

# Concentration inequalities from monotone couplings for graphs, walks, trees and branching processes

Tobias Johnson<sup>a,\*</sup>, Erol Peköz<sup>b</sup>

<sup>a</sup> College of Staten Island, United States of America

<sup>b</sup> Boston University, United States of America

Received 10 August 2021; received in revised form 14 April 2022; accepted 15 June 2022

Available online 18 June 2022

## Abstract

Generalized gamma distributions arise as limits in many settings involving random graphs, walks, trees, and branching processes. Peköz et al. (2016) exploited characterizing distributional fixed point equations to obtain uniform error bounds for generalized gamma approximations using Stein's method. Here we show how monotone couplings arising with these fixed point equations can be used to obtain sharper tail bounds that, in many cases, outperform competing moment-based bounds and the uniform bounds obtainable with Stein's method. Applications are given to concentration inequalities for preferential attachment random graphs, branching processes, random walk local time statistics and the size of random subtrees of uniformly random binary rooted plane trees.

© 2022 Elsevier B.V. All rights reserved.

MSC: 60E15; 60J80; 05C80

Keywords: Concentration inequality; Tail bound; Stein's method; Preferential attachment graph; Galton–Watson process

## 1. Introduction

Stein's method is used to obtain distributional approximation error bounds in a wide variety of settings in applied probability where normal, Poisson, gamma and other limits arise. The method was introduced by Stein for normal approximation [35]. Chen adapted it for Poisson approximation [13], and since then it has been developed in many directions. Introductions to Stein's method can be found in [6,14,36]; we also refer to the surveys [5,12,32].

One variant of the approach (see [22,30] and the references therein) starts with a distributional fixed point equation that the limit distribution satisfies and obtains error bounds in

\* Corresponding author.

E-mail addresses: [tobias.johnson@csi.cuny.edu](mailto:tobias.johnson@csi.cuny.edu) (T. Johnson), [pekoz@bu.edu](mailto:pekoz@bu.edu) (E. Peköz).

terms of how closely both sides of the fixed point equation can be coupled together, i.e., their Wasserstein distance. The fixed point equation often has a probabilistic interpretation that can be leveraged to achieve a close coupling. To illustrate, we define a generalization of the size-bias transform for a random variable.

**Definition 1.** If  $X$  is a nonnegative random variable with  $0 < \mathbf{E}[X^\beta] < \infty$ , we say that  $X^{(\beta)}$  is the  $\beta$ -power bias transform of  $X$  if

$$\mathbf{E}[X^\beta] \mathbf{E}[f(X^{(\beta)})] = \mathbf{E}[X^\beta f(X)]$$

holds for all for bounded measurable  $f$ .

If  $X^s \stackrel{d}{=} X^{(1)}$ , the familiar size-bias transform of  $X$ , it can be shown that  $X$  satisfies the distributional fixed-point equation  $X \stackrel{d}{=} X^s - 1$  if and only if  $X$  follows a Poisson distribution. Stein's method can be used to make an approximate version of this statement: the total variation distance between the law of  $X$  and a Poisson distribution can be bounded by the Wasserstein distance between  $X$  and  $X^s - 1$  [32, Theorem 4.13]. Similarly, the exponential distribution is characterized by the fixed-point equation  $X \stackrel{d}{=} UX^s$  for an independent Uniform(0, 1) variable  $U$ . Under the assumption that  $X$  has a finite second moment, Stein's method can be used to show that the Wasserstein distance between  $X$  and an exponential distribution is at most twice the Wasserstein distance between  $X$  and  $UX^s$  [29, Theorem 2.1].

The bounds above allow for accurate approximations of the law of  $X$  in the bulk of the distribution. Tail bounds can also be obtained when the two sides of the distributional fixed-point equation can be coupled together with some sort of monotonicity. This is carried out for the Poisson distribution in [1, 21]. For example, if there exists a coupling of  $X^s$  with  $X$  so that  $X^s - 1 \leq X$  a.s., then  $X$  satisfies a Poisson-like tail bound. [15] loosened this monotonicity condition to  $\mathbf{P}[X^s - 1 \leq X \mid X^s \geq x] \geq p$  and applied the results to bounding the second largest eigenvalue of the adjacency matrix of a random regular graph.

The monotonicity condition we study in the current article involves the following definition from [30].

**Definition 2.** For a nonnegative random variable  $X$  and  $\alpha, \beta > 0$ , the variable  $X^*$  is the  $(\alpha, \beta)$ -generalized equilibrium distribution of  $X$  if

$$X^* \stackrel{d}{=} V_\alpha X^{(\beta)},$$

where  $V_\alpha$  has density  $\alpha x^{\alpha-1} dx$  on  $[0, 1]$  and is independent of  $X^{(\beta)}$ .

Throughout the paper, we use the notation  $X^*$  to mean a random variable with the  $(\alpha, \beta)$ -generalized equilibrium distribution of  $X$ , with  $\alpha$  and  $\beta$  specified as necessary. The familiar equilibrium distribution  $X^e$  of  $X$  from renewal theory is the case where  $\alpha = \beta = 1$ . It is well known that the fixed points of the equilibrium transform, i.e., the distributions satisfying  $X^e \stackrel{d}{=} X$ , are the exponential distributions. Generalizing this fact, for any  $\alpha, \beta > 0$  let  $\text{GG}(\alpha, \beta)$  denote the  $(\alpha, \beta)$ -generalized gamma distribution, which has density function  $\beta x^{\alpha-1} e^{-x^\beta} / \Gamma(\alpha/\beta) dx$  for  $x > 0$ . The fixed points of the generalized equilibrium transform are the generalized gamma distributions, up to scaling. That is, for any choice of  $\alpha, \beta > 0$ , the random variable  $X$  satisfies  $X^* \stackrel{d}{=} X$  if and only if  $X \stackrel{d}{=} cZ$  where  $Z \sim \text{GG}(\alpha, \beta)$  [26, Theorem 5.1].

The main result of [30] is a bound on the Kolmogorov distance between a properly rescaled version of  $X$  and  $Z \sim \text{GG}(\alpha, \beta)$  when  $X$  and  $X^*$  are not identical but are close in the Lévy–Prokhorov metric. This yields a bound on  $|\mathbf{P}[cX > t] - \mathbf{P}[Z > t]|$  that can be used to estimate probabilities in the bulk of the distribution of  $X$ . But since this bound is uniform in  $t$ , it is too large to be useful for small tail probabilities in applications. This is the launching point for the current article.

Let  $X \leq Y$  denote that  $X$  is stochastically dominated by  $Y$  in the usual stochastic order. We introduce two stochastic orders as follows. For a constant  $p \in (0, 1]$ , we write  $X \leq_{\bar{p}} Y$  to denote that  $\mathbf{P}[X > t] \leq \mathbf{P}[Y > t]/p$  for all  $t \in \mathbb{R}$ . Similarly,  $X \leq_{\underline{p}} Y$  denotes that  $\mathbf{P}[X \leq t] \geq p\mathbf{P}[Y \leq t]$  for all  $t \in \mathbb{R}$ . In the  $p = 1$  case, both orders are the usual one. For  $p < 1$ , they represent two different relaxations of the usual order and have alternate characterizations given in Lemma 8.

Our two main result are that stochastic domination of  $X^*$  by  $X$  in these orders leads to upper and lower tail probability bounds:

**Theorem 3.** Suppose that  $X^* \leq_{\bar{p}} X$  for some  $p \in (0, 1]$  and  $\alpha, \beta > 0$ . Let  $\mu = \left(\frac{\beta}{\alpha} \mathbf{E}X^\beta\right)^{1/\beta}$ . If  $\alpha \neq \beta$ , then for any  $t \geq 1$ ,

$$\mathbf{P}[X \geq \mu t] \leq t^\alpha e^{2-pt^\beta}.$$

If  $\alpha = \beta$ , then for all  $t > 0$ ,

$$\mathbf{P}[X \geq \mu t] \leq e^{1-pt^\beta}.$$

**Theorem 4.** Suppose  $X^* \leq_{\underline{p}} X$  for some  $p \in (0, 1]$  and  $\alpha, \beta > 0$ . Let  $\mu = \left(\frac{\beta}{\alpha} \mathbf{E}X^\beta\right)^{1/\beta}$ . Then

$$\mathbf{P}[X \leq \mu t] \leq \left(\frac{\alpha}{\beta}\right)^{1-\alpha/\beta} \frac{t^\alpha}{\frac{\alpha}{\beta} p - t^\beta}, \quad (1)$$

so long as  $t^\beta < p\alpha/\beta$ . If  $\alpha = \beta$ , an improved bound holds for all  $t \geq 0$ :

$$\mathbf{P}[X \leq \mu t] \leq \frac{t^\alpha}{p}. \quad (2)$$

The factor  $\mu$  used to rescale  $X$  in these theorems is equal to 1 when  $X$  has the  $\text{GG}(\alpha, \beta)$  distribution.

We mention that the case  $\alpha = \beta = p = 1$  case of Theorem 3 was already proven by Mark Brown [10, Theorem 3.2]; see Section 2.2.

As an application of these concentration theorems, we prove several results for graphs, walks, trees and branching processes. We next define several quantities from such models and show that the above inequalities apply.

**Preferential attachment random graphs.** Consider a preferential attachment random graph model (see [4,30]) that starts with an initial “seed graph” consisting of one node, or a collection of nodes grouped together, having total weight  $w$ . Additional nodes are added sequentially and when a node is added it attaches  $l$  edges, one at a time, directed from it to either itself or to nodes in the existing graph according to the following rule: each edge attaches to a potential node with chance proportional to that node’s weight right before the moment of attachment, where incoming edges contribute weight one to a node and each node other than the initial node when added has initial weight one. The case where  $l = 1$  is the usual Barabási–Albert

tree with loops but started from a node with initial weight  $w$ . Let  $W$  be the total weight of the initial “seed graph” after an additional  $n$  edges have been added to the graph.

**Random binary rooted plane trees.** Let  $U$  be the number of vertices in the minimal spanning tree spanned by the root and  $k$  randomly chosen distinct leaves of a uniformly chosen binary, rooted plane tree with  $2n - 1$  nodes, that is, with  $n$  leaves and  $n - 1$  internal nodes.

**Random walk local times.** Consider the one-dimensional simple symmetric random walk  $S_n = (S_n(0), \dots, S_n(n))$  of length  $n$  starting at the origin. Define

$$L_n = \sum_{i=0}^n 1\{S_n(i) = 0\}$$

to be the number of times the random walk visits the origin by time  $n$ . Let

$$L_{2n}^b \sim [L_{2n} \mid S_{2n}(0) = S_{2n}(2n) = 0]$$

be the local time of a random walk bridge. Here we use the notation  $[X \mid E]$  to denote the distribution of a random variable  $X$  conditional on an event  $E$  with nonzero probability.

The next result shows that the above concentration inequalities hold for these quantities.

**Theorem 5.** *With the above definitions,*

- (a)  $W^* \leq W$  with  $\alpha = w, \beta = l + 1$
- (b)  $U^* \leq U$  with  $\alpha = 2k, \beta = 2$
- (c)  $(L_n)^* \leq L_n$  with  $\alpha = 1, \beta = 2$
- (d)  $(L_{2n}^b)^* \leq L_{2n}^b$  with  $\alpha = 2, \beta = 2$

and the conditions and conclusions of [Theorems 3](#) and [4](#) hold for  $W, V, L_n$ , and  $L_{2n}^b$  with  $p = 1$  and the corresponding values in (a)–(d) for  $\alpha, \beta$ .

We also give two tail bounds for Galton–Watson branching processes. See [Section 4.5](#) for a review of previous bounds.

**Theorem 6.** *Let  $Z_n$  be the size of the  $n$ th generation of a Galton–Watson process, and let  $\mu = \mathbf{E}Z_1$ , the mean of its child distribution. Consider  $Z_n^*$  with  $\alpha = \beta = 1$ . If  $Z_1^* \leq Z_1$ , then  $Z_n^* \leq Z_n$  for all  $n$ , and*

$$\mathbf{P}[Z_n \geq t\mu^n] \leq e^{1-t}, \quad (3)$$

and

$$\mathbf{P}[Z_n \leq t\mu^n] \leq t \quad (4)$$

for all  $t > 0$ .

[Theorem 6](#) requires the child distribution to be supported on  $\{1, 2, \dots\}$ , since the condition  $Z_1^* \leq Z_1$  fails if  $\mathbf{P}[Z_1 = 0] > 0$ . The next result relaxes this requirement and allows us to consider Galton–Watson trees with a nonzero extinction probability, with our concentration result applying conditional on nonextinction. We impose a condition on the child distribution that comes from reliability theory. For a random variable  $X$  taking values in the positive integers, we say that  $X$  is *D-IFR* (which stands for *discrete increasing failure rate*) if  $\mathbf{P}[X = k \mid X \geq k]$  is increasing in  $k$  for  $k \geq 1$ . As we will discuss in [Section 4.1](#), if  $X$  is D-IFR then  $X^* \leq X$  with  $\alpha = \beta = 1$ .

In the following theorem and onward, for a random variable  $X$  we use  $X^>$  to denote a random variable with distribution  $[X \mid X > 0]$ .

**Theorem 7.** Let  $Z_n$  be the size of the  $n$ th generation of a Galton–Watson tree, and suppose that  $Z_1^>$  is D-IFR. Then  $Z_n^* \leq Z_n^>$  with  $\alpha = \beta = 1$ , and with  $m(n) = \mathbf{E}[Z_n \mid Z_n > 0]$ ,

$$\mathbf{P}[Z_n \geq tm(n) \mid Z_n > 0] \leq e^{1-t}, \quad (5)$$

and

$$\mathbf{P}[Z_n \leq tm(n) \mid Z_n > 0] \leq t \quad (6)$$

for all  $t > 0$ .

This theorem holds in subcritical, critical, and supercritical cases alike. The difference comes only in the mean  $m(n)$ , which grows exponentially for a supercritical tree, grows linearly for a critical tree, and remains bounded for a subcritical tree.

Theorems 5–7 apply the  $p = 1$  cases of Theorems 3 and 4 with the usual stochastic order. We use the  $p < 1$  case in this paper only in Remark 28 to sketch a way to simplify the proof of Theorem 5 at the cost of a weaker concentration inequality. Nonetheless, we expect that the  $p < 1$  case will prove useful. For the analogous concentration bound in [15] based on the Poisson distributional fixed point  $X^s \stackrel{d}{=} X + 1$ , the  $p < 1$  case is essential to applications on random regular graphs (see [15, Proposition 2.3] and [40, Theorem 4.1]) and on interacting particle systems [24, Proposition 5.17].

