BOUNDDED SIZE BIAS COUPLING: A GAMMA FUNCTION BOUND, AND UNIVERSAL DICKMAN-FUNCTION BEHAVIOR

RICHARD ARRATIA AND PETER BAXENDALE

Abstract. Under the assumption that the distribution of a non-negative random variable $X$ admits a bounded coupling with its size biased version, we prove simple and strong concentration bounds. In particular the upper tail probability is shown to decay at least as fast the reciprocal of a Gamma function, guaranteeing a moment generating function that converges everywhere. The class of infinitely divisible distributions with finite mean, whose Lévy measure is supported on an interval contained in $[0, c]$ for some $c < \infty$, forms a special case in which this upper bound is logarithmically sharp. In particular the asymptotic estimate for the Dickman function, that $\rho(u) \approx u^{-u}$ for large $u$, is shown to be universal for this class.

A special case of our bounds arises when $X$ is a sum of independent random variables, each admitting a 1-bounded size bias coupling. In this case, our bounds are comparable to Chernoff–Hoeffding bounds; however, ours are broader in scope, sharper for the upper tail, and equal for the lower tail.

We discuss bounded and monotone couplings, give a sandwich principle, and show how this gives an easy conceptual proof that any finite positive mean sum of independent Bernoulli random variables admits a 1-bounded coupling with the same conditioned to be nonzero.

Contents

1. Introduction 2
2. Product bounds 2
3. Gamma bounds 5
4. Moment generating function 6
5. Examples admitting a 1-bounded size bias coupling 8
6. Lévy($[0, c]$), the infinitely divisible distributions with Lévy measure supported on $[0, c]$ 9
7. Bounded coupling, monotone coupling, and a sandwich principle 11

Date: June 11, 2013.
1. Introduction

For any random variable $X$, we write

$$F(x) := \mathbb{P}(X \leq x), \ G(x) := \mathbb{P}(X \geq x).$$

When $X$ has finite mean $a$, concentration inequalities refer to estimates on the upper tail probability $G(x) = \mathbb{P}(X \geq x)$ for $x \geq a$ and on the lower tail probability $F(x) = \mathbb{P}(X \leq x)$ for $x \leq a$. The remarkably effective idea of using bounded size bias couplings to prove concentration inequalities comes from Ghosh and Goldstein [8]; their proof is inspired by the $x \mapsto e^x$ is convex argument used to prove the Azuma–Hoeffding concentration bounds. Here we prove stronger bounds, under weaker hypotheses, and with a simpler proof, inspired by the easy proof, from [12], that Dickman’s function $\rho$ satisfies $\rho(u) \leq 1/\Gamma(u + 1)$, for $u \geq 0$. See [7, 8, 1, 4] for many examples of the application of concentration bounds derived from size bias couplings, for situations involving dependence.

We recall the notation, definition, and most basic fact about size bias: given a nonnegative random variable $X$ with $0 < \mathbb{E}X < \infty$, we say that the distribution of $Y$ is the size biased distribution of $X$, written $Y =^d X^*$, if the Radon-Nikodym derivative of the distribution of $Y$, with respect to the distribution of $X$, is given by $\mathbb{P}(Y \in dx)/\mathbb{P}(X \in dx) = x/\mathbb{E}X$. If $Y =^d X^*$, then for all bounded measurable $g$, $\mathbb{E}g(Y) = \mathbb{E}(Xg(X))/\mathbb{E}X$. This last fact will be applied in the proof of Lemma 2.1 with $g(z) = 1(z \geq x)$ and with $g(z) = 1(z \leq x)$. See [2], or [3, pp 78–80]. The main hypothesis of Lemma 2.1 below can be said as “The distribution of $X$ admits a $c$-bounded coupling with its size biased version,” or less formally, “$X$ admits a $c$-bounded size bias coupling.”

2. Product bounds

Lemma 2.1. Suppose that a nonnegative random variable $X$ has $\mathbb{E}X = a \in (0, \infty)$, that $c < \infty$, and that for $Y =^d X^*$, there exists a coupling in which

$$Y \leq X + c. \quad (1)$$

Then

$$\forall x > 0, \ G(x) \leq \frac{a}{x} G(x - c), \quad (2)$$
\[ \forall x, \quad F(x) \leq \frac{x + c}{a} F(x + c). \] (3)

**Proof.** To prove the upper bound on \( G(x) \), note that (1) implies that the event \( Y \geq x \) is a subset of the event \( X \geq x - c \). Hence for \( x > 0 \),

\[
xG(x) = x \mathbb{E} 1(X \geq x) \leq \mathbb{E}(X 1(X \geq x)) = a \mathbb{P}(Y \geq x) \leq a G(x - c).
\]

When \( x > 0 \) we can divide by \( x \) to get (2).

To prove the upper bound on \( F(x) \), note that (1) implies that the event \( Y \leq x \) is a superset of the event \( X \leq x - c \). Hence

\[
xF(x) = x \mathbb{E} 1(X \leq x) \geq \mathbb{E}(X 1(X \leq x)) = a \mathbb{P}(Y \leq x) \geq a F(x - c).
\]

This does not require that \( x \) be positive; for \( x < 0 \) it is the trivial inequality, that \( 0 \geq 0 \). Replacing \( x \) by \( x + c \) and dividing by \( a > 0 \) yields (3).

Given \( x > 0 \), the obvious strategy for obtaining good bounds is to iterate (2) or (3) for as long as the new value of \( x \), say \( x' = x \pm ic \), still gives a favorable ratio, \( a/x' \) in (2), or \( (x' + c)/a \) in (3), and using \( G(t) \leq 1 \) or \( F(t) \leq 1 \) as needed, to finish off.

