu)$ for small k^2/μ . This improves over previous uniform raw moment bounds by a factor exponential in k .

1 Introduction

Suppose we sample an urn of n balls, each coloured *red* with probability p and otherwise *blue*. What is the probability that a sample of k balls, with replacement, *from this urn* consists of only red balls? Such questions are of interest to sample-efficient statistics and the derandomisation of algorithms.

If $R \sim \text{Binomial}(n, p)$ denotes the number of red balls in the urn, the probability of drawing a single red ball from the urn is R/n . Thus, the probability that a sample of k balls from the urn is all red is given by $(R/n)^k$, or $P = E[(R/n)^k]$ when the probability is taken over both sample phases. Whenever the urn is large (n is large), R/n concentrates around p , so sampling from the urn is equivalent to sampling from the original distribution and $P \approx p^k$. Indeed, from Jensen's inequality, we can see that p^k is always a lower bound: $P = E[(R/n)^k] \geq E[(R/n)]^k = p^k$. Previous authors have shown a nearly matching upper bound of $C^k p^k$ in the range $k/(np) = O(1)$ for some constant $C > 1$. (See eq. (1) below for details.) In this note, we improve the upper bound to $P \leq p^k(1 + k/(2np))^k$, which shows that when $k = o(\sqrt{np})$, the factor C^k can be replaced by just $1 + o(1)$.

1.1 Related work

One direct approach to computing the Binomial moments expands them using the Stirling numbers of the second kind: $E[X^k] = \sum_{i=0}^k \left\{ \begin{smallmatrix} k \\ i \end{smallmatrix} \right\} n^i p^i$, where $n^i = n(n-1) \cdots (n-i+1)$. This equality can be derived as a sum of the much easier to compute “factorial moments”, $E[X^k] = n^k p^k$. See Knoblauch (2008) for details. Taking the leading two terms of the sum, one finds that $E[X^k] = (np)^k \left(1 + \binom{k}{2} \frac{1-p}{np} + O(1/n^2) \right)$ as $n \rightarrow \infty$. However, this

approach does not work when k is not constant with respect to n . Similarly, for the Poisson distribution, the moments can be expressed as the so-called Bell (or Touchard) polynomials in μ : $E[X^k] = \sum_{i=0}^k \{i\}^k \mu^i$. This sum gives a simple lower bound $E[X^k] \geq \{k\} \mu^k + \{k-1\} \mu^{k-1} = \mu^k (1 + \frac{k(k-1)}{2\mu})$, matching our upper bound asymptotically when $k = O(\sqrt{\mu})$. However, as in the Binomial case, the sum does not easily yield a uniform bound. We give the details of both lower bounds in Section 2.2.

A different approach uses the powerful results on moments of independent random variables by Latała (1997) and Pinelis (1995). In the case of Binomial and Poisson random variables, they yield:

$$\left(c \frac{k/\mu}{\log(1 + k/\mu)}\right)^k \leq E[(X/\mu)^k] \leq \left(C \frac{k/\mu}{\log(1 + k/\mu)}\right)^k \quad (1)$$

for some universal constants $c < 1 < C$. The bound is tight up to the factor $(C/c)^k$, which is negligible when the overall growth is $O(k^k)$. However, when $k/\mu \rightarrow 0$, we expect the upper bound to be 1, and so the factor C^k in the upper bound can be overwhelmingly large.

A third option is to use a Rosenthal bound, such as the following by Berend and Tassa (2010), (see also Johnson et al., 1985):

$$E[X^k] \leq B_k \max\{\mu, \mu^k\}. \quad (2)$$

Here, B_k is the k th Bell number, which Berend and Tassa show satisfies the uniform bound $B_k < \left(\frac{0.792k}{\log(k+1)}\right)^k$. For large k , a precise asymptotic bound, $B_k^{1/k} = \frac{k}{e \log k} (1 + o(1))$, is given by (e.g. de Bruijn, 1981; Ibragimov and Sharakhmetov, 1998). Unfortunately, the Rosenthal bound is incomparable to the other bounds in this paper when $\mu < 1$, as it grows with μ rather than μ^k . However, for $\mu \geq 1$ and integral, we show a matching asymptotic lower bound in the second half of Section 2.2. That indicates that the upper bound of this paper could be improved by a factor e^{-k} for large k .

