The Annals of Probability
2016, Vol. 44, No. 3, 1776–1816
DOI: 10.1214/15-AOP1010
© Institute of Mathematical Statistics, 2016

GENERALIZED GAMMA APPROXIMATION WITH RATES FOR URNS, WALKS AND TREES

BY EROL A. PEKÖZ[1], ADRIAN RÖLLIN[1] AND NATHAN ROSS[2]

Boston University, National University of Singapore and University of Melbourne

We study a new class of time inhomogeneous Pólya-type urn schemes and give optimal rates of convergence for the distribution of the properly scaled number of balls of a given color to nearly the full class of generalized gamma distributions with integer parameters, a class which includes the Rayleigh, half-normal and gamma distributions. Our main tool is Stein's method combined with characterizing the generalized gamma limiting distributions as fixed points of distributional transformations related to the equilibrium distributional transformation from renewal theory. We identify special cases of these urn models in recursive constructions of random walk paths and trees, yielding rates of convergence for local time and height statistics of simple random walk paths, as well as for the size of random subtrees of uniformly random binary and plane trees.

1. Introduction. Generalized gamma distributions arise as limits in a variety of combinatorial settings involving random trees [e.g., Janson (2006b), Meir and Moon (1978) and Panholzer (2004)], urns [e.g., Janson (2006a)], and walks [e.g., Chung (1976), Chung and Hunt (1949) and Durrett and Iglehart (1977)]. These distributions are those of gamma variables raised to a power and noteworthy examples are the Rayleigh and half-normal distributions. We show that for a family of time inhomogeneous generalized Pólya urn models, nearly the full class of generalized gamma distributions with integer parameters appear as limiting distributions, and we provide optimal rates of convergence to these limits. Apart from some special cases, both the characterizations of the limit distributions and the rates of convergence are new.

The result for our urn model (Theorem 1.2 below) follows from a general approximation result (Theorem 1.16 below) which provides a framework for bounding the distance between a generalized gamma distribution and a distribution of interest. This result is derived using Stein's method [see Ross (2011), Ross and

Received September 2013; revised February 2015.

[1]Supported in part by NUS Grant R-155-000-124-112.

[2]Supported in part by ARC Grant DP150101459, NSF Grants DMS-07-04159, DMS-08-06118, DMS-11-06999 and ONR Grant N00014-11-1-0140.

MSC2010 subject classifications. Primary 60F05, 60C05; secondary 60E10, 60K99.

Key words and phrases. Generalized gamma distribution, Pólya urn model, Stein's method, distributional transformations, random walk, random binary trees, random plane trees, preferential attachment random graphs.

Peköz (2007) and Chen, Goldstein and Shao (2011) for overviews] coupled with characterizing the generalized gamma distributions as unique fixed points of certain distributional transformations. Similar approaches to deriving approximation results have found past success for other distributions in many applications: the size-bias transformation for Poisson approximation by Barbour, Holst and Janson (1992), the zero-bias transformation for normal approximation by Goldstein and Reinert (1997, 2005) [and a discrete analog of Goldstein and Xia (2006)], the equilibrium transformation of renewal theory for both exponential and geometric approximation, and an extension to negative binomial approximation by Peköz and Röllin (2011), Peköz, Röllin and Ross (2013b) and Ross (2013), and a transformation for a class of distributions arising in preferential attachment graphs by Peköz, Röllin and Ross (2013a). Luk (1994) and Nourdin and Peccati (2009) developed Stein's method for gamma approximation, though the approaches there are quite different from ours. Theorem 1.16 is a significant generalization and embellishment of this previous work.

Using the construction of Rémy (1985) for generating uniform random binary trees, we find some of our urn distributions embedded in random subtrees of uniform binary trees and plane trees. Moreover, a well-known bijection between binary trees and Dyck paths yields analogous embeddings in some local time and height statistics of random walk. By means of these embeddings, we are able to prove convergence to generalized gamma distributions with rates for these statistics. These limits and in general the connection between random walks, trees and distributions appearing in Brownian motion are typically understood through classical bijections between trees and walks along with Donsker's invariance principle, or through the approach of Aldous' continuum random tree; see Aldous (1991). While these perspectives are both beautiful and powerful, the mathematical details are intricate and they do not provide rates of convergence. In this setting, our work can be viewed as a simple unified approach to understanding the appearance of these limits in the tree-walk context which has the added benefit of providing rates of convergence.

In the remainder of the Introduction, we state our urn, tree and walk results in detail.

1.1. *Generalized gamma distribution.* For $\alpha > 0$, denote by $\mathrm{G}(\alpha)$ the gamma distribution with shape parameter α having density $x^{\alpha-1}e^{-x}/\Gamma(\alpha)\,dx$, $x > 0$.

DEFINITION 1.1 (Generalized gamma distribution). For positive real numbers α and β, we say a random variable Z has the *generalized gamma distribution with parameters α and β* and write $Z \sim \mathrm{GG}(\alpha, \beta)$, if $Z \stackrel{\mathscr{D}}{=} X^{1/\beta}$, where $X \sim \mathrm{G}(\alpha/\beta)$.

The density of $Z \sim \mathrm{GG}(\alpha, \beta)$ is easily seen to be

$$\varphi_{\alpha,\beta}(x) = \frac{\beta x^{\alpha-1}e^{-x^\beta}}{\Gamma(\alpha/\beta)}\,dx, \qquad x > 0,$$

and for any real $p > -\alpha$, $\mathbb{E}Z^p = \Gamma((\alpha + p)/\beta)/\Gamma(\alpha/\beta)$; in particular $\mathbb{E}Z^\beta = \alpha/\beta$. The generalized gamma family includes the Rayleigh distribution, $\mathrm{GG}(2, 2)$, the absolute or "half" normal distribution, $\mathrm{GG}(1, 2)$, and the standard gamma distribution, $\mathrm{GG}(\alpha, 1)$.

1.2. *Pólya urn with immigration.* We now define a variation of Pólya's urn. An urn starts with black and white balls and draws are made sequentially. After each draw, the ball is replaced and another ball of the same color is added to the urn. Also, after every lth draw an additional black ball is added to the urn. Let $\mathcal{P}_n^l(b, w)$ 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. Note that for the case $l = 1$ the process is time homogeneous but for $l \geq 2$ it is time inhomogeneous. Define the *Kolmogorov distance* between two cumulative distribution functions P and Q (or their respective laws) as

$$d_{\mathrm{K}}(P, Q) = \sup_x \bigl|P(x) - Q(x)\bigr|.$$

The Kolmogorov metric is a standard and natural metric for random variables on the real line and is used for statistical inference, for example, in computing "p-values".

THEOREM 1.2. *Let $l, w \geq 1$ and let $N_n \sim \mathcal{P}_n^l(1, w)$. Then $\mathbb{E}N_n^k \asymp n^{kl/(l+1)}$ as $n \to \infty$ for any integer $k \geq 0$, and*

$$\mathbb{E}N_n^{l+1} \sim n^l w \left(\frac{l+1}{l}\right)^l.$$

Furthermore, there are constants $c = c_{l,w}$ and $C = C_{l,w}$, independent of n, such that

$$cn^{-l/(l+1)} \leq d_{\mathrm{K}}\bigl(\mathscr{L}(N_n/\mu_n), \mathrm{GG}(w, l+1)\bigr) \leq Cn^{-l/(l+1)}, \tag{1.1}$$

where

$$\mu_n = \mu_n(l, w) = \left(\frac{l+1}{w}\mathbb{E}N_n^{l+1}\right)^{1/(l+1)} \sim n^{l/(l+1)} \frac{(l+1)}{l^{l/(l+1)}}. \tag{1.2}$$

REMARK 1.3. A direct application of this result is to a preferential attachment random graph model [see Barabási and Albert (1999), Peköz, Röllin and Ross (2013a)] that initially has one node having weight w (thought of as the degree of that node or a collection of nodes grouped together). 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 at that exact moment, where incoming edges contribute weight one to a node and each node other than the initial node is born having initial weight one.

The case where $l = 1$ is the usual Barabasi–Albert tree with loops (though started from a node with initial weight w and no edges). A moment's thought shows that after an additional n edges have been added to the graph, the total weight of the initial node has distribution $\mathcal{P}_n^l(1, w)$. Peköz, Röllin and Ross (2014) extend the results of this paper in this preferential attachment context to obtain limits for joint distributions of the weights of nodes.

REMARK 1.4. Theorem 1.2 in the case when $l = 1$ is covered by Example 3.1 of Janson (2006a), but without a rate of convergence. The limit and rate for the two special cases where $w = l = 1$ and $l = 1$, $w = 2$ are stated in Theorem 1.1 of Peköz, Röllin and Ross (2013a); in fact the rate proved there is $n^{-1/2} \log n$ (there is an error in the last line of the proof of their Lemma 4.2), but our approach here yields the optimal rate claimed there.

REMARK 1.5. For $n \geq l$, it is clear that

$$\mathcal{P}_n^l(0, w) = \mathcal{P}_{n-l}^l(1, w + l), \tag{1.3}$$

since, if the urn is started without black balls, the progress of the urn is deterministic until the first immigration. $\mathcal{P}_n^l(1, w)$ is more natural in the context of the proof of Theorem 1.2 but in our combinatorial applications, $\mathcal{P}_n^l(0, w)$ can be easier to work with and so we will occasionally apply Theorem 1.2 directly to $\mathcal{P}_n^l(0, w)$ via (1.3). Further, in order to easily switch between these two cases without introducing unnecessary notation or case distinctions, we define, in accordance with (1.3), $\mathcal{P}_{-i}^l(1, w + l)$ to be a point mass at $w + l - i$ for all $0 \leq i \leq l$.

REMARK 1.6. Pólya urn schemes have a long history and large literature. In brief, the basic model, in which the urn starts with w white and b black balls and at each stage a ball is drawn at random and replaced with α balls of the same color, was introduced in Eggenberger and Pólya (1923) as a model for disease contagion. The proportion of white balls converges almost surely to a variable having beta distribution with parameters $(w/\alpha, b/\alpha)$. A well-known embellishment [see Friedman (1949)] is to replace the ball drawn along with α balls of the same color and β of the other color and here if $\beta \neq 0$ the proportion of white balls almost surely converges to $1/2$; and Freedman (1965) proves a Gaussian limit theorem for the fluctuation around this limit.

The general case can be encoded by $(\alpha, \beta; \gamma, \delta)_{b,w}$ where now the urn starts with b black and w white balls and at each stage a ball is drawn and replaced; if the ball drawn is black (white), then α (γ) black balls and β (δ) white balls are added. As suggested by the previous paragraph, the limiting behavior of the urn can vary wildly depending on the relationship of the six parameters involved and especially the Greek letters; even the first-order growth of the number of white balls is highly sensitive to the parameters.

A useful tool for analyzing the general case is to embed the urn process into a multitype branching process and use the powerful theory available there. This was first suggested and implemented by Athreya and Karlin (1968) and has found subsequent success in many further works; see Janson (2006a) and Pemantle (2007), and references therein. An alternative approach that is especially useful when α or δ are negative (under certain conditions this leads to a *tenable* urn) is the analytic combinatorics methods of Flajolet, Gabarró and Pekari (2005); see also the Introduction there for further references.

Note that all of the references of the previous paragraphs regard homogeneous urn processes and so do not directly apply to the model of Theorem 1.2 with $l \geq 2$. In fact, the extensive survey Pemantle (2007) has only a small section with a few references regarding time dependent urn models. Time inhomogeneous urn models do have an extensive statistical literature due to the their wide usage in the experimental design of clinical trials (the idea being that it is ethical to favor experimental treatments that initially do well over those that initially do not); see Zhang, Hu and Cheung (2006), Zhang et al. (2011) and Bai, Hu and Zhang (2002). This literature is concerned with models and regimes where the asymptotic behavior is Gaussian. As discussed in Janson (2006a), it is difficult to characterize nonnormal asymptotic distributions of generalized Pólya urns, even in the time homogeneous case.

REMARK 1.7. There are many possible natural generalizations of the model we study here, such as starting with more than one black ball or adding more than one black ball every lth draw. We have restricted our study to the $\mathcal{P}_n^l(1, w)$ urn because these variations lead to asymptotic distributions outside the generalized gamma class. For example, the case $\mathcal{P}_n^1(b, w)$ with integer $b \geq 1$ is studied in Peköz, Röllin and Ross (2013a), where it is shown for $b \geq 2$ the limits are powers of products of independent beta and gamma random variables. Our main purpose here is to study the generalized gamma regime carefully and to highlight the connection between these urn models and random walks and trees.

1.3. *Applications to sub-tree sizes in uniform binary and plane trees.* Denote by T_n^p a uniformly chosen rooted plane tree with n nodes, and denote by T_{2n-1}^b a uniformly chosen binary, rooted plane tree with $2n-1$ nodes, that is, with n leaves and $n-1$ internal nodes. It is well known that the number of such trees in both cases is the Catalan number $C_{n-1} = \binom{2n-2}{n-1}/n$ and that both families of random trees are instances of simply generated trees; see Examples 10.1 and 10.3 of Janson (2012).

For any rooted tree T let $\mathrm{sp}^k_{\mathrm{Leaf}}(T)$ be the number of vertices in the minimal spanning tree spanned by the root and k randomly chosen distinct *leaves* of T, and let $\mathrm{sp}^k_{\mathrm{Node}}(T)$ be the number of vertices in the minimal spanning tree spanned by the root and k randomly chosen distinct *nodes* of T.

THEOREM 1.8. *Let $\mu_n(1,w)$ be as in (1.2) of Theorem 1.2. Then, for any $k \geq 1$,*

(i) $d_{\mathrm{K}}(\mathscr{L}(\mathrm{sp}^k_{\mathrm{Leaf}}(T^b_{2n-1})/\mu_{n-k-1}(1,2k)), \mathrm{GG}(2k,2)) = \mathrm{O}(n^{-1/2})$,

(ii) $d_{\mathrm{K}}(\mathscr{L}(\mathrm{sp}^k_{\mathrm{Node}}(T^b_{2n-1})/\mu_{n-k-1}(1,2k)), \mathrm{GG}(2k,2)) = \mathrm{O}(n^{-1/2})$,

(iii) $d_{\mathrm{K}}(\mathscr{L}(2\,\mathrm{sp}^k_{\mathrm{Node}}(T^p_n)/\mu_{n-k-1}(1,2k)), \mathrm{GG}(2k,2)) = \mathrm{O}(n^{-1/2}\log n)$.

REMARK 1.9. The logarithms in (iii) of the theorem and in (iii) and (iv) of the forthcoming Theorem 1.11 are likely an artifact of our analysis, specifically in the use of Lemma 2.5.

REMARK 1.10. The limits in the theorem can also be seen using facts about the Brownian continuum random tree (CRT) due to Aldous (1991, 1993). Indeed, the trees T^b_{2n-1} and T^p_n can be understood to converge in a certain sense to the Brownian CRT. The limit of the subtrees we study having k leaves can be defined through the Poisson line-breaking construction as described following Theorem 7.9 of Pitman (2006):

> Let $0 < \Theta_1 < \Theta_2 < \cdots$ be the points of an inhomogeneous Poisson process on $\mathbb{R}_{>0}$ of rate $t\,dt$. Break the line $[0,\infty)$ at points Θ_k. Grow trees $\mathcal{T}_k$ by letting $\mathcal{T}_1$ be a segment of length Θ_1, then for $k \geq 2$ attaching the segment $(\Theta_{k-1}, \Theta_k]$ as a "twig" attached at a random point of the tree $\mathcal{T}_{k-1}$ formed from the first $k-1$ segments.

The length of this tree is just Θ_k which is the generalized gamma limit of the theorem (up to a constant scaling). In more detail, if we jointly generate the vector $\mathbf{U}_k(n) := (\mathrm{sp}^1_{\mathrm{Leaf}}(T^b_{2n-1}), \dots, \mathrm{sp}^k_{\mathrm{Leaf}}(T^b_{2n-1}))$ by first selecting k leaves uniformly at random from T^b_{2n-1}, then labeling the selected leaves $1,\dots,k$, and then setting $\mathrm{sp}^i_{\mathrm{Leaf}}(T^b_{2n-1})$ to be the number of nodes in the tree spanned by the root and the leaves labeled $1,\dots,i$, then the CRT theory implies $n^{-1/2}\mathbf{U}_k(n)$ converges in distribution to $(\Theta_1,\dots,\Theta_k)$; see also Peköz, Röllin and Ross (2014) for a proof of this fact with a rate of convergence.