**Theorem 2.2.** Under the hypotheses of Lemma 2.1, given \( x \) let

\[
k = \lfloor \frac{|x - a|}{c} \rfloor,
\]

so that \( k \) is a nonnegative integer, possibly zero. Then

for \( x \geq a \), \( G(x) \leq u(x, a, c) := \prod_{0 \leq i \leq k} \frac{a}{x - ic} \) (5)

and

for \( 0 \leq x \leq a \), \( F(x) \leq \ell(x, a, c) := \prod_{0 < i \leq k} \frac{x + ic}{a} \) (6)

**Proof.** Simply apply the strategy described in the paragraph preceding the statement of this theorem.

**Corollary 2.3.** Under the hypotheses of Lemma 2.1, the moment generating function of \( X \) is finite everywhere, that is, for all \( \beta \in \mathbb{R} \),

\[ M(\beta) := \mathbb{E} e^{\beta X} < \infty. \]
Proof. Since $X \geq 0$, trivially $M(\beta) \leq 1$ if $\beta \leq 0$, so assume $\beta > 0$. For motivation: as $x$ increases by $c$, $e^{\beta x}$ increases by a factor of $e^{\beta c}$, while the upper bound $u(x, a, c)$ on $G(x)$ decreases by a factor of $a/x$; hence for $x > x_0 := 2ae^{\beta c}$, the product $e^{\beta x}u(x, a, c)$ decreases by a factor of at least 2.

Writing

$$M(\beta) = \mathbb{E} \left[ e^{\beta X} 1(X < x_0) \right] + \sum_{i \geq 0} \mathbb{E} \left[ e^{\beta X} 1(x_0 + ic \leq X < x_0 + ic + c) \right]$$

$$\leq \mathbb{E} \left[ e^{\beta X} 1(X < x_0) \right] + \sum_{i \geq 0} e^{\beta(x_0 + ic + c)} u(x_0 + ic, a, c),$$

the series on the right side is bounded by a geometric series with ratio 1/2, and the net result will be $M(\beta) \leq \exp(\beta x_0) + 2 \exp(\beta(x_0 + c))$. □

Remark 2.4. Note the difference between the indexing in the products (5, 6): $u(x, a, c)$ includes the factor indexed by $i = 0$, while $\ell(x, a, c)$ excludes the factor indexed by $i = 0$. In case $x \in (a, a+c)$, which is equivalent to $a < x$ and $k = 0$, the bound in (5) has one factor, and simplifies to $G(x) \leq u(x, a, c) = a/x < 1$. In case $x \in (a-c, a)$, which is equivalent to $x < a$ and $k = 0$, the bound in (6) has no factors, and simplifies to the trivial observation $F(x) \leq \ell(x, a, c) = 1$.

2.1. Product bound combined with one-sided Chebyshev. The recursive nature of Lemma 2.1 allows the possibility of combining it with other information about $F(x)$ or $G(x)$. Here we pursue one possibility.

For a random variable $X$ with mean $a$ and variance $\sigma^2$, the one-sided Chebyshev inequality states that, for all $x \leq a$, $F(x) = \mathbb{P}(X \leq x)$ is upper bounded by $\sigma^2/(\sigma^2 + (a-x)^2)$. In our situation, $Y \overset{d}{=} X^*$ and $\mathbb{E} X = a$ yields $\mathbb{E} X^2 = a \mathbb{E} Y$, and $Y \leq X + c$ implies $\mathbb{E} Y \leq a + c$. Hence $\mathbb{E} X^2 = a \mathbb{E} Y \leq a(a+c)$, so that $\sigma^2 \leq ac$. Thus, under the hypotheses of Lemma 2.1, for all $x \leq a$,

$$\text{Var}X \leq ac, \quad F(x) \leq \frac{ac}{ac + (a-x)^2}.$$  \hspace{1cm} (7)

We want to improve on (6) by using one-sided Chebyshev in combination with iteration of (3). More precisely, given $x < a$ and any non-negative integer $j$ such that $x + jc \leq a$ we can iterate (3) from $x$ to $x + jc$ and then use the one-sided Chebyshev inequality at $x + jc$ to obtain

$$F(x) \leq \ell_j(x, a, c) := \frac{ac}{ac + (a-x-jc)^2} \prod_{0 < i \leq j} \frac{x + ic}{a}.$$  \hspace{1cm} (8)
Remark 2.5. Scaling. It is simplest, both for notation and concept, to work with the special case where the constant $c$ in the coupling (1) satisfies $c = 1$. The results derived for this special case easily transform into results for the general case, since if $Y = d X^*$ and $Y \leq X + c$, then

$$\frac{Y}{c} = d \frac{X}{c}^*, \quad \text{and} \quad \frac{Y}{c} \leq \frac{X}{c} + 1. \quad (9)$$

In particular the upper bounds in (5) and (6), for all $a, c > 0$ and $x \geq 0$, satisfy

$$u(x, a, c) = u(x/c, a/c, 1), \quad \ell(x, a, c) = \ell(x/c, a/c, 1). \quad (10)$$

Opportunities to use (10) are presented by (14) and (15).

3. Gamma bounds

In this section we restrict to the case $c = 1$. Results for general $c > 0$ can be recovered using equation (10), see Remark 2.5. For $c = 1$, making use of $\Gamma(z) = \Gamma(z + 1)$, the conclusions (5) and (6) in Theorem 2.2 can be rewritten as: for $x \geq a$ and $k = [x - a]$,

$$G(x) \leq u(x, a, 1) := \prod_{0 \leq i \leq k} \frac{a}{x - i} = \frac{a^{k+1} \Gamma(x - k)}{\Gamma(x + 1)}, \quad (11)$$

and for $0 \leq x \leq a$ and $k = [a - x]$,

$$F(x) \leq \ell(x, a, 1) := \prod_{0 < i \leq k} \frac{x + i}{a} = \frac{\Gamma(x + k + 1)}{a^k \Gamma(x + 1)}. \quad (12)$$

These upper and lower tail bounds might be viewed as too complicated; as $x$ varies they are not closed-form expressions, and are not analytic function. Here we replace them by simpler (but weaker) expressions which are analytic in $a$ and $x$.

Lemma 3.1. For $a > 0$ and $0 \leq f \leq 1$

$$a^{1-f} \Gamma(a + f) \leq \Gamma(a + 1), \quad (13)$$

with equality for $f = 0, 1$ and strict inequality for $f \in (0, 1)$.