The tail bounds we produce in this paper are sharp in many circumstances. The results of Theorems 3 and 4 in the  $\alpha = \beta$  case are sharp, which was known already in the previously proven case  $\alpha = \beta = p = 1$  for the upper tail [11]. When  $\alpha \neq \beta$ , we expect that the factor  $t^\alpha$  in the upper tail bound is not sharp. We discuss this further in Section 2.4. The applications of Theorems 3 and 4 given in Theorem 5 with  $\alpha = \beta$  seem likely to be sharp (see Remark 18), and our results on Galton–Watson processes are sharp as well (see Section 4.4).

Our main concentration results, Theorems 3 and 4, are proven in Section 2. In Section 3, we consider an urn model and prove that its counts  $N$  satisfy  $N^* \leq N$  with  $\alpha$  and  $\beta$  depending on the model’s parameters. The random variables  $W$ ,  $U$ ,  $L_n$ , and  $L_{2n}^b$  are expressed in terms of this urn model in [30], and Theorem 5 follows. One part of the proof, a regularity property for the urn model similar to log-concavity, is shown by a very technical argument in Appendix A. Section 4 gives the proofs of Theorems 6 and 7 on Galton–Watson trees. After some background material on reliability theory and on forming the equilibrium transform, Theorem 6 is easy to prove (and in fact is essentially proven in [39]). The proof of Theorem 7 is more difficult and requires us to establish some delicate properties of the D-IFR class of distributions. In Appendix B, we give some proofs that are well known in reliability theory but hard to find in the literature.

## 2. Proof of concentration theorems

Recall that  $X \leq_{\overline{p}} Y$  and  $X \leq_p Y$  denote that  $\mathbf{P}[X > t] \leq \mathbf{P}[Y > t]/p$  and  $\mathbf{P}[X \leq t] \geq p\mathbf{P}[Y \leq t]$ , respectively, for all  $t \in \mathbb{R}$ . We start by characterizing these orders in terms of couplings, along the same lines as the standard fact that  $X \leq Y$  if and only if there exists a coupling of  $X$  and  $Y$  so that  $X \leq Y$  a.s. We have not seen these stochastic orders defined before, but they are used in form (ii) in [15,16].

**Lemma 8.** *The following statements are equivalent:*

- (i) *there exists a coupling  $(X, Y)$  for which  $\mathbf{P}[X \leq Y \mid X] \geq p$  a.s.;*
- (ii) *there exists a coupling  $(X, Y)$  for which  $\mathbf{P}[X \leq Y \mid X \geq t] \geq p$  for all  $t \in \mathbb{R}$  where the conditional probability is defined;*
- (iii)  $X \leq_{\overline{p}} Y$ ;

as are the following statements:

- (i) *there exists a coupling  $(X, Y)$  for which  $\mathbf{P}[X \leq Y \mid Y] \geq p$  a.s.;*
- (ii) *there exists a coupling  $(X, Y)$  for which  $\mathbf{P}[X \leq Y \mid Y \leq t] \geq p$  for all  $t \in \mathbb{R}$  where the conditional probability is defined;*
- (iii)  $X \leq_p Y$ .

**Proof.** It is clear that (i) implies (ii) in both sets of statements. To go from (ii) to (iii) in the first set of statements, observe that for any  $s \in \mathbb{R}$ ,

$$\mathbf{P}[Y \geq s] \geq \mathbf{P}[X \leq Y \text{ and } X \geq s] \geq p\mathbf{P}[X \geq s],$$

applying (ii) in the second inequality. Now let  $s$  approach  $t$  downward to prove (iii). A similar argument proves that (ii) implies (iii) for the second set of statements.

Now we show that (iii)  $\implies$  (i) for the first set of statements. Let  $B \sim \text{Bernoulli}(p)$  be independent of  $X$ , and define a random variable  $X'$  taking values in  $[-\infty, \infty)$  by

$$X' = \begin{cases} X & \text{if } B = 1, \\ -\infty & \text{if } B = 0. \end{cases}$$

Then

$$\mathbf{P}[X' > t] = p\mathbf{P}[X > t] \leq \mathbf{P}[Y > t]$$

by (iii). Thus  $X'$  is stochastically dominated by  $Y$  in the standard sense, and therefore there exists a coupling of  $X'$  and  $Y$  such that  $X' \leq Y$  a.s. (The standard fact that  $\mathbf{P}[U > t] \leq \mathbf{P}[V > t]$  for all  $t$  is equivalent to the existence of a coupling for which  $U \leq V$  a.s. holds when  $U$  and  $V$  take values in  $[-\infty, \infty)$ , and in fact under considerably more general conditions [37, Theorem 11].) Under this coupling, given  $X$ , it holds with probability at least  $p$  that  $X = X'$ , in which case  $X \leq Y$ . Thus  $X \leq_{\overline{p}} Y$  implies (i).

To show that (iii)  $\implies$  (i) for the second set of statements, we use the same idea but define  $Y'$  to be equal to  $Y$  with probability  $p$  and equal to  $\infty$  with probability  $1 - p$ . Then  $X \leq Y'$ , yielding a coupling of  $X$  and  $Y'$  such that  $X \leq Y'$  a.s., and under this coupling given  $Y$  we have  $Y = Y' \geq X$  with probability at least  $p$ .  $\square$

Before we get started with our concentration estimates, we make an observation that allows us to rescale  $\alpha$  and  $\beta$ .

**Lemma 9.** *Let  $\alpha$ ,  $\beta$ , and  $\gamma$  be positive real numbers. If  $X^*$  has the  $(\alpha, \beta)$ -generalized equilibrium distribution of  $X$ , then  $(X^*)^\gamma$  has the  $(\frac{\alpha}{\gamma}, \frac{\beta}{\gamma})$ -generalized equilibrium distribution of  $X^\gamma$ .*

**Proof.** For  $a > 0$ , let  $V_a$  denote a random variable with density  $ax^{a-1}dx$ , and recall that  $X^* \stackrel{d}{=} V_\alpha X^{(\beta)}$ . Now observe that  $V_\alpha^\gamma \stackrel{d}{=} V_{\alpha/\gamma}$  and  $(X^{(\beta)})^\gamma \stackrel{d}{=} (X^\gamma)^{(\beta/\gamma)}$ .  $\square$

### 2.1. Upper tail bounds for $\alpha \neq \beta$

We start with a technical lemma.

**Lemma 10.** Suppose that  $X^* \leq_{\bar{p}} X$  and  $\mathbf{E}X^\beta = \alpha/\beta$ . Let  $G(t) = \mathbf{P}[X > t]$ . For all  $t > 0$ ,

$$\int_0^1 u^{\alpha-\beta-1} G(t/u) du \leq \frac{G(t)}{p\beta t^\beta}. \quad (7)$$

**Proof.** Since  $X^* \leq_{\bar{p}} X$ ,

$$G(t) \geq p\mathbf{P}[V_\alpha X^{(\beta)} > t] = p \int_0^1 \alpha u^{\alpha-1} \mathbf{P}[X^{(\beta)} > t/u] du.$$

By definition of the  $\beta$ -power bias,

$$\begin{aligned} p \int_0^1 \alpha u^{\alpha-1} \mathbf{P}[X^{(\beta)} > t/u] du &= p\beta \int_0^1 u^{\alpha-1} \mathbf{E}[X^\beta \mathbf{1}_{\{X > t/u\}}] du \\ &\geq p\beta t^\beta \int_0^1 u^{\alpha-\beta-1} \mathbf{P}[X > t/u] du \\ &= p\beta t^\beta \int_0^1 u^{\alpha-\beta-1} G(t/u) du. \quad \square \end{aligned}$$

Next, we apply this lemma to deduce bounds on  $\mathbf{E}[X \mid X > t]$  in the  $\beta - \alpha = 1$  case and on  $\mathbf{E}[X^{-1} \mid X > t]$  in the  $\beta - \alpha = -1$  case.

**Lemma 11.** Suppose that  $X^* \leq_{\bar{p}} X$  and  $\mathbf{E}X^\beta = \alpha/\beta$ . For all  $t \geq 0$  such that  $\mathbf{P}[X > t] > 0$ ,

$$\mathbf{E}[X - t \mid X > t] \leq \frac{1}{p\beta t^\alpha} \quad \text{if } \beta - \alpha = 1, \quad (8)$$

and

$$\mathbf{E}[X^{-1} - t^{-1} \mid X > t] \geq -\frac{1}{p\beta t^\alpha} \quad \text{if } \beta - \alpha = -1. \quad (9)$$

**Proof.** Let  $G(t) = \mathbf{P}[X > t]$ . When  $\beta - \alpha = 1$ , we apply [Lemma 10](#) and then switch the order of integration to obtain

$$\begin{aligned} \frac{G(t)}{p\beta t^\beta} &\geq \int_0^1 u^{-2} G(t/u) du = \mathbf{E} \int_0^1 u^{-2} \mathbf{1}_{\{X > t/u\}} du \\ &= \mathbf{E} \left[ \mathbf{1}_{\{X > t\}} \int_{t/X}^1 u^{-2} du \right] \\ &= \mathbf{E} \left[ \mathbf{1}_{\{X > t\}} (X/t - 1) \right] = G(t) (t^{-1} \mathbf{E}[X \mid X > t] - 1). \end{aligned}$$

Cancelling the  $G(t)$  factors and rearranging terms gives (8). When  $\beta - \alpha = -1$ , the same approach yields

$$\frac{G(t)}{p\beta t^\beta} \geq \int_0^1 G(t/u) du = \mathbf{E} \left[ \mathbf{1}_{\{X > t\}} (1 - t/X) \right] = G(t) (1 - t \mathbf{E}[X^{-1} \mid X > t]),$$

proving (9).  $\square$

To give a sense of the purpose of the preceding lemma, the *mean residual life* of a random variable  $X$  is the function  $m_X(t) = \mathbf{E}[X - t \mid X > t]$ , with  $m_X(t)$  defined as 0 if  $\mathbf{P}[X > t] = 0$ . In general, the distribution of a random variable can be recovered from its mean residual life function [25, Lemma 2]. In Lemma 11, we bound the mean residual life of  $X$  when  $\beta - \alpha = 1$  or of  $-X^{-1}$  when  $\beta - \alpha = -1$ . (A similar approach would bound the mean residual life of  $\log X$  when  $\beta = \alpha$ , but a different technique used in Section 2.2 gives better results in that case.) To prove Theorem 3 when  $\alpha \neq \beta$ , we will first use Lemma 9 to rescale  $\alpha$  and  $\beta$  so that  $\alpha - \beta = \pm 1$ , and then we apply the bounds on mean residual life from Lemma 11 to derive tail bounds on  $X$ .

**Proof of Theorem 3 for  $\alpha < \beta$ .** First, we prove the theorem under the assumption that  $\mu = 1$  (i.e.,  $\mathbf{E}X^\beta = \alpha/\beta$ ) and that  $\beta - \alpha = 1$ . Let  $t_0 = (\alpha/p\beta)^{1/\beta}$ , and let  $Z$  be the random variable supported on the interval  $(t_0, \infty)$  with

$$\mathbf{P}[Z > t] = \left(\frac{p\beta e}{\alpha}\right)^{\alpha/\beta} t^\alpha e^{-pt^\beta}. \quad (10)$$

We now compute the mean residual life function of  $Z$ . First, by Fubini's theorem,

$$\mathbf{E}[(Z - t)\mathbf{1}_{\{Z > t\}}] = \int_t^\infty \mathbf{P}[Z > u] du = \frac{1}{p\beta} \left(\frac{p\beta e}{\alpha}\right)^{\alpha/\beta} e^{-pt^\beta}. \quad (11)$$

Hence, for  $t \geq t_0$ ,

$$\mathbf{E}[Z - t \mid Z > t] = \frac{1}{p\beta t^\alpha}.$$

By Lemma 11, we have  $\mathbf{E}[X - t \mid X > t] \leq \mathbf{E}[Z - t \mid Z > t]$  for all  $t$  where  $\mathbf{P}[X > t] > 0$ . By [34, Theorem 4.A.26], we have  $\mathbf{E}\varphi(X) \leq \mathbf{E}\varphi(Z)$  for any increasing convex function  $\varphi(x)$ .

Now, we apply this statement with the right choice of  $\varphi(x)$  to obtain information about the tail of  $X$ . Fix  $t > 0$  and define  $\varphi(x) = \max(0, (x - t + \gamma)/\gamma)$  for  $\gamma$  to be chosen later satisfying  $0 < \gamma \leq t$ . Then  $\varphi$  is an increasing convex function, and hence  $\mathbf{E}\varphi(X) \leq \mathbf{E}\varphi(Z)$ . Observing that  $\mathbf{1}_{\{x \geq t\}} \leq \varphi(x)$ ,

$$\begin{aligned} \mathbf{P}[X \geq t] &\leq \mathbf{E}\varphi(X) \leq \mathbf{E}\varphi(Z) \\ &= \gamma^{-1} \mathbf{E}[(Z - (t - \gamma))\mathbf{1}_{\{Z > t - \gamma\}}] = \frac{1}{\gamma p\beta} \left(\frac{p\beta e}{\alpha}\right)^{\alpha/\beta} e^{-p(t - \gamma)^\beta} \end{aligned}$$

by (11), assuming that  $t - \gamma \geq t_0$ . Since  $\beta > 1$ , the function  $x^\beta$  is convex and its graph lies above its tangent lines. Hence  $x^\beta \geq t^\beta + (x - t)\beta t^{\beta-1}$ , and setting  $x = t - \gamma$  we have  $(t - \gamma)^\beta \geq t^\beta - \gamma\beta t^{\beta-1}$ . Using this inequality and then minimizing by setting  $\gamma = 1/p\beta t^\alpha$  yields

$$\begin{aligned} \mathbf{P}[X \geq t] &\leq \frac{1}{\gamma p\beta} \left(\frac{p\beta e}{\alpha}\right)^{\alpha/\beta} e^{-p(t^\beta - \gamma\beta t^{\beta-1})} \\ &= t^\alpha \left(\frac{p\beta e}{\alpha}\right)^{\alpha/\beta} e^{-pt^\beta + 1}, \end{aligned}$$

so long as  $t - 1/p\beta t^\alpha \geq t_0$ . To make this bound simpler, first observe that  $-x \log x \leq 1 - x$  for all  $x > 0$ , since  $-x \log x$  is concave and  $1 - x$  is its tangent line at  $x = 1$ . Exponentiating both sides of this inequality, setting  $x = \alpha/\beta p$ , and raising both sides to the  $p$ th power gives

$$\left(\frac{p\beta}{\alpha}\right)^{\alpha/\beta} \leq e^{p - \alpha/\beta}, \quad (12)$$



from which we obtain

$$\mathbf{P}[X \geq t] \leq t^\alpha e^{p-pt^\beta+1} \leq t^\alpha e^{2-pt^\beta}, \quad (13)$$

still assuming

$$t - 1/p\beta t^\alpha \geq t_0. \quad (14)$$

Now, we show that either (14) is satisfied or the right-hand side of (13) exceeds 1. Suppose  $t^\alpha e^{2-pt^\beta} < 1$ . Since we are assuming  $t \geq 1$ , we have  $t^\beta \geq 2/p$ . Thus our goal is to show that (14) holds when  $t^\beta \geq 2/p$ . Since the left-hand side of (14) is increasing in  $t$ , it suffices to prove that (14) holds when  $t = (2/p)^{1/\beta}$ . To obtain this, start with the inequality  $\log(1 - x/2) \geq -(\log 2)x$  for  $x \in (0, 1)$ , which holds since  $\log(1 - x/2)$  is concave and equals  $-(\log 2)x$  at  $x = 0, 1$ . Setting  $x = 1/\beta$  for  $\beta > 1$  and exponentiating, we obtain  $1 - 1/2\beta \geq 2^{-1/\beta}$ . Rearranging terms,  $2^{1/\beta}(1 - 1/2\beta) \geq 1$ . Hence,

$$2^{1/\beta} \left(1 - \frac{1}{2\beta}\right) \geq \left(\frac{\alpha}{\beta}\right)^{1/\beta},$$

since the right-hand side is smaller than 1. Multiplying both sides of this inequality by  $p^{-1/\beta}$  and substituting  $t = (2/p)^{1/\beta}$ ,

$$t - \frac{t}{2\beta} \geq \left(\frac{\alpha}{p\beta}\right)^{1/\beta}.$$

And this is exactly (14), since  $t/2\beta = 1/p\beta t^\alpha$  for  $t = (2/p)^{1/\beta}$  and  $\alpha = \beta - 1$ . This completes the proof assuming  $\mathbf{E}X^\beta = \alpha/\beta$  and  $\beta - \alpha = 1$ .