Panholzer [(2004), Theorem 6] provides local limit theorems for $\mathrm{sp}^k_{\mathrm{Node}}(T^b_{2n-1})$ and $\mathrm{sp}^k_{\mathrm{Node}}(T^p_n)$, from which the distributional convergence to the generalized gamma can be seen. It may be possible to obtain such (and other) local limits results using our Kolmogorov bounds and the approach of Röllin and Ross (2015), but in any case the convergence rates in the Kolmogorov metric in Theorem 1.8 appear to be new.

1.4. *Applications to occupation times and heights in random walk, bridge and meander.* 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

$$L_n = \sum_{i=0}^{n} \mathrm{I}[S_n(i) = 0]$$

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

$$L_{2n}^b \sim \mathscr{L}(L_{2n} | S_{2n}(0) = S_{2n}(2n) = 0)$$

be the local time of a random walk bridge, and define random walk excursion and meander by

$$S_{2n}^e \sim \mathscr{L}(S_{2n} | S_{2n}(0) = 0, S_{2n}(1) > 0, \dots, S_{2n}(2n-1) > 0, S_{2n}(2n) = 0),$$
$$S_n^m \sim \mathscr{L}(S_n | S_n(0) = 0, S_n(1) > 0, \dots, S_n(n) > 0).$$

THEOREM 1.11. *Let $\mu_n(1, w) = \mu_n$ be as in (1.2) of Theorem 1.2 and let K be uniformly distributed on $\{0, \dots, 2n\}$ and independent of $(S_{2n}^e(u))_{u=0}^{2n}$. Then*

(i) $d_{\mathrm{K}}(\mathscr{L}(L_n / \mu_{\lfloor n/2 \rfloor}(1, 1)), \mathrm{GG}(1, 2)) = \mathrm{O}(n^{-1/2})$,

(ii) $d_{\mathrm{K}}(\mathscr{L}(L_{2n}^b / \mu_{n-1}(1, 2)), \mathrm{GG}(2, 2)) = \mathrm{O}(n^{-1/2})$,

(iii) $d_{\mathrm{K}}(\mathscr{L}(2S_{2n}^e(K) / \mu_{n-2}(1, 2)), \mathrm{GG}(2, 2)) = \mathrm{O}(n^{-1/2} \log n)$,

(iv) $d_{\mathrm{K}}(\mathscr{L}(2S_n^m / \mu_{\lfloor (n-1)/2 \rfloor - 1}(1, 2)), \mathrm{GG}(2, 2)) = \mathrm{O}(n^{-1/2} \log n)$.

REMARK 1.12. An alternative viewpoint of the limits in Theorem 1.11 is that they are the analogous statistics of Brownian motion, bridge, meander and excursion which can be read from Chung (1976) and Durrett and Iglehart (1977); these Brownian fragments are the weak limits in the path space $C[0, 1]$ of the walk fragments we study; see Csáki and Mohanty (1981). For example, if B_t, $t \geq 0$, is a standard Brownian motion and $(L_t^x, t \geq 0, x \in \mathbb{R})$ its local time at level x up to time t, then Lévy's identity implies that L_1^0 is equal in distribution to the maximum of B_t up to time 1, which is equal in distribution to a half normal distribution; see also Borodin (1987).

To check the remaining limits of the theorem [which are Rayleigh, GG(2, 2)], we can use Pitman (1999), equation (1) [see also Borodin (1989)] which states that for $y > 0$ and $b \in \mathbb{R}$,

$$\begin{aligned} &\mathbb{P}[L_1^x \in dy, B_1 \in db] \\ &\quad = \frac{1}{\sqrt{2\pi}} (|x| + |b - x| + y) \exp\left(-\frac{1}{2}(|x| + |b - x| + y)^2\right) dy\, db. \end{aligned} \tag{1.4}$$

Roughly, for the local time of Brownian bridge at time 1 we set $b = x = 0$ in (1.4) and multiply by $\sqrt{2\pi}$ (due to conditioning $B_1 = 0$) to see the Rayleigh density. For the final time of Brownian meander, we set $x = y = 0$ in (1.4) and multiply by $\sqrt{\pi/2}$ (due to conditioning $L_1^0 = 0$), and note here that $b \in \mathbb{R}$ so by symmetry we restrict $b > 0$ and multiply by 2 to get back to the Rayleigh density. Finally, due to Vervaat's transformation [Vervaat (1979)], the height of standard Brownian excursion at a uniform random time has the same distribution as the maximum of Brownian bridge on [0, 1]. If we denote by M this maximum, then for $x > 0$ we apply (1.4) to obtain

$$\mathbb{P}[M > x] = \mathbb{P}[L_1^x > 0 | B_1 = 0] = \int_0^\infty (2x + y) \exp\left(-\frac{1}{2}(2x + y)^2\right) dy = e^{-2x^2},$$

which is the claimed Rayleigh distribution.

With the exception of the result for L_n, which can be read from Chung and Hunt (1949), inequality (1) or Döbler (2013), Theorem 1.2, the convergence rates appear to be new.

1.5. *A general approximation result via distributional transforms*. Theorem 1.2 follows from a general approximation result using Stein's method, a distributional transformation with a corresponding fixed point equation, which we describe now. We first generalize the size bias transformation used in Stein's method and appearing naturally in many places; see Arratia, Goldstein and Kochman (2013) and Brown (2006).

DEFINITION 1.13. Let $\beta > 0$ and let W be a nonnegative random variable with finite βth moment. We say a random variable $W^{(\beta)}$ has the *β-power bias distribution of W*, if

$$\mathbb{E}\{W^\beta f(W)\} = \mathbb{E}W^\beta \mathbb{E}f(W^{(\beta)}) \tag{1.5}$$

for all f for which the expectations exist.

In what follows, denote by $\mathrm{B}(a, b)$ the beta distribution with parameters $a, b > 0$.

DEFINITION 1.14. Let $\alpha > 0$ and $\beta > 0$ and let W be a positive random variable with $\mathbb{E}W^\beta = \alpha/\beta$. We say that W^* has the *(α, β)-generalized equilibrium distribution of W* if, for $V_\alpha \sim \mathrm{B}(\alpha, 1)$ independent of $W^{(\beta)}$, we have

$$W^* \stackrel{\mathscr{D}}{=} V_\alpha W^{(\beta)}. \tag{1.6}$$

REMARK 1.15. Pakes and Khattree (1992), Theorem 5.1 and Pitman and Ross (2012), Proposition 9 show that for a positive random variable W with $\mathbb{E}W^\beta = \alpha/\beta$, we have $W \sim \mathrm{GG}(\alpha, \beta)$ if and only if $W \stackrel{\mathscr{D}}{=} W^*$. The $(1, 2)$-generalized equilibrium distributional transformation is the nonnegative analog of the zero bias transformation of which the standard normal distributions are unique fixed points; see Chen, Goldstein and Shao (2011), Proposition 2.3, page 35, where the 2-power bias transformation is appropriately called "square" biasing; thus $\mathrm{GG}(1, 2)$ is the absolute normal distribution.

THEOREM 1.16. *Let W be a positive random variable with $\mathbb{E}W^\beta = \alpha/\beta$ for some integers $\alpha \geq 1$ and $\beta \geq 1$. Let W^* be a random variable constructed on the same probability space having the (α, β)-generalized equilibrium distribution of W. Then there is a constant $c > 0$ depending only on α and β such that, for all $0 < b \leq 1$,*

$$d_{\mathrm{K}}\bigl(\mathscr{L}(W), \mathrm{GG}(\alpha, \beta)\bigr) \leq c\bigl(b + \mathbb{P}\bigl[|W - W^*| > b\bigr]\bigr). \tag{1.7}$$

REMARK 1.17. Let X and Y be two random variables and let

$$d_{\mathrm{LP}}\bigl(\mathscr{L}(X), \mathscr{L}(Y)\bigr) = \inf\bigl\{b : \mathbb{P}[X \leq t] \leq \mathbb{P}[Y \leq t + b] + b \text{ for all } t \in \mathbb{R}\bigr\}$$

be the Lévy–Prokohorov distance between $\mathscr{L}(X)$ and $\mathscr{L}(Y)$. A theorem due to Strassen [see, e.g., Dudley (1968), Theorem 2] says that there is a coupling (X, Y) such that $\mathbb{P}[|X - Y| > \rho] \leq \rho$, where $\rho = d_{\mathrm{LP}}(\mathscr{L}(X), \mathscr{L}(Y))$. Hence, since (1.7) holds for *all* b and *all* couplings of W and W^*, it follows in particular that

$$d_{\mathrm{K}}\bigl(\mathscr{L}(W), \mathrm{GG}(k, r)\bigr) \leq 2c d_{\mathrm{LP}}\bigl(\mathscr{L}(W), \mathscr{L}\bigl(W^*\bigr)\bigr).$$

The paper is organized as follows. In Section 2, we embed our urn model into random trees via Rémy's algorithm and prove Theorem 1.8. In Section 3, we describe the various connections between trees and walk paths and then prove Theorem 1.11. In Section 4, we use Theorem 1.16 to prove Theorem 1.2, and finally in Section 5 we develop a general formulation of Stein's method for log concave densities and prove Theorem 1.16.

2. Random trees: Proof of Theorem 1.8.

2.1. *Rémy's algorithm for decorated binary trees.* Rémy (Rémy) introduced an elegant recursive algorithm to construct uniformly chosen decorated binary trees, where by "decorated" we mean that the leaves are labeled. This algorithm is the key ingredient to our approach as it relates to the urn schemes of Theorem 1.2. All trees are assumed to be plane trees throughout, and we will think of the tree as growing downward with the root at the top. We will refer to the "left" and "right" child of a node as seen from the readers point of view looking at the tree growing downward.

Rémy's algorithm for decorated binary trees (*see Figure* 1). Let $n \geq 1$ and assume that T^b_{2n-1} is a uniformly chosen decorated binary tree with n leaves, labeled from 1 to n. To obtain a uniformly chosen decorated binary tree T^b_{2n+1} with $n + 1$ leaves do the following:

Step 1. Choose a node uniformly at random; call it X. Remove X and its sub-tree, insert a new internal node at this position, call it Y, and attach X and its sub-tree, to Y.

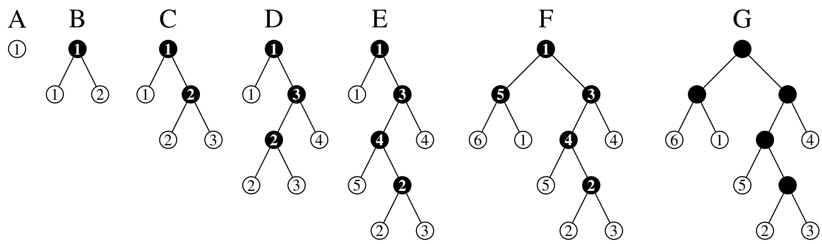


FIG. 1. *Illustration of Rémy's algorithm to construct decorated binary trees. Internal nodes are represented by black circles, and leaves by white circles. For the sake of clarity, we keep the internal nodes labeled, but these labels will be removed in the final step. We start with Tree A, the trivial tree. The step from Tree A to Tree B is* ①*L, where "*①*" indicates the node that was chosen, and "L" indicates that this node, along with its sub-tree, is attached to the new node as the left-child. Using this notation, the remaining steps to get to Tree F are* ②*L,* ❷*L,* ❷*R,* ①*R. Then remove the labels of the internal nodes to obtain Tree G, the final tree.*

Step 2. With probability $1/2$ each, do either of the following:

(a) Attach new leaf with label $n+1$ as the left-child to Y (making X the right-child of Y).

(b) Attach new leaf with label $n+1$ as the right-child to Y (making X the left-child of Y).

This recursive algorithm produces uniformly chosen decorated binary trees, since every decorated binary tree can be obtained in exactly one way, and since at every iteration every new tree is chosen with equal probability. By removing the labels, we obtain a uniformly chosen undecorated binary tree.

Figure 1 illustrates the algorithm by means of an example. We have labeled the internal nodes to make the procedure clearer, but it is important to note that these internal labels are not chosen uniformly among all such labelings and, therefore, have to be removed at the final step (to see this, note that Tree C in Figure 1 cannot be obtained through Rémy's algorithm if the labels of the two internal nodes are switched).

2.2. *Sub-tree sizes.*

Spanning trees in binary trees. Rémy's algorithm creates a direct embedding of a Pólya urn into a decorated binary tree. The following result is the key to our tree and walk results and is utilized via embeddings and bijections in this section and the following. The result is implicit in a construction of Pitman (2006), Exercise 7.4.11.

PROPOSITION 2.1. *For any $n \geq k \geq 1$,*

$$\mathrm{sp}^k_{\mathrm{Leaf}}(T^b_{2n-1}) \sim \mathcal{P}^1_{n-k}(0, 2k-1) = \mathcal{P}^1_{n-k-1}(1, 2k).$$

PROOF. Since the labeling is random, we may consider the tree spanned by the root and the leaves labeled 1 to k of a uniformly chosen decorated binary tree, rather than the tree spanned by the root and k uniformly chosen leaves of a random binary tree, cf. Pitman (2006), Exercise 7.4.11. Start with a uniformly chosen decorated binary tree T^b_{2k-1} with k leaves and note that the tree spanned by the root and leaves 1 to k is the whole tree. Now identify the $2k-1$ nodes of T^b_{2k-1} with $2k-1$ white balls in an urn that has no black balls. If the randomly chosen node in a given step in Rémy's algorithm is outside the current spanning tree, two nodes will be added outside the current spanning tree and we identify this as adding two black balls to the urn. If the randomly chosen node is in the current spanning tree, one node will be added to the current spanning tree and another outside of it, and we identify this as adding one black and one white ball to the urn.

Since we started with a tree of $2k-1$ nodes, we need $n-k$ steps to obtain a tree with $2n-1$ nodes. Hence, the size of the spanning tree is equal to the number of white balls in the urn, which follows the distribution $\mathcal{P}^1_{n-k}(0, 2k-1)$. $\square$

As a consequence of Proposition 2.1, whenever a quantity of interest can be coupled closely to $\mathrm{sp}^k_{\mathrm{Leaf}}(T^b_{2n-1})$, rates of convergence can be obtained if the closeness of the coupling can be quantified appropriately. In this section, we give two tree examples of this approach. Since the distribution $\mathcal{P}^1_{n-k}(0, 2k-1)$ will appear over and over again, we set $N^*_{j,k} \sim \mathcal{P}^1_j(0, 2k-1)$ in what follows. We use the notation $\mathrm{Ge}_0(p)$, $\mathrm{Ge}_1(p)$, $\mathrm{Be}(p)$, $\mathrm{Bi}(n, p)$ to, respectively, denote the geometric with supports starting at zero and one, Bernoulli and binomial distributions. For a nonnegative integer-valued random variable N, we also use the notation $X \sim \mathrm{Bi}(N, p)$ to denote that X is distributed as a mixture of binomial distributions such that $\mathscr{L}(X|N = n) = \mathrm{Bi}(n, p)$.

We now make a simple, but important observation about the edges in the spanning tree.