Finally, Ostrovsky and Sirota (2017) give another asymptotically sharp bound in a recent preprint. Using a technique based on moment generating functions, similar to this paper, they bound the Bell polynomial, which as discussed above is equivalent to bounding the moments of a Poisson random variable. The bound holds when $k \geq 2\mu$:

$$E[(X/\mu)^k]^{1/k} \leq \frac{k/\mu}{e \log(k/\mu)} \left(1 + C(\mu) \frac{\log \log(k/\mu)}{\log(k/\mu)}\right) \quad \text{if } k \geq 2\mu, \quad (3)$$

where $C(\mu) > 0$ is some “constant” depending only on μ . In the range $k < 2\mu$, Ostrovsky and Sirota only gives the bound $E[(X/\mu)^k] \leq 8.9758^k$, so similarly to the other bounds presented, it loses an exponential factor in k compared to Theorem 1 below, for smaller k .

2 Bounds

The theorem considers “sub-Poissonian” random variables, which are variables X , satisfying the requirement $E[\exp(tX)] \leq \exp(\mu(e^t - 1))$. Such sub-Poissonian include many

simple distributions, such as the Poisson or Binomial distribution. We give more examples in Section 3.

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

$$E[(X/\mu)^k] \leq \left(\frac{k/\mu}{e^{1-\alpha} \log(1 + \alpha k/\mu)} \right)^k.$$

The theorem has a free parameter, α , which is optimally set such that $1 + \alpha k/\mu = e^{W(k/\mu)}$, where W is the Lambert-W function, which is defined by $W(x)e^{W(x)} = x$.¹ In practice the following two corollaries may be easier to work with.

Corollary 1.

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

Proof. Set $\alpha = 1$ in Theorem 1. The second bound we use a standard logarithmic bound, $\frac{x}{\log(1+x)} \leq 1 + x/2$ (see e.g. Topsøe, 2007, eq. 6). The last inequality is the trivial $1 + x \leq \exp(x)$. \square

In the range $k = O(\sqrt{\mu})$ we show a matching lower bound of $1 + \Omega(k^2/\mu)$ in Section 2.2, eq. (9).

Corollary 2. *Let $x = k/\mu$, then*

$$E[(X/\mu)^k]^{1/k} \leq \frac{x e^{1/\log(e+x)}}{e \log(1 + x/\log(e+x))} = \frac{x}{e \log x} \left(1 + O\left(\frac{\log \log x}{\log x}\right) \right) \quad \text{as } x \rightarrow \infty. \quad (4)$$

Proof. Take $\alpha = 1/\log(e+x)$. For $x > 0$ we have $\log(e+x) > 0$ and so $\alpha > 0$ as required by Theorem 1. \square

Corollary 2 matches our lower bound in eq. (10), as well as Ostrovsky and Sirota in eq. (3), but without the restriction on the range of k/μ .

2.1 The proof

Technically our bound is shown using the moment-generating function and some new sharp inequalities involving the Lambert-W function. We will use the following lemma:

Lemma 1 (Hoorfar and Hassani, 2008). *For all $y > 1/e$ and $x > -1/e$,*

$$e^{W(x)} \leq \frac{x + y}{1 + \log y}. \quad (5)$$

¹The Lambert-W function has multiple branches. We always refer to the main one (sometimes called the 0th), in which $W(x)$ and x are both positive.

We perform an elementary proof of this fact for completeness:

Proof. Starting from $1 + t \leq e^t$, substitute $\log(y) - t$ for t to get $1 + \log y - t \leq ye^{-t}$. Multiplying by e^t we get $e^t(1 + \log y) \leq te^t + y$. Let $t = W(x)$ s.t. $te^t = x$. Rearranging, we get eq. (5). \square

Taking $y = e^{W(x)}$ in eq. (5) makes the two sides equal, so we can think of Lemma 1 as a way to turn a rough estimate into an upper bound. Hoorfar and Hassani make various substitutions, resulting in different bounds useful when $x \rightarrow \infty$. We will use the bound differently, focusing on having the right asymptotics as $x \rightarrow 0$.