Proof. The result is true (with equality) when $f = 0$ or $1$, so we can assume $0 < f < 1$. We use the integral formula $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$ for $x > 0$. Writing $t^{a+f-1} = (t^{a-1})^{1-f} (t^a)^f$ and using Hölder’s inequality (with $p = 1/(1 - f)$) gives

$$\Gamma(a + f) < [\Gamma(a)]^{1-f} [\Gamma(a + 1)]^f.$$

Since $a \Gamma(a) = \Gamma(a + 1)$ we get

$$a^{1-f} \Gamma(a + f) < [a^f \Gamma(a)]^{1-f} [\Gamma(a + 1)]^f = \Gamma[a + 1],$$

and we are done. \qed
Theorem 3.2. (i) For $0 < a \leq x$, the upper tail bound defined by (5) satisfies
\[ u(x,a,1) \leq \frac{a^{x-a} \Gamma(a+1)}{\Gamma(x+1)}, \tag{14} \]
with equality if and only if $x - a$ is an integer.

(ii) For $0 \leq x \leq a$ with $a > 0$, the lower tail bound defined by (6) satisfies
\[ \ell(x,a,1) \leq \frac{\Gamma(a+1)}{a^{a-x} \Gamma(x+1)}, \tag{15} \]
with equality if and only if $a - x$ is an integer.

Proof. (i) Let $k = \lfloor x - a \rfloor$ and $f = x - a - k \in [0,1)$. Since $c = 1$ and $x \geq a$, this is consistent with the notation in (4). Note that $x - k = a - f$, and combine (11) with (13).

(ii) Let $k = \lfloor a - x \rfloor$ and $f = a - x - k \in [0,1)$. Since $c = 1$ and $a \geq x$, this is consistent with the notation in (4). Replacing $f$ by $1 - f$ in (13) gives $a f \Gamma(a - f + 1) \leq \Gamma(a+1)$. Combining this with (12) and noting that $x + k = a - f$ gives the result. \qed

4. Moment generating function

We recall our notation, $M(\beta) := \mathbb{E} e^{\beta X}$ for the moment generating function of $X$. Recall that every random variable has $M(0) = 1$. We observe that if $X$ is Poisson with parameter $a$, then $\log M(\beta) = a(e^\beta - 1)$, and $X$ admits a $c$-bounded size bias coupling with $c = 1$, so that in this case, the inequality (16) holds as an equality for all $\beta$.

Proposition 4.1. Under the hypotheses of Lemma 2.1 the moment generating function $M(\beta)$ for $X$ satisfies
\[ \log M(\beta) \leq \frac{a}{c} (e^{\beta c} - 1) \tag{16} \]
for all $\beta \in \mathbb{R}$.

Proof. We know from Corollary 2.3 that the moment generating function $M(\beta)$ is finite for all $\beta \in \mathbb{R}$. It follows that $M$ is continuously differentiable and $M'(\beta) = \mathbb{E} (X e^{\beta X})$. Moreover, since $Y$ is the size biased version of $X$, we have $\mathbb{E} (X e^{\beta X}) = a \mathbb{E} (e^{\beta Y})$. Together we have
\[ M'(\beta) = a \mathbb{E} (e^{\beta Y}). \]

For $\beta \geq 0$ we have $e^{\beta Y} \leq e^{\beta (X+c)} = e^{\beta c} e^{\beta X}$ so that
\[ M'(\beta) = a e^{\beta c} \mathbb{E} (e^{\beta X}) \leq a e^{\beta c} \mathbb{E} (e^{\beta X}) = a e^{\beta c} M(\beta). \]

Then
\[ (\log M)'(\beta) \leq a e^{\beta c} \]
so that

\[ \log M(\beta) = \log M(\beta) - \log M(0) \leq \int_0^\beta ae^{uc} du = \frac{a}{c} (e^{\beta c} - 1). \]

for all \( \beta \geq 0 \).

For \( \beta \leq 0 \) we have \( e^{\beta Y} \geq e^{\beta(X+c)} = e^{\beta c}e^{\beta X} \) so that

\[ M'(\beta) = aE(e^{\beta Y}) \geq ae^{\beta c}E(e^{\beta X}) = ae^{\beta c}M(\beta). \]

Then

\[ (\log M)'(\beta) \geq ae^{\beta c} \]

so that

\[ -\log M(\beta) = \log M(0) - \log M(\beta) \geq \int_0^\beta ae^{uc} du = \frac{a}{c} (1 - e^{\beta c}), \]

and the proof is complete. \( \square \)

We can now obtain some different upper tail and lower tail bounds, using the standard “large deviation upper bound” method together with the information about \( M(\beta) \) in Proposition 4.1.

With the convention that \( (1/0)^0 = \lim_{x \to 0^+} (1/x)^x = 1 \), we can regard \( (a/x)^{x/c} \) as being well-defined and taking the value 1 when \( x = 0 \). This is the interpretation to use for the right side of (18) when \( x = 0 \), as well as elsewhere in the paper.

**Theorem 4.2.** Under the hypotheses of Lemma 2.1,

\[ \mathbb{P}(X \geq x) \leq \left( \frac{a}{x} \right)^{x/c} e^{(x-a)/c} \quad \text{for all } x \geq a \]  

(17)

and

\[ \mathbb{P}(X \leq x) \leq \left( \frac{a}{x} \right)^{x/c} e^{(x-a)/c} \quad \text{whenever } 0 \leq x \leq a. \]  

(18)

**Proof.** Suppose first \( x \geq a \). For any \( \beta \geq 0 \) we have

\[ \mathbb{P}(X \geq x) = \mathbb{P}(e^{\beta X} \geq e^{\beta x}) \leq M(\beta)/e^{\beta x} \leq \exp \left\{ \frac{a}{c} (e^{\beta c} - 1) - \beta x \right\}. \]

Choosing \( \beta = (1/c) \log(x/a) \geq 0 \) we get

\[ \mathbb{P}(X \geq x) \leq \exp \left\{ \frac{a}{c} \left( \frac{x}{a} - 1 \right) - \frac{x}{c} \log \left( \frac{x}{a} \right) \right\} = e^{(x-a)/c} \left( \frac{a}{x} \right)^{x/c}. \]