Now, we drop the assumption  $\mathbf{E}X^\beta = \alpha/\beta$  and relax the assumption  $\beta - \alpha = 1$  to  $\beta - \alpha > 0$ . Let  $Y = (X/\mu)^{\beta-\alpha}$  and let  $Y^\dagger = (X^*/\mu)^{\beta-\alpha}$ , and observe that  $(X/\mu)^*$  is given by  $X^*/\mu$ . By Lemma 9, the random variable  $Y^\dagger$  has the  $(\alpha', \beta')$ -generalized equilibrium distribution of  $Y$ , where  $\alpha' = \alpha/(\beta - \alpha)$  and  $\beta' = \beta/(\beta - \alpha)$ . Now  $\beta' - \alpha' = 1$  and  $\mathbf{E}Y^{\beta'} = \mathbf{E}(X/\mu)^\beta = \alpha/\beta$ . We now apply the special case of the theorem already proven to obtain

$$\mathbf{P}[X \geq \mu t] = \mathbf{P}[Y \geq t^{\beta-\alpha}] \leq t^{(\beta-\alpha)\alpha'} e^{2-pt^{(\beta-\alpha)\beta'}} = t^\alpha e^{2-pt^\beta}$$

for any  $t \geq 1$ .  $\square$

The proof when  $\alpha > \beta$  follows the same structure, differing only in some analytical details.

**Proof of Theorem 3 for  $\alpha > \beta$ .** As in the  $\alpha < \beta$  case, it suffices to prove the theorem under the assumptions  $\mu = 1$  (or equivalently  $\mathbf{E}X^\beta = \alpha/\beta$ ) and  $\alpha - \beta = 1$ . Let  $t_0 = (\alpha/p\beta)^{1/\beta}$  and  $Z$  be the random variable supported on  $(t_0, \infty)$  defined by (10). The mean residual life function of  $Z$  computed in (11) does not hold in the  $\alpha - \beta = 1$  case, since the computation of the integral in (11) relies on  $\beta - \alpha = 1$ , but we can instead observe that

$$\mathbf{P}[Z^{-1} \leq t] = \left(\frac{p\beta e}{\alpha}\right)^{\alpha/\beta} t^{-\alpha} e^{-pt^{-\beta}}$$

for  $0 < t < t_0^{-1}$  and then compute

$$\mathbf{E}[(Z^{-1} - t)\mathbf{1}_{\{Z^{-1} < t\}}] = - \int_0^t \mathbf{P}[Z \leq u] du = - \frac{1}{p\beta} \left(\frac{p\beta e}{\alpha}\right)^{\alpha/\beta} e^{-pt^{-\beta}}.$$

Hence,

$$\mathbf{E}[Z^{-1} - t \mid Z^{-1} \leq t] = -\frac{t^\alpha}{p\beta}.$$

We then have

$$\mathbf{E}[X^{-1} - t \mid X^{-1} \leq t] \geq \mathbf{E}[X^{-1} - t \mid X^{-1} < t] \geq \mathbf{E}[Z^{-1} - t \mid Z^{-1} \leq t]$$

by Lemma 11. By [34, Theorem 4.A.27], it holds for all decreasing convex functions  $\varphi(x)$  that  $\mathbf{E}\varphi(X^{-1}) \leq \mathbf{E}\varphi(Z^{-1})$ . Taking  $\varphi(x) = \max(1 + (t - x)/\gamma, 0)$  for  $\gamma > 0$  to be chosen later and observing that  $\varphi(x) \geq \mathbf{1}\{x \leq t\}$ , we obtain

$$\begin{aligned} \mathbf{P}[X^{-1} \leq t] &\leq \mathbf{E}\varphi(X^{-1}) \\ &\leq \mathbf{E}\varphi(Z^{-1}) = \frac{1}{\gamma} \mathbf{E}[(\gamma + t - Z^{-1})\mathbf{1}\{Z^{-1} \leq \gamma + t\}] \\ &= \frac{1}{p\beta\gamma} \left(\frac{p\beta e}{\alpha}\right)^{\alpha/\beta} e^{-p(\gamma+t)^{-\beta}}, \end{aligned}$$

assuming  $\gamma + t < t_0^{-1}$ . Applying the inequality  $(\gamma + t)^{-\beta} \geq t^{-\beta} - \beta\gamma t^{-\beta-1}$  which holds by convexity of  $x^{-\beta}$  and then optimizing by setting  $\gamma = t^\alpha/p\beta$ ,

$$\begin{aligned} \mathbf{P}[X^{-1} \leq t] &\leq \frac{1}{p\beta\gamma} \left(\frac{p\beta e}{\alpha}\right)^{\alpha/\beta} e^{-p(t^{-\beta} + \beta\gamma t^{-\alpha})} \\ &= \left(\frac{p\beta e}{\alpha}\right)^{\alpha/\beta} t^{-\alpha} e^{1-pt^{-\beta}}. \end{aligned}$$

Thus, under the assumption

$$1/p\beta t^\alpha + t^{-1} < (p\beta/\alpha)^{1/\beta}, \quad (15)$$

we obtain

$$\mathbf{P}[X \geq t] = \mathbf{P}[X^{-1} \leq t^{-1}] \leq \left(\frac{p\beta e}{\alpha}\right)^{\alpha/\beta} t^\alpha e^{1-pt^\beta} \leq t^\alpha e^{p+1-pt^\beta} \leq t^\alpha e^{2-pt^\beta}, \quad (16)$$

applying (12) in the last step.

Finally, we show that either (15) holds or the right-hand side of (16) exceeds 1. Let  $f(t) = 2 - pt^\beta + (\beta + 1)\log t$ , so that the right-hand side of (16) is equal to  $e^{f(t)}$ . Let  $g(t) = 1/p\beta t^{\beta+1} + t^{-1}$ , the left-hand side of (15). Our goal is to show that for all  $t \geq 1$  that if  $f(t) < 0$ , then  $g(t) < (p\beta/(\beta + 1))^{1/\beta}$ . It is easy to check that for  $t \geq 1$ , the function  $f(t)$  increases and then decreases, with  $f(1) > 0$ . Since  $g(t)$  is decreasing, it suffices to find a value  $s \geq 1$  so that  $f(s) \geq 0$  and  $g(s) \leq (p\beta/(\beta + 1))^{1/\beta}$ . We claim that this holds for

$$s = \left(\frac{2(\beta + 1)}{p\beta}\right)^{1/\beta}.$$

To confirm this, we compute

$$f(s) = 2 + \frac{\beta + 1}{\beta} \left(\log\left(\frac{2(\beta + 1)}{p\beta}\right) - 2\right) \geq 2 + \frac{\beta + 1}{\beta} \left(\log\left(\frac{2(\beta + 1)}{\beta}\right) - 2\right).$$

A bit of calculus shows that the right-hand side of this inequality is minimized when  $(\beta + 1)/\beta = e/2$  and is equal to  $2 - e/2 > 0$  in this case. Finally, we must confirm that

$g(s) \leq (p\beta/(\beta+1))^{1/\beta}$ . We compute

$$g(s) = \left( \frac{p\beta}{2(\beta+1)} \right)^{1/\beta} \left( \frac{2\beta+3}{2\beta+2} \right).$$

Cancelling a factor of  $(\beta/(\beta+1))^{1/\beta}$ , we must show that

$$2^{-1/\beta} \left( \frac{2\beta+3}{2\beta+2} \right) \leq 1. \quad (17)$$

Using the inequality  $\log(x+1) - \log x \leq 1/x$ ,

$$\log(2\beta+3) - \log(2\beta+2) \leq \frac{1}{2\beta+2} \leq \frac{1}{2\beta} \leq \frac{\log 2}{\beta},$$

and exponentiating both sides of this inequality confirms (17).  $\square$

## 2.2. Upper tail bounds for $\alpha = \beta$

As we mentioned in the introduction, [Theorem 3](#) in the case  $\alpha = \beta = p = 1$  was first proven in [\[10\]](#). The general  $\alpha = \beta$  case with  $p = 1$  then follows by an application of [Lemma 9](#), and it is not difficult to modify Brown's argument to allow  $p < 1$ . To save the reader the effort of going back and forth between Brown's paper and this one, and to highlight his elegant proof, we present the full argument here.

Let  $Z(u)$  be a random variable with the distribution  $[X^{(\beta-\alpha)} \mid X^{(\beta-\alpha)} \geq u]$ . If  $U \sim \mu$  is a random variable, let  $Z(U)$  denote a random variable whose distribution is the mixture governed by  $\mu$ , i.e.,

$$\mathbf{P}[Z(U) \in B] = \int \mathbf{P}[X^{(\beta-\alpha)} \in B \mid X^{(\beta-\alpha)} \geq u] \mu(du).$$

**Proposition 12.** *Let  $\pi$  be the law of  $X$ , and define  $(U, W)$  as the random variables with joint density*

$$\mathbf{1}\{0 \leq u \leq w\} \frac{\alpha u^{\alpha-1} w^{\beta-\alpha}}{\mathbf{E}X^\beta} du d\pi(w).$$

Then

- (a)  $U \stackrel{d}{=} X^*$ ;
- (b)  $W \stackrel{d}{=} X^{(\beta)}$ ;
- (c) for any Borel set  $\mathcal{B} \subset \mathbb{R}$ ,

$$\mathbf{P}[U \in \mathcal{B} \mid W] = \mathbf{P}[V_\alpha W \in \mathcal{B} \mid W] \text{ a.s.,}$$

where  $V_\alpha$  has density  $\alpha x^{\alpha-1}$  on  $[0, 1]$  and is independent of  $W$ ;

- (d) for any Borel set  $\mathcal{B} \subset \mathbb{R}$ ,

$$\mathbf{P}[W \in \mathcal{B} \mid U] = \mathbf{P}[Z(U) \in \mathcal{B} \mid U] \text{ a.s.,}$$

- (e)  $Z(X^*) \stackrel{d}{=} X^{(\beta)}$ .

**Proof.** The density of  $W$  is  $(w^\beta/\mathbf{E}X^\beta) d\pi(w)$ , proving (b). Looking at the conditional density of  $U$  given  $W = w$ , we see that (c) holds. Taking expectations in (c), we have  $U \stackrel{d}{=} V_\alpha W$ , and

together with (b) this implies (a). Fact (d) is proven by observing that the conditional density of  $W$  given  $U = u$  is

$$\mathbf{1}\{w \geq u\} \frac{w^{\beta-\alpha}}{\mathbf{E}[X^{\beta-\alpha} \mathbf{1}\{X \geq u\}]} d\pi(w),$$

which is the density of  $Z(u)$ , the  $(\beta - \alpha)$ -power bias transform of  $[X \mid X \geq u]$ . Finally, taking expectations in (d) gives  $W \stackrel{d}{=} Z(U)$ , and then (a) and (b) prove (e).  $\square$

**Lemma 13.** For any  $p \in (0, 1]$ ,

$$X^* \preceq_{\bar{p}} X \implies X^{(\beta)} \preceq_{\bar{p}} Z(X),$$

and

$$X^* \preceq_p X \implies X^{(\beta)} \preceq_p Z(X).$$

**Proof.** First, we observe that  $Z(u)$  is stochastically increasing in  $u$ . Thus we can couple the random variables  $(Z(u))_{u \geq 0}$  so that  $Z(u) \leq Z(v)$  whenever  $u \leq v$  (for example, by coupling all  $Z(u)$  to the same Uniform(0, 1) random variable by the inverse probability transform). Under such a coupling,  $X^* \leq X$  implies that  $Z(X^*) \leq Z(X)$ .

Now, suppose  $X^* \preceq_{\bar{p}} X$ . By Lemma 8, there exists a coupling  $(X, X^*)$  under which  $\mathbf{P}[X^* \leq X \mid X^*] \geq p$  a.s. Hence,

$$\mathbf{P}[Z(X^*) \leq Z(X) \mid X^*] \geq \mathbf{P}[X^* \leq X \mid X^*] \geq p \text{ a.s.}$$

Thus  $\mathbf{P}[Z(X^*) \leq Z(X) \mid Z(X^*)] \geq p$  a.s. as well. Since  $Z(X^*) \stackrel{d}{=} X^{(\beta)}$  by Proposition 12(e), this proves the lemma when  $X^* \preceq_{\bar{p}} X$ . The proof when  $X^* \preceq_p X$  is identical except we take conditional expectations given  $X$  rather than  $X^*$ .  $\square$

The next lemma will be used to prove a tail estimate for  $Z(X)$  when  $\alpha = \beta$ .