We are now ready to prove the main theorem of the paper:

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

$$\mathbb{E}[X^k] \leq m(t)(k/(et))^k, \quad (6)$$

which holds for all $k \geq 0$ and $t > 0$. This follows from the basic inequality $1 + z \leq e^z$, where we substitute $tz/k - 1$ for z to get $tz/k \leq e^{tz/k-1} \implies z^k \leq e^{tz}(k/(et))^k$. Letting $z = X$ and taking expectations, we get eq. (6).

We now define $x = k/\mu$ and take t such that $te^t = x$. In the notation of the Lambert-W function, this means $t = W(x)$. We note that $t > 0$ whenever $x > 0$. We proceed to bound the moments of X/μ using eq. (6):

$$\begin{aligned} \mathbb{E}[(X/\mu)^k] &\leq m(t)(k/(et))^k \mu^{-k} \\ &\leq \exp(\mu(e^t - 1)) \left(\frac{k}{e\mu t} \right)^k \\ &= \exp(\mu(x/t - 1)) \left(\frac{e^t}{e} \right)^k \end{aligned} \quad (7)$$

$$\begin{aligned} &= \exp((k/x)(x/t - 1) + k(t - 1)) \\ &= \exp(kf(x)), \end{aligned} \quad (8)$$

where we define $f(x) := 1/t - 1/x + t - 1$. Here eq. (7) came from the simple rewriting of the definition of t , $1/t = e^t/x$

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

$$f(x) = \frac{1}{W(x)} + W(x) - 1 - \frac{1}{x} \leq \log\left(\frac{x}{\log(1+x)}\right)$$

for all $x > 0$, where $W(x)$ is the Lambert-W function. The proof uses the identities $W(X) = \log x - \log(W(x))$ and $\frac{1}{W(x)} = \frac{1}{x} \exp(W(x))$ which are simple rewritings of the definition $W(x)e^{W(x)} = x$. The main idea is to introduce a new variable $z > 0$, to be determined later, which allows us to control the effect of applying the logarithmic inequality $\log x \geq 1 - 1/x$. We also use Lemma 1 which introduces another new variable $y > 1$ to be determined.

$$\frac{1}{W(x)} + W(x) = \frac{1}{W(x)} + \log x - \log(W(x))$$

$$\begin{aligned}
&= \frac{1}{W(x)} + \log\left(\frac{x}{z}\right) - \log\left(\frac{W(x)}{z}\right) \\
&\leq \frac{1}{W(x)} + \log\left(\frac{x}{z}\right) - \left(1 - \frac{z}{W(x)}\right) \\
&= \frac{1+z}{W(x)} - 1 + \log\left(\frac{x}{z}\right) \\
&= e^{W(x)} \frac{1+z}{x} - 1 + \log\left(\frac{x}{z}\right) \\
&\leq \frac{x+y}{1+\log(y)} \frac{1+z}{x} - 1 + \log\left(\frac{x}{z}\right) \\
&= \frac{y}{x} + \log\left(\frac{x}{\log y}\right).
\end{aligned}$$

Here the last two steps come from the inequality eq. (5) in its general form, and the substitution $z = \log y$. Note that we have equality all the way through if we let $y = e^{W(x)}$.

We have to make sure z is non-negative, so we set $y = 1 + \alpha x$. That finally gives us the bound

$$f(x) = \frac{1}{W(x)} + W(x) - 1 - \frac{1}{x} \leq \alpha - 1 + \log\left(\frac{x}{\log(1 + \alpha x)}\right).$$

Backing up, we have shown $E[(X/\mu)^k] \leq \exp(kf(x)) \leq (\frac{x}{e^{1-\alpha}\log(1+\alpha x)})^k$, which finishes the proof. \square

2.2 Lower bound

As mentioned in the introduction, the expansion for the Poisson moments $E[X^k] = \sum_{i=0}^k \{i\}^k \mu^i$ gives a simple lower bound by taking the two highest terms. We note that $\{k\}^k = 1$ and $\{k-1\}^k = \binom{k}{2}$ to get

$$E[X^k] \geq \mu^k \left(1 + \frac{k(k-1)}{2\mu}\right), \quad (9)$$

matching Theorem 1 asymptotically for $k = O(\sqrt{\mu})$.

