Proceedings of the 2022 Winter Simulation Conference
B. Feng, G. Pedrielli, Y. Peng, S. Shashaani, E. Song, C.G. Corlu, L.H. Lee, E.P. Chew, T. Roeder, and P. Lendermann, eds.

EFFICIENT RARE EVENT ESTIMATION FOR MAXIMA OF BRANCHING RANDOM WALKS

Michael Conroy

Department of Mathematics
University of Arizona
617 N Santa Rita Avenue
Tucson, AZ 85721, USA

Mariana Olvera-Cravioto

Department of Statistics & Operations Research
University of North Carolina
318 Hanes Hall, CB #3260
Chapel Hill, NC 27599, USA

ABSTRACT

We present a hybrid importance sampling estimator that is strongly efficient for tail probabilities of the all-time maximum of a branching random walk, where the increments satisfy a Cramer-Lundberg condition. The estimator uses conditional Monte Carlo in combination with the population dynamics algorithm to compute an expression for the tail of the distribution obtained from a spine change of measure. It has computational complexity (measured by the number of input random vectors required) that is independent of the offspring distribution, allowing for fast computation even when the mean number of offspring is very large. We remark on consistency of this estimator and give numerical examples.

1 INTRODUCTION

We consider the all-time maximum W of a branching random walk with offspring distribution N and increments $X_1, \dots, X_N$. That is, starting from 0, at each time step all existing nodes produce a random number N offspring with increment sizes determined by the random variables $\{X_i\}$. The random vector $\boldsymbol{\psi} = (N, \{X_i\})$ is allowed to have arbitrary dependence. An explicit construction of W is provided in Section 1.2.

It is known that when there exists $\alpha > 0$ such that

$$E\left[\sum_{i=1}^{N} e^{\alpha X_i}\right] = 1 \qquad \text{and} \qquad 0 < E\left[\sum_{i=1}^{N} X_i e^{\alpha X_i}\right] < \infty, \tag{1}$$

as well as some $\beta \in (0, \alpha)$ such that

$$\rho_\beta := E\left[\sum_{i=1}^{N} e^{\beta X_i}\right] < 1, \tag{2}$$

the maximum W satisfies the tail asymptotic

$$P(W > t) \sim H e^{-\alpha t}, \qquad t \to \infty,$$

for some constant $H > 0$ with various representations (Basrak et al. 2022; Jelenković and Olvera-Cravioto 2012a; Jelenković and Olvera-Cravioto 2012b; Jelenković and Olvera-Cravioto 2015). However, in applications it is important to estimate $P(W > t)$ itself, rather than relying on its asymptotic expression, and this can be challenging. Although there exist efficient methods to generate samples from the law of W (e.g., the population dynamics algorithm (Olvera-Cravioto 2019)), this is not enough to compute tail probabilities. Specifically, given a sample from the law of W, the naive Monte Carlo estimator $1(W > t)$

has unbounded relative error, i.e.,

$$\frac{\mathrm{Var}\left(1(W>t)\right)}{P(W>t)^2} = \frac{P(W>t)P(W\le t)}{P(W>t)^2} \to \infty \qquad \text{as} \qquad t\to\infty.$$

An importance sampling approach to estimating $P(W>t)$ was developed in Basrak et al. (2022) based on a representation of the tail event under a likelihood ratio change of measure. The result is an estimator $Z(t)$ for $P(W>t)$ that is unbiased under the new probabiliy measure $\widetilde{P}$ and that is strongly efficient, i.e.,

$$\limsup_{t\to\infty} \frac{\widetilde{\mathrm{Var}}\left(Z(t)\right)}{P(W>t)^2} < \infty, \tag{3}$$

where $\widetilde{\mathrm{Var}}$ denotes variance with respect to $\widetilde{P}$.

The change of measure $\widetilde{P}$ is based on a "spine decomposition" of the branching process, which selects one lineage at random to which it applies an exponential tilt, and is similar to the changes of measure used in Lyons (2018) and Lyons et al. (1995). Under the condition (2), the paths of the walk have negative drift since

$$E\left[\max_{1\le i\le N} X_i\right] \le \beta^{-1}\log E\left[\max_{1\le i\le N} e^{\beta X_i}\right] \le \beta^{-1}\log\rho_\beta < 0$$

by Jensen's inequality. Under $\widetilde{P}$, the chosen path is tilted to have *positive* drift $\mu>0$, while the drift on the other branches remains unchanged. This chosen path along with its immediate offspring are referred to as the *spine*. This generalizes the classical exponential change of measure used in renewal theory and importance sampling for the (nonbranching) random walk on $\mathbb{R}$. See Section 2 for more details.

The estimator $Z(t)$ from Basrak et al. (2022) is generated by simulating the branching random walk until the first time that any path exceeds the value t. While allowing for unbiased estimation, it is computationally very costly when either $E[N]$ or t is very large. Throughout, we measure the computational complexity of an algorithm by the number of branching vectors $\boldsymbol{\psi}$ that need to be generated. By tilting only one path of the branching random walk, the measure $\widetilde{P}$ virtually guarantees that the algorithm will terminate there. If $\tau(t)$ denotes the first passage time of t along that path, then it is shown in Basrak et al. (2022) that $\tau(t)\sim t/\mu$ as $t\to\infty$ $\widetilde{P}$-a.s., and hence the complexity of generating a single copy of $Z(t)$ is at least $(E[N])^{t/\mu}$. In practice, this cannot be done in a reasonable amount of time for large t when $E[N]$ is significantly larger than 1.

Herein we propose an estimator for $P(W>t)$ that takes advantage of the fact that the algorithm will almost always terminate on the spine to circumvent the computational complexity induced by N. The idea is to *only* simulate the spine of the process by first accounting for the contribution of all other paths using conditional Monte Carlo, and using population dynamics to numerically approximate this estimator. The computational complexity of the population dynamics algorithm does not depend on $E[N]$, and hence neither does that for the hybrid approach presented here. Furthermore, it depends on t only linearly.