**Lemma 14.** Let  $\mu$  be a probability measure on  $[0, \infty)$ . Then for any  $t > 0$ ,

$$\int_{[0,t)} \frac{1}{\mu[x, \infty)} \mu(dx) \leq -\log \mu[t, \infty).$$

**Proof.** Assume without loss of generality  $\mu[t, \infty) > 0$ . For some partition  $0 = x_0 < \dots < x_n = t$ , let  $\varphi(x)$  be the step function taking value  $1/\mu[x_i, \infty)$  on interval  $[x_i, x_{i+1})$  for  $i = 0, \dots, n-1$ . Then

$$\int_{[0,t)} \varphi(x) \mu(dx) = \sum_{i=0}^{n-1} \frac{\mu[x_i, x_{i+1})}{\mu[x_i, \infty)} \leq \sum_{i=0}^{n-1} -\log \frac{\mu[x_{i+1}, \infty)}{\mu[x_i, \infty)} = -\log \mu[t, \infty), \quad (18)$$

with the inequality holding because  $x \leq -\log(1-x)$  for  $x \in [0, 1)$ .

Now, consider any sequence  $\varphi_n(x)$  of such step functions where each partition refines the last and the mesh size of the partition goes to zero. Then  $\varphi_n(x)$  converges upward to  $1/\mu[x, \infty)$  for all  $x \in [0, t)$ , and

$$\int_{[0,t)} \varphi_n(x) \mu(dx) \rightarrow \int_{[0,t)} \frac{1}{\mu[x, \infty)} \mu(dx)$$

by the monotone convergence theorem. This proves the lemma by (18).  $\square$

**Lemma 15.** For  $\alpha = \beta$  and any random variable  $X$ ,

$$\mathbf{P}[Z(X) \geq t] \leq \mathbf{P}[X \geq t](1 - \log \mathbf{P}[X \geq t]).$$

**Proof.** Let  $G(t) = \mathbf{P}[X \geq t]$ . For any  $u \geq 0$ ,

$$\mathbf{P}[Z(u) \geq t] = \mathbf{1}\{u \geq t\} + \mathbf{1}\{u < t\} \frac{G(t)}{G(u)}.$$

Hence,

$$\mathbf{P}[Z(X) \geq t] = G(t) + G(t)\mathbf{E}\left[\frac{\mathbf{1}\{X < t\}}{G(X)}\right],$$

and by Lemma 14 we have

$$\mathbf{P}[Z(X) \geq t] \leq G(t) + G(t)(-\log G(t)). \quad \square$$

**Proof of Theorem 3 for  $\alpha = \beta$ .** As in the  $\alpha \neq \beta$  cases, it suffices to prove the theorem under the assumption  $\mu = 1$ , or equivalently  $\mathbf{E}X^\beta = \alpha/\beta = 1$ . From the definition of  $X^{(\beta)}$ ,

$$pt^\beta \mathbf{P}[X \geq t] \leq p\mathbf{E}[X^\beta \mathbf{1}\{X \geq t\}] = p\mathbf{P}[X^{(\beta)} \geq t].$$

By Lemma 13 followed by Lemma 15,

$$p\mathbf{P}[X^{(\beta)} \geq t] \leq \mathbf{P}[Z(X) \geq t] \leq \mathbf{P}[X \geq t](1 - \log \mathbf{P}[X \geq t]).$$

We have now shown that  $pt^\beta \mathbf{P}[X \geq t] \leq \mathbf{P}[X \geq t](1 - \log \mathbf{P}[X \geq t])$ . Assuming  $\mathbf{P}[X \geq t] > 0$ , we can divide to obtain  $pt^\beta \leq 1 - \log \mathbf{P}[X \geq t]$ , demonstrating that  $\mathbf{P}[X \geq t] \leq e^{1-pt^\beta}$ .  $\square$

### 2.3. Lower tail bounds

**Proof of Theorem 4.** As in the proof of Theorem 3, by rescaling it suffices to prove the theorem when  $\mu = 1$ , i.e.,  $\mathbf{E}X^\beta = \alpha/\beta$ . From the definition of  $\underline{p}$ ,

$$\begin{aligned} p\mathbf{P}[X \leq t] &\leq \mathbf{P}[X^* \leq t] = \int_0^1 \alpha u^{\alpha-1} \mathbf{P}[X^{(\beta)} \leq t/u] du \\ &= \frac{\beta}{\alpha} \mathbf{E}\left[\int_0^1 \alpha u^{\alpha-1} X^\beta \mathbf{1}\{X \leq t/u\} du\right] \\ &= \frac{\beta}{\alpha} \mathbf{E}\left[X^\beta \int_0^{\min(t/X, 1)} \alpha u^{\alpha-1} du\right] \\ &= \frac{\beta}{\alpha} \mathbf{E}\left[X^\beta \min((t/X)^\alpha, 1)\right] \\ &= \frac{\beta}{\alpha} \mathbf{E}\left[t^\alpha X^{\beta-\alpha} \mathbf{1}\{X > t\} + X^\beta \mathbf{1}\{X \leq t\}\right]. \end{aligned}$$

If  $\alpha = \beta$ , then the expectation is bounded by  $t^\alpha$ , proving (2). Otherwise, we bound the two terms in the expectation separately to get

$$p\mathbf{P}[X \leq t] \leq \frac{\beta}{\alpha} \left( t^\alpha \mathbf{E}X^{\beta-\alpha} + t^\beta \mathbf{P}[X \leq t] \right).$$

By Jensen's inequality,  $\mathbf{E}X^{\beta-\alpha} \leq (\mathbf{E}X^\beta)^{(\beta-\alpha)/\beta} = (\alpha/\beta)^{1-\alpha/\beta}$ . Applying this bound and rearranging terms yields (1).  $\square$

#### 2.4. Sharpness of bounds

**Theorems 3** and **4** are nearly optimal when  $\alpha = \beta$  but seem to be missing a factor of  $t^{-\beta}$  when  $\alpha \neq \beta$ . First, let us assume that  $p = 1$ . In [11], it is proven that **Theorem 3** is sharp when  $p = \alpha = \beta = 1$ , including the constant factor of  $e$ . By **Lemma 9**, the theorem is sharp whenever  $\alpha = \beta$ . For the reader's convenience, we present a family of examples demonstrating that **Theorem 3** cannot be improved in this case; it is a discrete counterpart to the example given in [11]. Choose integers  $\mu$  and  $n$  and let  $p = 1/\mu$ , and let  $X$  have a capped version of the geometric distribution with success probability  $p$  as follows:

$$\mathbf{P}[X = k] = \begin{cases} (1-p)^{k-1}p & \text{if } 1 \leq k \leq n, \\ (1-p)^n & \text{if } k = n + \mu, \\ 0 & \text{otherwise.} \end{cases}$$

Then  $X^* \leq X$  with  $\alpha = \beta = 1$  (easy to check with **Proposition 19(b)**), and  $\mathbf{E}X = \mu$ . Now, set  $n = (t-1)\mu$  for some integer  $t \geq 2$ , and we have

$$\mathbf{P}[X \geq \mu t] = \mathbf{P}[X \geq n + \mu] = (1 - \mu^{-1})^{(t-1)\mu}.$$

This converges to  $e^{1-t}$  as  $m \rightarrow \infty$ , confirming that there exist examples in which  $\mathbf{P}[X \geq \mu t]$  comes arbitrarily close to  $e^{1-t}$ .

When  $\alpha \neq \beta$ , one would hope for an upper tail bound of  $O(t^{\alpha-\beta}e^{-pt^\beta})$  rather than the  $O(t^\alpha e^{-pt^\beta})$  achieved in **Theorem 3**, which would match the tail of the generalized gamma distribution. But the best tail bound via moments for the generalized gamma distribution loses a factor of  $t^{-\beta/2}$  (the calculation is similar to the one carried out in **Remark 18**), and the Chernoff approach of bounding the moment generating function used in the proof of **Theorem 3** in the  $\alpha \neq \beta$  case is always inferior to the moment bound [31]. Thus a new approach would be needed for the proof if the optimal tail bound is to be achieved. Perhaps Brown's proof for the  $\alpha = \beta$  case could be adapted when  $\alpha \neq \beta$ , though we are not sure what the replacement for **Lemma 15** would be.

As for lower tail bounds, if  $X$  has the  $(\alpha, \beta)$ -generalized gamma distribution, then **Theorem 4** applies to  $X$  with  $\mu = p = 1$ . Up to constants, the  $O(t^\alpha)$  bound for  $\mathbf{P}[X \leq t]$  shown in **Theorem 4** matches the true tail behavior of  $X$  as  $t \rightarrow 0$ .

Now, we show that the dependence on  $p$  in the theorem is nearly optimal. Suppose that  $X^* \leq X$  for some  $\alpha, \beta > 0$ , and define

$$Y = \begin{cases} X & \text{with probability } p, \\ 0 & \text{with probability } 1-p, \end{cases}$$

for some  $0 < p \leq 1$ . Then

$$Y^* \stackrel{d}{=} X^* \leq_p Y.$$

Let  $\mu_X = (\frac{\beta}{\alpha}\mathbf{E}X^\beta)^{1/\beta}$  and  $\mu_Y = (\frac{\beta}{\alpha}\mathbf{E}Y^\beta)^{1/\beta} = p^{1/\beta}\mu_X$ . For any  $t \geq 1$ , applying the  $p = 1$  case of **Theorem 3** to  $X$ ,

$$\mathbf{P}[Y \geq \mu_Y t] = p\mathbf{P}[X \geq p^{1/\beta}t\mu_X] \leq \begin{cases} p^{1+\alpha/\beta}t^\alpha e^{2-pt^\beta}, & \text{if } \alpha \neq \beta, \\ e^{1-pt^\beta}, & \text{if } \alpha = \beta. \end{cases}$$

Meanwhile the bound from applying **Theorem 3** to  $Y$  is

$$\mathbf{P}[Y \geq \mu_Y t] \leq \begin{cases} t^\alpha e^{2-pt^\beta}, & \text{if } \alpha \neq \beta, \\ e^{1-pt^\beta}, & \text{if } \alpha = \beta. \end{cases}$$

Thus, the bound on  $Y$  is equally sharp as the bound on  $X$  provided by the  $p = 1$  case of [Theorem 3](#), besides losing a factor of  $p^{1+\alpha/\beta}$  when  $\alpha \neq \beta$ .

For an example showing optimal dependence on  $p$  in the lower tail bound, for the sake of simplicity take  $\alpha = \beta = 1$ . Choose some  $b < 1 < a$ , and define  $X$  as the mixture

$$X \stackrel{d}{=} \begin{cases} \text{Exp}(a) & \text{with probability } \frac{a(1-b)}{a-b}, \\ \text{Exp}(b) & \text{with probability } \frac{(a-1)b}{a-b}, \end{cases}$$

and observe that  $\mathbf{E}X = 1$ . We can compute directly that

$$\mathbf{P}[X \leq t] = \frac{1}{a-b}(a(1-b)(1-e^{-at}) + (a-1)b(1-e^{-bt})) = (a+b-ab)t - O(t^2). \quad (19)$$

The equilibrium transform of a mixture is the mixture of the equilibrium transforms, with the new mixture governed by the old governor reweighted by expectation (see [Lemma 21](#)). Together with the fact exponential distributions are fixed points of the equilibrium transform, this yields

$$X^* \stackrel{d}{=} \begin{cases} \text{Exp}(a) & \text{with probability } \frac{1-b}{a-b}, \\ \text{Exp}(b) & \text{with probability } \frac{a-1}{a-b}. \end{cases}$$

With a bit of work, one can show that  $X^* \leq_p X$  with  $p = 1/(a+b-ab)$ . Thus [Theorem 4](#) yields

$$\mathbf{P}[X \leq t] \leq (a+b-ab)t,$$

matching (19).

### 3. Concentration for urns, graphs, walks, and trees

Each of random variables  $W$ ,  $U$ ,  $L_n$ , and  $L_{2n}^b$  in [Theorem 5](#) can be expressed in terms of an urn model that we describe now. An urn starts with black and white balls and draws are made sequentially. After a ball is drawn, it is replaced and another ball of the same color is added to the urn. Also, after every  $l$ th draw an additional black ball is added to the urn for some  $l \geq 1$ . As defined in Section 1.2 of [\[30\]](#), let  $P_n^l(b, w)$  denote the distribution of the number of white balls in the urn after  $n$  draws have been made when the urn starts with  $b \geq 0$  black balls and  $w > 0$  white balls, and let  $N_n(b, w) \sim P_n^l(b, w)$ .

Let  $N_n^{[r]}(b, w)$  be a rising factorial biased version of  $N_n(b, w)$ , as defined in Lemma 4.2 of [\[30\]](#), so that

$$\mathbf{P}[N_n^{[r]}(b, w) = k] = \left( c \prod_{i=0}^{r-1} (k+i) \right) \mathbf{P}[N_n(b, w) = k]$$

for some  $c$ . There it is shown that  $N_n^{[r]}(b, w) + r \stackrel{d}{=} N_n(b, w+r)$ . We will use this fact to prove concentration for  $N_n(1, w)$ , but first we relate the rising factorial bias to the power bias  $N_n^{(l+1)}(b, w)$ .

**Lemma 16.** *For all  $1 \leq l \leq n$  we have  $N_{n-l}^{[l+1]}(b, w) + l \geq N_n^{(l+1)}(b, w)$ .*

**Proof.** We will show that  $\mathbf{P}[N_{n-l}^{[l+1]}(b, w) + l = k] / \mathbf{P}[N_n^{(l+1)}(b, w) = k]$  is increasing in  $k$  on the union of the support of  $N_{n-l}^{[l+1]}(b, w) + l$  and  $N_n^{(l+1)}(b, w)$ , which is  $\{w, w+1, \dots, w+n\}$ .

We have

$$\frac{\mathbf{P}[N_{n-l}^{[l+1]}(b, w) + l = k]}{\mathbf{P}[N_n^{[l+1]}(b, w) = k]} = \frac{C(k-l) \cdots k \mathbf{P}[N_{n-l}(b, w) = k-l]}{k^{l+1} \mathbf{P}[N_n(b, w) = k]},$$

where  $C$  is a value that does not depend on  $k$ . The expression  $(k-l) \cdots k / k^{l+1}$  is increasing in  $k$ , as is

$$\frac{\mathbf{P}[N_{n-l}(b, w) = k-l]}{\mathbf{P}[N_n(b, w) = k]} = \prod_{j=n-l}^{n-1} \frac{\mathbf{P}[N_j(b, w) = j-n+k]}{\mathbf{P}[N_{j+1}(b, w) = j-n+k+1]}$$

for  $w \leq k \leq w+n$ , since it is 0 for  $w \leq k < w+l$  and each factor on the right-hand side in the product is increasing in  $k$  by Lemma 25 for  $w+l \leq k \leq w+n$  (in this range of  $k$ , the denominators of the fractions in the product are all nonzero). This proves the desired stochastic domination by [34, Theorem 1.C.1].  $\square$

**Proposition 17.** For all  $l, w, n \geq 1$ ,

$$N_n^*(1, w) \preceq N_n(1, w),$$

where  $N_n^*(1, w)$  is the  $(w, l+1)$ -generalized equilibrium transform of  $N_n(1, w)$ .