The expansion for Binomial moments $E[X^k] = \sum_{i=0}^k \{i\}^k n^i p^i$ yields a similar lower bound

$$\begin{aligned}
E[X^k] &\geq n^k p^k + \binom{k}{2} n^{k-1} p^{k-1} \\
&= (np)^k \left(\frac{n^k}{n^k}\right) \left(1 + \binom{k}{2} \frac{1}{(n-k+1)p}\right) \\
&= (np)^k \left(\prod_{i=0}^{k-1} 1 - \frac{i}{n}\right) \left(1 + \binom{k}{2} \frac{1}{(n-k+1)p}\right) \\
&\geq (np)^k \left(1 - \binom{k}{2} \frac{1}{n}\right) \left(1 + \binom{k}{2} \frac{1}{np}\right)
\end{aligned}$$

$$= (np)^k \left(1 + \binom{k}{2} \frac{1-p}{np} \left(1 - \binom{k}{2} \frac{1}{n} \right) \right),$$

which matches Theorem 1 for $k = O(\sqrt{\mu})$ and p not too close to 1.

We will investigate some more precise lower bounds as k/μ gets large. As mentioned briefly in the introduction, there is a correspondence between the moments of a Poisson random variable and the Bell polynomials defined by $B(k, \mu) = \sum_i \{i^k\} \mu^i$. In particular, $E[X^k] = B(k, \mu)$, if μ is the mean of the Poissonian random variable. The Bell polynomials are so named because $B(k, 1)$ is the k th Bell number. By Dobiński's formula $B(k, 1) = \frac{1}{e} \sum_{i=0}^{\infty} \frac{i^k}{i!}$ the Bell numbers are generalised for real k . We write these as $B_x = B(x, 1)$.

We give a lower bound for $E[(X/\mu)^k]$ by showing the following simple connection between the Bell polynomials and Bell numbers:

Theorem 2. *Let k be a positive real number and $\mu \geq 1$ be an integer. Then*

$$B(k, \mu)/\mu^k \geq B_{k/\mu}^\mu.$$

While the proof below assumes μ is an integer, we will conjecture Theorem 2 to be true for any $\mu \geq 1$. Now by de Bruijn's (1981) asymptotic expression for the Bell numbers:

$$E[(X/\mu)^k]^{1/k} \geq B_{k/\mu}^{\mu/k} = \frac{k/\mu}{e \log(k/\mu)} \left(1 + \Theta \left(\frac{\log \log(k/\mu)}{\log(k/\mu)} \right) \right) \quad \text{as } k/\mu \rightarrow \infty. \quad (10)$$

matching our upper bound, eq. (4), the upper bound of Ostrovsky and Sirota, eq. (3), for large k , as well as Latała's uniform lower bound with a different constant.

Proof of Theorem 2. Let X, X_1, \dots, X_μ be i.i.d. Poisson variables with mean 1, then $S = \sum_{i=1}^\mu X_i$ is Poisson with mean μ . We write $\|X\|_k = E[X^k]^{1/k}$. Then by the AG inequality:

$$\|S/\mu\|_k = \left\| \frac{1}{\mu} \sum_{i=1}^\mu X_i \right\|_k \geq \left\| \left(\prod_{i=1}^\mu X_i \right)^{1/\mu} \right\|_k = \left\| \prod_{i=1}^\mu X_i \right\|_{k/\mu}^{1/\mu} = \left(\prod_{i=1}^\mu \|X_i\|_{k/\mu} \right)^{1/\mu} = \|X\|_{k/\mu}. \quad (11)$$