LEMMA 2.2. *Let $1 \le k \le n$ and T^b_{2n-1} be a uniformly chosen binary tree with n leaves and consider the tree spanned by the root and k uniformly chosen distinct leaves. Let $M_{k,n}$ be the number of edges in this spanning tree that connect a node to its left-child ("left-edges"). Conditional on the spanning tree having $N^*_{n-k,k}$ nodes,*

$$M_{k,n} - (k-1) \sim \mathrm{Bi}\bigl(N^*_{n-k,k} - (2k-1), 1/2\bigr). \tag{2.1}$$

PROOF. We use Rémy's algorithm and induction over n. Fix $k \ge 1$. For $n = k$ note that the spanning tree is the whole tree with $N^*_{0,k} = 2k-1$ nodes and $2(k-1)$ edges. Since half of the edges must connect a node to the left-child, $M_{k,k} = k-1$ which is (2.1) and this proves the base case. Assume now that (2.1) is true for some $n \ge k$. Two things can happen when applying Rémy's algorithm: either the current spanning tree is not changed, in which case $N^*_{n-k+1,k} = N^*_{n-k,k}$ and $M_{k,n+1} = M_{k,n}$, and hence (2.1) holds by the induction hypothesis, or one node and one edge are inserted into the spanning tree, in which case $N^*_{n-k+1,k} = N^*_{n-k,k} + 1$ and $M_{k,n+1} = M_{k,n} + J$ with $J \sim \mathrm{Be}(1/2)$ independent of all else. In the latter case, using the induction hypothesis, $M_{k,n} + J - (k-1) \sim \mathrm{Bi}(N^*_{n-k,k} - (2k-1) + 1, 1/2) = \mathrm{Bi}(N^*_{n-k+1,k} - (2k-1), 1/2)$, which is again (2.1). This concludes the induction step. $\square$

PROPOSITION 2.3. *Let $n \geq k \geq 1$ and let $N^*_{j,k} \sim \mathcal{P}^1_j(0, 2k-1)$. There exist nonnegative, integer-valued random variables $Y_1, \dots, Y_k$ such that, for each i,*

$$\mathbb{P}[Y_i > m | N^*_{n-k,k}] \leq 2^{-m} \qquad \text{for all } m \geq 0, \tag{2.2}$$

and such that for

$$X_{n,k} := N^*_{n-k,k} - \sum_{i=1}^{k} Y_k \tag{2.3}$$

we have

$$d_{\mathrm{TV}}\big(\mathscr{L}\big(\mathrm{sp}^k_{\mathrm{Node}}(T^b_{2n-1})\big), \mathscr{L}(X_{n,k})\big) \leq \frac{k}{2n} + \frac{(k-1)^2}{2n-k+1}, \tag{2.4}$$

where d_{TV} denotes total variation distance.

For $k = 1$ we have the explicit representation

$$\mathscr{L}\big(\mathrm{sp}^1_{\mathrm{Node}}(T^b_{2n-1})\big) = \mathscr{L}(X_{n,1} | X_{n,1} > 0), \tag{2.5}$$

where $Y_1 \sim \mathrm{Ge}_0(1/2)$ is independent of $N^*_{n-1,1}$.

PROOF. We first prove (2.5). We start by regarding T^b_{2n-1} as being "planted", that is, we think of the root node as being the left-child of a "ground node" (which itself has no right-child). We also think of the ground node as being internal. Furthermore, we think of the minimal spanning tree between the ground node and the root node as being empty, hence its size as being 0. We first construct a pairing between leaves and internal nodes as follows (see Figure 2). Pick a leaf and follow the path from that leaf toward the ground node and pair the leaf with the first parent of a left-child encountered in that path. Equivalently, pick an internal node and, in direction away from the ground node, first follow the left child of that internal node, and then keep following the right child until reaching a leaf. In particular, with this algorithm, if a selected leaf is a left-child it is assigned directly to its parent and the right-most leaf is assigned to the ground node. The fact that this description is indeed a pairing follows inductively by considering the left and right subtrees connected to the root, whereby the left subtree uses the root of the tree as its ground node.

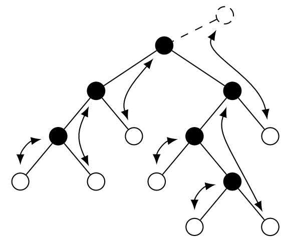


FIG. 2. *Pairing up leaves and internal nodes in planted binary trees. Note that the right-most leaf in the tree is paired up with the "ground node".*

Recall that we are considering the case $k = 1$. Now, instead of choosing a node uniformly at random among the $2n$ nodes of the planted tree (the ground node included), we may equivalently choose Leaf 1 with probability $1/2$, or choose the internal node paired with Leaf 1 with probability $1/2$. Denote by X_n the number of nodes in the path from the chosen node to the root, denote by J the indicator of the event that we choose an internal node, and denote by $N^*_{n-1,1}$ the number of nodes in the path from Leaf 1 up to the root. From Proposition 2.1 with $k = 1$, we have that $N^*_{n-1,1} \sim \mathcal{P}^1_{n-1}(0, 1)$. If $J_1 = 0$, then $X_{n,1} = N^*_{n-1,1}$. If $J_1 = 1$, the number of nodes in the path to the root is that of Leaf 1 minus the number of nodes until the first parent of a left-child in the path is encountered. Considering Lemma 2.2, given $N^*_{n-1,1}$, the number of left-edges are $N^*_{n-1,1}$ independent coin tosses with success probability $1/2$, hence, if $\tilde{Y}_1$ is the time until the first parent of a left-child is encountered, we have $\tilde{Y}_1 \sim \mathrm{Ge}_1(1/2)$, truncated at $N^*_{n-1,1}$. Thus, if $J_1 = 1$, we have $X_{n,1} = N^*_{n-1,1} - \tilde{Y}_1 \wedge N^*_{n-1,1}$. Putting the two cases together we obtain the representation $X_n = N^*_{n-1,1} - (J_1\tilde{Y}_1) \wedge N^*_{n-1,1}$, which has the same distribution as

$$N^*_{n-1,1} - Y_1 \wedge N^*_{n-1,1}, \tag{2.6}$$

since $J_1\tilde{Y}_1 \sim \mathrm{Ge}_0(1/2)$. As $X_{n,1}$ is zero if and only if the ground node was paired with Leaf 1 (i.e., Leaf 1 being the right most leaf) and $J_1 = 1$, conditioning on $X_{n,1}$ being positive is equivalent to conditioning on choosing any node apart from the ground node, which concludes (2.5).

Now, let k be arbitrary. In a first step, instead of choosing k distinct nodes at random, choose k distinct leaves at random and, for each leaf, toss a fair coin J_i, $i = 1, \dots, k$, to determine whether to choose the leaf or its internal partner, similar to the case $k = 1$. Denote by $N^*_{n-k,k}$ the number of nodes in the minimal spanning tree spanned by Leaves 1 to k and the root, and denote by $X_{n,k}$ the number of nodes in the minimal spanning tree spanned by the leaves or paired nodes and the root (if one of the chosen nodes is the ground node, then ignore that node in determining the minimal spanning tree). It is easy to see through two coupling arguments that choosing the nodes in this different way introduces a total variation error of at most

$$\left[1-\binom{2n-1}{k}\Big/\binom{2n}{k}\right]+\left[1-\prod_{i=1}^{k-1}\left(1-\frac{i}{2n-i}\right)\right];$$

the first term stems from the possibility of choosing the ground node, and the second term from restricting the k nodes to be from different pairings. From this, (2.4) easily follows.

It remains to show (2.2) and (2.3). For (2.3), for each $i=1,\dots,k$, let $N_{n-1}^{(i)}$ be the number of nodes in the path from leaf i up to the root, and let $Y_i'=J_i\tilde{Y}_i$ be the geometric random variable from the representation (2.6). With $Y_i=Y_i'\wedge N_{n-1}^{(i)}$, we hence have

$$X_{n,k}=N_{n-k,k}^{*}-\sum_{i=1}^{k}Y_i'\wedge N_{n-1}^{(i)}=N_{n-k,k}^{*}-\sum_{i=1}^{k}Y_i.$$

It is not difficult to check that Y_i and $N_{n-k,k}^{*}-N_{n-1}^{(i)}$ are conditionally independent given $N_{n-1}^{(i)}$. For (2.2), notice that $\mathbb{P}[Y_i>m|N_{n-1}^{(i)}]=\mathrm{I}[m<N_{n-1}^{(i)}]2^{-m}$. Hence,

$$\begin{aligned}\mathbb{P}[Y_i>m|N_{n-k,k}^{*}]&=\mathbb{E}\{\mathbb{P}[Y_i>m|N_{n-k,k}^{*},N_{n-1}^{(i)}-N_{n-k,k}^{*}]|N_{n-k,k}^{*}\}\\&=\mathbb{E}\{\mathbb{P}[Y_i>m|N_{n-1}^{(i)}]|N_{n-k,k}^{*}\}\le 2^{-m}.\end{aligned}$$ □

Uniform plane tree. It is well known that there are $n!C_{n-1}$ decorated binary trees of size $2n-1$ as well as labeled plane trees of size n nodes, where $C_1,C_2,\dots$ are the Catalan numbers. There are various ways to describe bijections between the two sets. We first give a direct algorithm to construct a plane tree from a binary tree; see Figure 3.

Given a binary tree, we do a depth-first exploration, starting from the root and exploring left-child before right-child. We construct the plane tree as we explore

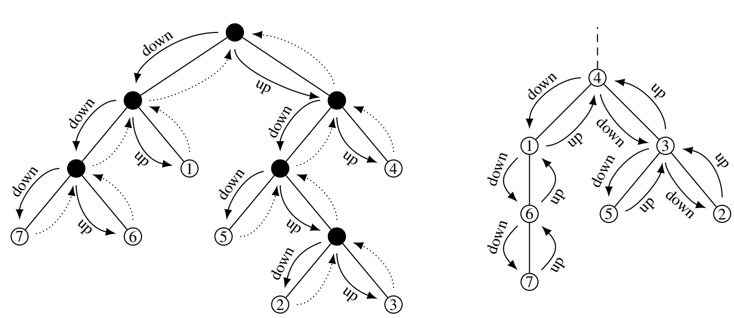


FIG. 3. *Bijection between a decorated binary tree of size* $2n-1$ (*on the left*), *and a rooted labeled plane tree of size* n (*on the right*).

the binary tree, starting with an unlabeled root node. Whenever a left-edge in the binary tree is visited for the first time, we add one new unlabeled child to the current node in the plane tree to the right of all existing children of that node, and move to that new child. If a right-edge is visited for the first time, we move back to the parent of the current node in the plane tree. Whenever we encounter a leaf in the binary tree, we copy that label to the node in the plane tree.

Another way to describe the bijection, initially described between unlabeled objects, is by means of *Dyck paths* of length $2(n-1)$. These are syntactically valid strings of $n-1$ nested bracket pairs. To go from a Dyck path to a binary tree, we parse the string from left-to-right and at the same time do a depth-first construction of the binary tree. Start with one active node. Any opening bracket corresponds to adding a left-child to the currently active node and then making that child the active node, whereas a closing bracket corresponds to adding a right-child as sibling of the left-child that belongs to the opening bracket of the current closing bracket, and then making that right child the active node. The labeling is added by inserting $n-1$ of the n leaf labels in front of the $n-1$ closing brackets, as well as one label at the end of the string in any of the $n!$ possible orderings. When converting the labeled Dyck path into a binary tree, every time a label is encountered that label is copied to the currently active node in the tree. The Dyck path corresponding to the tree in Figure 3 would be "((()))(()())", respectively, with the labeling, "(((7)6)1)((5)(2)3)4".

To obtain a labeled plane tree from a labeled Dyck path, again do a depth-first construction, starting with one active node. An opening bracket corresponds to adding a new child to the currently active node to the right of all already present siblings and then making that child the new active node, whereas a closing bracket represents making the parent of the currently active node the new active node. If a label is encountered in the string, the label is copied to the currently active node.

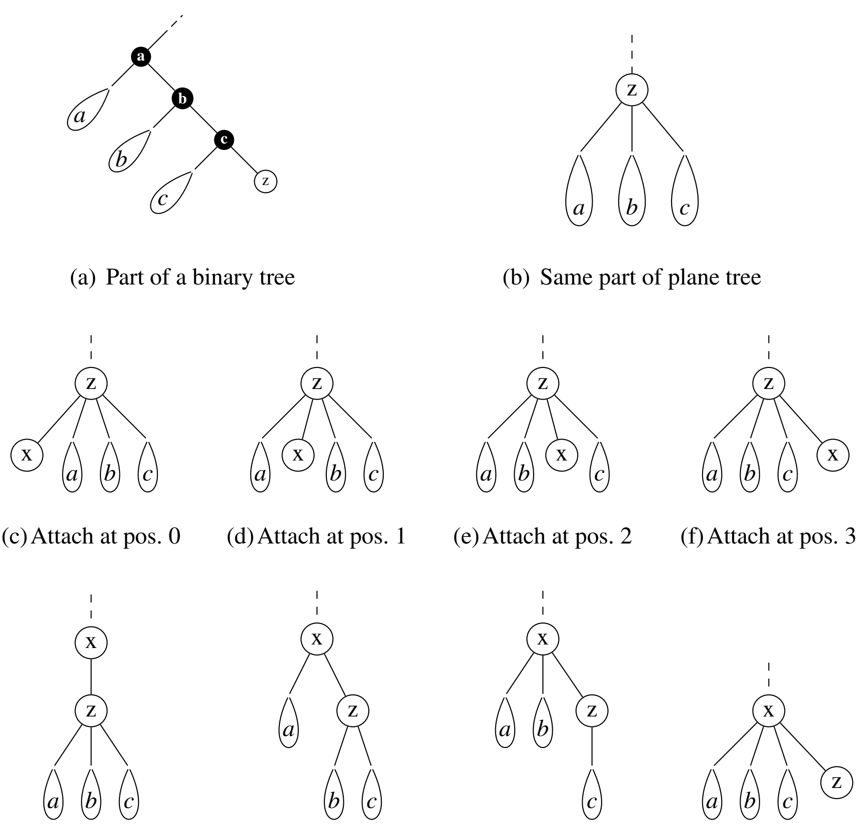
PROPOSITION 2.4. *Let $n \geq k \geq 1$ and $N^*_{j,k} \sim \mathcal{P}^1_j(0, 2k-1)$. Assume that $X_{n,k} \sim \mathrm{Bi}(N^*_{n-k,k} - (2k-1), 1/2)$. Then*

$$\mathrm{sp}^k_{\mathrm{Node}}(T^p_n) \stackrel{\mathscr{D}}{=} X_{n,k} + k.$$

PROOF. We use the bijection between binary and plane trees. The number of edges in the spanning tree of k nodes in the plane tree is equal to the number of left-edges in the spanning tree of the corresponding k leaves in the binary tree (note that in the spanning tree of the binary tree, we count left-edges both between internal nodes as well as between internal nodes and leaves). This is because only left-edges in the binary tree contribute to the number of edges in the plane tree. The proof is now a simple consequence of Proposition 2.1 and Lemma 2.2 and the fact that the number of nodes in any spanning tree is equal to one plus the number of edges in that spanning tree. $\square$

It is illuminating to see how Rémy's algorithm acts on plane trees by means of the bijection described above (see Figure 4). Apart from adding new edges to existing nodes, we also observe an operation that "cuts" existing nodes. The trees T_n^p and T_{2n-1}^b are special cases of Galton–Watson trees (respective offspring distributions geometric and uniform on $\{0, 2\}$) conditioned to have n and $2n-1$ nodes, respectively. As noted by Janson (2006c), such conditioned trees cannot in general be grown by only adding edges. Hence, it is tempting to speculate whether there is a wider class of offspring distributions for which conditional Galton–Watson trees can be grown using only local operations on trees such as those in Figure 4.

Before proceeding with the proof of Theorem 1.8, we need an auxiliary lemma used to transfer rates from our urn model to the distributions in Propositions 2.3 and 2.4. Here and below $\|\cdot\|$ denotes the essential supremum norm.



(a) Part of a binary tree

(b) Same part of plane tree

(c) Attach at pos. 0

(d) Attach at pos. 1

(e) Attach at pos. 2

(f) Attach at pos. 3

(g) Cut at pos. 0

(h) Cut at pos. 1

(i) Cut at pos. 2

(j) Cut at pos. 3

FIG. 4. *Rémy's algorithm acting on plane trees by means of the bijection given in Figure 3. We leave it to the reader to find the operations in the binary tree as given in* (a) *that correspond to the operations* (c)–(j).