**Proof.** Let  $Q_w(n)$  have the distribution of the number of white balls in a regular Polya urn after  $n$  draws starting with 1 black and  $w$  white balls. For now, assume that  $l \leq n$ . We argue that

$$\begin{aligned} N_n(1, w) &\stackrel{d}{=} Q_w(N_n(0, w+1) - w - 1) \\ &\stackrel{d}{=} Q_w(N_{n-l}(1, w+1+l) - w - 1) \\ &\stackrel{d}{=} Q_w(N_{n-l}^{[l+1]}(1, w) + l + 1 - w - 1) \\ &\geq (N_{n-l}^{[l+1]}(1, w) + l) V_w, \end{aligned} \tag{20}$$

where  $V_w$  has density  $w x^{w-1} dx$  on  $[0, 1]$  and is independent of  $N_{n-l}^{[l+1]}(1, w)$ . The first line is Lemma 4.5 from [30]. In the second line, we use the trivial relation

$$N_n(0, w+1) \stackrel{d}{=} N_{n-l}(1, w+1+l).$$

In the third line, we use

$$N_n(1, w+r) \stackrel{d}{=} N_n^{[r]}(1, w) + r,$$

which is the statement of Lemma 4.2 of [30]. The final line uses

$$Q_w(n) \geq (n+w) V_w, \tag{21}$$

which follows from the fact, taken from the proof of Lemma 4.4 of [30], that for independent and identically distributed Uniform(0, 1) variables  $U_1, U_2, \dots, U_{w-1}$  we can write

$$Q_w(n) \stackrel{d}{=} \max_{i=0,1,\dots,w-1} (i + \lceil (n+w-i)U_i \rceil)$$

and this implies

$$Q_w(n) \geq \max_{i=0,1,\dots,w-1} (n+w)U_i = (n+w)V_w.$$



We now apply [Lemma 16](#) to (20) to obtain

$$N_n(1, w) \geq V_w N_n^{(l+1)}(1, w) \stackrel{d}{=} N_n^*(1, w)$$

when  $l \leq n$ .

When  $l > n$ , the quantity  $N_n(1, w)$  is the number of white balls in a regular Polya urn after  $n$  draws starting with 1 black and  $w$  white balls. Using (21) and observing that the maximum of  $N_n(1, w)$  is  $n + w$ ,

$$N_n(1, w) \stackrel{d}{=} Q_w(n) \geq (n + w)V_w \geq N_n^{(l+1)}(1, w)V_w \stackrel{d}{=} N_n^*(1, w). \quad \square$$

**Proof of Theorem 5.** We have  $W \sim P_n^l(1, w)$ ,  $U \sim P_{n-k-1}^1(1, 2k)$ , and  $L_{2n} \sim P_n^1(1, 1)$  respectively from Remark 1.3, Proposition 2.1, and Proposition 3.4 in [30]. From Proposition 3.2 in [30], we have  $L_{2n}^b \sim P_n^1(0, 1)$ , and  $P_n^1(0, 1)$  is the same distribution as  $P_{n-1}^1(1, 2)$ . The result then follows from [Proposition 17](#) and then noting that the conditions of [Theorems 3](#) and [4](#) hold.  $\square$

**Remark 18.** The rising factorial moments of  $N_n(1, w)$  are explicitly computed in [30, Lemma 4.1]. When  $w = l + 1$ , the concentration bound given in [Theorem 5](#) is better than the one obtained from these moments via Markov's inequality. For the sake of simplicity, we illustrate with the case  $l = 1$ ,  $w = 2$ . The result of Part (a) of [Theorem 5](#) along with [Theorem 3](#) shows that

$$\mathbf{P}[N_n(1, w) \geq \gamma_n t] \leq e^{1-t^2}, \quad (22)$$

where  $\gamma_n^2 = \mathbf{E}N_n(1, w)^2$ . From [30, Theorem 1.2], we know that  $\gamma_n \sim 2\sqrt{n}$ .

Now, we compute the concentration inequality given by the rising factorial moments of  $N_n(1, w)$ . Using the notation  $x^{[n]} = x(x+1) \cdots (x+n-1)$ , from [30, Lemma 4.1] we have

$$\mathbf{E}(N_n(1, w))^{[2m]} = 2^{[2m]} \prod_{i=0}^{n-1} \left(1 + \frac{2m}{2i+3}\right) = 2^{[2m]} \prod_{i=0}^{m-1} \frac{2n+2i+3}{2i+3}.$$

The bound given by applying Markov's inequality to this is

$$\begin{aligned} \mathbf{P}[N_n(1, w) \geq \gamma_n t] &\leq \frac{\mathbf{E}(N_n(1, w))^{[2m]}}{(\gamma_n t)^{[2m]}} = \frac{2^{[2m]}}{(\gamma_n t)^{[2m]}} \prod_{i=0}^{m-1} \frac{2n+2i+3}{2i+3} \\ &= \prod_{i=0}^{m-1} \frac{(2i+2)(2n+2i+3)}{(\gamma_n t + 2i)(\gamma_n t + 2i + 1)}. \end{aligned} \quad (23)$$

Let  $m^*$  be the minimizing choice of  $m$  in (23). Some algebra shows that the multiplicand in this expression is bounded by 1 if and only if

$$i \leq \frac{\frac{\gamma_n^2}{n} t^2 + \frac{\gamma_n}{n} t - 4 - \frac{6}{n}}{4 + \frac{8}{n} - \frac{4\gamma_n t}{n}}.$$

Take  $t$  to be fixed with respect to  $n$ . From the asymptotics for  $\gamma_n$ , the right-hand side of this inequality converges to  $t^2 - 1$  as  $n \rightarrow \infty$ . Hence either  $m^* = \lceil t^2 \rceil - 1$  or  $m^* = \lceil t^2 \rceil$  when  $n$  is sufficiently large. The optimal tail bound obtained from the rising factorial moments is then

$$\prod_{i=0}^{m^*-1} \frac{(2i+2)(2 + \frac{2i+3}{n})}{\left(\frac{\gamma_n}{\sqrt{n}} t + \frac{2i}{\sqrt{n}}\right)\left(\frac{\gamma_n}{\sqrt{n}} t + \frac{2i+1}{\sqrt{n}}\right)},$$

which converges as  $n \rightarrow \infty$  to

$$\prod_{i=0}^{m^*-1} \frac{(2i+2)(2)}{4t^2} = \frac{(m^*)!}{(t^2)^{m^*}} = \Omega(te^{-t^2}),$$

applying Stirling's approximation to obtain the last estimate. Thus this bound is worse than (22) by a factor of  $t$ , when  $n$  and  $t$  are large.

A more involved calculation for the general case  $w = \alpha$ ,  $l = \beta - 1$  shows that the tail bound from moments is on the order of  $t^{\alpha-\beta/2}e^{-t^\beta}$ . Outside of the  $\alpha = \beta$  case, our bound is on the order of  $t^\alpha e^{-t^\beta}$  and is outperformed by the moment bound.

#### 4. Concentration for Galton–Watson processes

We adopt the terminology from reliability theory that a random variable satisfying  $X^* \leq X$  with  $\alpha = \beta = 1$  is *NBUE*, which stands for “new better than used in expectation” (see Proposition 19(b) for the source of this name). Since we will often be applying the equilibrium transform to discrete random variables (e.g., the child distribution of a Galton–Watson tree), we will use the notation  $X^e$  to denote the discrete version of the  $\alpha = \beta = 1$  equilibrium transform, which we can define by setting  $X^e = \lceil X^* \rceil$  with  $X^*$  the standard  $\alpha = \beta = 1$  equilibrium transform. Equivalently, we can define  $X^e$  to be chosen uniformly at random from  $\{1, 2, \dots, X^s\}$ , where  $X^s$  is the size-bias transform of  $X$ . Observe that for a random variable  $X$  taking values in the nonnegative integers, it is a consequence of the coupling interpretation of stochastic dominance that  $X^e \leq X$  if and only if  $X^* \leq X$ .

##### 4.1. Some concepts from reliability theory

We consider three classes of discrete probability distributions; we will state the relationship between the three classes and give some characterizations of them. All are standard in the reliability theory literature, sometimes with varying notation.

To define the first class, the *log-concave* distributions, we first define a sequence  $t_0, t_1, \dots$  as *log-concave* if

- (i)  $t_n^2 \geq t_{n-1}t_{n+1}$  for all  $n \geq 1$ , and
- (ii)  $t_0, t_1, \dots$  has no internal zeroes (i.e., if  $t_i > 0$  and  $t_k > 0$  for some  $i < k$ , then  $t_j > 0$  for all  $i < j < k$ ).

For  $X$  taking nonnegative integer values, we say that  $X$  is *log-concave* if the sequence  $\mathbf{P}[X = k]$  for  $k \geq 0$  is log-concave.

Next, we recall the class of distributions on the positive integers with the *D-IFR* property, which stands for *discrete increasing failure rate*. As we defined in the introduction, the distribution of a positive integer-valued random variable  $X$  is in this class if  $\mathbf{P}[X = k \mid X \geq k]$  is increasing for  $k \geq 1$ , and in that case we say that  $X$  is D-IFR. Sometimes in the literature, such random variables are just said to be IFR, with it understood to use the above definition rather than the continuous version when considering a discrete distribution. Sometimes the notation DS-IFR is used to refer to a random variable  $X$  on the nonnegative integers for which  $\mathbf{P}[X = k \mid X \geq k]$  is increasing for  $k \geq 0$ ; see for example [27].

As we mentioned in the introduction, a nonnegative random variable  $X$  is said to be NBUE if  $X^* \leq X$  with  $\alpha = \beta = 1$ . In the reliability theory literature, a random variable  $X$  taking

positive integer values is sometimes said to be D-NBUE if

$$\frac{1}{\mathbf{E}X} \sum_{k=0}^{\infty} \mathbf{P}[X > n+k] \leq \mathbf{P}[X > n] \quad \text{for all } n \in \{0, 1, 2, \dots\}. \quad (24)$$

But since the left-hand side of (24) is equal to  $\mathbf{P}[X^e > n]$  (see (39)), Eq. (24) is equivalent to the assertion that  $X^e \leq X$ , which holds for  $X$  taking positive integer values if and only if  $X$  is NBUE.

**Proposition 19.** *For a positive integer-valued random variable  $X$ :*

- (a)  $X$  is D-IFR if and only if the distributions  $[X - k \mid X > k]$  are stochastically decreasing for integers  $k \geq 0$ ;
- (b)  $X$  is NBUE if and only if  $\mathbf{E}[X - k \mid X > k] \leq \mathbf{E}X$  for all integers  $k \geq 1$ ; and
- (c)  $X$  is log-concave  $\implies X$  is D-IFR  $\implies X$  is NBUE.

These properties are often stated in the reliability theory literature (see [27, Fig. 2] and [33, Lemma 2]). Since we have had trouble digging up proofs of some of them, we have provided them in [Appendix B](#).

Now, we introduce a class of distributions on the nonnegative integers that we call NBUEZT, with ZT standing for *zero-truncated*. For  $X$  taking nonnegative integer values, we say that  $X$  is NBUEZT if  $X^>$  is NBUE, or equivalently if  $X^e \leq X^>$  (recall from the introduction that  $X^>$  denotes a random variable with the distribution  $[X \mid X > 0]$ ).

In the language defined here, [Theorem 6](#) states that if the child distribution of a Galton–Watson process is NBUE, then all generations are NBUE. This is a simple consequence of the statement that a random sum of NBUE-many i.i.d. NBUE summands is NBUE, which was proven in [39, Corollary 2.2] (though we provide a more conceptual proof). [Theorem 7](#) states that with  $L$  the child distribution, if  $L^>$  is D-IFR then all generations of the process are NBUEZT. This raises a number of questions – for example, if  $L$  is only assumed to be NBUEZT, then are successive generations NBUEZT? – that we address in [Section 4.4](#).

#### 4.2. Forming the equilibrium transform

First, we give a recipe for forming the equilibrium transform of a sum:

**Lemma 20.** *Let  $X_1, \dots, X_n$  be i.i.d. nonnegative random variables, and let  $S = X_1 + \dots + X_n$ . Then*

$$S^* \stackrel{d}{=} \sum_{k=1}^{I-1} X_k + X_I^*, \quad (25)$$

where  $I$  is chosen uniformly at random from  $\{1, \dots, n\}$ , independent of all else, and  $X^*$  denotes the  $\alpha = \beta = 1$  equilibrium transform of  $X$ . If  $X_1, \dots, X_n$  are integer-valued, then

$$S^e \stackrel{d}{=} \sum_{k=1}^{I-1} X_k + X_I^e. \quad (26)$$

**Proof.** Eq. (25) is the special case of [28, Theorem 4.1] in which  $X_1, \dots, X_n$  are i.i.d. When  $X_1, \dots, X_n$  are integer-valued, applying the ceiling function to both sides of (25) gives (26).  $\square$

Next, we consider the equilibrium transform of a mixture. To give notation for a mixture, let  $h$  be a probability measure on the real numbers. Suppose that for each  $b$  in the support of  $h$ , we have a random variable  $X_b$  with distribution  $\nu_b$  and mean  $m_b \in [0, \infty)$ . Also assume that  $b \mapsto \mu_b$  is measurable. The random variable  $X$  is the *mixture of  $(X_b)$  governed by  $h$*  if for all bounded measurable functions  $g$ ,

$$\mathbf{E}g(X) = \int \mathbf{E}g(X_b) dh(b).$$

The basic recipe for the equilibrium transform  $X^e$  is that it is a mixture of the equilibrium transforms  $X_b^e$ , governed by a biased version of  $h$ . The analogous recipe works for forming the size-bias transform of a mixture, and this result follows from that.

**Lemma 21.** *Let  $X$  be the mixture of  $(X_b)$  governed by  $h$  as described above, and assume that  $\mathbf{E}X < \infty$ . Define the measure  $h^s$  by its Radon–Nikodym derivative:*

$$\frac{dh^s(b)}{dh} = \frac{m(b)}{\mathbf{E}X}.$$

*Then the distribution of  $X^e$  is the mixture of  $(X_b^e)$  governed by  $h^s$ .*

**Proof.** By [2, Lemma 2.4], the size-bias transform  $X^s$  is distributed as the mixture of  $X_b^s$  governed by  $h^s$ . With  $U \sim \text{Uniform}(0, 1)$  independent of all else, the equilibrium transform  $[UX^s]$  is thus the mixture of  $[UX_b^s]$  governed by  $h^s$ .  $\square$

#### 4.3. Proofs of the concentration theorems for galton–watson trees

**Theorem 6** is a simple consequence of the following statement that an NBUE quantity of i.i.d. NBUE summands is NBUE. This fact was previously proven in [27, Corollary 2.2]. We include our proof here, as it takes a very different approach from theirs.