Now suppose \( 0 \leq x \leq a \). For any \( \beta \leq 0 \) we have

\[ \mathbb{P}(X \leq x) = \mathbb{P}(e^{\beta X} \geq e^{\beta x}) \leq M(\beta)/e^{\beta x} \leq \exp \left\{ \frac{a}{c} (e^{\beta c} - 1) - \beta x \right\}. \]

If \( 0 < x \leq a \) take \( \beta = (1/c) \log(x/a) \leq 0 \) to get

\[ \mathbb{P}(X \leq x) \leq \exp \left\{ \frac{a}{c} \left( \frac{x}{a} - 1 \right) - \frac{x}{c} \log \left( \frac{x}{a} \right) \right\} = e^{(x-a)/c} \left( \frac{a}{x} \right)^{x/c}, \]
while if \( x = 0 \) let \( \beta \to -\infty \) to get \( \mathbb{P}(X \leq 0) \leq e^{-a/c} \).

See Sections 8 and 9 for results relating (14, 18) to our earlier bounds.

5. Examples admitting a 1-bounded size bias coupling

Here we give some examples of random variables \( X \) which satisfy the hypotheses of Lemma 2.1 with \( c = 1 \). Examples with general \( c > 0 \) can be obtained by scaling, see Remark 2.5.

**Example 5.1.** If \( X \) has a Poisson distribution, it can be verified directly that \( X + 1 \overset{d}{=} X^* \).

**Remark 5.2** (Sharpness of the bounds). Suppose \( X \) is Poisson distributed with mean \( a \in (0, \infty) \). Taking \( x = 0 \) in the lower tail bound (18) from Theorem 4.2 we get \( F(0) \leq e^{-a} \) whereas the exact value is \( F(0) = e^{-a} \). Therefore in this setting the lower tail bound (18) from Theorem 4.2 is sharp.

Now suppose further that \( a \in (0, 1] \). When \( x = n \), the upper tail bound (5) in Theorem 2.2 simplifies, with \( k = n - 1 \), to \( G(x) \leq u(n, a, 1) = a^n/n! \), while for large \( n \), \( G(n) \sim \mathbb{P}(X = n) = e^{-a}a^n/n! \). Hence, for large \( x \) with \( x \) an integer, the upper bound (5) is sharp up to a factor of approximately \( e^a \). Letting \( a \to 0 \) so that \( e^a \to 1 \), one sees that the upper bound (5), for large \( x \), is sharp up to a factor arbitrarily close to 1.

**Example 5.3.** Lévy[0,1], the infinitely divisible distributions with Lévy measure supported on \([0, 1]\). This is the case \( c = 1 \) of the Lévy[0,\( c \)] distributions, discussed in detail in Section 6.

**Example 5.4.** Let \( X \) be a random variable with values in \([0, 1]\), with \( \mathbb{E}X > 0 \). This includes, but is not restricted to, Bernoulli random variables. The size biased version \( Y \), say, of \( X \) takes values in \([0, 1]\) also, and hence \( Y \leq 1 \leq 1 + X \), so that \( X \) admits a 1-bounded size bias coupling.

**Example 5.5.** The uniform distribution on \([0, b]\) admits a 1-bounded size bias coupling for any \( b \in (0, 4) \).

**Proof.** Suppose that \( X \) is uniformly distributed on \([0, 1]\), and that \( Y \) is the size biased version of \( X \). On \([0, 1]\), the density of \( X \) is \( f_X(x) = 1 \), the density of \( Y \) is \( f_Y(x) = 2x \), the cumulative distribution functions are \( F_X(t) = t \) and \( F_Y(t) = t^2 \), and the inverse cumulative distribution functions are \( F_X^{-1}(-1)(u) = u \) and \( F_Y^{-1}(-1)(u) = \sqrt{u} \) for \( 0 \leq u \leq 1 \). Observe that \( \max_{0 \leq u \leq 1}(\sqrt{u} - u) = 1/4 \), achieved at \( u = 1/4 \). Hence, using the quantile transform as in the proof of Lemma 7.2, there is a \( c \)-bounded
size bias coupling for the standard uniform with \( c = \frac{1}{4} \). By scaling, as in Remark 2.5, the uniform distribution on \([0, b]\) admits a \( b/4\)-bounded size bias coupling. □

**Proposition 5.6.** Suppose \( X = \sum X_i \) is the sum of finitely or countably many independent non-negative random variables \( X_i \) with \( 0 < \mathbb{E} X = \sum \mathbb{E} X_i < \infty \). If each \( X_i \) admits a \( 1\)-bounded size bias coupling, then so does \( X \).

**Proof.** For each \( i \) there is a coupling of \( X_i \) with its size biased version \( Y_i \), say, so that \( Y_i \leq X_i + 1 \). Let \( I \) denote an index \( i \) chosen (independently of all the \( X_j \) and \( Y_j \)) with probability \( \frac{\mathbb{E} X_i}{\mathbb{E} X} \). Then \( Y := X - X_I + Y_I \) is the size biased version of \( X \), see [10, Lemma 2.1] or [2, Sect. 2.4]. Since \( Y_i \leq X_i + 1 \) for all \( i \), it follows that \( Y \leq X + 1 \). □

**Example 5.7.** By Proposition 5.6 if \( X \) is Binomial, or more generally if \( X \) is the sum of (possibly infinitely many) independent Bernoulli random variables (with possibly different parameters) with \( 0 < \mathbb{E} X < \infty \) then \( X \) has a \( 1\)-bounded size bias coupling.