Throughout, we assume the following on the random vector $\boldsymbol{\psi}$. Let $D=\sum_{i=1}^N e^{\alpha X_i}$, where α satisfies (1).

Condition 1 There exists $\alpha>0$ such that (1) holds. Additionally,

(a) The probability measure $E\left[\sum_{i=1}^N e^{\alpha X_i}1(X_i\in dx)\right]$ on $\mathbb{R}$ is nonarithmetic,
(b) $P(D>0)=1$,
(c) $\rho_\beta<1$ for some $\beta\in(0,\alpha)$, and
(d) $E[N]<\infty$ and $E[ND]<\infty$.

Part (a) of Condition 1 will ensure that certain renewal-theoretic results from Basrak et al. (2022) hold. Part (b) ensures that the change of measure defined in Section 2 is well defined. Note that (b) implies that $P(N\ge 1)=1$, namely that no lineage of the branching random walk goes extinct. The condition

that $E[ND] < \infty$ is needed for the mean of N under the change of measure to be finite, something that is convenient for simulation. We also assume the following (see Jelenković and Olvera-Cravioto (2015)).

Condition 2 $E\left[\left(\sum_{i=1}^N e^{X_i}\right)^\alpha\right] < \infty$ if $\alpha > 1$ and $E\left[\left(\sum_{i=1}^N e^{\alpha X_i/(1+\varepsilon)}\right)^{1+\varepsilon}\right] < \infty$ for some $\varepsilon \in (0,1)$ if $\alpha \leq 1$.

Conditions 1 and 2 provide the only assumed dependencies among the components of the branching vector $\boldsymbol{\psi} = (N, \{X_i\})$.

1.1 The High-Order Lindley Equation

The law of the all time maximum of the branching random walk is a solution to the stochastic fixed point equation

$$W \stackrel{\mathscr{D}}{=} \left(\max_{1\leq i\leq N}(X_i + W_i)\right)^+, \tag{4}$$

where $\stackrel{\mathscr{D}}{=}$ denotes equality in distribution, and the $\{W_i\}$ are i.i.d. copies of W independent of $(N, \{X_i\})$. (4) is a special case of the more general *high order Lindley equation*

$$W \stackrel{\mathscr{D}}{=} \max\left\{Y, \max_{1\leq i\leq N}(X_i + W_i)\right\},$$

when the random variable Y is identically 0.

Recursion (4) was studied in Olvera-Cravioto and Ruiz-Lacedelli (2021) (see also Karpelevich et al. (1994)) in the context of queueing networks with n servers, whererin jobs are split into a random number N fragments to be processed in a synchronized fashion. When $X_i = \chi_i - \tau_i$ is the difference of the limiting service time χ_i of a fragment and the limiting interarrival time τ_i of a job, W gives the stationary waiting time distribution of jobs as $n \to \infty$. Note that when N is identically 1, (4) is simply the classical Lindley equation, and the queueing network becomes the G/G/1 queue (Asmussen 2003, Section III.7). It is useful in the context of the more general queueing systems to have efficient means to estimate the probability that the limiting waiting time exceeds some large value. The importance sampling approach to estimating stationary waiting time distributions in the simpler G/G/1 queue is part of the classical literature, for which we refer to Siegmund (1976) and Chapter VI in Asmussen (2003).

1.2 The Marked Galton-Watson Process

Here we define the branching random walk and its all-time maximum W through a marked Galton-Watson process that lives on a tree $\mathscr{T}$ constructed as follows.

Let $U = \bigcup_{k=0}^\infty \mathbb{N}_+^k$ be the set of all finite sequences $\mathbf{i} = (i_1, \ldots, i_k)$, $k \geq 0$. We take the convention $\mathbb{N}_+^0 = \{\emptyset\}$, where $\emptyset$ denotes the null sequence, which will be the root node of our tree $\mathscr{T}$. For $\mathbf{i} = (i_1, \ldots, i_k) \in U$, we will write $\mathbf{i}|r = (i_1, \ldots, i_r)$ when $r \leq k$ to denote truncation, and we will write $(\mathbf{i}, j) = (i_1, \ldots i_k, j)$ for $j \in \mathbb{N}_+$ to denote concatenation. When needed, we use the convention $(\mathbf{i}|0) = \emptyset$ and $(\emptyset, j) = j$. We will also write $|\mathbf{i}| = k$ to denote length, and we order U according to length-lexicographic order, denoted by $\prec$. That is, for $\mathbf{i}, \mathbf{j} \in U$, $\mathbf{i} \prec \mathbf{j}$ if either $|\mathbf{i}| < |\mathbf{j}|$ or $|\mathbf{i}| = |\mathbf{j}|$ and for some $r \leq |\mathbf{i}|$, $i_n = j_n$ for $n = 1, \ldots, r-1$ and $i_r < j_r$.

Now let $\{\boldsymbol{\psi}_\mathbf{i} = (N_\mathbf{i}, X_{(\mathbf{i},1)}, X_{(\mathbf{i},2)}, \ldots) : \mathbf{i} \in U\}$ be i.i.d. copies of $\boldsymbol{\psi}$. For convenience, we will often denote $\boldsymbol{\psi}_\emptyset = \boldsymbol{\psi}$. The structure of $\mathscr{T}$ is defined as follows. Let A_n denote the set of nodes in the nth generation, where $A_0 = \{\emptyset\}$,

$$\begin{aligned} A_1 &= \{i \in \mathbb{N}_+ : 1 \leq i \leq N_\emptyset\}, \qquad \text{and} \\ A_n &= \{(\mathbf{i}, j) \in U : \mathbf{i} \in A_{n-1}, 1 \leq j \leq N_\mathbf{i}\}, \qquad n \geq 2. \end{aligned}$$