**Proposition 22.** *Let  $X, X_1, X_2, \dots$  be i.i.d. nonnegative random variables, and let  $L$  be a positive integer-valued random variable independent of all else. Suppose that  $X$  and  $L$  are NBUE. Then the random sum  $S = \sum_{k=1}^L X_k$  is NBUE as well.*

**Proof.** We construct  $S^*$  using **Lemmas 20** and **21**, as was done in [29, Theorem 3.1]. Let  $S_n = \sum_{k=1}^n X_k$ , so that  $S = S_L$ , and let  $T_k \stackrel{d}{=} S_k^*$ . By **Lemma 20**,

$$T_k \stackrel{d}{=} X_1 + \dots + X_{I_k-1} + X_{I_k}^*,$$

where  $I_k$  is chosen uniformly at random from  $\{1, \dots, k\}$ .

By **Lemma 21**, the equilibrium transform of  $S$  is a mixture of  $T_k$  governed by a distribution whose Radon–Nikodym derivative with respect to  $L$  is

$$\frac{\mathbf{E}S_k}{\mathbf{E}S} = \frac{k\mathbf{E}X}{\mathbf{E}L\mathbf{E}X} = \frac{k}{\mathbf{E}L}, \quad k \in \mathbb{N},$$

which is exactly the Radon–Nikodym distribution of  $L^s$  with respect to  $L$ . Hence,

$$\begin{aligned} S^* &\stackrel{d}{=} T_{L^s} = X_1 + \dots + X_{I_{L^s}-1} + X_{I_{L^s}}^*, \\ &\leq X_1 + \dots + X_{I_{L^s}-1} + X_{I_{L^s}}, \end{aligned}$$

with the second line following because  $X$  is NBUE. Since  $I_{L^s}$  is a uniform selection from  $\{1, \dots, L^s\}$ , it is the discrete equilibrium transform of  $L$ . Hence  $I_{L^s} \leq L$ , and

$$S^* \leq X_1 + \dots + X_L = S. \quad \square$$

**Proof of Theorem 6.** Let  $L$  be a random variable whose distribution is the child distribution of the tree. Each generation of the Galton–Watson process is the sum of  $L$  independent copies of the previous generation, i.e.,

$$Z_{n+1} \stackrel{d}{=} \sum_{j=1}^L Z_n^{(j)},$$

where  $Z_n^{(j)}$  for  $j \geq 1$  denote independent copies of  $Z_n$  and  $L$  is independent of  $(Z_n^{(j)}, j \geq 1)$ . Observing that  $Z_1^{(j)} \stackrel{d}{=} L$  is NBUE, we can apply Proposition 22 inductively to conclude that  $Z_n$  is NBUE for all  $n$ . The concentration inequalities (3) and (4) then follow from Theorems 3 and 4 with  $\alpha = \beta = 1$ .  $\square$

For Theorem 7, it would be nice to argue that an NBUEZT quantity of NBUEZT summands remain NBUEZT, but we have not been able to prove or disprove this (see Section 4.4). But we can show the following weaker statement. For a random variable  $X$  taking nonnegative integer values, we write  $\text{Bin}(X, p)$  to denote the distribution obtained by thinning  $X$  by  $p$  (i.e., the sum of  $X$  independent Bernoulli( $p$ ) random variables).

**Proposition 23.** *Let  $X_1, X_2, \dots$  be i.i.d. and NBUEZT. Let  $p = \mathbf{P}[X_i \geq 1]$ , and let  $L$  be a random variable taking nonnegative integer values, independent of  $X_1, X_2, \dots$ , such that  $\text{Bin}(L, p)$  is NBUEZT. Then  $X_1 + \dots + X_L$  is NBUEZT.*

**Proof.** Let  $S = X_1 + \dots + X_L$ , and let  $M$  be the number of the random variables  $X_1, \dots, X_L$  that are nonzero. Then

$$S \stackrel{d}{=} \sum_{k=1}^M X_k^>, \quad (27)$$

where  $M \sim \text{Bin}(L, p)$  is independent of  $(X_i^>)_{i \geq 1}$ . Since  $S$  is then a sum of  $M$  many strictly positive random variables, it is positive if and only if  $M$  is positive. Hence

$$S^> \stackrel{d}{=} \sum_{k=1}^{M^>} X_k^>.$$

Since  $M$  and  $X_k$  are NBUEZT, their conditioned versions  $M^>$  and  $X_k^>$  are NBUE. Thus  $S^>$  is NBUE by Proposition 22, and hence  $S$  is NBUEZT.  $\square$

To apply Proposition 23, the NBUEZT property for  $L$  must be preserved under thinning. We now show that this holds when  $L^>$  is D-IFR.

**Lemma 24.** *Let  $L$  be a random variable taking nonnegative integer values. If  $L^>$  is D-IFR, then  $\text{Bin}(L, p)^>$  is D-IFR for all  $0 < p \leq 1$ .*

**Proof.** Let  $(B_k)_{k \geq 1}$  be i.i.d.-Bernoulli( $p$ ) for arbitrary  $p \in (0, 1)$ , and let  $M = B_1 + \dots + B_L$ , so that  $M \sim \text{Bin}(L, p)$ . Our goal is to show that  $\mathbf{P}[M = n \mid M \geq n]$  is increasing for  $n \geq 1$ .

Define

$$\varphi(t) = \mathbf{E}[(1-p)^{L-t} \mid L \geq t],$$

which is the conditional probability that  $B_{t+1} = \dots = B_L = 0$  given that  $L \geq t$ . Let  $T_n$  be the smallest index  $t$  such that  $B_1 + \dots + B_t = n$ . We make the following claims:

- (i)  $\mathbf{P}[M = n \mid M \geq n] = \mathbf{E}[\varphi(T_n) \mid L \geq T_n]$ ;
- (ii) the function  $\varphi(t)$  is increasing for integers  $t \geq 1$ ;
- (iii) the distributions  $[T_n \mid L \geq T_n]$  are stochastically increasing in  $n$ .

To prove (i), we start by observing that  $M = n$  holds if and only if  $L \geq T_n$  and  $B_{T_n+1}, \dots, B_L$  are all zero. Thus,

$$\begin{aligned} \mathbf{P}[M = n] &= \mathbf{P}[B_{T_n+1} = \dots = B_L = 0 \text{ and } L \geq T_n] \\ &= \sum_{t, \ell} \mathbf{P}[T_n = t, L = \ell, B_{t+1} = \dots = B_\ell = 0] \mathbf{1}\{\ell \geq t\} \\ &= \sum_{t, \ell} \mathbf{P}[T_n = t, L = \ell] (1-p)^{\ell-t} \mathbf{1}\{\ell \geq t\} \\ &= \sum_t \mathbf{P}[T_n = t] \sum_\ell \mathbf{P}[L = \ell] (1-p)^{\ell-t} \mathbf{1}\{\ell \geq t\} \\ &= \sum_t \mathbf{P}[T_n = t] \mathbf{E}[(1-p)^{L-t} \mathbf{1}\{L \geq t\}] = \sum_t \mathbf{P}[T_n = t] \varphi(t) \mathbf{P}[L \geq t]. \end{aligned}$$

Since  $L$  and  $T_n$  are independent, we obtain

$$\begin{aligned} \mathbf{P}[M = n] &= \sum_t \mathbf{P}[T_n = t, L \geq t] \varphi(t) \\ &= \sum_{t, \ell} \mathbf{P}[T_n = t, L = \ell] \varphi(t) \mathbf{1}\{\ell \geq t\} = \mathbf{E}[\varphi(T_n) \mathbf{1}\{L \geq T_n\}]. \end{aligned}$$

Observing that the events  $\{M \geq n\}$  and  $\{L \geq T_n\}$  are the same and dividing both sides of the above equation by its probability yields (i).

Now we prove (ii). Since  $L$  is D-IFR, the distributions  $[L - t \mid L \geq t]$  are stochastically decreasing by Proposition 19(a). Thus  $\varphi(t)$  is obtained by taking the expectation of the decreasing function  $x \mapsto (1-p)^x$  under a stochastically decreasing sequence of distributions, showing that  $\varphi(t)$  is increasing in  $t$ .

To prove (iii), it suffices (see [34, Theorem 1.C.1]) to show that

$$\frac{\mathbf{P}[T_{n+1} = k \mid L \geq T_{n+1}]}{\mathbf{P}[T_n = k \mid L \geq T_n]} \quad \text{is increasing in } k \text{ for } k \geq n. \quad (28)$$

Thus we consider

$$\begin{aligned} \frac{\mathbf{P}[T_{n+1} = k \mid L \geq T_{n+1}]}{\mathbf{P}[T_n = k \mid L \geq T_n]} &= \frac{\mathbf{P}[T_{n+1} = k, L \geq k]}{\mathbf{P}[L \geq T_{n+1}]} \cdot \frac{\mathbf{P}[L \geq T_n]}{\mathbf{P}[T_n = k, L \geq k]} \\ &= \frac{\mathbf{P}[T_{n+1} = k]}{\mathbf{P}[T_n = k]} \cdot \frac{\mathbf{P}[L \geq T_n]}{\mathbf{P}[L \geq T_{n+1}]}, \end{aligned}$$

with the second line following from the independence of  $T_n$  and  $T_{n+1}$  from  $L$ . The final bit is to compute probabilities for the negative binomial distribution:

$$\frac{\mathbf{P}[T_{n+1} = k]}{\mathbf{P}[T_n = k]} = \frac{\binom{k-1}{n} p^{n+1} (1-p)^{k-n-1}}{\binom{k-1}{n-1} p^n (1-p)^{k-n}} = \frac{(k-n)p}{n(1-p)}.$$

This is increasing in  $k$  for  $k \geq n$ , which proves (28).

Now, statements (i)–(iii) combine to prove the lemma: the quantity  $\mathbf{E}[\varphi(T_n) \mid L \geq T_n]$  is increasing in  $n$  by (ii) and (iii), and hence  $M^>$  is D-IFR by (i).  $\square$

**Proof of Theorem 7.** Let  $L$  be the child distribution of the tree. By Proposition 19(c) and Lemma 24, all thinnings of  $L$  are NBUEZT. Hence Proposition 23 applies and shows that

$$\sum_{k=1}^L X_k$$

is NBUEZT whenever  $(X_k)_k \geq 1$  are an i.i.d. family of NBUEZT random variables. Applying this inductively to each generation  $Z_n$  of the Galton–Watson process shows that  $Z_n$  is NBUEZT for all  $n$ . Therefore Theorems 3 and 4 apply to  $Z_n^>$  and prove (5) and (6).  $\square$

#### 4.4. On the sharpness and optimality of these results

Consider a Galton–Watson process whose child distribution is geometric with success probability  $p$  on  $\{1, 2, \dots\}$ , which is NBUE. The size of the  $n$ th generation is geometric with success probability  $p^n$ . Then  $Z_n/\mu^n \rightarrow \text{Exp}(1)$  in law as  $n \rightarrow \infty$ . Hence both upper and lower bounds in Theorem 6 are sharp, besides the extra factor of  $e$  in the upper bound.

If  $Z_n$  is the  $n$ th generation of a critical Galton–Watson tree with  $m(n) = \mathbf{E}[Z_n \mid Z_n > 0]$ , then  $Z_n/m(n) \rightarrow \text{Exp}(1)$  in law as  $n \rightarrow \infty$  [3, Theorem I.9.2]. Thus Theorem 7 provides optimal bounds in this case, again besides the extra factor of  $e$  in the upper bound.

We can also ask whether the conditions of Theorems 6 and 7 could be weakened and more broadly what properties of the child distribution are preserved for all generations. Theorem 6 states that if the child distribution of a Galton–Watson process is NBUE, then all its generations are NBUE. Log-concave and D-IFR distributions are not preserved in this way. For a counterexample, consider a child distribution placing probability  $1/8$  on 1, probability  $49/64$  on 2, and probability  $7/64$  on 3. This distribution is log-concave and D-IFR, but the size of the second generation is neither.

Lemma 24 states that if  $L^>$  is D-IFR, then  $\text{Bin}(L, p)^>$  is D-IFR. The NBUEZT property is not preserved under thinning in this way. Let

$$L = \begin{cases} 1 & \text{with probability } 89/100, \\ 2 & \text{with probability } 109/1000, \\ 3 & \text{with probability } 9/10000, \\ 4 & \text{with probability } 1/11250, \\ 5 & \text{with probability } 1/90000. \end{cases}$$

We leave it as an exercise that this distribution is NBUEZT (in fact, NBUE) but that all of its thinnings fail to be NBUEZT.

It seems more natural (and would be a weaker condition) to assume only that the child distribution  $L$  is NBUEZT in Theorem 7, rather than that  $L^>$  is D-IFR. But since the NBUEZT

property is not preserved by thinning, our proof of [Theorem 7](#) does not go through with this change. In fact, we are truly unsure the theorem holds with the weaker condition.

#### 4.5. Previous concentration results for Galton–Watson processes

Let  $Z_n$  be the  $n$ th generation of a Galton–Watson process whose child distribution has mean  $\mu > 1$ . Let  $W$  be the almost sure limit of  $Z_n/\mu^n$ , which exists and is nondegenerate when  $\mathbf{E}[Z_1 \log Z_1] < \infty$ . In [Theorems 6](#) and [7](#), properties of the child distribution continue to hold for  $Z_n/\mu^n$  at all generations. This is in a similar spirit to many results linking properties of the child distribution to those of  $W$ . For example, for  $\alpha > 1$  it holds that  $\mathbf{E}Z_1^\alpha$  is finite if and only if  $\mathbf{E}W^\alpha$  is finite [\[9\]](#). Similarly,  $Z_1$  has a regularly varying distribution with index  $\alpha > 1$  if and only if  $W$  does [\[17\]](#).