LEMMA 2.5. *Let $\alpha \geq 1$ and $\beta \geq 1$. There is a constant $C = C_{\alpha,\beta}$, such that for any positive random variable X and any real-valued random variable ξ,*

$$\begin{aligned} &d_{\mathrm{K}}\big(\mathscr{L}(X+\xi), \mathrm{GG}(\alpha,\beta)\big) \\ &\qquad \leq C\big(d_{\mathrm{K}}\big(\mathscr{L}(X), \mathrm{GG}(\alpha,\beta)\big) + \big\|\mathbb{E}(\xi^2|X)\big\|^{1/2}\big). \end{aligned} \tag{2.7}$$

If X and ξ satisfy

$$\mathbb{P}\big[|\xi| \geq t|X\big] \leq c_1 e^{-c_2 t^2/X} \tag{2.8}$$

for some constants $c_1 > 0$ and $c_2 > 1$, then

$$\begin{aligned} &d_{\mathrm{K}}\big(\mathscr{L}(X+\xi), \mathrm{GG}(\alpha,\beta)\big) \\ &\qquad \leq C\left(d_{\mathrm{K}}\big(\mathscr{L}(X), \mathrm{GG}(\alpha,\beta)\big) + \frac{1+c_1+\log c_2}{\sqrt{c_2}}\right). \end{aligned} \tag{2.9}$$

PROOF. The proofs of (2.7) and (2.9) follow along the lines of the proof of Lemma 1 of Bolthausen (1982). Once one observes that $\mathrm{GG}(\alpha,\beta)$ has bounded density, the modifications needed to prove (2.7) are straightforward, and hence omitted. The modifications to prove (2.9), however, are more substantial, hence we give a complete proof for this case. Let $Z \sim \mathrm{GG}(\alpha,\beta)$, and let

$$\begin{aligned} F(t) &= \mathbb{P}[X \leq t], & F^*(t) &= \mathbb{P}[X+\xi \leq t], \\ G(t) &= \mathbb{P}[Z \leq t], & \delta &= \sup_{t>0}\big|F(t) - G(t)\big|. \end{aligned}$$

If $t > \varepsilon > 0$, then

$$\begin{aligned} F^*(t) &= \mathbb{E}\big\{\mathbb{P}[\xi \leq t - X|X]\big\} \geq \mathbb{E}\big\{\mathrm{I}[X \leq t-\varepsilon]\mathbb{P}[\xi \leq t - X|X]\big\} \\ &= F(t-\varepsilon) - \mathbb{E}\big\{\mathrm{I}[X \leq t-\varepsilon]\mathbb{P}[\xi > t - X|X]\big\}. \end{aligned}$$

Let $t_0 = \log c_2$ and $\varepsilon = \frac{\log c_2}{\sqrt{c_2}}$, and observe that, since $c_2 > 1$, we have $t_0 > \varepsilon > 0$. Also note that one can find a constant c_3 such that $1 - G(t) \leq c_3 e^{-t/2}$. Using (2.8) and setting $M_{\alpha,\beta}$ the maximum of the density of $\mathrm{GG}(\alpha,\beta)$ (defined explicitly in Lemma 5.12 below),

$$\begin{aligned} &\mathbb{E}\big\{\mathrm{I}[X \leq t-\varepsilon]\mathbb{P}[\xi > t - X|X]\big\} \\ &\qquad \leq \mathbb{E}\big\{\mathrm{I}[X \leq t\wedge t_0 - \varepsilon]\mathbb{P}[\xi > t\wedge t_0 - X|X]\big\} + \mathbb{P}[X > t_0 - \varepsilon] \\ &\qquad \leq c_1\mathbb{E}\big\{\mathrm{I}[X \leq t\wedge t_0 - \varepsilon]e^{-c_2(t\wedge t_0 - X)^2/X}\big\} + \mathbb{P}[Z > t_0 - \varepsilon] + \delta \\ &\qquad \leq c_1 e^{-c_2\varepsilon^2/t_0} + \mathbb{P}[Z > t_0] + \delta + \varepsilon M_{\alpha,\beta} \\ &\qquad \leq c_1 e^{-\log c_2} + \delta + \frac{c_3 + M_{\alpha,\beta}\log c_2}{\sqrt{c_2}} \leq \delta + \frac{c_1 + c_3 + M_{\alpha,\beta}\log c_2}{\sqrt{c_2}}. \end{aligned}$$

Therefore,

$$\begin{aligned}F^*(t) - G(t) &\geq F(t-\varepsilon) - G(t-\varepsilon) - \varepsilon M_{\alpha,\beta} - \delta - \frac{c_1 + c_3 + M_{\alpha,\beta}\log c_2}{\sqrt{c_2}} \\ &\geq -2\delta - \frac{c_1 + c_3 + 2M_{\alpha,\beta}\log c_2}{\sqrt{c_2}}.\end{aligned}$$

On the other hand,

$$F^*(t) \leq F(t+\varepsilon) + \mathbb{E}\{\mathrm{I}[t+\varepsilon < X \leq t_0]\mathbb{P}[\xi \leq t - X|X]\} + \mathbb{P}[X > t_0].$$

Since

$$\begin{aligned}&\mathbb{E}\{\mathrm{I}[t+\varepsilon < X \leq t_0]\mathbb{P}[\xi \leq t - X|X]\} \\ &\qquad \leq c_1\mathbb{E}\{\mathrm{I}[t+\varepsilon < X \leq t_0]e^{-c_2(t-X)^2/X}\} \leq c_1 e^{-c_2\varepsilon^2/t_0} \leq \frac{c_1}{c_2}\end{aligned}$$

and $\mathbb{P}[X > t_0] \leq \delta + c_3/\sqrt{c_2}$, by a similar reasoning as above,

$$\begin{aligned}F^*(t) - G(t) &\leq F(t+\varepsilon) + G(t-\varepsilon) + \varepsilon M_{\alpha,\beta} + \delta + \frac{c_1 + c_3}{\sqrt{c_2}} \\ &\leq 2\delta + \frac{c_1 + c_3 + M_{\alpha,\beta}\log c_2}{\sqrt{c_2}}.\end{aligned}$$

Hence,

$$|F^*(t) - G(t)| \leq 2\delta + \frac{c_1 + c_3 + 2M_{\alpha,\beta}\log c_2}{\sqrt{c_2}}. \tag{2.10}$$

From this, one easily obtains (2.9). □

PROOF OF THEOREM 1.8. *Case* (i). This follows directly from Proposition 2.1 and (1.1) of Theorem 1.2.

Case (ii). Let $W_n = \mathrm{sp}^k_{\mathrm{Node}}(T^b_{2n-1})/\nu_n$ with $\nu_n = \mu_{n-k-1}(1, 2k)$, let $X_{n,k}$ be as in Proposition 2.3. Applying the triangle inequality, we obtain

$$\begin{aligned}&d_{\mathrm{K}}(\mathscr{L}(W_n), \mathrm{GG}(2k, 2)) \\ &\qquad \leq d_{\mathrm{K}}(\mathscr{L}(W_n), \mathscr{L}(X_{n,k}/\nu_n)) + d_{\mathrm{K}}(\mathscr{L}(X_{n,k}/\nu_n), \mathrm{GG}(2k, 2)).\end{aligned} \tag{2.11}$$

Since the total variation distance is an upper bound on the Kolmogorov distance, (2.4) yields that the first term in (2.11) is of order $\mathrm{O}(n^{-1})$. To bound the second term in (2.11), let $N^*_{n-k,k}$ and $Y_1, \ldots, Y_k$ be as in Proposition 2.3; set $X := N^*_{n-k,k}/\nu_n$ and $\xi := (Y_1 + \cdots + Y_k)/\nu_n$. From (2.2) and recalling that $(\sum_{i=1}^k Y_k)^2 \leq k\sum_{i=1}^k Y_i^2$, it is easy to see that $\mathbb{E}(\xi^2|X) \leq 6k/\nu_n$ almost surely. Applying (2.7) from Lemma 2.5, we hence obtain that

$$d_{\mathrm{K}}(\mathscr{L}(X_{n,k}/\nu_n), \mathrm{GG}(2k, 2)) \leq C(d_{\mathrm{K}}(\mathscr{L}(N^*_{n-k,k}/\nu_n), \mathrm{GG}(2k, 2)) + \nu_n^{-1/2}).$$

Combining this with Theorem 1.2 and (2.11), the claim follows.

Case (iii). Let $N^*_{n-k,k}$ and $X_{n,k}$ be as in Proposition 2.4 and let again $\nu_n = \mu_{n-k-1}$. We may consider $2X_{n,k}/\mu_n$ in place of $2\,\mathrm{sp}^k_{\mathrm{Node}}(T^p_n)/\nu_n$, since by (1.2) of Theorem 1.2, the constant shift $2k/\nu_n$ is of order $n^{-1/2}$, which, by Lemma 2.5, translates into an error of order at most $n^{-1/2}$. Let $X := N^*_{n-k,k}/\nu_n$ and $\xi := (2X_n - N^*_{n-k,k})/\nu_n$ and note that $2X_n/\nu_n = X + \xi$. From Chernoff's inequality, it follows that (2.8) holds with $c_1 = 2$ and $c_2 = \nu_n^2/4$. For n large enough, $c_2 > 1$ [again using (1.2)] and applying (2.9) from Lemma 2.5 and (1.1) from Theorem 1.2, the claim follows. □

3. Random walk: Proof of Theorem 1.11. That random walks and random trees are intimately connected has been observed in many places; see, for example, Aldous (1991) and Pitman (2006). The specific bijections between binary trees and random walk, excursion, bridge and meander which we will make use of were sketched by Marchal (2003) and see also the references therein. It is clear that for each such bijection Rémy's algorithm can be translated to recursively create random walk, excursion, bridge and meander of arbitrary lengths.

Random walk excursion. The simplest bijection is that between a binary tree of size $2n-1$ and a (positive) random walk excursion of length $2n$, as illustrated in Figure 5. Note first that the first and last step of the excursion must be $+1$ and -1, respectively, that is, $S^e_{2n}(1) = S^e_{2n}(2n-1) = 1$. To map the tree to the path from 1 to $2n-1$, we do a left-to-right depth-first exploration of the tree (i.e., counterclockwise): starting from the root, each time an edge is visited the first time (out of a total of two times that each edge is visited), the excursion will go up by one if the edge is a left-edge and go down by one if the edge is a right-edge. By means of the Dyck path representation of the binary tree, we conclude that in this exploration process, the number of explored left edges ("opening brackets") is always larger than the number of explored right edges ("closing brackets"), hence

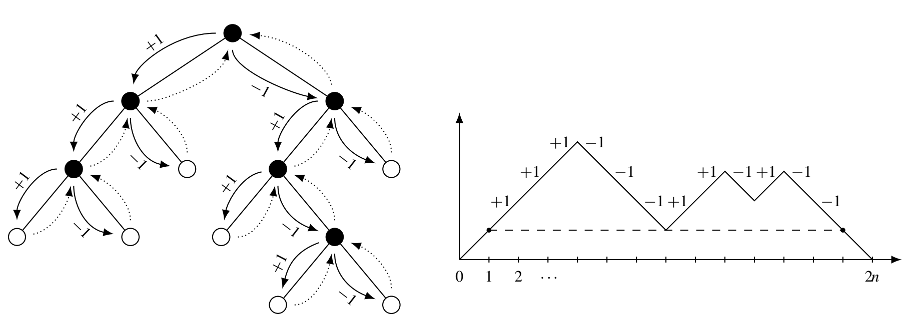


FIG. 5. *Illustration of the bijection between a binary tree with n leaves (on the left), and random walk excursions of length* $2n$ *(on the right).*

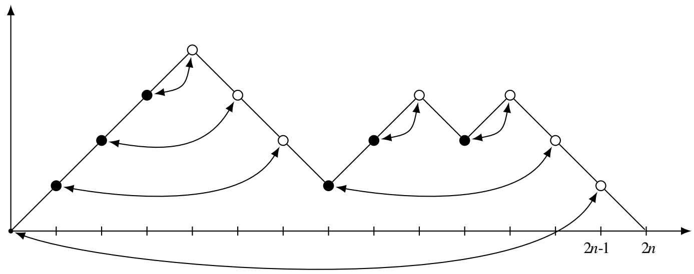


FIG. 6. *Pairing up the points in the random walk excursion. Note that we pair up time point* $2n-1$ *with time point* 0, *whereas time point* $2n$ *is left without a partner.*

the random walk stays positive. Furthermore, since the number of left- and right-edges is equal, the final height is the same as the starting height. It is not hard to see that the height of a time point in the excursion corresponds to one plus the number of left-edges from the corresponding point in the binary tree up to the root.

Furthermore, the pairing between leaves and internal nodes in the (planted) binary tree induces a pairing between the time points in the random walk excursion (the pairing in Figure 2, by means of the bijection in Figure 5, results in the pairing in Figure 6). Note that all time points can be paired except for the final time point $2n$ for which, however, we know the height.

PROPOSITION 3.1 (Height of an excursion at a random time). *If* $n \geq 1$, $N_{n-1}^* \sim \mathcal{P}_{n-1}^1(0,1)$ *and* $K' \sim \mathrm{U}\{0,1,\dots,2n-1\}$ *independent of* N_{n-1}^* *and the excursion* $(S_{2n}^e(u))_{u=0}^{2n}$, *then*

$$S_{2n}^e(K') \sim \mathrm{Bi}(N_{n-1}^*, 1/2).$$

PROOF. Mapping the pairing of leaves and internal nodes from the planted binary tree to the excursion, we have that the heights in each pair differ by exactly one because, by definition of the pairing, each leaf has one more left edge in its path up to the root as compared to the internal node it is paired with.

Let $J \sim \mathrm{Be}(1/2)$ independent of all else. Instead of choosing a random time point K', we may as well choose with probability $1/2$ the time point corresponding to Leaf 1 ($J=0$), and choose with probability $1/2$ the time point paired with the time point given by Leaf 1 ($J=1$). Recall that the height of a time point corresponding to a leaf is just one plus the number of left-edges $M_{1,n}$ in the path to the root in the corresponding binary tree. From Lemma 2.2 with $k=1$, we have $M_{1,n} \sim \mathrm{Bi}(N_{n-1}^*-1, 1/2)$. Let X_n be the height of the excursion at the time point corresponding to the node chosen in the binary tree; we have $X_n = 1 + M_{1,n} - J$. Since J is independent of the tree and since $1-J \sim \mathrm{Be}(1/2)$, we have $X_n \sim \mathrm{Bi}(N_{n-1}^*, 1/2)$, which proves the claim. $\square$

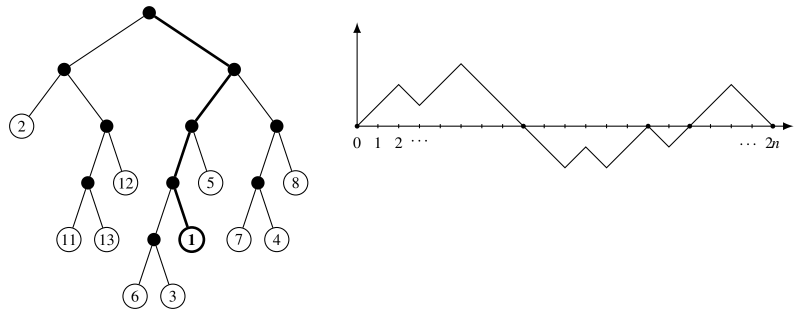


FIG. 7. *Illustration of the bijection between a decorated binary tree of size* $2n+1$ *with a spine and random walk bridge of length* $2n$. *Note that within sub-trees that grow to the left of the spine, the depth-first exploration is done counterclockwise, whereas within sub-trees that grow to the right it is done clockwise.*

Random walk bridge. We now discuss the bijection between decorated binary trees and random walk bridges; see Figure 7 for an example. We first mark the path from Leaf 1 to the root. We call all the internal nodes along this path, including the root, the *spine* (the trivial tree of size one has no internal node and, therefore, an empty spine). As before, a left edge represents "$+1$" and a right-edge represents "-1". The exploration starts at the root. Whenever a spine node is visited, explore first the child (and its subtree) that is not part of the spine, and then the child that is next in the spine. Also, if the right child of a spine node is being explored and if that child is not itself a spine node do the exploration *clockwise*, until the exploration process is back to the spine. This makes each sub-tree to the left of the spine a positive excursion and each sub-tree to the right a negative excursion; cf. Pitman (2006), Exercise 7.4.14.

PROPOSITION 3.2 (Occupation time of bridge). *If* $n \geq 0$, *then*

$$L^b_{2n} \sim \mathcal{P}^1_n(0,1).$$

PROOF. The proof is straightforward by observing that the number of visits to the origin L^b_{2n} is exactly the number of nodes in the path from Leaf 1 to the root and then applying Proposition 2.1 with $k=1$ and n replaced by $n+1$. □

Random walk meander. We use a well-known bijection between random walk bridges of length $2n$ and meanders of length $2n+1$; see Figure 8. Start the meander with one positive step. Then, follow the absolute value of the bridge, except that the last step of every negative excursion is flipped. Alternatively, consider the random walk bridge difference sequence. Leave all the steps belonging to positive excursions untouched, and multiply all steps belonging to negative excursions by -1, except for the last step of each respective negative excursion (which must necessarily be a "$+1$"). Now, start the meander with one positive step and then follow the new difference sequence.