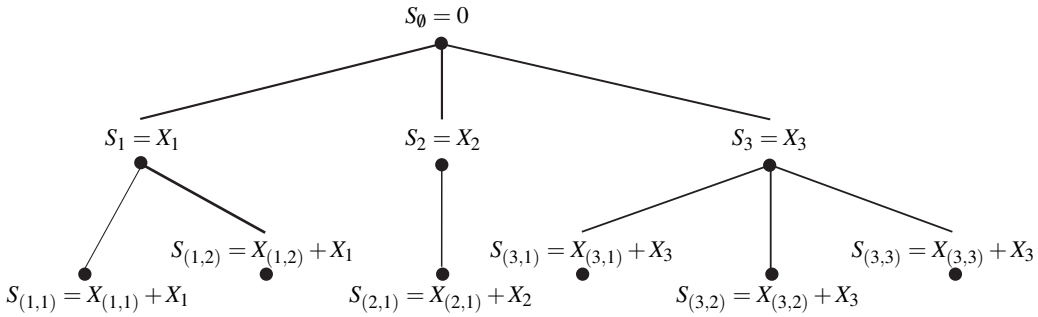

Figure 1: A weighted branching tree.

To each node $\mathbf{i}$, we assign the increment $X_\mathbf{i}$ and the value of the branching random walk $S_\mathbf{i}$, where

$$S_\emptyset = 0, \qquad S_\mathbf{i} = \sum_{k=1}^{n} X_{\mathbf{i}|k}, \ \ \mathbf{i} \in A_n, \ \ n \geq 1.$$

This construction is visualized in Figure 1. Note that $S_\mathbf{i}$ is the sum of independent but not necessarily identically distributed random variables. Then,

$$W := \sup_{\mathbf{i} \in \mathscr{T}} S_\mathbf{i}.$$

2 CHANGE OF MEASURE ALONG THE SPINE

We start from the root node of $\mathscr{T}$ and recursively construct the chosen path $\{\mathbf{J}_k : k \geq 0\}$ by choosing one offspring at each step to continue the path with probability proportional to its exponentiated increment to the power of α. We begin with $\mathbf{J}_0 = \emptyset$ and for $k \geq 1$ we let

$$\mathbf{J}_k = (\mathbf{J}_{k-1}, i) \qquad \text{with probability} \qquad \frac{e^{\alpha X_{(\mathbf{J}_{k-1},i)}}}{D_{\mathbf{J}_{k-1}}}, \qquad 1 \leq i \leq N_{\mathbf{J}_{k-1}},$$

where $D_\mathbf{i} = \sum_{i=1}^{N_\mathbf{i}} e^{\alpha X_{(\mathbf{i},i)}}$ for each $\mathbf{i} \in \mathscr{T}$.

Then, define the process $\{L_k : k \geq 0\}$ by

$$L_0 = 1, \qquad L_k = \prod_{r=0}^{k-1} D_{\mathbf{J}_r}, \ \ k \geq 1,$$

and the filtration $\{\mathscr{G}_k : k \geq 0\}$ by

$$\mathscr{G}_0 = \sigma(\varnothing), \qquad \mathscr{G}_k = \sigma(\boldsymbol{\psi}_\mathbf{i}, \mathbf{i} \in A_r, r < k; \mathbf{J}_r, r \leq k), \ \ k \geq 1.$$

It follows that L_k is a mean-1 positive martingale with respect to $\mathscr{G}_k$, and therefore we can define the measure $\widetilde{P}$ on $\sigma\left(\bigcup_{k=0}^{\infty} \mathscr{G}_k\right)$ induced by

$$\widetilde{P}(A) = E\left[1(A) L_k\right] \qquad \text{for} \qquad A \in \mathscr{G}_k.$$

As mentioned in the Introduction, this change of measure affects only the spine and leaves all other paths unchanged. This observation is formalized as follows.

Lemma 1 (Lemma 2.5 in Basrak et al. (2022)) For $k \geq 0$, $\mathbf{i} \in A_k$, and measurable $B \subset \mathbb{N}_+ \times \mathbb{R}^\infty$,

$$\widetilde{P}(\, \boldsymbol{\psi}_\mathbf{i} \in B | \, \mathbf{i} \neq \mathbf{J}_k) = P(\boldsymbol{\psi} \in B), \qquad \text{and}$$
$$\widetilde{P}(\, \boldsymbol{\psi}_\mathbf{i} \in B | \, \mathbf{i} = \mathbf{J}_k) = E\left[1(\boldsymbol{\psi} \in B) D\right].$$

Moreover, under $\widetilde{P}$ the vectors $\{\boldsymbol{\psi}_\mathbf{i} : \mathbf{i} \in A_k\}$ are conditionally independent given $\mathscr{G}_{k-1}$.

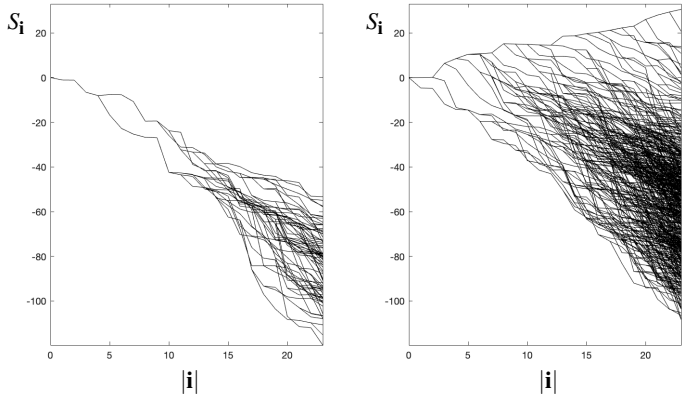

Figure 2: A branching random walk simulated under both P (left) and $\widetilde{P}$ (right).