One line of results is on the right tail when the child distribution is bounded. Let  $d$  be its maximum value, and let  $\gamma = \log d / \log \mu > 1$ . Biggins and Bingham [\[8\]](#) used a classic result of Harris [\[23\]](#) to show that

$$-\log \mathbf{P}[W \geq x] = x^{\gamma/(\gamma-1)} N(x) + o(x^{\gamma/(\gamma-1)}), \quad (29)$$

where  $N(x)$  is a continuous, multiplicatively periodic function. Hence, in the limit the tail of  $Z_n/\mu^n$  decays faster than exponentially. Fleischmann and Wachtel give a more precise version of this result [\[20, Remark 3\]](#), showing that the tail of  $W$  decays as

$$N_2(x) x^{-\gamma/2(\gamma-1)} \exp\left(-x^{\gamma/(\gamma-1)} N(x)\right), \quad (30)$$

where  $N(x)$  and  $N_2(x)$  are continuous, multiplicatively periodic function. Biggins and Bingham give a version of their result that applies directly to  $Z_n$  rather than its limit, and more detailed results on the right tail of  $Z_n$  in this situation can also be obtained from combinatorial results of Flajolet and Odlyzko [\[19, Theorem 1\]](#).

Results on the right tail are also available when the child distribution is heavy-tailed. When  $Z_1$  satisfies

$$\sup_x \frac{\mathbf{P}[Z_1 > x/2]}{\mathbf{P}[Z_1 > x]} < \infty,$$

the tails of  $Z_n$  satisfy

$$c_1 \mathbf{P}[Z_1 > x] \leq \mathbf{P}[Z_n > \mu^n x] \leq c_2 \mathbf{P}[Z_1 > x]$$

for constants  $c_1 > 0$  and  $c_2 < \infty$  independent of  $x$  and  $n$  [\[38, Theorem 1\]](#). This result applies, for instance, when the tail of  $Z_1$  has polynomial decay. A similar result [\[38, Theorem 3\]](#) holds when the tail of  $Z_1$  behaves like  $e^{-x^\alpha}$  for  $0 < \alpha < 1$ .

For the left tail of  $Z_n$ , the behavior depends on the weight that the child distribution places on 0 and 1. It is known as the *Schröder case* when positive weight is placed on those values and as the *Böttcher case* when it is not. Roughly speaking, the left tail in the Schröder case behaves similarly to the right tail in the heavy-tailed case, while the left tail in the Böttcher case behaves similarly to the right tail in the bounded child distribution case. For example, suppose that the child distribution places no weight on 0 and weight  $p_1$  on 1. In the Schröder case, where  $p_1 > 0$ , let  $\alpha = -\log p_1 / \log \mu$ . Then  $\mathbf{P}[W \leq x]$  behaves like  $x^\alpha$  as  $x \rightarrow 0$  [\[18\]](#). Note that  $\alpha = 1$  for a geometric child distribution, coinciding with the lower tail bound we prove in [Theorem 6](#). For the Böttcher case, where  $p_1 = 0$ , let  $d \geq 2$  be the minimum value taken by child distribution, and let  $\beta = \log d / \log \mu \in (0, 1)$ . Then  $-\log \mathbf{P}[W \leq x]$  behaves



like  $x^{-\beta/(1-\beta)}$ . A result like (29) is shown in [8, Theorem 3], and finer asymptotics along the lines of (30) are given in [20, Theorem 1].

Our results apply best to distributions that are unbounded but have exponential tails, a case that seems poorly covered by the existing literature. Our bound is also more explicit than any we have encountered, with no limits or unspecified constants.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

The initial portion of this work was conducted at the meeting *Stein's method and applications in high-dimensional statistics* held at the American Institute of Mathematics in August 2018. We would also like to express our gratitude to John Fry and staff Estelle Basor, Brian Conrey, and Harpreet Kaur at the American Institute of Mathematics for the generosity and excellent hospitality in hosting this meeting at the Fry's Electronics corporate headquarters in San Jose, CA, and Jay Bartroff, Larry Goldstein, Stanislav Minsker and Gesine Reinert for organizing such a stimulating meeting. T.J. received support from National Science Foundation grant DMS-1811952 and PSC-CUNY Award #62628-00 50.

## Appendix A

Recall that  $P_n^l(b, w)$  is the distribution of the number of white balls after  $n$  draws in the urn model defined in Section 3, and let  $N_n(b, w) \sim P_n^l(b, w)$ . This appendix is dedicated to proving the following result, which is used in Lemma 16 to compare the rising factorial bias transform of these distributions to their power bias transforms. The property proven in the following lemma is something like log-concavity of the sequence  $\mathbf{P}[N_n(b, w) = k]$  for fixed  $n$  (which is proven along the way, in Lemma 26), but it involves varying both  $k$  and  $n$ . We cannot give much intuition for the proof; it seems to us to be a technical fact that happens to be true and can be proven by pushing symbols around in the right way.

**Lemma 25.** *For all  $n, k \geq 0$ ,*

$$\mathbf{P}[N_n(b, w) = k - 1] \mathbf{P}[N_{n+1}(b, w) = k + 1] \leq \mathbf{P}[N_n(b, w) = k] \mathbf{P}[N_{n+1}(b, w) = k]. \quad (31)$$

We will in fact prove Lemma 25 for a slightly generalized version of the urn process. As with that process, start with  $b \geq 1$  black balls and  $w \geq 1$  white balls, and after each draw add an extra ball with the same color as the ball drawn. Instead of adding an additional black ball after every  $l$ th draw, we allow black balls to be added at arbitrary but predetermined times. Thus the number of balls in the urn after  $n$  draws, denoted by  $B_n$ , is an arbitrary but deterministic strictly increasing sequence with  $B_0 = b + w$ . Let  $N_n$  be the number of white balls in the urn after  $n$  draws. Let  $t_k^{(n)} = \mathbf{P}[N_n = k]$ . The dynamics of the urn process gives

$$t_k^{(n+1)} = \left( \frac{k-1}{B_n} \right) t_{k-1}^{(n)} + \left( 1 - \frac{k}{B_n} \right) t_k^{(n)}. \quad (32)$$

First, we show that  $t_k^{(n)}$  is log-concave in  $k$  for each fixed  $n$ :

**Lemma 26.** For all  $n \geq 0$  and all  $k$ ,

$$(t_k^{(n)})^2 \geq t_{k-1}^{(n)} t_{k+1}^{(n)}. \quad (33)$$

**Proof.** We prove this by induction. For the base case, we have  $t_k^{(0)} = \mathbf{1}\{k = w\}$ , and hence the right-hand side of (33) is always zero when  $n = 0$ . Now, we expand  $(t_k^{(n+1)})^2 - t_{k-1}^{(n+1)} t_{k+1}^{(n+1)}$  using (32) as  $(A_1 + A_2 + A_3)/B_n^2$  for

$$A_1 = (k-1)^2 t_{k-1}^2 - (k-2)k t_{k-2} t_k, \quad (34)$$

$$A_2 = (B_n - k)^2 t_k^2 - (B_n - k + 1)(B_n - k - 1) t_{k-1} t_{k+1}, \quad (35)$$

$$A_3 = 2(k-1)(B_n - k) t_{k-1} t_k - (k-2)(B_n - k - 1) t_{k-2} t_{k+1} - k(B_n - k + 1) t_{k-1} t_k, \quad (36)$$

where we have simplified notation by writing  $t_k$  for  $t_k^{(n)}$ . Applying the inductive hypothesis to (34) and (35) gives

$$A_1 \geq ((k-1)^2 - (k-2)k) t_{k-1}^2 = t_{k-1}^2,$$

and

$$A_2 \geq ((B_n - k)^2 - (B_n - k + 1)(B_n - k - 1)) t_k^2 = t_k^2.$$

To bound  $A_3$ , we note that the inductive hypothesis implies  $t_{k-2} t_{k+1} \leq t_{k-1} t_k$ , which together with (36) gives

$$A_3 \geq (2(k-1)(B_n - k) - (k-2)(B_n - k - 1) - k(B_n - k + 1)) t_{k-1} t_k = -2 t_{k-1} t_k.$$

Hence

$$A_1 + A_2 + A_3 \geq t_{k-1}^2 + t_k^2 - 2 t_{k-1} t_k = (t_{k-1} - t_k)^2 \geq 0,$$

thus extending the induction.  $\square$

Next, we establish a variant of log-concavity with a similar but more complicated proof.

**Lemma 27.** For all  $n \geq 0$  and all  $k$ ,

$$(B_n - k)(t_k^{(n)})^2 - (B_n - k - 1)t_{k-1}^{(n)} t_{k+1}^{(n)} - t_{k-1}^{(n)} t_k^{(n)} \geq 0. \quad (37)$$

**Proof.** We proceed by induction. Let  $E_k^{(n)}$  be the left-hand side of (37). Since  $t_k^{(0)} = \mathbf{1}\{k = w\}$  and  $B_0 = w + b$ , we have  $E_k^{(0)} = b \mathbf{1}\{k = w\}$ , demonstrating (37) when  $n = 0$ . Now we assume  $E_k^{(n)} \geq 0$  for all  $k$ , and we show  $E_k^{(n+1)} \geq 0$  for all  $k$ . It suffices to prove  $E_k^{(n+1)} \geq 0$  under the assumption that  $B_{n+1} = B_n + 1$ , because  $B_{n+1}$  is at least this large, and we can see that  $E_k^{(n+1)}$  is increasing in  $B_{n+1}$  by writing it as

$$E_k^{(n+1)} = (B_{n+1} - k - 1) \left[ (t_k^{(n+1)})^2 - t_{k-1}^{(n+1)} t_{k+1}^{(n+1)} \right] + (t_k^{(n+1)})^2 - t_{k-1}^{(n+1)} t_k^{(n+1)}$$

and applying Lemma 26.

For the sake of readability, we write  $t_k$  for  $t_k^{(n)}$  and  $B$  for  $B_n$  in this proof. We apply (32) to obtain

$$\begin{aligned} (t_k^{(n+1)})^2 &= \frac{1}{B^2} \left( (k-1)^2 t_{k-1}^2 + 2(k-1)(B-k) t_{k-1} t_k + (B-k)^2 t_k^2 \right), \\ t_{k-1}^{(n+1)} t_{k+1}^{(n+1)} &= \frac{1}{B^2} \left( (k-2)k t_{k-2} t_k + (k-2)(B-k-1) t_{k-2} t_{k+1} \right. \end{aligned}$$

$$\begin{aligned}
& + k(B - k + 1)t_{k-1}t_k + (B - k - 1)(B - k + 1)t_{k-1}t_{k+1}), \\
t_{k-1}^{(n+1)}t_k^{(n+1)} &= \frac{1}{B^2} \left( (k-2)(k-1)t_{k-2}t_{k-1} + (k-2)(B-k)t_{k-2}t_k \right. \\
& \quad \left. + (k-1)(B-k+1)t_{k-1}^2 + (B-k)(B-k+1)t_{k-1}t_k \right).
\end{aligned}$$

Now, under the assumption that  $B_{n+1} = B + 1$ , we expand  $E_k^{(n+1)}$  as  $(A_1 + A_2 + A_3)/B^2$ , where

$$\begin{aligned}
A_1 &= (k-2) \left( (B-k+1)(k-1)t_{k-1}^2 - (k+1)(B-k)t_{k-2}t_k - (k-1)t_{k-2}t_{k-1} \right) \\
&= (k-2) \left( (k-1)E_{k-1}^{(n)} - 2(B-k)t_{k-2}t_k \right) \geq -2(k-2)(B-k)t_{k-2}t_k,
\end{aligned}$$

and

$$\begin{aligned}
A_2 &= (B-k+1)(B-k) \left( (B-k)t_k^2 - (B-k-1)t_{k-1}t_{k+1} \right) \\
&= (B-k+1)(B-k) \left( E_k^{(n)} + t_{k-1}t_k \right) \geq (B-k+1)(B-k)t_{k-1}t_k,
\end{aligned}$$

and

$$\begin{aligned}
A_3 &= (B-k+1)(B-k)(k-3)t_{k-1}t_k - (k-2)(B-k)(B-k-1)t_{k-2}t_{k+1} \\
&= (B-k) \left( (k-2) \left[ (B-k+1)t_{k-1}t_k - (B-k-1)t_{k-2}t_{k+1} \right] - (B-k+1)t_{k-1}t_k \right) \\
&= (B-k) \left( (k-2) \left[ \frac{E_k^{(n)}t_{k-2} + E_{k-1}^{(n)}t_k}{t_{k-1}} + 2t_{k-2}t_k \right] - (B-k+1)t_{k-1}t_k \right) \\
&\geq (B-k) \left( 2(k-2)t_{k-2}t_k + (B-k+1)t_{k-1}t_k \right),
\end{aligned}$$

where we have applied the inductive hypothesis in each final step. Combining these bounds,

$$\begin{aligned}
E_k^{(n+1)} &\geq \frac{(B-k)}{B^2} \left( -2(k-2)t_{k-2}t_k + (B-k+1)t_{k-1}t_k \right. \\
&\quad \left. + 2(k-2)t_{k-2}t_k - (B-k+1)t_{k-1}t_k \right) \\
&= 0. \quad \square
\end{aligned}$$

**Proof of Lemma 25.** First, we dispense with the case that any of  $t_{k-1}^{(n)}$ ,  $t_{k+1}^{(n+1)}$ ,  $t_k^{(n)}$ , or  $t_k^{(n+1)}$  are equal to zero. If either of  $t_{k-1}^{(n)}$  or  $t_{k+1}^{(n+1)}$  equals zero, then the left-hand side of (31) is zero and the inequality holds. If  $t_k^{(n)}$  or  $t_k^{(n+1)}$  equals 0 then  $t_{k+1}^{(n+1)} = 0$ , since the support of  $t_k^{(n)}$  is  $\{w, \dots, w+n\}$ ; in this case both sides of (31) are zero. Thus we assume from now on that these four terms are all nonzero.