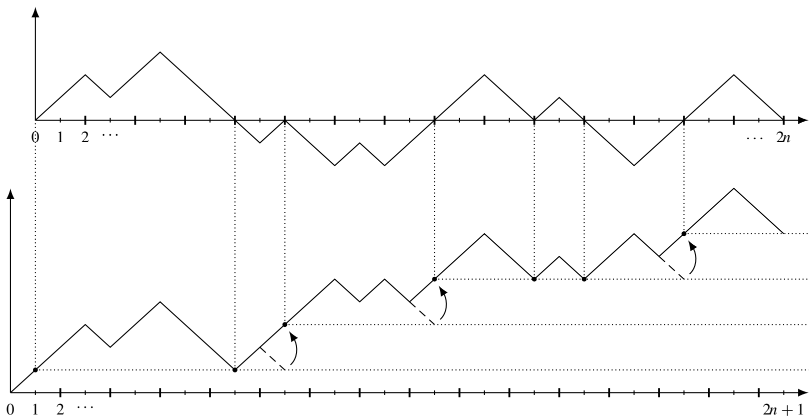


FIG. 8. *Illustration of the bijection between a random walk bridge of length* $2n$ (*above*) *and a meander of length* $2n+1$ (*below*).

PROPOSITION 3.3 (Final height of meander). *If* $n \geq 0$, $N_n^* \sim \mathcal{P}_n^1(0,1)$, $X_n \sim \mathrm{Bi}(N_n^* - 1, 1/2)$ *and* $Y_n \sim \mathrm{Bi}(N_n^*, 1/2)$, *then*

$$S_{2n+1}^m(2n+1) \sim \mathscr{L}(2X_n+1), \qquad S_{2n+2}^m(2n+2) \sim \mathscr{L}(2Y_n | Y_n > 0).$$

PROOF. It is clear that every negative excursion in the random walk will increase the final height of the meander by two. Since the number of negative excursions equals the number of left-edges in the spine of the corresponding binary tree, the first identity follows directly from Lemma 2.2 for $k=1$. To obtain a meander of length $2n+2$, proceed as follows. First, consider a meander of length $2n+1$, let $2X_n+1$ be its final height, and then add one additional time step to the meander by means of an independent fair coin toss. The resulting process is a simple random walk, conditioned to be positive from time steps 1 to $2n+1$. The height of this process at time $2n+2$ has distribution $2Y_n$, where we can take $Y_n = X_n + J$ and where $J \sim \mathrm{Be}(1/2)$ independent of X_n. However, the final height of this process may now be zero. Hence, conditioning on the path being positive results in a meander of length $2n+2$. This proves the second identity. □

Unconditional random walk. In order to represent an unconditional random walk of length $2n+1$, we use two decorated binary trees, the first tree representing the bridge part of the random walk (i.e., the random walk until the last return to the origin) and the second tree representing the meander part (i.e., the random walk after the last return to the origin); see Figure 9. Note that every random walk of odd length has a meander part. First, with equal probability, start either with the two trivial trees ① and ⊕ or with the two trivial trees ① and ⊖ [representing the random walk S_1 with $S_1(1) = 1$, resp., $S_1(1) = -1$]. Then, perform Rémys

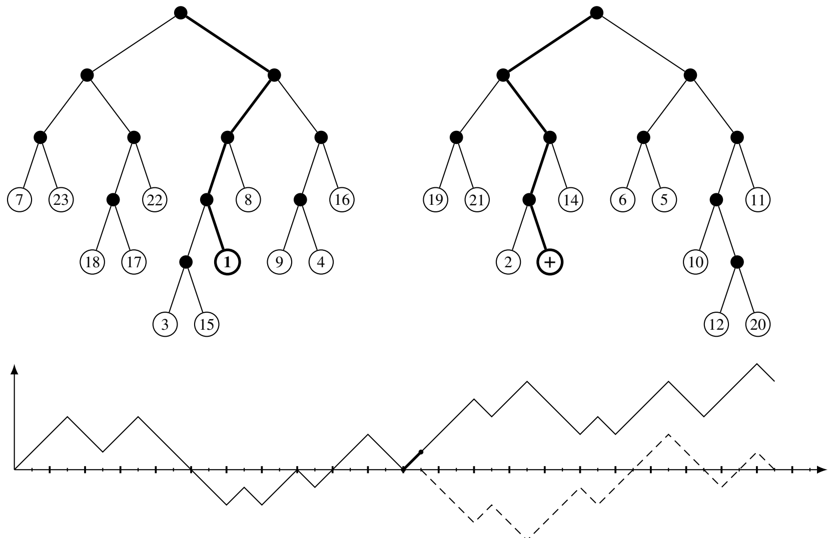


FIG. 9. *Bijection between a pair of decorated binary trees with a total size of* $2n + 2$ *and an unconditional random walk of length* $2n + 1$. *The meander part of the walk is constructed through a random walk bridge, which is plotted in dashed lines.*

algorithm in exactly the same way as for a single tree. That is, at each time step, a random node is chosen uniformly among all nodes of the two trees and then an internal node as well as a new leaf are inserted. From these two trees, the random walk is constructed in a straight forward manner: the first tree represents the bridge part, whereas the second tree represents the meander part (if the initial second tree was $\ominus$, then the whole meander is first constructed as illustrated in Figure 8 and is then flipped to become negative).

PROPOSITION 3.4 (Occupation time of random walk). *If* $n \geq 0$, *then*

$$L_{2n} \sim \mathcal{P}_n^1(1, 1), \qquad L_{2n+1} \sim \mathcal{P}_n^1(1, 1).$$

PROOF. Note that the number of visits to the origin is exactly the number of nodes in the path from Leaf 1 (which is always in the first tree) to the root. Hence, we can use a similar urn embedding as for Proposition 2.1 with $k = 1$, except that at the beginning the urn contains one black ball and one white ball (the black ball representing the leaf of the second tree).

This proves the second identity of the proposition. To obtain the first identity, take a random walk of length $2n + 1$ and remove the last time step, obtaining a random walk of length $2n$. Since the number of visits to the origin cannot be changed in this way, the first identity follows. $\square$

REMARK 3.5. Proposition 3.2 is implicitly used in Pitman (2006), Exercise 7.4.14. The other propositions do not appear to have been stated explicitly in the literature.

PROOF OF THEOREM 1.11. Cases (i) and (ii) are immediate from Theorem 1.2 in combination with Proposition 3.4 and Proposition 3.2, respectively. Using Proposition 3.1, case (iii) is proved in essentially the same way as case (iii) of Theorem 1.8, also noting that the total variation error introduced by using K instead of K' is of order $\mathrm{O}(n^{-1})$. Using Proposition 3.3, case (iv) for odd n is also proved in essentially the same way as case (iii) of Theorem 1.8.

In order to prove case (iv) for even n, note that the total variation distance between $\mathscr{L}(Y_n)$ and $\mathscr{L}(Y_n|Y_n > 0)$ is $\mathbb{P}[Y_n = 0] = \mathbb{E}2^{-N_n^*}$. Let $Z \sim \mathrm{GG}(2, 2)$; using Theorem 1.2,

$$\begin{aligned}\mathbb{E}2^{-N_n^*} &\le \mathbb{P}\big[N_n < \tfrac{1}{2}\log_2 n\big] + 2^{-1/2\log_2 n}\\ &\le \mathbb{P}\big[Z < \tfrac{1}{2}\mu_n^{-1}\log_2 n\big] + d_{\mathrm{K}}\big(\mathscr{L}(N_n/\mu_n), \mathscr{L}(Z)\big) + n^{-1/2} = \mathrm{O}\big(n^{-1/2}\big).\end{aligned}$$

Now, estimating $d_{\mathrm{K}}(\mathscr{L}(2Y_n), \mathrm{GG}(2, 2))$ again follows the proof of case (iii) of Theorem 1.8. $\square$

4. Proof of urn Theorem 1.2. In order to prove Theorem 1.2, we need a few lemmas.

LEMMA 4.1. *Let $b \ge 0$, $w > 0$, $N_n = N_n(b, w) \sim \mathcal{P}_n^l(b, w)$ and let $n_i = n_i(b, w) = w + b + i + \lfloor i/l \rfloor$ be the total number of balls in the $\mathcal{P}_n^l(b, w)$ urn after the ith draw. If $m \ge 1$ is an integer and $D_{n,m}(b, w) := \prod_{i=0}^{m-1}(i + N_n(b, w))$, then*

$$\mathbb{E}D_{n,m}(b, w) = \prod_{j=0}^{m-1}(w + j)\prod_{i=0}^{n-1}\big(1 + m/n_i(b, w)\big) \tag{4.1}$$

and for some positive values $c := c(b, w, l, m)$ and $C := C(b, w, l, m)$ not depending on n we have

$$cn^{ml/(l+1)} < \mathbb{E}\big[N_n(b, w)^m\big] < Cn^{ml/(l+1)}. \tag{4.2}$$

PROOF. Fix b, w and write $D_{n,m} = D_{n,m}(b, w)$. We first prove (4.1). Conditioning on the contents of the urn after draw and replacement $n - 1$, and noting that at each step, the number of white balls in the urn either stay the same or increase by exactly one, we have

$$\begin{aligned}\mathbb{E}\{D_{n,m}|N_{n-1}\} &= \frac{N_{n-1}}{n_{n-1}}\frac{D_{n-1,m}(N_{n-1} + m)}{N_{n-1}} + \frac{n_{n-1} - N_{n-1}}{n_{n-1}}D_{n-1,m}\\ &= (1 + m/n_{n-1})D_{n-1,m},\end{aligned}$$

which when iterated yields (4.1).

By the definition of n_i,

$$i + w + b - 1 + i/l \le n_i \le i + w + b + i/l,$$

and now setting $x = l/(l+1)$ and $y = (w+b-1)l/(l+1)$, we find for some constants c, C not depending on n that

$$cn^{mx} \le c\frac{\Gamma(mx+y+x+n)}{\Gamma(y+x+n)} \le \mathbb{E}D_{n,m} \le C\frac{\Gamma(mx+y+n)}{\Gamma(y+n)} \le Cn^{mx}. \tag{4.3}$$

The upper bound follows from this and the easy fact that $\mathbb{E}N_n^m \le \mathbb{E}D_{n,m}$. The lower bound follows from (4.3) and the following inequality which follows from Jensen's inequality $\mathbb{E}N_n^m = \mathbb{E}D_{n,1}^m \ge (\mathbb{E}D_{n,1})^m$. □

Our next result implies that biasing the distribution $\mathcal{P}_n^l(b, w)$ against the r rising factorial is the same as adding r white balls to the urn before starting the process, and then removing r white balls at the end. We will only use the lemma for $r = l+1$, but state and prove it for general r because it is an interesting result in its own right.

LEMMA 4.2. *Let $N_n(b, w)$ and $D_{n,m}(b, w)$ be as in Lemma 4.1 and let $r \ge 2$. If $N_n^{[r]} = N_n^{[r]}(b, w)$ is a random variable such that*

$$\mathbb{P}[N_n^{[r]} = k] = \frac{[\prod_{i=0}^{r-1}(k+i)]\mathbb{P}[N_n(b, w) = k]}{\mathbb{E}D_{n,r}(b, w)}, \tag{4.4}$$

then

$$N_n(b, w+r) \stackrel{\mathscr{D}}{=} N_n^{[r]}(b, w) + r. \tag{4.5}$$

PROOF. Since $N_n(b, w+r)$ and $N_n^{[r]}(b, w) + r$ are bounded variables, the lemma follows by verifying their factorial moments are equal. With $n_i(b, w)$ as in Lemma 4.1, for any $m \ge 1$ we have

$$\begin{aligned}
\mathbb{E}\prod_{i=0}^{m-1}\bigl(N_n^{[r]}(b, w) + r + i\bigr) &= \frac{\mathbb{E}D_{n,m+r}(b, w)}{\mathbb{E}D_{n,r}(b, w)} \\
&= \prod_{j=0}^{m-1}(w+r+j)\prod_{i=1}^{n}\frac{n_{i-1}(b, w) + m + r}{n_{i-1}(b, w) + r} \\
&= \mathbb{E}D_{n,m}(b, w+r) = \mathbb{E}\prod_{i=0}^{m-1}\bigl(i + N_n(b, w+r)\bigr);
\end{aligned}$$

the second and third equalities follow by (4.1) and the definition of $n_i(b, w)$, and the last follows from the definition of $D_{n,m}(b, w)$. □

LEMMA 4.3. *For $N_n(1,w) \sim \mathcal{P}_n^l(1,w)$ and $l \geq 1$, there is a coupling of $N_n^{(l+1)}(1,w)$, a random variable having the $(l+1)$-power bias distribution of $N_n(1,w)$, with a variable $N_{n-l}(1,w+l+1) \sim \mathcal{P}_{n-l}^l(1,w+l+1)$ such that for some constant $C := C(w,l)$,*

$$\mathbb{P}\big[\big|N_{n-l}(1,w+l+1) - N_n^{(l+1)}(1,w)\big| > 2l+1\big] \leq Cn^{-l/(l+1)}.$$

PROOF. Obviously, we can couple $N_n(1,w+l+1) \sim \mathcal{P}_n^l(1,w+l+1)$ with $N_{n-l}(1,w+l+1)$ so that

$$\big|N_{n-l}(1,w+l+1) - N_n(1,w+l+1)\big| \leq l,$$

and then Lemma 4.2 implies that we may couple $N_n(1,w+l+1)$ with $N_n^{[l+1]}(1,w)$ [with distribution defined at (4.4)] so that almost surely

$$\begin{aligned}
&\big|N_{n-l}(1,w+l+1) - N_n^{[l+1]}(1,w)\big| \\
&\quad \leq \big|N_{n-l}(1,w+l+1) - \big(N_n^{[l+1]}(1,w)+l+1\big)\big| + l+1 \\
&\quad = \big|N_{n-l}(1,w+l+1) - N_n(1,w+l+1)\big| + l + 1 \leq 2l+1.
\end{aligned}$$

And we show

$$d_{\mathrm{TV}}\big(\mathscr{L}\big(N_n^{[l+1]}(1,w)\big), \mathscr{L}\big(N_n^{(l+1)}(1,w)\big)\big) \leq Cn^{-l/(l+1)}, \tag{4.6}$$

where d_{TV} is the total variation distance, which for integer-valued variables X and Y can be defined in two ways:

$$d_{\mathrm{TV}}\big(\mathscr{L}(X), \mathscr{L}(Y)\big) = \frac{1}{2}\sum_{z\in\mathbb{Z}}\big|\mathbb{P}[X=z] - \mathbb{P}[Y=z]\big| = \inf_{(X,Y)} \mathbb{P}[X \neq Y];$$

here, the infimum is taken over all possible couplings of X and Y. Due to the latter definition, (4.6) will imply the lemma since

$$\begin{aligned}
&\mathbb{P}\big[\big|N_{n-l}(1,w+l+1) - N_n^{(l+1)}(1,w)\big| > 2l+1\big] \\
&\quad = \mathbb{P}\big[\big|N_{n-l}(1,w+l+1) - N_n^{(l+1)}(1,w)\big| \\
&\qquad > 2l+1, N_n^{[l+1]}(1,w) \neq N_n^{(l+1)}(1,w)\big] \\
&\quad \leq \mathbb{P}\big[N_n^{[l+1]}(1,w) \neq N_n^{(l+1)}(1,w)\big].
\end{aligned}$$

Let $\nu_m = \mathbb{E}N_n^m(1,w)$ and note that we can write $\prod_{i=0}^l (x+i) = \sum_{i=0}^{l+1} a_i x^i$ for nonnegative coefficients a_i with $a_{l+1} = 1$ (these coefficients are the unsigned Stirling numbers). Also note that for nonnegative integers k and $0 \leq i \leq l+1$, we have $k^i \leq k^{l+1}$, and hence $\nu_i \leq \nu_{l+1}$. Thus,

$$\begin{aligned}
&2d_{\mathrm{TV}}\big(\mathscr{L}\big(N_n^{[l+1]}(1,w)\big), \mathscr{L}\big(N_n^{(l+1)}(1,w)\big)\big) \\
&\quad = \sum_{k\geq 0}\big|\mathbb{P}\big[N_n^{[l+1]}(1,w) = k\big] - \mathbb{P}[N_n^{(l+1)}(1,w) = k)\big|
\end{aligned}$$

$$
\begin{aligned}
&= \sum_k \left| \frac{\prod_{i=0}^l (k+i)}{\mathbb{E}D_{n,l+1}(1,w)} - \frac{k^{l+1}}{\nu_{l+1}} \right| \mathbb{P}\bigl[N_n(1,w)=k\bigr] \\
&= \sum_k \left| \left(k^{l+1} + \sum_{i=0}^{l} a_i k^i \right) \nu_{l+1} - k^{l+1} \left(\nu_{l+1} + \sum_{i=0}^{l} a_i \nu_i \right) \right| \frac{\mathbb{P}[N_n(1,w)=k]}{\nu_{l+1}\mathbb{E}D_{n,l+1}(1,w)} \\
&\le C\nu_l / \mathbb{E}D_{n,l+1}(1,w) \le C n^{-l/(1+l)},
\end{aligned}
$$

where the last line follows from (4.2) of Lemma 4.1. This proves the lemma. □

Below let $\mathcal{P}_n(b,w)$ be the distribution of the number of white balls in the classical Pólya urn started with b black balls and w white balls after n draws. Recall that in the classical Pólya urn balls are drawn and returned to the urn along with an additional ball of the same color [the notation is to suggest $\mathcal{P}_n^\infty(b,w) = \mathcal{P}_n(b,w)$].