To identify the random walk along $\{\mathbf{J}_k\}$, define $\hat{X}_k = X_{\mathbf{J}_k}$ and let

$$V_0 = 0, \qquad V_k = S_{\mathbf{J}_k} = \hat{X}_1 + \cdots + \hat{X}_k, \ k \geq 1.$$

As the following lemma shows, $\{V_k : k \geq 0\}$ is a positive-drift random walk with i.i.d. increments under $\widetilde{P}$. Throughout, we use $\widetilde{E}$ to denote expectation with respect to $\widetilde{P}$.

Lemma 2 (Lemma 2.6 in Basrak et al. (2022)) For $k \geq 1$ and $x_1, \ldots, x_k \in \mathbb{R}$,

$$\widetilde{P}\left(\hat{X}_1 \leq x_1, \ldots, \hat{X}_k \leq x_k\right) = \prod_{j=1}^{k} G(x_j),$$

where

$$G(x) = E\left[\sum_{i=1}^{N} e^{\alpha X_i} 1(X_i \leq x)\right].$$

Furthermore, $\widetilde{E}\left[|\hat{X}_1|\right] < \infty$ and

$$\mu := \widetilde{E}\left[\hat{X}_1\right] = E\left[\sum_{i=1}^{N} X_i e^{\alpha X_i}\right] \in (0, \infty).$$

Figure 2 shows a branching random walk simulated under both P and $\widetilde{P}$, when the $\{X_i\}$ are i.i.d. with $X_1 = \chi - \tau$ with $\chi \sim \text{Exponential}(5)$ independent of $\tau \sim \text{Exponential}(1/4)$ (see Section 1.1). In the simulation, $N \sim \text{Bernoulli}(1/4) + 1$ under P independent of the $\{X_i\}$. $\widetilde{P}$ puts more weight on larger values on N, as can be seen along the spine in the figure. Indeed, when N is independent of i.i.d. $\{X_i\}$, the first equality in (1) becomes $E[N]E\left[e^{\alpha X_1}\right] = 1$, and as a consequence of Lemma 1,

$$\widetilde{P}(N = n) = E\left[1(N = n)D\right] = nP(N = n)E\left[e^{\alpha X_1}\right] = \frac{nP(N = n)}{E[N]}.$$

That is, N has its size-biased distribution under $\widetilde{P}$. In Figure 2, N has the size-biased distribution $\text{Bernoulli}(6/7) + 1$ along the spine.

A common challenge in importance sampling procedures is how to simulate under the change of measure. It turns out that in this case the law of $\boldsymbol{\psi}$ under $\widetilde{P}$ is relatively easy to simulate from since it admits a mixture representation that essentially reduces it to simulating from a single exponentially tilted distribution.

Lemma 3 (Example 4.6 in Basrak et al. (2022)) For each $n \in \mathbb{N}_+$ and $1 \le i \le n$, let $f_{\{X_j: j\neq i\}|X_i,N}(\cdot|x,n)$ denote the joint density of $\{X_j : j \neq i\}$ under P conditional on $\{X_i = x, N = n\}$. For each n, let $\{p_{i,n} : 1 \le i \le n\}$ be the discrete probability distribution given by

$$p_{i,n} = \frac{E\left[e^{\alpha X_i}\middle| N = n\right]}{\sum_{j=1}^n E\left[e^{\alpha X_j}\middle| N = n\right]}.$$

Then, the joint density of $\{X_i\}$ under $\widetilde{P}$ conditional on $N = n$ satisfies

$$\widetilde{f}_{\{X_i\}|N}(x_1,\ldots,x_n|n) = \sum_{i=1}^n p_{i,n} f_{\{X_j: j\neq i\}|X_i,N}(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n|x_i,n)\widetilde{g}_{i,n}(x_i),$$

where

$$\widetilde{g}_{i,n}(x) = \frac{e^{\alpha x} f_{X_i|N}(x|n)}{E\left[e^{\alpha X_i}|N = n\right]} \tag{5}$$

is the marginal density of X_i conditional on $N = n$ after an exponential tilt. Therefore, after sampling $N = n$, $\{X_i\}$ can be sampled under $\widetilde{P}$ by first picking $i \in \{1,\ldots,n\}$ according to $\{p_{i,n} : 1 \le i \le n\}$, sampling X_i according to the tilted distribution (5), then sampling $\{X_j : j \neq i\}$ under P conditional on X_i and N.

3 THE HYBRID ESTIMATOR

Let $\gamma(t)$ denote the first node in $\mathscr{T}$ (in length-lexicographic ordering) at which $S_{\mathbf{i}}$ exceeds t, and similarly let $\tau(t)$ be the first passage time of t for the random walk V_k. Namely,

$$\gamma(t) := \inf\{\mathbf{i} \in \mathscr{T} : S_{\mathbf{i}} > t\} \qquad \text{and} \qquad \tau(t) := \inf\{k \ge 0 : V_k > t\}.$$

The estimator in Basrak et al. (2022) is based on the representation

$$P(W > t) = \widetilde{E}\left[1(\gamma(t) = \mathbf{J}_{\tau(t)})e^{-\alpha V_{\tau(t)}}\right] \tag{6}$$

(see Remark 2.7(a) therein). The indicator function in the above display is 1 when the first node at which $S_{\mathbf{i}} > t$ occurs is on the chosen path. Notably, if $N \equiv 1$, the indicator is always 1, and the representation reduces to that commonly used in renewal theory and importance sampling for random walks (see, for instance, Iglehart (1972), Siegmund (1976), and Chapter VIII in Asmussen (2003)). The display (6) naturally suggests the unbiased importance sampling estimator

$$Z(t) := 1(\gamma(t) = \mathbf{J}_{\tau(t)})e^{-\alpha V_{\tau(t)}}$$

for $P(W > t)$ sampled under $\widetilde{P}$. It was shown in Basrak et al. (2022), Lemma 4.3 that under Conditions 1 and 2, $Z(t)$ also has bounded relative error in the sense of (3). As mentioned previously, generating a single copy of $Z(t)$ requires evaluating the indicator $1(\gamma(t) = \mathbf{J}_{\tau(t)})$ and thus generating the entire branching random walk until $\tau(t)$ occurs.