6. **Lévy([0, c]), the infinitely divisible distributions with Lévy measure supported on \([0, c]\).**

Among the distributions satisfying the hypotheses of Lemma 2.1 are those with characteristic function of the form

\[
\phi_X(u) := \mathbb{E} e^{iuX} = \exp \left( a \left( iu \alpha(\{0\}) + \int_{[0,c]} \frac{e^{iy} - 1}{y} \alpha(dy) \right) \right) \tag{19}
\]

where \( a \in (0, \infty) \) and \( \alpha \) is the probability distribution of a nonnegative nonzero random variable \( D \), with \( \mathbb{P}(D \in [0, c]) = 1 \). Given this characteristic function, the random variable \( X \) has \( a = \mathbb{E} X \), and, with \( X, D \) independent, \( X^* = dX + D \). See [2]. Special cases include:

1. \( c = 1, \mathbb{P}(D = 1) = 1; X \) is Poisson with mean \( a \).
2. \( c = 1, a = 1, D \) is uniformly distributed on \((0, 1)\); \( X \) has density \( f(x) = e^{-\gamma} \rho(x) \) where \( \rho \) is Dickman’s function and \( \gamma \) is Euler’s constant.

6.1. **Universal Dickman function like behavior.**

**Theorem 6.1.** Suppose \( X \) has distribution given by (19) for \( c > 0 \), and that for every \( \varepsilon > 0 \), the probability measure \( \alpha \) is not supported on \([0, c - \varepsilon]\). Then \( G(x) := \mathbb{P}(X \geq x) \) satisfies, as \( x \to \infty \),

\[
G(x) \approx x^{-x/c}, \text{ that is, } \frac{\log G(x)}{(-x/c) \log x} \to 1. \tag{20}
\]
Proof. The upper bound on $G(x)$ follows directly and easily from (5) in Theorem 2.2; see also Theorem 3.2 or the first inequality in Proposition 8.2.

For the lower bound, let $\varepsilon > 0$ be given, with $\varepsilon < c$. The characteristic function in (19) can also be expressed as

$$\phi_X(u) = \exp\left(iau\alpha_0 + \int_{(0,c]} (e^{iuy} - 1) \gamma(dy)\right). \quad (21)$$

Here $\gamma$ is a nonnegative measure on $(0,\infty)$, with $\gamma(dy)/(dy) = a/y$, and may be called the Lévy measure of the infinitely divisible random variable $X$; see Sato [15, Section 51], Bertoin [5], or [2]. The random variable $X$ can be realized as the constant $a\alpha_0$ plus the sum of the arrivals in the Poisson process $\mathcal{X}$ on $(0, c]$ with intensity measure $\gamma$. Let $Z$ be the number of arrivals of $\mathcal{X}$ in $[c - \varepsilon, c]$. We have $X \geq (c - \varepsilon)Z$. This yields

$$G(x) := \mathbb{P}(X \geq x) \geq \mathbb{P}(Z \geq \frac{x}{c - \varepsilon}).$$

Finally, $Z$ is a Poisson random variable with mean $\lambda = \gamma([c - \varepsilon, c]) \geq (a/c)\alpha([c - \varepsilon, c]) > 0$. We recall an elementary calculation: for integers $k \to \infty$, the Poisson ($\lambda$) distribution for $Z$ has $\mathbb{P}(Z \geq k) \approx \mathbb{P}(Z = k)$ with

$$\log \mathbb{P}(Z = k) = -\lambda + k \log \lambda - \log k! \sim -\log k! \sim -k \log k. \quad (22)$$

We use this with $k = \lfloor x/(c - \varepsilon) \rfloor \sim x/(c - \varepsilon)$ and $\log k \sim \log x$. \qed

Remark 6.2. Most probabilists are familiar with the $\approx$ notation of (20), with $a_n \approx b_n$ defined to mean $\log a_n \sim \log b_n$, for use in the context where $a_n$ grows or decays exponentially. The standard example is the large deviation statement that for i.i.d. sums, $\mathbb{P}(S_n \geq an) \approx \exp(-nI(a))$, and the $\approx$ relation hides factors with a slower than exponential order of growth or decay, in this case $1/\sqrt{n}$. When both $a_n$ and $b_n$ grow or decay even faster, for example with $a_n \sim n^{\pm n/c}$, an unfamiliar phenomenon arises, with $\approx$ hiding factors which grow or decay exponentially fast. One example appears in the conclusion (20) of Theorem 6.1, where $x^{-x/c} \approx (x/c)^{-x/c} \approx 1/\Gamma(x/c)$ as $x \to \infty$ — with the last expression being relevant because it corresponds to the Gamma function upper bound in Theorem 3.2. A second example occurs in (22), where at first it appears strange that the parameter $\lambda$ does not appear on the right hand side — this reflects the fact that for any fixed $\lambda, \lambda' > 0$, when $Z$ and $Z'$ are Poisson with parameters $\lambda, \lambda'$ respectively, $\mathbb{P}(Z = k) \approx \mathbb{P}(Z' = k)$ as $k \to \infty$. The first example hides the exponentially growing factor $c^{x/c}$, and the second hides the factor $(\lambda'/\lambda)^k$. 


7. Bounded coupling, monotone coupling, and a sandwich principle

Suppose that the distributions of random variables $X, Y$ have been specified, with cumulative distribution functions $F_X, F_Y$ respectively. We will clarify the relations between hypotheses of the form

\[ \exists \text{ a coupling, } \mathbb{P}(Y \leq X + c) = 1, \]  
\[ \exists \text{ a coupling, } \mathbb{P}(|Y - X| \leq c) = 1, \]  
\[ \exists \text{ a coupling, } \mathbb{P}(Y \in [X, X + c]) = 1, \]  
\[ \exists \text{ a coupling, } \mathbb{P}(X \leq Y) = 1. \]

It is well-known that if $Y = d X^*$, then (26) holds. We observe that (23) is the hypothesis for our Lemma 2.1 and Theorem 2.2, while (24) and (25) are used as hypotheses for the first results on concentration via size bias couplings, in [8].

**Proposition 7.1.** Given that (26) holds, all of (23) – (25) are equivalent.

The proof of Proposition 7.1 will be given later in this section.

7.1. One sided bounds. Among the four hypotheses (23) – (26), only (26) is standard. The relation of stochastic domination, that $X$ lies below $Y$, written $X \preceq Y$, is usually defined by the condition

\[ \forall t, F_X(t) \geq F_Y(t), \]

and it is well known that (27) is equivalent to (26). See for example Lemma 7.2. Stochastic domination is often considered in the more general context where $X, Y$ are random elements in a partially ordered set, see for example [14].