LEMMA 4.4. *There is a coupling* $(Q_w(n), nV_w)_{n\ge 1}$ *with* $Q_w(n) \sim \mathcal{P}_n(1,w)$ *and* $V_w \sim \mathrm{B}(w,1)$ *such that* $|Q_w(n) - nV_w| \le w+1$ *for all* n *almost surely.*

PROOF. Using Feller (1968), equation (2.4), page 121, for $w \le t \le w+n$ we obtain

$$
\mathbb{P}\bigl[Q_w(n) \le t\bigr] = \prod_{i=0}^{w-1} \frac{t-i}{n+w-i}. \tag{4.7}
$$

For $U_0, U_1, \ldots, U_{w-1}$ i.i.d. uniform (0, 1) variables, we may set

$$
Q_w(n) = \max_{i=0,1,\ldots,w-1} \bigl(i + \lceil (n+w-i)U_i \rceil\bigr),
$$

since it is not difficult to verify that this gives the same cumulative distribution function as in (4.7). By a well-known representation of the beta distribution, we can take $V_w = \max(U_0, \ldots, U_{w-1})$, and with this coupling the claim follows. □

LEMMA 4.5. *If* $N_n(0, w+1) \sim \mathcal{P}_n^l(0, w+1)$ *then*

$$
\mathcal{P}_n^l(1,w) = \mathcal{P}_{N_n(0,w+1)-w-1}(1,w).
$$

PROOF. Consider an urn with 1 black ball and w white balls. Balls are drawn from the urn and replaced as follows. After the mth ball is drawn, it is replaced in the urn along with another ball of the same color plus, if m is divisible by l, an additional green ball. If H is the number of times a nongreen ball is drawn in n draws, the number of white balls in the urn after n draws is distributed as $\mathcal{P}_H(1,w)$. The lemma follows after noting $H+w+1$ is distributed as $\mathcal{P}_n^l(0,w+1)$ [which by definition is the distribution of $N_n(0,w+1)$] and the number of white balls in the urn after n draws has distribution $\mathcal{P}_n^l(1,w)$. □

PROOF OF THEOREM 1.2. The asymptotic $\mathbb{E}N_n^k \asymp n^{kl/(l+1)}$ is (4.2) of Lemma 4.1. We now show that

$$\lim_{n\to\infty} \frac{\mathbb{E}N_n^{l+1}}{n^l} = w\left(\frac{l+1}{l}\right)^l.$$

The asymptotic $\mathbb{E}N_n^k \asymp n^{kl/(l+1)}$ implies that

$$\frac{\mathbb{E}N_n^{l+1}}{n^l} = \frac{\mathbb{E}\prod_{i=0}^{l}(i+N_n)}{n^l} + \mathrm{o}(1).$$

The numerator in the fraction on the right-hand side of the equality can be written using (4.1) from Lemma 4.1 with $b=1$, $w=w$ and $m=l+1$ as

$$\mathbb{E}\prod_{i=0}^{l}(i+N_n) = \frac{\Gamma(w+l+1)}{\Gamma(w)} \prod_{i=0}^{n-1+\lfloor (n-1)/l\rfloor} \frac{w+1+i+l+1}{w+1+i} \times \prod_{k=1}^{\lfloor (n-1)/l\rfloor} \frac{w+1+kl+k-1}{w+1+kl+k+l},$$

and simplifying, especially noting the telescoping product in the final part of the term (which critically depends on having taken $m=l+1$), we have

$$\begin{aligned}\mathbb{E}\prod_{i=0}^{l}(i+N_n) &= \frac{\Gamma(w+l+1)}{\Gamma(w)} \frac{\Gamma(w+2+l+n+\lfloor (n-1)/l\rfloor)\Gamma(w+1)}{\Gamma(w+l+2)\Gamma(w+1+n+\lfloor (n-1)/l\rfloor)} \\ &\quad \times \frac{w+1+l}{w+l+1+\lfloor (n-1)/l\rfloor(l+1)} \\ &= w\frac{\Gamma(w+1+l+n+\lfloor (n-1)/l\rfloor)}{\Gamma(w+1+n+\lfloor (n-1)/l\rfloor)} \\ &\quad \times \frac{w+1+l+n+\lfloor (n-1)/l\rfloor}{w+l+1+\lfloor (n-1)/l\rfloor(l+1)}.\end{aligned}$$

The asymptotic for $\mathbb{E}N_n^{l+1}$ now follows by taking the limit as $n\to\infty$, using the well-known fact that, for $a>0$, $\lim_{x\to\infty}\frac{\Gamma(x+a)}{\Gamma(x)x^a}=1$ with $x=w+1+n+\lfloor\frac{n-1}{l}\rfloor$.

The claimed asymptotic for μ_n follows directly from that of $\mathbb{E}N_n^{l+1}$, and with the order of the scaling μ_n in hand, the lower bound of Theorem 1.2 follows from Peköz, Röllin and Ross (2013a), Lemma 4.1, which says that for a sequence of scaled integer valued random variables (a_nN_n), if $a_n\to 0$ and ν is a distribution with density bounded away from zero on some interval, then there is a positive constant c such that $d_{\mathrm{K}}(\mathscr{L}(a_nN_n),\nu)\ge ca_n$.

To prove the upper bound we will invoke Theorem 1.16 and so we want to closely couple variables having marginal distributions equal to those of N_n/μ_n

and $N^* = V_w N_n^{(l+1)}/\mu_n$. Lemma 4.4 implies there is a coupling of variables $(Q_w(n))_{n\geq 1}$ with corresponding marginal distributions $(\mathcal{P}_n(1,w))_{n\geq 1}$ satisfying

$$\bigl|V_w N_n^{(l+1)} - Q_w\bigl(N_n^{(l+1)}\bigr)\bigr| \leq w+1 \qquad \text{almost surely.}$$

Further, by Lemma 4.3 we can construct a variable $N_{n-l}(1, w+l+1) \sim \mathcal{P}_{n-l}^{l}(1, w+l+1)$ such that

$$\mathbb{P}\bigl[\bigl|Q_w\bigl(N_{n-l}(1, w+l+1)\bigr) - Q_w\bigl(N_n^{(l+1)}\bigr)\bigr| > 2l+1\bigr] \leq Cn^{-l/(l+1)};$$

here we used that $|Q_w(s) - Q_w(t)| \leq |s-t|$. Recalling that $P_{n-l}^l(1, w+l+1) = P_n^l(0, w+1)$, Lemma 4.5 says that we can set $N_n = Q_w(N_{n-l}(1, w+l+1) - w - 1)$ and it is immediate that

$$\bigl|Q_w\bigl(N_{n-l}(1, w+l+1) - w - 1\bigr) - Q_w\bigl(N_{n-l}(1, w+l+1)\bigr)\bigr| \leq w+1 \qquad \text{almost surely.}$$

Thus, if we set $b = (2w+2l+3)/\mu_n$ then using the couplings above we find

$$\mathbb{P}\bigl[\bigl|N_n/\mu_n - V_w N_n^{(l+1)}/\mu_n\bigr| > b\bigr] \leq C\mu_n^{-1} \leq Cn^{-l/(l+1)},$$

where the last inequality follows from (4.2) of Lemma 4.1 which also implies $b \leq Cn^{-l/(l+1)}$. Using these couplings and the value of b in Theorem 1.16 completes the proof. □

5. Stein's method and proof of Theorem 1.16. We first provide a general framework to develop Stein's method for log-concave densities. The generalized gamma is a special case of this class. We use the density approach which is due to Charles Stein [see Reinert (2005)]. This approach has already been discussed in other places in greater generality; see, for example, Chatterjee and Shao (2011), Chen, Goldstein and Shao (2011) and Döbler (2012). However, it seems to have gone unnoticed, at least explicitly, that the approach can be developed much more directly for log-concave densities.

5.1. *Density approach for* log-*concave distributions*. Let B be a function on the interval (a,b) where $-\infty \leq a < b \leq \infty$. Assume also B is absolutely continuous on (a,b), $C_B = \int_a^b e^{-B(z)}\,dz < \infty$ and $B(a) := \lim_{x\to a^+} B(x)$ and $B(b) := \lim_{x\to b^-} B(x)$ exist as values in $\mathbb{R}\cup\{\infty\}$ and we use these to extend the domain of B to $[a,b]$. Assume that B has a left-continuous derivative on (a,b), denoted by B'. From B, we can construct a distribution P_B with probability density function

$$\varphi_B(x) = C_B e^{-B(x)}, \qquad a < x < b \qquad \text{where } C_B^{-1} = \int_a^b e^{-B(z)}\,dz.$$

Let $L^1(P_B)$ be the set of measurable functions h on (a,b) such that

$$\int_a^b |h(x)| e^{-B(x)}\,dx < \infty.$$

The distribution P_B is log-concave if and only if B is convex. However, before dealing with this special case, we state a few more general results.

PROPOSITION 5.1. *If $Z \sim P_B$, we have*

$$\mathbb{E}\{f'(Z) - B'(Z)f(Z)\} = 0$$

for all functions f for which the expectations exists and for which

$$\lim_{x \to a^+} f(x)e^{-B(x)} = \lim_{x \to b^-} f(x)e^{-B(x)} = 0.$$

PROOF. Integration by parts. We omit the straightforward details. □

Now, for $h \in L^1(P_B)$ and $Z \sim P_B$, let

$$\tilde{h}(x) = h(x) - \mathbb{E}h(Z)$$

and, for $x \in (a, b)$,

$$f_h(x) = e^{B(x)} \int_a^x \tilde{h}(z)e^{-B(z)}\, dz = -e^{B(x)} \int_x^b \tilde{h}(z)e^{-B(z)}\, dz. \tag{5.1}$$

The key fact is that f_h satisfies the differential (Stein) equation

$$f_h'(x) - B'(x) f_h(x) = \tilde{h}(x), \qquad x \in (a, b). \tag{5.2}$$

Define the Mills's-type ratios

$$\kappa_a(x) = e^{B(x)} \int_a^x e^{-B(z)}\, dz, \qquad \kappa_b(x) = e^{B(x)} \int_x^b e^{-B(z)}\, dz. \tag{5.3}$$

From (5.1) and (5.2), we can easily deduce the following nonuniform bounds.

LEMMA 5.2. *If $h \in L^1(P_B)$ is bounded, then for all $x \in (a, b)$,*

$$|f_h(x)| \le \|\tilde{h}\|\big(\kappa_a(x) \wedge \kappa_b(x)\big), \tag{5.4}$$

$$|f_h'(x)| \le \|\tilde{h}\|\big\{1 + |B'(x)|\big(\kappa_a(x) \wedge \kappa_b(x)\big)\big\}. \tag{5.5}$$

In the case of convex functions, we can easily adapt the proof of Stein (1986) to obtain the following uniform bounds.

LEMMA 5.3. *If B is convex on (a, b) with unique minimum $x_0 \in [a, b]$, then for any $h \in L^1(P_B)$,*

$$\|f_h\| \le \|\tilde{h}\| \frac{e^{B(x_0)}}{C_B}, \qquad \|B' f_h\| \le \|\tilde{h}\|, \qquad \|f_h'\| \le 2\|\tilde{h}\|. \tag{5.6}$$

PROOF. By convexity, we clearly have

$$x_0 \le x \le z \le b \quad \Longrightarrow \quad B(x) \le B(z) \quad \text{and} \quad B'(x) \le B'(z). \tag{5.7}$$

This implies that for $x > x_0$

$$\int_x^b e^{-B(z)}\,dz \le \int_x^b \frac{B'(z)}{B'(x)} e^{-B(z)}\,dz = \frac{e^{-B(x)} - e^{-B(b)}}{B'(x)} \le \frac{e^{-B(x)}}{B'(x)},$$

where in the last bound we use (5.7) which implies $B'(x) > 0$. So

$$B'(x)\kappa_b(x) \le 1. \tag{5.8}$$

Now, from this we have for $x > x_0$

$$\kappa_b'(x) = -1 + B'(x)\kappa_b(x) \le 0.$$

Similarly, we have

$$a \le z \le x \le x_0 \quad \Longrightarrow \quad B(z) \ge B(x) \quad \text{and} \quad \left|B'(z)\right| \ge \left|B'(x)\right|. \tag{5.9}$$

So, using (5.9), for $x < x_0$,

$$\int_a^x e^{-B(z)}\,dz \le \int_a^x \frac{|B'(z)|}{|B'(x)|} e^{-B(z)}\,dz = \frac{e^{-B(x)} - e^{-B(a)}}{|B'(x)|} \le \frac{e^{-B(x)}}{|B'(x)|},$$

thus

$$\left|B'(x)\right|\kappa_a(x) \le 1, \tag{5.10}$$

and so for $x < x_0$

$$\kappa_a'(x) = 1 + B'(x)\kappa_a(x) \ge 0.$$

From (5.4), we obtain

$$\|f\| \le \|\tilde{h}\| \sup_x \begin{cases} \kappa_a(x), & \text{if } x < x_0, \\ \kappa_b(x), & \text{if } x \ge x_0. \end{cases}$$

Hence, having an increasing bound on $x < x_0$ and a decreasing bound on $x > x_0$, implies that there is a maximum at x_0 and

$$\|f\| \le \|\tilde{h}\| \bigl(\kappa_a(x_0) \vee \kappa_b(x_0)\bigr).$$

The first bound of (5.6) now follows from the fact that $\kappa_a(x_0) \vee \kappa_b(x_0) \le \kappa_a(x_0) + \kappa_b(x_0)$. The second bound of (5.6) follows from (5.4) in combination with (5.8) and (5.10). Using (5.5), the third bound of (5.6) follows in the same way. $\square$

REMARK 5.4. Lemma 5.3 applies to the standard normal distribution in which case $B(x) = x^2/2$, $x_0 = 0$, and $C_B = (2\pi)^{-1/2}$ and (5.6) implies

$$\|f_h\| \le \|\tilde{h}\|\sqrt{2\pi}, \qquad \|f_h'\| \le 2\|\tilde{h}\|.$$

The best-known bounds are given in Chen, Goldstein and Shao (2011), Lemma 2.4, which improve the first bound by a factor of 2 and match the second. In the special case of the form $h(\cdot) = \mathrm{I}[\cdot \le t]$, Chen, Goldstein and Shao (2011), Lemma 2.3, matches the bound of Lemma 5.3 of $|xf_h(x)| \le \|\tilde{h}\|$.

Though not used below explicitly, we record the following theorem summarizing the utility of the lemmas above.