Now let $W^{(k)} := \max_{|\mathbf{i}|\le k} S_{\mathbf{i}}$ for each $k \ge 0$, i.e., the maximum W truncated at the kth generation of the branching random walk, and denote its CDF with respect to P by

$$F_k(x) := P\left(W^{(k)} \le x\right).$$

Note in particular that $W^{(0)} \equiv 0$. Additionally, for each k define the nodes in the spine at generation k to the left and the right of the chosen node as

$$B_k^{\prec} = \{\mathbf{i} \in A_k : \mathbf{i}|k-1 = \mathbf{J}_{k-1}, \mathbf{i} \prec \mathbf{J}_k\} \qquad \text{and} \qquad B_k^{\succ} = \{\mathbf{i} \in A_k : \mathbf{i}|k-1 = \mathbf{J}_{k-1}, \mathbf{i} \succ \mathbf{J}_k\}.$$

Then the behavior of $\mathscr{T}$ under $\widetilde{P}$ given by Lemma 1 gives the following.

Theorem 1 For any $t > 0$,

$$P(W > t) = \widetilde{E}\left[e^{-\alpha V_{\tau(t)}} \prod_{i=1}^{\tau(t)} \prod_{\mathbf{i} \in B_i^{\prec}} F_{\tau(t)-i}(t - S_{\mathbf{i}}) \prod_{\mathbf{j} \in B_i^{\succ}} F_{\tau(t)-i-1}(t - S_{\mathbf{j}}) \right],$$

with the conventions $F_{-1}(x) \equiv 1$ and $\prod_{i=1}^{0} x_i \equiv 1$ for any values $\{x_i\}$.

Proof. Start by writing the event $\{\mathbf{J}_{\tau(t)} = \gamma(t)\}$ in the representation (6) as

$$\{\mathbf{J}_{\tau(t)} = \gamma(t)\} = \left\{ \max_{\mathbf{i} \prec \mathbf{J}_{\tau(t)}} S_{\mathbf{i}} \le t \right\} = \left\{ \max_{0 \le k < \tau(t)-1} V_k + U_k^{(\tau(t)-k)} \le t, \max_{\mathbf{i} \in B_{\tau(t)}^{\prec}} S_{\mathbf{i}} \le t \right\}, \tag{7}$$

where $U_k^{(r)} = \left(\max_{\mathbf{i} \in B_{k+1}^{\prec}} (S_{\mathbf{i}} - V_k + W_{\mathbf{i}}^{(r-1)}) \right) \vee \left(\max_{\mathbf{i} \in B_{k+1}^{\succ}} (S_{\mathbf{i}} - V_k + W_{\mathbf{i}}^{(r-2)}) \right)$, and

$$W_{\mathbf{i}}^{(r)} = \max_{0 \le k \le r} \max_{(\mathbf{i},\mathbf{j}) \in A_{|\mathbf{i}|+k}} (S_{(\mathbf{i},\mathbf{j})} - S_{\mathbf{i}}), \qquad \mathbf{i} \ne \mathbf{J}_{k+1},$$

are i.i.d. with CDF F_r under both P and $\widetilde{P}$, for each $r \ge 2$. Next, define the filtration

$$\mathscr{H}_n = \sigma\left(\mathbf{J}_k, \{X_{(\mathbf{J}_k,i)}\} : 0 \le k \le n \right),$$

and note that

$$P(W > t) = \widetilde{E}\left[1(\mathbf{J}_{\tau(t)} = \gamma(t)) e^{-\alpha V_{\tau(t)}} \right] = \widetilde{E}\left[\widetilde{P}\left(\mathbf{J}_{\tau(t)} = \gamma(t) \middle| \mathscr{H}_{\tau(t)} \right) e^{-\alpha V_{\tau(t)}} \right].$$

Since the $W_{\mathbf{i}}^{(r)}$ are independent of $\mathscr{H}_n$ for all n, from (7) we have

$$\begin{aligned}
\widetilde{P}\left(\mathbf{J}_{\tau(t)} = \gamma(t) \middle| \mathscr{H}_{\tau(t)} \right) &= \prod_{k=0}^{\tau(t)-2} \widetilde{P}\left(V_k + U_k^{(\tau(t)-k)} \le t \middle| \mathscr{H}_{\tau(t)} \right) \prod_{\mathbf{i} \in B_{\tau(t)}^{\prec}} \widetilde{P}\left(S_{\mathbf{i}} \le t \middle| \mathscr{H}_{\tau(t)} \right) \\
&= \prod_{k=0}^{\tau(t)-1} \prod_{\mathbf{i} \in B_{k+1}^{\prec}} \widetilde{P}\left(S_{\mathbf{i}} + W_{\mathbf{i}}^{(\tau(t)-k-1)} \le t \middle| \mathscr{H}_{\tau(t)} \right) \\
&\quad \times \prod_{j=0}^{\tau(t)-2} \prod_{\mathbf{j} \in B_{k+1}^{\succ}} \widetilde{P}\left(S_{\mathbf{j}} + W_{\mathbf{j}}^{(\tau(t)-j-2)} \le t \middle| \mathscr{H}_{\tau(t)} \right) \\
&= \prod_{k=1}^{\tau(t)} \prod_{\mathbf{i} \in B_k^{\prec}} F_{\tau(t)-k}(t - S_{\mathbf{i}}) \prod_{\mathbf{j} \in B_k^{\succ}} F_{\tau(t)-k-1}(t - S_{\mathbf{j}}).
\end{aligned}$$