7.2. Two sided bounds: historical perspective. Dudley [6, Prop 1 and Thm 2], following earlier work of Strassen [16], proved the following result. Given distributions $\mu$ and $\nu$ respectively for random variables $X$ and $Y$ taking values in a Polish metric space $(S, d)$ and $\alpha > 0$, write $A^\alpha = \{x : d(x, A) \leq \alpha\}$ to denote the closed $\alpha$ neighborhood of the set $A \subset S$. Then the following are equivalent:

For all closed sets $A$, $\mu(A) \leq \nu(A^\alpha)$ \hspace{1cm} (28)

For all closed sets $A$, $\nu(A) \leq \mu(A^\alpha)$ \hspace{1cm} (29)

There exists a coupling under which $\mathbb{P}(d(X, Y) \leq \alpha) = 1.$ (30)
When \( X, Y \) are real valued random variables, and \( \alpha = c \), condition (23) is trivially equivalent to (30), and (28) or (29) can be shown to be equivalent to \( \forall t, F_Y(t - c) \leq F_X(t) \leq F_Y(t + c) \).

7.3. Proof of Proposition 7.1. It is clear that (25) implies (24) implies (23). It remains to show that (26) and (23) together imply (25), and this follows immediately by using the implication i) implies iii) with \( b = 0 \) in the following Lemma.

**Lemma 7.2.** For random variables \( X, Y \) with cumulative distribution functions \( F_X, F_Y \) respectively and \( b, c \in \mathbb{R} \), the following are equivalent:

i) There exist a coupling such that \( P(X \leq Y + b) = 1 \) and a coupling such that \( P(Y \leq X + c) = 1 \).

ii) For all \( t, F_Y(t - b) \leq F_X(t) \leq F_Y(t + c) \).

iii) There exists a coupling such that \( P(X - b \leq Y \leq X + c) = 1 \).

iv) The exists a coupling such that \( X - b \leq Y \leq X + c \) for all \( \omega \).

**Proof.** To prove that i) implies ii) we use the coupling in which \( P(X \leq Y + b) = 1 \) and calculate

\[
F_Y(t - b) = P(Y \leq t - b) \leq P(X \leq t) + P(X > Y + b) = P(X \leq t) = F_X(t)
\]

and similarly for \( F_X(t) \leq F_Y(t + c) \).

To show ii) implies iv) we may use the quantile transform to couple \( X \) and \( Y \) to a single random variable \( U \), uniformly distributed in \((0, 1)\). This transform is written informally as

\[
X(\omega) := F_X^{(-1)}(U(\omega)), \quad Y(\omega) := F_Y^{(-1)}(U(\omega))
\]

and more formally as \( X(\omega) := \inf\{t : F_X(t) \geq U(\omega)\}, Y(\omega) := \inf\{t : F_Y(t) \geq U(\omega)\} \). Under this coupling, \( F_Y(t - b) \leq F_X(t) \) for all \( t \) implies \( X(\omega) \leq Y(\omega) + b \) for all \( \omega \), and \( F_X(t) \leq F_Y(t + c) \) for all \( t \) implies \( Y(\omega) \leq X(\omega) + c \) for all \( \omega \). Together ii) implies \( X(\omega) - b \leq Y(\omega) \leq X(\omega) + c \) holds for all \( \omega \).

Finally, iv) implies iii) trivially, and iii) implies i) a fortiori. \( \square \)

7.4. A sandwich principle.

**Corollary 7.3. Sandwich principle.** Suppose \( X, Y, Z \) are random variables with cumulative distribution functions \( F_X, F_Y, F_Z \) satisfying \( F_Z(t) \leq F_Y(t) \leq F_X(t) \) for all \( t \). If there is a \( c \)-bounded coupling of \( X \) and \( Z \), as defined via (24), then there exists a \( c \)-bounded monotone coupling of \( X \) and \( Y \), as defined by (25).

**Proof.** Using Lemma 7.2 with \( Z \) in place of \( Y \), the \( c \)-bounded coupling of \( X \) and \( Z \) implies \( F_X(t) \leq F_Z(t + c) \) for all \( t \). Then \( F_Y(t) \leq F_X(t) \leq F_Z(t + c) \leq F_Y(t + c) \) for all \( t \), and Lemma 7.2 gives the existence of the \( c \)-bounded monotone coupling of \( X \) and \( Y \). \( \square \)
Application. As an illustration of the sandwich principle, we prove the following corollary. The special case where $X$ is Binomial$(m, p)$ for $p \in (0, 1)$ and $m \geq 1$ was proved in [9, Lemmas 3.2, 3.3] by an explicit calculation. The sandwich principle enables a short proof for the general case, with no calculation.

Corollary 7.4. Suppose $X$ satisfies the hypotheses of Lemma 2.1, with $c = 1$, and the distribution of $Y$ is defined to be that of $X$, conditional on $X > 0$. Then there is a coupling in which $Y - X \in [0, 1]$ for all $\omega$.

Proof. Trivially, $Y$ stochastically dominates $X$. Take the distribution of $Z$ to be the size biased distribution of $X$. By assumption, there is a coupling in which $P(Z - X \in [0, 1]) = 1$. Trivially, the distribution of $Z$, initially defined as the size-biased distribution of $X$, is also the size-biased distribution of $Y$. Hence $Z$ dominates $Y$, and the sandwich principle applies with $c = 1$. □

This result may be applied to any of the random variables $X$ discussed in Section 5, and in particular to those obtained using Proposition 5.6. If $X$ is (non-negative) integer valued, for example, a Binomial random variable or a sum of independent Bernoulli random variables, then $Y$ is also integer valued and the conclusion of Corollary 7.4 is easily strengthened to $Y - X \in \{0, 1\}$ for all $\omega$.

8. Analysis of the upper tail and lower tail bounds

In this section we obtain results enabling us to compare the upper and lower tail bounds in Theorem 3.2 with those in Theorem 4.2.