THEOREM 5.5. *Let B be convex on (a,b) with unique minimum x_0, $Z \sim P_B$, and W be a random variable on (a,b). If $\mathcal{F}$ is the set of functions on (a,b) such that for $f \in \mathcal{F}$*

$$\|f\| \le \frac{e^{B(x_0)}}{C_B}, \qquad \|B'f\| \le 1, \qquad \|f'\| \le 2,$$

then

$$\sup_{t\in(a,b)} \bigl|\mathbb{P}[Z \le t] - \mathbb{P}[W \le t]\bigr| \le \sup_{f\in\mathcal{F}} \bigl|\mathbb{E}\{f'(W) - B'(W)f(W)\}\bigr|.$$

PROOF. For $t \in (a,b)$, if $h_t(x) = \mathrm{I}[x \le t]$, then taking the expectation in (5.2) implies that

$$\mathbb{P}[W \le t] - \mathbb{P}[Z \le t] = \mathbb{E}\{f_t'(W) - B'(W)f_t(W)\}, \tag{5.11}$$

where f_t satisfies (5.2) with $h = h_t$. Taking the absolute value and the supremum over $t \in (a,b)$ on both sides of (5.11), we find

$$\sup_{t\in(a,b)} \bigl|\mathbb{P}[W \le t] - \mathbb{P}[Z \le t]\bigr| = \sup_{t\in(a,b)} \bigl|\mathbb{E}\{f_t'(W) - B'(W)f_t(W)\}\bigr|.$$

The result follows since $h_t(x) \in [0,1]$ implies $\|\tilde{h}\| \le 1$, and so by Lemma 5.3, $f_t \in \mathcal{F}$ for all $t \in (a,b)$. □

Finally, we will need the following two lemmas to develop Stein's method. The proofs are standard, and can be easily adopted from the normal case; see, for example, Chen and Shao (2005) and Raič (2003).

LEMMA 5.6 (Smoothing inequality). *Let B be convex on (a,b) with unique minimum x_0 and let $Z \sim P_B$. Then, for any random variable W taking values in (a,b) and for any $\varepsilon > 0$, we have*

$$d_{\mathrm{K}}\bigl(\mathscr{L}(W), \mathscr{L}(Z)\bigr) \le \sup_{a<s<b} \bigl|\mathbb{E}h_{s,\varepsilon}(W) - \mathbb{E}h_{s,\varepsilon}(Z)\bigr| + C_B e^{-B(x_0)}\varepsilon,$$

where

$$h_{s,\varepsilon}(x) = \frac{1}{\varepsilon}\int_0^{\varepsilon} \mathrm{I}[x \le s+u]\,du. \tag{5.12}$$

LEMMA 5.7 (Bootstrap concentration inequality). *Let B be convex on (a,b) with unique minimum x_0 and let $Z \sim P_B$. Then, for any random variable W taking values in (a,b), for any $a < x < b$, and for any $\varepsilon > 0$, we have*

$$\mathbb{P}[s \le W \le s+\varepsilon] \le C_B e^{-B(x_0)}\varepsilon + 2d_{\mathrm{K}}\bigl(\mathscr{L}(W), \mathscr{L}(Z)\bigr).$$

5.2. *Application to the generalized gamma distribution.* We use the general results of Section 5.1 to prove the following more explicit statement of Theorem 1.16 for the generalized gamma distribution.

THEOREM 5.8. *Let $Z \sim \mathrm{GG}(\alpha, \beta)$ for some $\alpha \geq 1$, $\beta \geq 1$ and let W be a nonnegative random variable with $\mathbb{E}W^\beta = \alpha/\beta$. Let W^* have the (α, β)-generalized equilibrium transformation of Definition 1.14. If $\beta = 1$ or $\beta \geq 2$, then for all $0 < b \leq 1$,*

$$
\begin{aligned}
&d_{\mathrm{K}}\bigl(\mathscr{L}(W), \mathscr{L}(Z)\bigr)\\
&\quad\leq b\bigl[10M_{\alpha,\beta} + 2\beta(\beta-1)\bigl(1 + 2^{\beta-2}(\mathbb{E}W^{\beta-1} + b^{\beta-1})\bigr)M'_{\alpha,\beta} + 4\beta\mathbb{E}W^{\beta-1}\bigr]\\
&\qquad + 4\bigl(2 + (\beta+\alpha-1)M'_{\alpha,\beta}\bigr)\mathbb{P}\bigl[|W - W^*| > b\bigr],
\end{aligned}
$$

where here and below

$$
\begin{aligned}
M_{\alpha,\beta} &:= \alpha^{1-1/\beta}\beta^{1/\beta}e^{-4/9+1/(6((\alpha-1)/\beta)+9/4)}\biggl(2\frac{\alpha-1}{\beta} + 1\biggr)^{-1/2}\\
&\leq e^{1/e}\alpha^{1-1/\beta}\biggl(2\frac{\alpha-1}{\beta} + 1\biggr)^{-1/2},\\
M'_{\alpha,\beta} &:= \sqrt{2\pi}e^{-1/(6((\alpha-1)/\beta)+9/4)}\biggl(\frac{\alpha-1}{\beta} + 1/2\biggr)^{1/2}\biggl(\frac{\alpha-1}{\beta} + 1\biggr)^{1/\beta}\alpha^{-1}\\
&\leq \sqrt{2\pi}\biggl(\frac{\alpha-1}{\beta} + 1/2\biggr)^{1/2}\biggl(\frac{\alpha-1}{\beta} + 1\biggr)^{1/\beta}\alpha^{-1}.
\end{aligned}
$$

If $1 < \beta < 2$, then for all $0 < b \leq 1$,

$$
\begin{aligned}
&d_{\mathrm{K}}\bigl(\mathscr{L}(W), \mathscr{L}(Z)\bigr)\\
&\quad\leq b\bigl(10M_{\alpha,\beta} + 4\beta\mathbb{E}W^{\beta-1}\bigr) + 2\beta b^{\beta-1}M'_{\alpha,\beta}\\
&\qquad + 4\bigl(2 + (\beta+\alpha-1)M'_{\alpha,\beta}\bigr)\mathbb{P}\bigl[|W - W^*| > b\bigr].
\end{aligned}
$$

REMARK 5.9. For a given α and β, the constants in the theorem may be sharpened. For example, the case $\alpha = \beta = 1$ of the theorem is the exponential approximation result (2.5) of Theorem 2.1 of Peköz and Röllin (2011), but here with larger constants. These larger constants come from three sources: first, below we bound some maximums of nonnegative numbers by sums for the sake of simple formulas (only if all but one of the terms in the maximum is positive is there any hope of optimality in the constants). Second, $M_{\alpha,\beta}$ and $M'_{\alpha,\beta}$ arise from bounds on the generalized gamma density, achieved by using both sides of the inequalities in Theorems 5.10 and 5.11 below. These inequalities are not optimal at the same value for each side, so some precision could be gained by using the appropriate exact bounds on the density which in principle are recoverable from the work below, but not particularly informative. Finally, in special cases more information about the Stein solution may be obtained. For example, in Peköz and Röllin (2011) the term $|g(W) - g(W^*)|$ that appears in the proof of Theorem 5.8 is there bounded by 1, whereas following Lemma 5.16, our general bound specializes to $2\|g\| \le 4.3$.

In the notation of Section 5.1, for the generalized gamma distribution we have $\varphi_{\alpha,\beta}(x) = Ce^{-B(x)}$, $x > 0$ with $a = 0$ and $b = \infty$, and

$$B(x) = x^\beta - (\alpha - 1)\log x, \qquad C = \frac{\beta}{\Gamma(\alpha/\beta)}.$$

If $\alpha \ge 1$ and $\beta \ge 1$, then B has nonnegative second derivative and is thus convex. Since

$$B'(x) = \beta x^{\beta-1} - \frac{(\alpha - 1)}{x},$$

B has a unique minimum at $x_0 = (\frac{\alpha-1}{\beta})^{1/\beta}$. Hence,

$$B(x_0) = \psi\left(\frac{\alpha - 1}{\beta}\right) \qquad \text{with } \psi(x) = x - x\log(x),\ \psi(0) = 0,$$

and

$$Ce^{-B(x_0)} = Ce^{-\psi((\alpha-1)/\beta)}. \tag{5.13}$$

In order to apply Lemmas 5.6 and 5.7, we need to bound (5.13), for which we use the following two results about the gamma function.

THEOREM 5.10 [Batir (2008), Corollary 1.2]. *For all* $x \ge 0$,

$$\sqrt{2}e^{4/9} \le \frac{\Gamma(x+1)}{x^x e^{-x-1/(6x+9/4)}\sqrt{x+1/2}} \le \sqrt{2\pi}.$$

THEOREM 5.11 [Wendel (1948), (7)]. *If* $x > 0$ *and* $0 \le s \le 1$, *then*

$$\left(\frac{x}{x+s}\right)^{1-s} \le \frac{\Gamma(x+s)}{x^s\Gamma(x)} \le 1.$$

LEMMA 5.12. *If* C, B, *and* x_0 *are as above for the generalized gamma distribution and* $\alpha \ge 1$, $\beta \ge 1$, *then* $Ce^{-B(x_0)} \le M_{\alpha,\beta}$

PROOF. Using Theorem 5.10 with $x = (\alpha - 1)/\beta$ in the inequality below implies

$$\begin{aligned} e^{-B(x_0)} &= \left(\frac{\alpha-1}{\beta}\right)^{(\alpha-1)/\beta} e^{-(\alpha-1)/\beta} \\ &\le \Gamma\left(\frac{\alpha-1}{\beta}+1\right)e^{-4/9+1/(6((\alpha-1)/\beta)+9/4)}\left(2\frac{\alpha-1}{\beta}+1\right)^{-1/2}. \end{aligned} \tag{5.14}$$

Since $C = \beta/\Gamma(\alpha/\beta)$, Theorem 5.11 with $x = \alpha/\beta$ and $s = 1 - 1/\beta$ yields

$$C\Gamma\left(\frac{\alpha-1}{\beta}+1\right) \le \alpha^{1-1/\beta}\beta^{1/\beta},$$

and combining this with (5.14), the lemma follows. $\square$

We can also now prove the following lemma which is used in applying Lemma 5.3.

LEMMA 5.13. *If B and x_0 are as above for the generalized gamma distribution and $\beta \ge 1, \alpha \ge 1$, then $e^{B(x_0)}\Gamma(\alpha/\beta)/\beta \le M'_{\alpha,\beta}$.*

PROOF. Using Theorem 5.10 with $x = (\alpha-1)/\beta$ in the following inequality, we find

$$\begin{aligned} e^{B(x_0)} &= \left(\frac{\alpha-1}{\beta}\right)^{-(\alpha-1)/\beta} e^{(\alpha-1)/\beta} \\ &\le \sqrt{2\pi}e^{-1/(6((\alpha-1)/\beta)+9/4)}\left(\frac{\alpha-1}{\beta}+1/2\right)^{1/2}\Gamma\left(\frac{\alpha-1}{\beta}+1\right)^{-1}. \end{aligned} \tag{5.15}$$

Now, Theorem 5.11 with $x = \alpha/\beta$ and $s = 1 - 1/\beta$ yields

$$\frac{\Gamma(\alpha/\beta)}{\Gamma(((\alpha-1)/\beta)+1)} \le \frac{r}{\alpha}\left(\frac{\alpha-1}{\beta}+1\right)^{1/\beta},$$

and combining this with (5.15), the lemma follows. $\square$

Before proving Theorem 5.8, we collect properties of the Stein solution for the generalized gamma distribution, which, according to (5.1) and (5.2) satisfies

$$\begin{aligned} f(x) := f_h(x) &= x^{1-\alpha}e^{x^\beta}\int_0^x \tilde{h}(z)z^{\alpha-1}e^{-z^\beta}\,dz, \\ f'(x) + \left(\frac{\alpha-1}{x} - \beta x^{\beta-1}\right)f(x) &= \tilde{h}(x). \end{aligned} \tag{5.16}$$

First, we record a straightforward application of Lemmas 5.3 and 5.13.

LEMMA 5.14. *If f is given by (5.16), then*

$$\|f\| \le \|\tilde{h}\|M'_{\alpha,\beta}, \qquad \|f'\| \le 2\|\tilde{h}\|.$$

LEMMA 5.15. *If f is given by (5.16), $x > 0$, $|t| \le b \le 1$, and $x + t > 0$, then for $\beta = 1$ and $\beta \ge 2$,*

$$\begin{aligned} &\left|(x+t)^{\beta-1}f(x+t) - x^{\beta-1}f(x)\right| \\ &\quad \le \|\tilde{h}\|b\left[(\beta-1)\left(1+2^{\beta-2}\left(x^{\beta-1}+b^{\beta-1}\right)\right)M'_{\alpha,\beta} + 2x^{\beta-1}\right] =: \|\tilde{h}\|C_{b,\alpha,\beta}(x). \end{aligned}$$

For $1 < \beta < 2$, *we have*

$$\begin{aligned}&\left|(x+t)^{\beta-1} f(x+t) - x^{\beta-1} f(x)\right| \\ &\quad \le \|\tilde{h}\|\left(b^{\beta-1} M'_{\alpha,\beta} + 2bx^{\beta-1}\right) =: \|\tilde{h}\| C_{b,\alpha,\beta}(x).\end{aligned}$$

PROOF. Observe that for all $\beta \ge 1$,

$$\begin{aligned}&\left|(x+t)^{\beta-1} f(x+t) - x^{\beta-1} f(x)\right| \\ &\quad \le \left|(x+t)^{\beta-1} - x^{\beta-1}\right|\left|f(x+t)\right| + x^{\beta-1}\left|f(x+t) - f(x)\right| \\ &\quad \le \left|(x+t)^{\beta-1} - x^{\beta-1}\right|\|f\| + bx^{\beta-1}\|f'\|.\end{aligned} \tag{5.17}$$

In all cases, we use Lemma 5.14 to bound the norms appearing in (5.17). For the remaining term, if $\beta = 1$, then $|(x+t)^{\beta-1} - x^{\beta-1}| = 0$ and the result follows.

If $\beta \ge 2$, then the mean value theorem implies

$$\left|(x+t)^{\beta-1} - x^{\beta-1}\right| \le |t|(\beta-1)\left(x+|t|\right)^{\beta-2} \le b(\beta-1)(x+b)^{\beta-2}.$$

Since $\beta \ge 2$,

$$(x+b)^{\beta-2} \le \max\left\{1, (x+b)^{\beta-1}\right\} \le \max\left\{1, 2^{\beta-2}\left(x^{\beta-1} + b^{\beta-1}\right)\right\},$$

where the last inequality is Hölder's, and the result in this case follows by bounding the maximum by the sum.

For $1 < \beta < 2$, since $x^{\beta-1}$ is concave and increasing, $|(x+t)^{\beta-1} - x^{\beta-1}|$ is maximized when $x = 0$ and $t = b$ in which case it equals $b^{\beta-1}$. $\square$

LEMMA 5.16. *If f is given by* (5.16), *and we define*

$$g(x) = f'(x) + \frac{\alpha - 1}{x} f(x), \qquad x > 0, \tag{5.18}$$

then

$$g(x) = \tilde{h}(x) + \beta x^{\beta-1} f(x), \tag{5.19}$$

and for $\beta \ge 1$,

$$\|g\| \le \|\tilde{h}\| \max\left\{2 + (\alpha-1)M'_{\alpha,\beta}, 1 + \beta M'_{\alpha,\beta}\right\} \le \|\tilde{h}\|\left(2 + (\beta+\alpha-1)M'_{\alpha,\beta}\right).$$

PROOF. The fact that (5.18) equals (5.19) is a simple rearrangement of the second equality of (5.16).

For the bounds, if $x \ge 1$, then (5.18) implies

$$\left|g(x)\right| \le \|f'\| + (\alpha-1)\|f\|,$$

and if $x \le 1$, then (5.19) implies

$$\left|g(x)\right| \le \|\tilde{h}\| + \beta\|f\|,$$

so that

$$\|g\| \le \max\{\|f'\| + (\alpha - 1)\|f\|, \|\tilde{h}\| + \beta\|f\|\},$$

and the result follows from Lemma 5.14. □

The purpose of introducing g in Lemma 5.16 is illustrated in the following lemma.