□

Were the CDFs $\{F_k : k \ge 0\}$ known, out of the above theorem would fall a conditional Monte Carlo algorithm for estimating $P(W > t)$ based on an unbiased estimator that depends only on $\{F_k\}$ and the values of $S_{\mathbf{i}}$ on the spine. It is natural then to consider an approximation of this estimator in terms of some estimators $\{\hat{F}_k : k \ge 0\}$ for the $\{F_k\}$. If such estimates exist, then the proposed estimator is

$$\hat{Z}(t) := e^{-\alpha V_{\tau(t)}} \prod_{i=1}^{\tau(t)} \prod_{\mathbf{i} \in B_i^{\prec}} \hat{F}_{\tau(t)-i}(t - S_{\mathbf{i}}) \prod_{\mathbf{j} \in B_i^{\succ}} \hat{F}_{\tau(t)-i-1}(t - S_{\mathbf{j}}). \tag{8}$$

The *population dynamics algorithm*, described in the following section, gives an efficient way to produce these estimates.

3.1 The Population Dynamics Algorithm

The population dynamics algorithm (Mezard and Montanari 2009; Olvera-Cravioto 2019) provides an efficient way to generate samples of a given size m approximately from the distribution of $W^{(k)}$ based on bootstrapping. Using the fact that W is the solution to (4), the algorithm starts with a sample of $W^{(0)}$ (all zeros), then approximates a sample from $W^{(1)}$ by, for each $1 \le i \le m$, generating an independent copy of $\boldsymbol{\psi}$ and setting

$$\hat{W}_i^{(1,m)} = \left(\max_{1 \le j \le N} X_j \right)^+.$$

It then proceeds by successively, for each $1 \le i \le m$, generating another copy of $\boldsymbol{\psi}$ and sampling $\{\hat{W}_{(i,1)}^{(k-1,m)}, \ldots, \hat{W}_{(i,N)}^{(k-1,m)}\}$ from $\{\hat{W}_1^{(k-1,m)}, \ldots, \hat{W}_m^{(k-1,m)}\}$ uniformly with replacement to produce

$$\hat{W}_i^{(k,m)} = \left(\max_{1 \le j \le N} \left(X_j + \hat{W}_{(i,j)}^{(k-1,m)} \right) \right)^+.$$

For some pre-selected values $K, m \in \mathbb{N}_+$, the result is a collection of samples

$$\left\{ \hat{W}_1^{(k,m)}, \ldots, \hat{W}_m^{(k,m)} \right\}, \quad k \le K,$$

generated with a computational complexity of $K \cdot m$. Estimates of $\{F_k\}$ can then be made as the empirical CDFs

$$\hat{F}_{k,m}(x) := \frac{1}{m} \sum_{i=1}^{m} 1\left(\hat{W}_i^{(k,m)} \le x \right).$$

Of course, the estimator (8) requires the use of estimates $\hat{F}_0, \ldots, \hat{F}_{\tau(t)}$ for a value of $\tau(t)$ that is a priori unknown. The idea, then, is to choose a value of K large enough, generate $\{\hat{F}_{k,m} : k \le K\}$, and use these estimates to sample $\hat{Z}(t)$ by inputting $\hat{F}_{K,m}$ when $k > K$. By Basrak et al. (2022), Lemma 4.2, $\tau(t) \sim t/\mu$ as $t \to \infty$ with $\widetilde{P}$-probability one, and so picking K at least as large as t/μ is a safe choice. The following result gives the sense in which the estimates $\{\hat{F}_{k \wedge K,m} : k \ge 0\}$ are consistent as $K, m \to \infty$.

Lemma 4 Suppose $\beta \in (0, \alpha)$ is such that $\rho_\beta < 1$. Let

$$F_k^{(\beta)}(x) = F_k\left(\beta^{-1} \log x\right) \qquad \text{and} \qquad \hat{F}_{k,m}^{(\beta)}(x) = \hat{F}_{k,m}\left(\beta^{-1} \log x\right),$$

for $k \ge 0$, so that, for example, $F_k^{(\beta)}$ is the CDF of $e^{\beta W^{(k)}}$. Then, there exists a constant $C \in (0, \infty)$ such that

$$\sup_{k \ge 0} E\left[d_1\left(F_k^{(\beta)}, \hat{F}_{k \wedge K,m}^{(\beta)} \right) \right] \le C\left(\rho_\beta^K + m^{-1/2} \right),$$

where d_1 denotes the Wasserstein-1 distance.

Proof. Combine Theorem 2.5, Lemma 2.6, and Theorem 2.8 in Olvera-Cravioto (2019). □

3.2 Properties of the Algorithm

The proposed algorithm to estimate $P(W > t)$ proceeds by picking K and m, generating $\{\hat{F}_{k \wedge K,m} : k \ge 0\}$ by the population dynamics algorithm, then generating a sample of some size n of the estimator

$$\hat{Z}_{K,m}(t) = e^{-\alpha V_{\tau(t)}} \prod_{i=1}^{\tau(t)} \prod_{\mathbf{i} \in B_i^{\prec}} \hat{F}_{(\tau(t)-i) \wedge K,m}(t - S_{\mathbf{i}}) \prod_{\mathbf{j} \in B_i^{\succ}} \hat{F}_{(\tau(t)-i-1) \wedge K,m}(t - S_{\mathbf{j}}), \tag{9}$$

using the same values $\{\hat{F}_{k \wedge K,m}\}$ on each iteration. $P(W > t)$ can then be estimated as a sample average of the n copies of $\hat{Z}_{K,m}(t)$. Table 1 gives the pseudocode for simulating a single copy.

Table 1: The hybrid importance sampling algorithm.