Now, proving the lemma is equivalent to showing  $t_k^{(n+1)}/t_{k+1}^{(n+1)} - t_{k-1}^{(n)}/t_k^{(n)} \geq 0$ . We compute

$$\begin{aligned}
\frac{t_k^{(n+1)}}{t_{k+1}^{(n+1)}} - \frac{t_{k-1}^{(n)}}{t_k^{(n)}} &= \frac{t_k^{(n+1)} - \frac{t_{k-1}^{(n)}}{t_k^{(n)}}t_{k+1}^{(n+1)}}{t_{k+1}^{(n+1)}} \\
&= \frac{1}{t_{k+1}^{(n+1)}} \left( \left( \frac{k-1}{B_n} \right) t_{k-1}^{(n)} + \left( 1 - \frac{k}{B_n} \right) t_k^{(n)} \right. \\
&\quad \left. - \frac{t_{k-1}^{(n)}}{t_k^{(n)}} \left[ \left( \frac{k}{B_n} \right) t_k^{(n)} + \left( 1 - \frac{k+1}{B_n} \right) t_{k+1}^{(n)} \right] \right) \\
&= \frac{1}{B_n t_{k+1}^{(n+1)} t_k^{(n)}} \left( -t_{k-1}^{(n)} t_k^{(n)} + (B_n - k)(t_k^{(n)})^2 - (B_n - k - 1)t_{k-1}^{(n)} t_{k+1}^{(n)} \right),
\end{aligned}$$

which is nonnegative by [Lemma 27](#).  $\square$

**Remark 28.** It is possible to avoid all the work of this appendix, at the cost of a slightly inferior concentration bound for  $N_n(1, w)$ . The result of this appendix ([Lemma 25](#)) is used to prove that  $N_{n-l}^{[l+1]}(b, w) + l \geq N_n^{(l+1)}(b, w)$  ([Lemma 16](#)), which is then applied in the proof of [Proposition 17](#). An alternate path is to invoke the following stochastic inequality between the factorial and power bias transformations, which holds for any nonnegative random variable:

$$X^{(l+1)} \leq_{\bar{p}} X^{[l+1]}, \quad (38)$$

where

$$p = \frac{\mathbf{E}X^{l+1}}{\mathbf{E}[X(X+1)\cdots(X+l)]}.$$

Modifying the derivation in [Proposition 17](#) slightly, we get

$$\begin{aligned} N_n(1, w) &\stackrel{d}{=} Q_w(N_{n-l}(1, w+1+l) - w - 1) \\ &\geq Q_w(N_n(1, w+1+l) - l - w - 1), \end{aligned}$$

with the second line holding since at most  $l$  white balls can be added from steps  $n-l$  to  $n$ . Then following the same steps as in [Proposition 17](#),

$$N_n(1, w) \geq Q_w(N_n^{[l+1]}(1, w) - w) \geq V_w N_n^{[l+1]}(1, w).$$

Finally, invoking (38), we have

$$N_n(1, w) \geq_{\bar{p}} V_w N_n^{(l+1)}(1, w) \stackrel{d}{=} N_n^*(1, w).$$

The concentration bounds obtained from this are worse because of the factor of  $p$  in the exponent, but it does illustrate how the  $p < 1$  versions of our concentration bounds can be used.

## Appendix B

In this appendix, we prove [Proposition 19](#) for the convenience of the reader. See also [7, Chapters 4 and 6] for more background material.

**Proof of Proposition 19(a).** Suppose that  $X$  is D-IFR and let  $p_n = \mathbf{P}[X = n \mid X \geq n]$ . To show that  $[X - k \mid X > k]$  is stochastically decreasing in  $k$ , construct a random variable  $T$  by the following procedure: Fix some  $k \geq 0$ . Start at 1 and halt with probability  $p_{k+1}$ ; otherwise advance to 2 and halt with probability  $p_{k+2}$ ; otherwise advance to 3, and continue like this, letting  $T$  be the value where we halt. It is evident that  $T \sim [X - k \mid X > k]$ . Since  $p_n$  is increasing, we are more likely to halt at each step when  $k$  is increased. By a simple coupling, this demonstrates that  $[X - k \mid X > k]$  is stochastically decreasing in  $k$ .

Conversely, suppose that  $[X - k \mid X > k]$  is stochastically decreasing in  $k$ . Then

$$\mathbf{P}[X - k + 1 \leq 1 \mid X > k - 1] \leq \mathbf{P}[X - k \leq 1 \mid X > k]$$

by the definition of stochastic dominance, which proves that

$$\mathbf{P}[X = k \mid X \geq k] \leq \mathbf{P}[X = k + 1 \mid X \geq k + 1]. \quad \square$$

**Proof of Proposition 19(b).** From the definition of the discrete equilibrium transform,

$$\mathbf{P}[X^e = n] = \sum_{k=n}^{\infty} \frac{1}{k} \mathbf{P}[X^s = k] = \sum_{k=n}^{\infty} \frac{1}{\mathbf{E}X} \mathbf{P}[X = k] = \frac{1}{\mathbf{E}X} \mathbf{P}[X \geq n].$$

Hence

$$\begin{aligned} \mathbf{P}[X^e > k] &= \frac{1}{\mathbf{E}X} \sum_{n=k+1}^{\infty} \mathbf{P}[X \geq n] \\ &= \frac{1}{\mathbf{E}X} \mathbf{E}[(X - k) \mathbf{1}\{X > k\}] = \mathbf{P}[X > k] \frac{\mathbf{E}[X - k \mid X > k]}{\mathbf{E}X}. \end{aligned} \quad (39)$$

Therefore  $\mathbf{P}[X^e > k] \leq \mathbf{P}[X > k]$  holds for all  $k \geq 1$  if and only if  $\mathbf{E}[X - k \mid X > k] \leq \mathbf{E}X$  for all  $k \geq 1$ .  $\square$

**Proof of Proposition 19(c).** Suppose  $X$  takes values in the positive integers and is log-concave. Let  $p_n = \mathbf{P}[X = n]$ , and let  $N$  be the highest value such that  $p_N > 0$ , with  $N = \infty$  a possibility. From the definition of log-concave,

$$\frac{p_{n-1}}{p_n} \leq \frac{p_n}{p_{n+1}}$$

for all  $1 \leq n \leq N$ . This implies that for any fixed  $k$ , the ratio

$$\frac{\mathbf{P}[X - k = n \mid X > k]}{\mathbf{P}[X - k - 1 = n \mid X > k + 1]} = \frac{p_{n+k}}{p_{n+k+1}} \cdot \frac{\mathbf{P}[X > k + 1]}{\mathbf{P}[X > k]}$$

is increasing in  $n$ , and this condition implies that  $[X - k \mid X > k]$  stochastically dominates  $[X - k - 1 \mid X > k + 1]$  (see [34, Theorem 1.C.1]). Hence  $X$  is D-IFR by Proposition 19(a).

Now, suppose that  $X$  is D-IFR. It follows from Proposition 19(a) that  $\mathbf{E}[X - k \mid X > k]$  is decreasing in  $k$  for integers  $k \geq 0$ , proving that

$$\mathbf{E}[X - k \mid X > k] \leq \mathbf{E}[X - 0 \mid X > 0] = \mathbf{E}X.$$

By Proposition 19(b), this shows that  $X$  is NBUE.  $\square$

## References

- [1] Richard Arratia, Peter Baxendale, Bounded size bias coupling: A Gamma function bound, and universal dickman-function behavior, *Probab. Theory Related Fields* 162 (3–4) (2015) 411–429.
- [2] Richard Arratia, Larry Goldstein, Fred Kochman, Size bias for one and all, *Probab. Surv.* 16 (2019) 1–61, [3896143](#).
- [3] K.B. Athreya, P.E. Ney, *Branching Processes*, Dover Publications, Inc., Mineola, NY, 2004, p. xii+287, Reprint of the 1972 original [Springer, New York; MR0373040], [2047480](#).
- [4] Albert-László Barabási, Réka Albert, Emergence of scaling in random networks, *Science* 286 (5439) (1999) 509–512, [2091634](#).
- [5] Andrew D. Barbour, Louis H.Y. Chen, Stein’s (magic) method, 2014, available at [arXiv:1411.1179](#).
- [6] A.D. Barbour, Lars Holst, Svante Janson, *Poisson Approximation*, in: *Oxford Studies in Probability*, vol. 2, The Clarendon Press Oxford University Press, New York, 1992, p. x+277, Oxford Science Publications, <http://www.ams.org/mathscinet-getitem?mr=1163825> (93g:60043)[1163825 (93g:60043)].
- [7] Richard E. Barlow, Frank Proschan, *Statistical Theory of Reliability and Life Testing*, in: *International Series in Decision Processes*, Holt, Rinehart and Winston, Inc., New York-Montreal, Que.-London, 1975, p. xiii+290, Probability models, [0438625](#).
- [8] J.D. Biggins, N.H. Bingham, Large deviations in the supercritical branching process, *Adv. Appl. Probab.* 25 (4) (1993) 757–772, [1241927](#).

- [9] N.H. Bingham, R.A. Doney, Asymptotic properties of supercritical branching processes. I. The Galton-Watson process, *Adv. Appl. Probab.* 6 (1974) 711–731, [362525](#).
- [10] Mark Brown, Exploiting the waiting time paradox: applications of the size-biasing transformation, *Probab. Engrg. Inform. Sci.* 20 (2) (2006) 195–230, [2261286](#).
- [11] Mark Brown, Sharp bounds for NBUE distributions, *Ann. Oper. Res.* 208 (2013) 245–250, [3100632](#).
- [12] Sourav Chatterjee, A short survey of Stein’s method, in: *Proceedings of the International Congress of Mathematicians—SEoul 2014*. Vol. IV, Kyung Moon Sa, Seoul, 2014, pp. 1–24, [3727600](#).
- [13] Louis H.Y. Chen, Poisson approximation for dependent trials, *Ann. Probab.* 3 (3) (1975) 534–545, <http://www.ams.org/mathscinet-getitem?mr=0428387> (55 #1408)[0428387 (55 #1408)].
- [14] Louis H.Y. Chen, Larry Goldstein, Qi-Man Shao, Normal Approximation By Stein’s Method, in: *Probability and its Applications* (New York), Springer, Heidelberg, 2011, p. xii+405, [2732624](#).
- [15] Nicholas Cook, Larry Goldstein, Tobias Johnson, Size biased couplings and the spectral gap for random regular graphs, *Ann. Probab.* 46 (1) (2018) 72–125, [3758727](#).
- [16] Fraser Daly, Oliver Johnson, Relaxation of monotone coupling conditions: Poisson approximation and beyond, *J. Appl. Probab.* 55 (3) (2018) 742–759, [3877881](#).
- [17] A. De Meyer, On a theorem of bingham and doney, *J. Appl. Probab.* 19 (1) (1982) 217–220, [644434](#).
- [18] M. Serge Dubuc, La densité de la loi-limite d’un processus en cascade expansif, *Z. Wahrscheinlichkeitstheorie Verwandte Geb.* 19 (1971) 281–290, [300353](#).
- [19] P. Flajolet, A.M. Odlyzko, Limit distributions for coefficients of iterates of polynomials with applications to combinatorial enumerations, *Math. Proc. Camb. Philos. Soc.* 96 (2) (1984) 237–253, [757658](#).
- [20] Klaus Fleischmann, Vitali Wachtel, On the left tail asymptotics for the limit law of supercritical galton-watson processes in the böttcher case, *Ann. Inst. Henri Poincaré Probab. Stat.* 45 (1) (2009) 201–225, [2500235](#).
- [21] Subhankar Ghosh, Larry Goldstein, Concentration of measures via size-biased couplings, *Probab. Theory Related Fields* 149 (1–2) (2011) 271–278, [2773032](#).
- [22] Larry Goldstein, Gesine Reinert, Stein’s method and the zero bias transformation with application to simple random sampling, *Ann. Appl. Probab.* 7 (4) (1997) 935–952, [1484792](#).
- [23] T.E. Harris, Branching processes, *Ann. Math. Statist.* 19 (1948) 474–494, [27465](#).
- [24] Tobias Johnson, Matthew Junge, Hanbaek Lyu, David Sivakoff, Particle density in diffusion-limited annihilating systems, 2020, available at [arXiv:2005.06018v1](https://arxiv.org/abs/2005.06018).
- [25] Isaac Meilijson, Limiting properties of the mean residual lifetime function, *Ann. Math. Stat.* 43 (1972) 354–357, [0315763](#).
- [26] Anthony G. Pakes, Ravindra Khattree, Length-biasing, characterizations of laws and the moment problem, *Aust. J. Stat.* 34 (2) (1992) 307–322, [1193781](#).
- [27] Kristina P. Pavlova, Jun Cai, Gordon E. Willmot, The preservation of classes of discrete distributions under convolution and mixing, *Insur. Math. Econom.* 38 (2) (2006) 391–405, [2212536](#).
- [28] Erol Peköz, Adrian Röllin, Exponential approximation for the nearly critical galton-watson process and occupation times of Markov chains, *Electron. J. Probab.* 16 (2011) no. 51, 1381–1393, [2827464](#).
- [29] Erol A. Peköz, Adrian Röllin, New rates for exponential approximation and the theorems of Rényi and Yaglom, *Ann. Probab.* 39 (2) (2011) 587–608, [2789507](#).
- [30] Erol A. Peköz, Adrian Röllin, Nathan Ross, Generalized gamma approximation with rates for urns, walks and trees, *Ann. Probab.* 44 (3) (2016) 1776–1816, [3502594](#).
- [31] Thomas K. Philips, Randolph Nelson, The moment bound is tighter than Chernoff’s bound for positive tail probabilities, *Amer. Statist.* 49 (2) (1995) 175–178, [1347726](#).
- [32] Nathan Ross, Fundamentals of stein’s method, *Probab. Surv.* 8 (2011) 210–293, [2861132](#).
- [33] Sheldon M. Ross, J. George Shanthikumar, Zegang Zhu, On increasing-failure-rate random variables, *J. Appl. Probab.* 42 (3) (2005) 797–809, [2157521](#).
- [34] Moshe Shaked, J. George Shanthikumar, *Stochastic Orders*, in: *Springer Series in Statistics*, Springer, New York, 2007, p. xvi+473, <http://www.ams.org/mathscinet-getitem?mr=2265633> (2008g:60005)[2265633 (2008g:60005)].
- [35] Charles Stein, A bound for the error in the normal approximation to the distribution of a sum of dependent random variables, in: *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability Theory, Univ. California Press, Berkeley, Calif., 1972, pp. 583–602, <http://www.ams.org/mathscinet-getitem?mr=0402873> (53 #6687)[0402873 (53 #6687)].
- [36] Charles Stein, Approximate Computation of Expectations, in: *Institute of Mathematical Statistics Lecture Notes—Monograph Series*, vol. 7, Institute of Mathematical Statistics, Hayward, CA, 1986, p. iv+164, [882007](#).

- [37] V. Strassen, The existence of probability measures with given marginals, *Ann. Math. Stat.* 36 (1965) 423–439, [177430](#).
- [38] V.I. Vakhtel', D.È. Denisov, D.A. Korshunov, On the asymptotics of the tail of distribution of a supercritical galton-watson process in the case of heavy tails, *Tr. Mat. Inst. Steklova* 282 (Vetvyashchiesya Protsessy, Sluchaïnye Bluzhdaniya, i Smezhnye Voprosy) (2013) 288–314, English version published in *Proc. Steklov Inst. Math.* 282 (2013) no. 1, 273–297, [3308596](#).
- [39] Gordon E. Willmot, Steve Drekic, Jun Cai, Equilibrium compound distributions and stop-loss moments, *Scand. Actuar. J.* (1) (2005) 6–24, [2118522](#).
- [40] Yizhe Zhu, On the second eigenvalue of random bipartite biregular graphs, 2022, available at [arXiv:2005.08103v4](#).