LEMMA 5.17. *If f is a bounded function on $[0, \infty)$ with bounded derivative such that $f(0) = 0$, $W \ge 0$ is a random variable with $\mathbb{E}W^\beta = \alpha/\beta$, and W^* has the (α, β)-generalized equilibrium distribution of W as in Definition 1.14, then for $g(x) = f'(x) + (\alpha - 1)x^{-1} f(x)$,*

$$\mathbb{E}g(W^*) = \beta\mathbb{E}W^{\beta-1} f(W).$$

PROOF. If $V_\alpha \sim \mathrm{B}(\alpha, 1)$ is independent of $W^{(\beta)}$ having the β-power bias distribution of W, then we can set $W^* = V_\alpha W^{(\beta)}$ and

$$\begin{aligned}\mathbb{E}f'(W^*) &= \mathbb{E}f'(V_\alpha W^{(\beta)}) = \frac{\beta}{\alpha}\mathbb{E}W^\beta f'(V_\alpha W)\\ &= \beta\mathbb{E}W^\beta \int_0^1 u^{\alpha-1} f'(uW)\,du.\end{aligned} \tag{5.20}$$

The case $\alpha = 1$ easily follows from performing the integration in (5.20), keeping in mind that $f(0) = 0$. If $\alpha > 1$, similar to the computation of (5.20),

$$(\alpha - 1)\mathbb{E}\frac{f(W^*)}{W^*} = \beta(\alpha - 1)\mathbb{E}W^{\beta-1} \int_0^1 u^{\alpha-2} f(uW)\,du. \tag{5.21}$$

Applying integration by parts in (5.21) and noting $f(0) = 0$ yields

$$(\alpha - 1)\mathbb{E}\frac{f(W^*)}{W^*} = \beta\mathbb{E}\left\{W^{\beta-1}\left(f(W) - W\int_0^1 u^{\alpha-1} f'(uW)\,du\right)\right\}, \tag{5.22}$$

and adding the right-hand sides of (5.20) and (5.22) yields the lemma. □

We are now in a position to prove our generalized gamma approximation result.

PROOF OF THEOREM 5.8. Let $\delta = d_{\mathrm{K}}(\mathscr{L}(W), \mathscr{L}(Z))$ and let $h_{s,\varepsilon}$ be the smoothed indicators defined at (5.12) in Lemma 5.6. From Lemmas 5.6 and 5.12, we have for every $\varepsilon > 0$,

$$\delta \le \sup_{s>0}\left|\mathbb{E}h_{s,\varepsilon}(W) - \mathbb{E}h_{s,\varepsilon}(Z)\right| + M_{\alpha,\beta}\varepsilon. \tag{5.23}$$

Fix ε and s, let f solve the Stein equation given explicitly by (5.16) with $h := h_{s,\varepsilon}$ and let g be as in Lemma 5.16. By Lemma 5.17,

$$\begin{aligned}\mathbb{E}h(W) - \mathbb{E}h(Z) &= \mathbb{E}\{f'(W) - B'(W)f(W)\}\\ &= \mathbb{E}\left\{f'(W) - \left(\beta W^{\beta-1} - \frac{\alpha-1}{W}\right)f(W)\right\}\\ &= \mathbb{E}\{g(W) - \beta W^{\beta-1}f(W)\} = \mathbb{E}\{g(W) - g(W^*)\}.\end{aligned}$$

And we want to bound this last term since in absolute value it is equal to the first part of the bound in (5.23). With $I_1 = \mathrm{I}[|W - W^*| \le b]$,

$$\begin{aligned}&\left|\mathbb{E}\{g(W) - g(W^*)\}\right|\\ &\quad\le 2\|g\|\mathbb{P}\left[|W - W^*| > b\right] + \left|\mathbb{E}\{I_1(g(W) - g(W^*))\}\right|.\end{aligned}\tag{5.24}$$

Note from the representation (5.19) of g, if $x > 0$, $|t| \le b \le 1$, and $x + t > 0$,

$$g(x+t) - g(x) = h(x+t) - h(x) + \beta\left((x+t)^{\beta-1}f(x+t) - x^{\beta-1}f(x)\right)$$

and since $|h(x+t) - h(x)| \le \varepsilon^{-1}\int_{t\wedge 0}^{t\vee 0}\mathrm{I}[s < x + u \le s + \varepsilon]\,du$, we apply Lemma 5.15 to find

$$\begin{aligned}&\left|\mathbb{E}\{I_1(g(W) - g(W^*))\}\right|\\ &\quad\le \frac{1}{\varepsilon}\sup_{s\ge 0}\int_0^b \mathbb{P}[s < W + u \le s + \varepsilon]\,du + C_{b,\alpha,\beta},\end{aligned}\tag{5.25}$$

where $C_{b,\alpha,\beta} := \mathbb{E}C_{b,\alpha,\beta}(W)$ and $C_{b,\alpha,\beta}(x)$ is defined in Lemma 5.15; and observe that for $1 < \beta < 2$, $C_{b,\alpha,\beta}$ is bounded since $\mathbb{E}W^{\beta-1} \le (\mathbb{E}W^{\beta})^{(\beta-1)/\beta}$.

Now using Lemmas 5.7 and 5.12 to find

$$\mathbb{P}[s < W + u \le s + \varepsilon] \le M_{\alpha,\beta}\varepsilon + 2\delta$$

and combining (5.23), (5.24) and (5.25), we have

$$\delta \le M_{\alpha,\beta}\varepsilon + 2\|g\|\mathbb{P}\left[|W - W^*| > b\right] + C_{b,\alpha,\beta} + bM_{\alpha,\beta} + 2\varepsilon^{-1}b\delta.$$

Applying Lemma 5.16 to bound $\|g\|$, setting $\varepsilon = 4b$, and solving for δ now yields the bounds of the theorem. □

Acknowledgments. A portion of the work for this project was completed when Nathan Ross was at University of California, Berkeley. Erol A. Peköz would like to thank the Department of Statistics and Applied Probability, National University of Singapore, for its hospitality. We also thank Shaun McKinlay for pointing us to the reference Pakes and Khattree (1992), Larry Goldstein for the suggestion to study Rayleigh limits in bridge random walks, Louigi Addario-Berry for helpful discussions, Jim Pitman for detailed pointers to the literature connecting trees, walks and urns, and the Associate Editor and referee for their many detailed and useful comments which have greatly improved this work.

REFERENCES

ALDOUS, D. (1991). The continuum random tree. I. *Ann. Probab.* **19** 1–28. MR1085326

ALDOUS, D. (1993). The continuum random tree. III. *Ann. Probab.* **21** 248–289. MR1207226

ARRATIA, R., GOLDSTEIN, L. and KOCHMAN, F. (2013). Size bias for one and all. Preprint. Available at arXiv:1308.2729.

ATHREYA, K. B. and KARLIN, S. (1968). Embedding of urn schemes into continuous time Markov branching processes and related limit theorems. *Ann. Math. Statist.* **39** 1801–1817. MR0232455

BAI, Z. D., HU, F. and ZHANG, L.-X. (2002). Gaussian approximation theorems for urn models and their applications. *Ann. Appl. Probab.* **12** 1149–1173. MR1936587

BARABÁSI, A.-L. and ALBERT, R. (1999). Emergence of scaling in random networks. *Science* **286** 509–512. MR2091634

BARBOUR, A. D., HOLST, L. and JANSON, S. (1992). *Poisson Approximation. Oxford Studies in Probability* **2**. Oxford Univ. Press, New York. MR1163825

BATIR, N. (2008). Inequalities for the gamma function. *Arch. Math.* (*Basel*) **91** 554–563. MR2465874

BOLTHAUSEN, E. (1982). Exact convergence rates in some martingale central limit theorems. *Ann. Probab.* **10** 672–688. MR0659537

BORODIN, A. N. (1987). On the distribution of random walk local time. *Ann. Inst. Henri Poincaré Probab. Stat.* **23** 63–89. MR0877385

BORODIN, A. N. (1989). Brownian local time. *Uspekhi Mat. Nauk* **44** 7–48. MR0998360

BROWN, M. (2006). Exploiting the waiting time paradox: Applications of the size-biasing transformation. *Probab. Engrg. Inform. Sci.* **20** 195–230. MR2261286

CHATTERJEE, S. and SHAO, Q.-M. (2011). Nonnormal approximation by Stein's method of exchangeable pairs with application to the Curie–Weiss model. *Ann. Appl. Probab.* **21** 464–483. MR2807964

CHEN, L. H. Y., GOLDSTEIN, L. and SHAO, Q.-M. (2011). *Normal Approximation by Stein's Method*. Springer, Heidelberg. MR2732624

CHEN, L. H. Y. and SHAO, Q.-M. (2005). Stein's method for normal approximation. In *An Introduction to Stein's Method. Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.* **4** 1–59. Singapore Univ. Press, Singapore. MR2235448

CHUNG, K. L. (1976). Excursions in Brownian motion. *Ark. Mat.* **14** 155–177. MR0467948

CHUNG, K. L. and HUNT, G. A. (1949). On the zeros of $\sum_1^n \pm 1$. *Ann. of Math.* (2) **50** 385–400. MR0029488

CSÁKI, E. and MOHANTY, S. G. (1981). Excursion and meander in random walk. *Canad. J. Statist.* **9** 57–70. MR0638386

DÖBLER, C. (2012). Stein's method of exchangeable pairs for absolutely continuous, univariate distributions with applications to the polya urn model. Preprint.

DÖBLER, C. (2012). Stein's method for the half-normal distribution with applications to limit theorems related to simple random walk. Preprint. Available at arXiv:1303.4592.

DUDLEY, R. M. (1968). Distances of probability measures and random variables. *Ann. Math. Statist.* **39** 1563–1572. MR0230338

DURRETT, R. T. and IGLEHART, D. L. (1977). Functionals of Brownian meander and Brownian excursion. *Ann. Probab.* **5** 130–135. MR0436354

EGGENBERGER, F. and PÓLYA, G. (1923). Über die Statistik verketteter Vorgänge. *Z. angew. Math Mech.* **3** 279–289.

FELLER, W. (1968). *An Introduction to Probability Theory and Its Applications. Vol. I*, 3rd ed. Wiley, New York. MR0228020

FLAJOLET, P., GABARRÓ, J. and PEKARI, H. (2005). Analytic urns. *Ann. Probab.* **33** 1200–1233. MR2135318

FREEDMAN, D. A. (1965). Bernard Friedman's urn. *Ann. Math. Statist* **36** 956–970. MR0177432

FRIEDMAN, B. (1949). A simple urn model. *Comm. Pure Appl. Math.* **2** 59–70. MR0030144

GOLDSTEIN, L. and REINERT, G. (1997). Stein's method and the zero bias transformation with application to simple random sampling. *Ann. Appl. Probab.* **7** 935–952. MR1484792

GOLDSTEIN, L. and REINERT, G. (2005). Zero biasing in one and higher dimensions, and applications. In *Stein's Method and Applications*. *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.* **5** 1–18. Singapore Univ. Press, Singapore. MR2201883

GOLDSTEIN, L. and XIA, A. (2006). Zero biasing and a discrete central limit theorem. *Ann. Probab.* **34** 1782–1806. MR2271482

JANSON, S. (2006a). Limit theorems for triangular urn schemes. *Probab. Theory Related Fields* **134** 417–452. MR2226887

JANSON, S. (2006b). Random cutting and records in deterministic and random trees. *Random Structures Algorithms* **29** 139–179. MR2245498

JANSON, S. (2006c). Conditioned Galton–Watson trees do not grow. Technical report, Fourth Colloquium on Mathematics and Computer Science Algorithms, Trees Combinatorics and Probability.

JANSON, S. (2012). Simply generated trees, conditioned Galton–Watson trees, random allocations and condensation. *Probab. Surv.* **9** 103–252. MR2908619

LUK, H. M. (1994). Stein's method for the Gamma distribution and related statistical applications. Ph.D. thesis, Univ. Southern California. MR2693204

MARCHAL, P. (2003). Constructing a sequence of random walks strongly converging to Brownian motion. In *Discrete Random Walks* (*Paris*, 2003). *Discrete Math. Theor. Comput. Sci. Proc.*, *AC* 181–190 (electronic). Assoc. Discrete Math. Theor. Comput. Sci., Nancy. MR2042386

MEIR, A. and MOON, J. W. (1978). On the altitude of nodes in random trees. *Canad. J. Math.* **30** 997–1015. MR0506256

NOURDIN, I. and PECCATI, G. (2009). Stein's method on Wiener chaos. *Probab. Theory Related Fields* **145** 75–118. MR2520122

PAKES, A. G. and KHATTREE, R. (1992). Length-biasing, characterizations of laws and the moment problem. *Austral. J. Statist.* **34** 307–322. MR1193781

PANHOLZER, A. (2004). The distribution of the size of the ancestor-tree and of the induced spanning subtree for random trees. *Random Structures Algorithms* **25** 179–207. MR2076338

PEKÖZ, E. A. and RÖLLIN, A. (2011). New rates for exponential approximation and the theorems of Rényi and Yaglom. *Ann. Probab.* **39** 587–608. MR2789507

PEKÖZ, E. A., RÖLLIN, A. and ROSS, N. (2013a). Degree asymptotics with rates for preferential attachment random graphs. *Ann. Appl. Probab.* **23** 1188–1218. MR3076682

PEKÖZ, E. A., RÖLLIN, A. and ROSS, N. (2013b). Total variation error bounds for geometric approximation. *Bernoulli* **19** 610–632. MR3037166

PEKÖZ, E., RÖLLIN, A. and ROSS, N. (2014). Joint degree distributions of preferential attachment random graphs. Preprint. Available at arXiv:1402.4686.

PEMANTLE, R. (2007). A survey of random processes with reinforcement. *Probab. Surv.* **4** 1–79. MR2282181

PITMAN, J. (1999). The distribution of local times of a Brownian bridge. In *Séminaire de Probabilités*, *XXXIII*. *Lecture Notes in Math.* **1709** 388–394. Springer, Berlin. MR1768012

PITMAN, J. (2006). *Combinatorial Stochastic Processes*. *Lecture Notes in Math.* **1875**. Springer, Berlin. MR2245368

PITMAN, J. and ROSS, N. (2012). Archimedes, Gauss, and Stein. *Notices Amer. Math. Soc.* **59** 1416–1421. MR3025901

RAIČ, M. (2003). Normal approximation with Stein's method. In *Proceedings of the Seventh Young Statisticians Meeting*. Metodoloski zvezki, Ljubljana.

REINERT, G. (2005). Three general approaches to Stein's method. In *An Introduction to Stein's Method*. *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.* **4** 183–221. Singapore Univ. Press, Singapore. MR2235451

RÉMY, J.-L. (1985). Un procédé itératif de dénombrement d'arbres binaires et son application à leur génération aléatoire. *RAIRO Inform. Théor.* **19** 179–195. MR0803997

RÖLLIN, A. and ROSS, N. (2015). Local limit theorems via Landau–Kolmogorov inequalities. *Bernoulli* **21** 851–880. MR3338649

ROSS, N. (2011). Fundamentals of Stein's method. *Probab. Surv.* **8** 210–293. MR2861132

ROSS, N. (2013). Power laws in preferential attachment graphs and Stein's method for the negative binomial distribution. *Adv. in Appl. Probab.* **45** 876–893. MR3102476

ROSS, S. and PEKÖZ, E. (2007). A second course in probability. www.ProbabilityBookstore.com, Boston, MA.

STEIN, C. (1986). *Approximate Computation of Expectations. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **7**. IMS, Hayward, CA. MR0882007

VERVAAT, W. (1979). A relation between Brownian bridge and Brownian excursion. *Ann. Probab.* **7** 143–149. MR0515820

WENDEL, J. G. (1948). Note on the gamma function. *Amer. Math. Monthly* **55** 563–564. MR0029448

ZHANG, L.-X., HU, F. and CHEUNG, S. H. (2006). Asymptotic theorems of sequential estimation-adjusted urn models. *Ann. Appl. Probab.* **16** 340–369. MR2209345

ZHANG, L.-X., HU, F., CHEUNG, S. H. and CHAN, W. S. (2011). Immigrated urn models—theoretical properties and applications. *Ann. Statist.* **39** 643–671. MR2797859

E. A. PEKÖZ
QUESTROM SCHOOL OF BUSINESS
BOSTON UNIVERSITY
595 COMMONWEALTH AVENUE
BOSTON, MASSACHUSETTS 02215
USA
E-MAIL: pekoz@bu.edu

A. RÖLLIN
DEPARTMENT OF STATISTICS
AND APPLIED PROBABILITY
NATIONAL UNIVERSITY OF SINGAPORE
6 SCIENCE DRIVE 2
SINGAPORE 117546
E-MAIL: adrian.roellin@nus.edu.sg

N. ROSS
SCHOOL OF MATHEMATICS AND STATISTICS
UNIVERSITY OF MELBOURNE
PARKVILLE VIC, 3010
AUSTRALIA
E-MAIL: nathan.ross@unimelb.edu.au