Input: $t > 0$ and Pop. Dyn. estimators $\{\hat{F}_{k\wedge K,m} : k \geq 0\}$
Output: A single copy of $\hat{Z}_{K,m}(t)$
Generate $(N, X_1, \ldots, X_N) \sim E[1(\boldsymbol{\psi} \in \cdot)D]$, i.e., from the tilted distribution of $\boldsymbol{\psi}$
Choose $j \in \{1, \ldots, N\}$ w.p. $e^{\alpha X_j}/D$ and set $\mathbf{J}_1 \leftarrow j$, i.e., pick the 'chosen' node in gen. 1
Set $S_j \leftarrow X_j$ for $j = 1, \ldots, N$
Initialize $V_0 \leftarrow 0$, $V_1 \leftarrow S_{\mathbf{J}_1}$, $k \leftarrow 0$, $\mathbf{J}_0 \leftarrow \emptyset$
while $V_k \leq t$ **do**
    Update $k \leftarrow k+1$
    Generate $(N_{\mathbf{J}_k}, X_{(\mathbf{J}_k,1)}, \ldots, X_{(\mathbf{J}_k,N_{\mathbf{J}_k})}) \sim E[1(\boldsymbol{\psi} \in \cdot)D]$
    Choose $j \in \{1, \ldots, N_{\mathbf{J}_k}\}$ w.p. $e^{\alpha X_{(\mathbf{J}_k,j)}}/D_{\mathbf{J}_k}$ and set $\mathbf{J}_{k+1} \leftarrow (\mathbf{J}_k, j)$
    Set $S_{(\mathbf{J}_k,j)} \leftarrow S_{\mathbf{J}_k} + X_{(\mathbf{J}_k,j)}$ for $j = 1, \ldots, N_{\mathbf{J}_k}$ and $V_{k+1} \leftarrow S_{\mathbf{J}_{k+1}}$
end while
Compute $\hat{F}_{(k-j)\wedge K,m}(t - S_{\mathbf{i}})$ for $\mathbf{i} \in B_j^{\prec}$, $j = 1, \ldots, k$
Compute $\hat{F}_{(k-j-1)\wedge K,m}(t - S_{\mathbf{i}})$ for $\mathbf{i} \in B_j^{\succ}$, $j = 1, \ldots, k-1$
Set $\hat{Z}_{K,m}(t) \leftarrow e^{-\alpha V_k} \prod_{j=1}^{k} \prod_{\mathbf{i}\in B_j^{\prec}} \hat{F}_{(k-j)\wedge K,m}(t - S_{\mathbf{i}}) \prod_{\mathbf{j}\in B_j^{\succ}} \hat{F}_{(k-j-1)\wedge K,m}(t - S_{\mathbf{j}})$, i.e., compute (9)
Output $\hat{Z}_{K,m}(t)$

Remark 1 The computational complexity of generating $\{\hat{F}_{k\wedge K,m} : k \geq 0\}$ is Km, and since $\tau(t) \sim t/\mu$ $\widetilde{P}$-a.s., the complexity of generating a sample of size n of copies of $\hat{Z}_{K,m}(t)$ is asymptotically of order

$$Km + \frac{nt}{\mu}.$$

Notably, this does not depend on the distribution of N and grows only linearly in t.

While Lemma 4 suggests consistency of $\hat{Z}_{K,m}(t)$ as $K, m \to \infty$ in a distributional sense, an exact rate of convergence will be provided in a forthcoming article. A main question is how the rate of convergence varies with t. Since $\hat{Z}_{K,m}(t)$ is an estimate for a rare event probability, the relevant quantity to study is the *relative bias*

$$\text{Rel. Bias}\left(\hat{Z}_{K,m}(t)\right) := \left|\frac{\widetilde{E}\left[\hat{Z}_{K,m}(t)\right] - P(W > t)}{P(W > t)}\right|.$$

In the numerical examples of the following section, we give some estimates of this quantity. In that particular example, the relative bias appears to not be growing much, if at all, in t. We anticipate being able to show that the realtive bias has a rate of converegence that is uniform in $t > 0$.

4 NUMERICAL EXAMPLES

We finish by showing the results of two numerical experiments. The first compares the hybrid approach discussed here with the unbiased algorithm from Basrak et al. (2022) for a relatively small branching rate $E[N] = 2.5$, and the second emphasizes the novelty of the hybrid algorithm by choosing $E[N] = 50$.

For the first, we choose $P(N = 2) = P(N = 3) = 1/2$, and independent of N, the $\{X_i\}$ are i.i.d. with $X_1 = \chi - \tau$, where $\chi \sim \text{Exponential}(5)$ independent of $\tau \sim \text{Exponential}(1/4)$. We calculate $\alpha = 4.32$ and $\mu = 1.24$. Under the change of measure, $\widetilde{P}(N = 2) = 2/5$ and $\widetilde{P}(N = 3) = 3/5$, and according to Lemma 3, we simulate from the law $\widetilde{P}(\{X_i\} \in \cdot | N = n)$ by choosing $i \in \{1, \ldots, n\}$ uniformly at random, sampling $X_i \stackrel{\mathscr{D}}{=} \widetilde{\chi} - \widetilde{\tau}$, where $\widetilde{\chi} \sim \text{Exponential}(5 - \alpha)$ is independent of $\widetilde{\tau} \sim \text{Exponential}(1/4 + \alpha)$, then sampling $X_j \stackrel{\mathscr{D}}{=} \chi - \tau$ independently for $j \neq i$.