Since X has mean 1 we have $\|X\|_{k/\mu} = B_{k/\mu}^{\mu/k}$, and as S has mean μ we have $\|S/\mu\|_k = B(k, \mu)^{1/k}/\mu$. Thus, taking k th powers, eq. (11) is what we wanted to show. \square

For small k/μ this bound is less interesting since $B_x \rightarrow 0$ as $x \rightarrow 0$, rather than 1 as our upper bound. However, it is pretty tight, as we conjecture by the following matching upper bound in terms of the Bell numbers:

Conjecture 1. *For all $k > 0$ and $\mu \geq 1$,*

$$B_{k/\mu}^{1/(k/\mu)} \leq \frac{B(k, \mu)^{1/k}}{\mu} \leq B_{k/\mu+1}^{1/(k/\mu+1)}.$$

Furthermore, for $0 < \mu \leq 1$, $\frac{B(k, \mu)^{1/k}}{\mu} \leq B_{k/\mu}^{1/(k/\mu)}.$

While the upper bound appears true numerically, it can't follow from our moment-generating function bound eq. (8), since it drops below that for k/μ bigger than 40. The conjectured upper bound is even incomparable with our Theorem 1, since it is slightly above $\frac{k/\mu}{\log(1+k/\mu)}$ for very small k/μ . In the region $k < 2$ and $\mu < 1$, the conjectured bound is weaker than eq. (2) by Berend and Tassa (2010), but for all other parameters, it is substantially tighter.

3 Sub-Poissonian Random Variables

We call a non-negative random variable X sub-Poissonian if $E[X] = \mu$ and the moment-generating function, mgf., $E[\exp(tX)] \leq \exp(\mu(e^t - 1))$ for all $t > 0$. We will briefly show that this notion includes all sums of bounded random variables, such as the Binomial distribution.

If X_1, \dots, X_n are sub-Poissonian with mgf. $m_1(t), \dots, m_n(t)$ and mean μ_1, \dots, μ_n respectively, then $\sum_i X_i$ is sub-Poissonian as well, since

$$E\left[\exp\left(t \sum_i X_i\right)\right] = \prod_i m_i(t) \leq \prod_i \exp(\mu_i(e^t - 1)) = \exp\left(\left(\sum_i \mu_i\right)(e^t - 1)\right).$$

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

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

Hence if $X = X_1 + \dots + X_n$ where each $X_i \in [0, 1]$ we have $\mu = E[X] = \sum_i E[X_i]$ 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 geometric distribution with mean μ . This has moment generating function $m(t) = \frac{1}{1-\mu(e^t-1)}$, which is larger than $\exp(\mu(e^t - 1))$ for all $t > 0$. However, likely, similar methods to those in the proof of Theorem 1 will still apply to bound its moments.

4 Acknowledgements

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

References

- 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.
- Nicolaas Govert de Bruijn. *Asymptotic methods in analysis*, volume 4. Courier Corporation, 1981.

- 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.
- Rustam Ibragimov and Sh Sharakhmetov. On an Exact Constant for the Rosenthal Inequality. *Theory of Probability & Its Applications*, 42(2):294–302, 1998.
- 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*, 13(1):234 – 253, 1985.
- Andreas Knoblauch. Closed-Form Expressions for the Moments of the Binomial Probability Distribution. *SIAM Journal on Applied Mathematics*, 69(1):197–204, 2008.
- Rafał Łatała. Estimation of moments of sums of independent real random variables. *The Annals of Probability*, 25(3):1502–1513, 1997.
- Eugene Ostrovsky and Leonid Sirota. Non-asymptotic estimation for Bell function, with probabilistic applications. *arXiv preprint arXiv:1712.08804*, 2017.
- Iosif Pinelis. Optimum bounds on moments of sums of independent random vectors. *Siberian Adv. Math*, 5(3):141–150, 1995.
- Flemming Topsøe. Some bounds for the logarithmic function. *Inequality theory and applications*, 4(01), 2007.
- Jacques Touchard. Sur les cycles des substitutions. *Acta Mathematica*, 70(1):243–297, 1939.