Lemma 8.1. For $0 < u \leq v$

$$\frac{\Gamma(v + 1/2)}{\Gamma(u + 1/2)} \geq \frac{v^v}{e^{v-u}u^u}. \tag{31}$$

Proof. For $x > 0$, using Gauss’ formula, see [17, Sect 12.3],

$$(\log \Gamma)'(x + 1/2) = \int_0^\infty \left( \frac{e^{-t}}{t} - \frac{e^{-t(x+1/2)}}{1 - e^{-t}} \right) dt$$

together with

$$\log x = \int_0^\infty \left( \frac{e^{-t} - e^{-tx}}{t} \right) dt$$
we obtain
\[
(\log \Gamma)'(x + 1/2) - \log x = \int_0^\infty \left( \frac{e^{-tx}}{t} - \frac{e^{-t(x+1/2)}}{1 - e^{-t}} \right) dt = \int_0^\infty \frac{e^{-t(x+1/2)}(2 \sinh(t/2) - t)}{t(1 - e^{-t})} dt \geq 0.
\]

Therefore for $0 < u \leq v$,
\[
\frac{\Gamma(u + 1/2)}{\Gamma(v + 1/2)} \geq \exp \left\{ \int_u^v \log x \, dx \right\} = \exp \left\{ \left[ x \log x - x \right]_u^v \right\} = \frac{v^v}{e^{v-u}u^u}.
\]

\textbf{Proposition 8.2.} For $0 < a \leq x$,
\[
\frac{a^{x-a} \Gamma(a+1)}{\Gamma(x+1)} \leq \frac{(a + \frac{1}{2})^{a+1/2} (ae)^{x-a}}{(x + \frac{1}{2})^{x+1/2}} \leq \left( a + \frac{1}{2} \right)^{1/2} \left( \frac{a}{x} \right)^x e^{x-a}.
\]

\textit{Proof.} Taking $u = a + 1/2$ and $v = x + 1/2$ in (31), we get
\[
\frac{\Gamma(x + 1)}{\Gamma(a+1)} \geq \frac{(x + \frac{1}{2})^{x+1/2}}{e^{x-a}(a + \frac{1}{2})^{a+1/2}},
\]
giving (32). A simple calculus argument shows that $t \mapsto (a+t)^a/(x+t)^x$ is decreasing for $t \in [0, \infty)$. In particular $(a + \frac{1}{2})^a / (x + \frac{1}{2})^x \leq a^a/x^x$, and this gives (33).

\textbf{Proposition 8.3.} For $0 < x < a$,
\[
\frac{\Gamma(a+1)}{a^{a-x} \Gamma(x+1)} \geq \frac{(a + \frac{1}{2})^{a+1/2}}{(ae)^{a-x} (x + \frac{1}{2})^{x+1/2}} \geq \left( a + \frac{1}{2} \right)^{1/2} \left( \frac{a}{x} \right)^x e^{x-a}.
\]

\textit{Proof.} Essentially the same as for Proposition 8.2, but with the roles of $x$ and $a$ switched. \hfill $\square$
9. Comparison of the bounds

9.1. Relative strength of our upper and lower tail bounds. Here we compare the elementary “product bounds” (5, 6) given in Theorem 2.2 with the Gamma function bounds (14, 15) given in Theorem 3.2 and the bounds (17, 18) obtained in Theorem 4.2 from the moment generating function estimate.

It is clear from Theorem 3.2 (together with the scaling relationship (10)) that the product bounds (5, 6) are at least as strong (i.e. small) as the corresponding Gamma bounds (14, 15). However the relationship between the Gamma bounds and the bounds (17, 18) in Theorem 4.2 is more complicated. See Propositions 8.2 and 8.3. For the upper tail, with \( a \leq x \), the Gamma bound (14) is sharper than the bound (17) from Theorem 4.2 by a factor \( \sqrt{(a + c/2)/(x + c/2)} \). However, the situation is reversed for the lower tail, with \( 0 \leq x \leq a \), where now the bound (18) from Theorem 4.2 is sharper than the Gamma bound (15) by a factor \( \sqrt{(x + c/2)/(a + c/2)} \). Since the Gamma bound (15) and the product bound (6) agree whenever \( a - x \) is an integer, this suggests (but does not prove) that the bound (18) is the best of the three for the lower tail.

Numerical investigations suggest that (18) is in fact the best estimate to use for the lower tail, and beats the simple product rule \( \ell(x, a, c) \) of (6). Recall however, from Section 2.1 the lower tail bounds derived from (3) in combination with the one-sided Chebyshev inequality (7), in particular the functions \( \ell_j(x, a, c) \) defined in (8). Numerical investigations suggest that for all \( a, c > 0 \) with sufficiently large \( a/c \) there exists \( x \in (0, a) \) such that the bound given in (18) is less than the product bound (6) and is less than the one-sided Chebyshev estimate (7), but is greater than \( \ell_j(x, a, c) \) for some nonnegative integer \( j \leq (a - x)/c \).

9.2. Results of Ghosh and Goldstein. The paper [8] proved the inequalities \( G(x) \leq e^{-(x-a)^2/(ca+cx)} \) for \( x \geq a \) and \( F(x) \leq e^{-(a-x)^2/(2ca)} \) for \( x \leq a \) for a random variable which admits a \( c \)-bounded size bias coupling. The following lemmas show that our bounds given in Theorem 4.2 outperform those given by [8].

**Lemma 9.1.** Suppose \( 0 < a \leq x \) and \( c > 0 \). Then

\[
\left( \frac{a}{x} \right)^{x/c} e^{(x-a)/c} \leq e^{-(x-a)^2/(ca+cx)}.
\]

**Proof.** Jensen’s inequality applied to the function \( f(x) = 1/x \) on the interval \([1, 1+u]\) gives \( \log(1+u) \geq 2u/(2+u) \) for \( u > 0 \). Taking
\( u = (x - a)/a \) we get \( \log(x/a) \geq 2(x - a)/(x + a) \) and so
\[
x \log(a/x) + x - a \leq \frac{-2x(x - a)}{x + a} + x - a = \frac{(x - a)^2}{x + a}.
\]
Dividing by \( c \) and applying the exponential function to both sides gives (36).