Table 2: Numerical results for the unbiased vs. hybrid experiment.

t	$Z(t)$	Time	$\hat{Z}_{K,m}(t)$	Time	Total time	Rel. bias
1	1.8304e-03	5.95	1.8719e-03	2.13	6.20	0.0227
2	2.4038e-05	76.93	2.4425e-05	2.19	6.25	0.0161
3	3.1800e-07	133.84	3.3133e-07	2.81	6.88	0.0419
4	4.4749e-09	700.01	4.3597e-09	3.14	7.21	0.0257
5	5.7474e-11	1643.12	5.6500e-11	3.60	7.67	0.0170
6	7.3680e-13	5761.24	7.6879e-13	4.06	8.13	0.0434
7	1.0007e-14	46447.66	1.0447e-14	4.85	8.92	0.0440

The results of this experiment are displayed in Table 2 for a range of t values, and they include the estimates and run times (in seconds) for both the biased and hybrid procedures. Both the unbiased estimates $Z(t)$ and the hybrid estimates $\hat{Z}_{K,m}(t)$ are based on a sample size of 5000. For the population dynamics algorithm in the hybrid procedure, $K = 20$ and $m = 5000$ are chosen, and the same empirical CDFs are used for each value of t in the table. For the hybrid estimates, two times are listed: the time in seconds to generate 5000 copies of $\hat{Z}_{K,m}(t)$ after the CDFs are already generated (column 5) and that time added to the time it took to generate the empirical CDFs, which was 4.07 seconds (column 6). Notably, when $t = 1$, the procedures do not differ by much in time, but on the other extreme of $t = 7$, the hybrid estimator can be generated in under 10 seconds while $Z(t)$ takes over 12 hours. The final column of Table 2 gives estimates of the relative bias in the hybrid estimates, which is calculated

$$\text{Rel. Bias}\left(\hat{Z}_{K,m}(t)\right) \approx \frac{\left|\hat{Z}_{K,m}(t) - Z(t)\right|}{Z(t)},$$

with sample averages from the 5000 iterations used. As can be seen, despite the drastic reduction in simulation time, the hybrid estimates are still quite accurate.

We conclude with an example where the offspring distribution is chosen to have very large mean. Again, N and $\{X_i\}$ are taken to be independent with $\{X_i\}$ i.i.d. Normal$(-5,1)$ and $N \sim \text{Uniform}\{1,2,\ldots,99\}$, so that $E[N] = 50$. In this case, $\alpha = 9.14$ and $\mu = 4.14$. Under the change of measure, N has mass function $\widetilde{P}(N = n) = n/4950$, and the $\{X_i\}$ are simulated according to Lemma 3 with all but one generated i.i.d. from a Normal$(-5,1)$ distribution and one chosen at random generated from the tilted distribution Normal$(\mu,1)$. Estimates for a range of t values based on a sample size of 5000 are listed in Table 3 along with the computation time (in seconds) for each estimate apart from sampling the CDFs and the total computation time (i.e., with the time to generate the CDFs included, which was 19.21 seconds). Again, $K = 20$ and $m = 5000$, and the same empirical CDFs are used for each value of t.

Table 3: Numerical results for $E[N] = 50$.

t	$\hat{Z}_{K,m}(t)$	Time	Total time
1	3.6419e-08	17.25	36.46
2	6.1537e-11	15.80	35.01
3	3.1472e-14	23.89	43.10
4	5.4582e-18	33.05	52.26
5	4.0702e-22	33.89	53.10

REFERENCES

Asmussen, S. 2003. *Applied Probability and Queues*. New York: Springer.

Basrak, B., M. Conroy, M. Olvera-Cravioto, and Z. Palmowski. 2022. "Importance Sampling for Maxima on Trees". *Stochastic Processes and Their Applications* 148:139–179.

Iglehart, D. L. 1972. "Extreme Values in the GI/G/1 Queue". *Annals of Mathematical Statistics* 43(2):627–635.

Jelenković, P., and M. Olvera-Cravioto. 2012a. "Implicit Renewal Theory and Power Tails on Trees". *Advances in Applied Probability* 44(2):528–561.

Jelenković, P., and M. Olvera-Cravioto. 2012b. "Implicit Renewal Theory for Trees with General Weights". *Stochastic Processes and Their Applications* 122(9):3209–3238.

Jelenković, P., and M. Olvera-Cravioto. 2015. "Maximums on Trees". *Stochastic Processes and Their Applications* 125:217–232.

Karpelevich, F., M. Kelbert, and Y. Suhov. 1994. "Higher-order Lindley Equations". *Stochastic Processes and Their Applications* 53:65–96.

Lyons, R. 2018. "A Simple Path to Biggins' Martingale Convergence for Branching Random Walk". In *Classical and Modern Branching Processes*, 217–221. New York: Springer.

Lyons, R., R. Pemantle, and Y. Peres. 1995. "Conceptual Proofs of L log L Criteria for Mean Behavior of Branching Processes". *Annals of Probability* 23(3):1125–1138.

Mezard, M., and A. Montanari. 2009. *Information, Physics, and Computation*. New York: Oxford University Press, Inc.

Olvera-Cravioto, M. 2019. "Convergence of the Population Dynamics Algorithm in the Wasserstein Metric". *Electronic Journal of Probability* 24(61):1–27.

Olvera-Cravioto, M., and O. Ruiz-Lacedelli. 2021. "Stationary Waiting Time in Parallel Queues with Synchronization". *Mathematics of Operations Research* 46(1):1–27.

Siegmund, D. 1976. "Importance Sampling in the Monte Carlo Study of Sequential Tests". *Annals of Statistics* 3(1):673–684.

AUTHOR BIOGRAPHIES

MICHAEL CONROY is a Postdoctoral Research Associate in the Department of Mathematics at the University of Arizona. He earned his Ph.D. in Statistics and Operations Research at the University of North Carolina, Chapel Hill in 2021. He is interested in stochastic simulation and rare event estimation, as well as large deviations, stochastic control, and interacting particle systems. His e-mail address is michaelconroy@math.arizona.edu and his website is https://sites.google.com/math.arizona.edu/michaelconroy/.

MARIANA OLVERA-CRAVIOTO is an Associate Professor in the Department of Statistics and Operations Research at the University of North Carolina, Chapel Hill. She earned her Ph.D. from the Department of Management Science and Engineering at Stanford University. Her research is in applied probability. In particular, she works on problems involving complex networks and heavy-tailed phenomena. Her e-mail address is molvera@email.unc.edu and her website is https://molvera.web.unc.edu/.