**Lemma 9.2.** Suppose \( 0 \leq x \leq a \) and \( a, c > 0 \). Then
\[
\left( \frac{a}{x} \right)^{x/c} e^{(x-a)/c} \leq e^{-(x-a)^2/(2ca)}.
\] (37)

**Proof.** Integrating the inequality \( u^{-1} \leq (1 + u^{-2})/2 \) from 1 to \( 1/\delta > 1 \) gives \( \log(1/\delta) \leq (1/\delta - \delta)/2 \). Putting \( \delta = x/a \) gives
\[
x \log(a/x) - a + x \leq \frac{x}{2} \left( \frac{a}{x} - \frac{x}{a} \right) - a + x = \frac{(x - a)^2}{2a}.
\]
Dividing by \( c \) and applying the exponential function to both sides gives (37). \( \square \)

9.3. **Hoeffding bounds.** Suppose \( X = X_1 + \cdots + X_n \) where the \( X_i \) are independent and take values in \([0, 1]\), and let \( a = \mathbb{E} X \). Clearly \( a < \infty \) and to avoid trivialities we assume \( a > 0 \). Hoeffding [13, Thm 1] proved that for \( a \leq x < n \)
\[
\mathbb{P}(X \geq x) \leq \left( \frac{a}{x} \right)^x \left( \frac{n-a}{n-x} \right)^{n-x}.
\]
The inequality \( \left( \frac{n-a}{n-x} \right)^{n-x} \leq e^{x-a} \)
for \( 0 < x < n \), (see for example [11]), and the fact that \( \mathbb{P}(X \geq x) = 0 \) for \( x > n \), together give the upper tail bound
\[
\mathbb{P}(X \geq x) \leq \left( \frac{a}{x} \right)^x e^{-x-a} \quad \text{for all } x \geq a.
\] (38)
A similar argument with \( X_i \) replaced by \( 1 - X_i \) and \( X \) replaced by \( n - X \) gives the lower tail bound
\[
\mathbb{P}(X \leq x) \leq \left( \frac{a}{x} \right)^x e^{x-a} \quad \text{whenever } 0 \leq x \leq a.
\] (39)
Notice that the right sides of (38) and (39) do not depend on the number \( n \) of summands in \( X \). Other related inequalities, also referred to as Hoeffding or Chernoff–Hoeffding bounds, involve the parameter \( n \).
From Proposition 5.6 we know that any random variable $X$ of the form above admits a 1-bounded size bias coupling. Therefore the Hoeffding bounds (38, 39) are a special case of the bounds (17, 18) in our Theorem 4.2. Our best upper tail bound, given by (1) is smaller than the Hoeffding upper tail bound (38) by a factor \[ \left( \frac{a + 1/2}{x + 1/2} \right)^{1/2} \]
Moreover our results have a broader scope, applying to any random variable $X$ which admits a 1-bounded size bias coupling. In particular it applies to sums of independent nonnegative random variables such as Uniform[0,4] and Lévy[0,1] (as discussed in Examples 5.5 and 5.3).

REFERENCES

[1] R. Arratia, S. Garibaldi, and J. Killian. Asymptotic distribution for the birthday problem with multiple coincidences, via an embedding of the collision process. Preprint. 2013.
[2] R. Arratia, L. Goldstein, and F. Kochman. Size bias for one and all. Preprint. 2013.
[3] Richard Arratia, A. D. Barbour, and Simon Tavaré. Logarithmic combinatorial structures: a probabilistic approach. EMS Monographs in Mathematics. European Mathematical Society (EMS), Zürich, 2003.
[4] J. Bartroff, L. Goldstein, and Ü Işlak. Bounded size biased couplings for log concave distributions and concentration of measure for occupancy models. Preprint. 2013.
[5] Jean Bertoin. Subordinators: examples and applications. In Lectures on probability theory and statistics (Saint-Flour, 1997), volume 1717 of Lecture Notes in Math., pages 1–91. Springer, Berlin, 1999.
[6] R. M. Dudley. Distances of probability measures and random variables. Ann. Math. Statist., 39:1563–1572, 1968.
[7] Subhankar Ghosh and Larry Goldstein. Applications of size biased couplings for concentration of measures. Electron. Commun. Probab., 16:70–83, 2011.
[8] Subhankar Ghosh and Larry Goldstein. Concentration of measures via size-biased couplings. Probab. Theory Related Fields, 149(1-2):271–278, 2011.
[9] Larry Goldstein and Mathew D. Penrose. Normal approximation for coverage models over binomial point processes. Ann. Appl. Probab., 20(2):696–721, 2010.
[10] Larry Goldstein and Yosef Rinott. Multivariate normal approximations by Stein’s method and size bias couplings. J. Appl. Probab., 33(1):1–17, 1996.
[11] Torben Hagerup and Christine Rüb. A guided tour of Chernoff bounds. Inform. Process. Lett., 33(6):305–308, 1990.
[12] Adolf Hildebrand and Gérald Tenenbaum. Integers without large prime factors. J. Théor. Nombres Bordeaux, 5(2):411–484, 1993.
[13] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. J. Amer. Statist. Assoc., 58:13–30, 1963.
[14] Thomas M. Liggett. Interacting particle systems, volume 276 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, New York, 1985.
[15] Ken-iti Sato. *Lévy processes and infinitely divisible distributions*, volume 68 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1999. Translated from the 1990 Japanese original, Revised by the author.

[16] V. Strassen. The existence of probability measures with given marginals. *Ann. Math. Statist.*, 36:423–439, 1965.

[17] E. T. Whittaker and G. N. Watson. *A course of modern analysis. An introduction to the general theory of infinite processes and of analytic functions: with an account of the principal transcendental functions*. Fourth edition. Reprinted. Cambridge University Press, New York, 1962.

(Richard Arratia) Department of Mathematics, University of Southern California, Los Angeles CA 90089.  
*E-mail address*: rarratia@usc.edu

(Peter Baxendale) Department of Mathematics, University of Southern California, Los Angeles CA 90089.  
*E-mail address*: baxendal@usc.edu