Importance Splitting for Finite-Time Rare Event Simulation

Guangxin Jiang, Michael C. Fu

Abstract

In this note, a general framework is proposed for using importance splitting to estimate rare event probabilities with finite time constraints. We prove that the splitting estimator is unbiased and characterize the optimal splitting curves. A new unbiased estimator with truncated sample paths is proposed to improve computational efficiency, and a pilot algorithm is provided to determine the optimal truncation and splitting curves. Numerical examples illustrate the optimality of the splitting curves and the effectiveness of the new estimator.

Index Terms

rare event simulation; importance splitting; truncated sample paths; variance reduction

I. Introduction

Estimating rare event probabilities via simulation, which has important applications in various types of networks, electrical power systems, and financial industries (e.g., [1], [2] and [3]), is challenging because the relative error of consistent or unbiased estimators diverges to infinity as the rare event probability tends to zero; see [4] and [5] for more details. As a result, variance reduction techniques are essential for making rare event estimation computationally practical. Importance splitting is a variance reduction technique that attempts to generate more paths with the potential to hit the rare event set; see [6], [4], [7], [8], [9], and a related method RESTART [10]. Recent work includes [1], [5], [11], [12], [13], [14], [15], [16] and [17]. Importance splitting has broad applicability, since less distributional information for the model is needed compared with importance sampling, although the latter is generally more efficient.

Most work on importance splitting focuses on estimating the probability of hitting a set without time constraints, where the underlying process is a discrete state space Markov process. When a simulated path hits a splitting level, which is determined by a so-called importance function, the path splits a given number of times. [6] and [4] investigated the asymptotic properties of the splitting method in the fixed splitting setting. [7] provided more details in choosing the importance function and implementation in both fixed splitting and fixed effort settings. [5] proposed a fixed number of successes setting and provided a way to estimate the importance function. The efficiency of such methods is highly dependent on the choice of the importance function, which may be difficult to determine. The setting in this note considers the probability of hitting a set *prior* to a given constant time, for which the classical methods for choosing the optimal splitting levels would require state augmentation, leading to a higher-dimensional problem. The finite-time setting has many applications. For example, in risk management, this finite time may be the date at which a portfolio loss risk needs to be reported to shareholders. This note is the first work to explicitly address importance splitting and truncation in the finite-time setting directly without state augmentation. We propose a framework for Markov processes to estimate rare event probabilities, and provide new conditions for optimal splitting and truncation *curves* in the finite-time setting. Variants of splitting estimators for the classical settings include fixed splitting [4], fixed effort [7], and fixed number of successes [5]. In this note, we consider both fixed splitting and fixed effort estimators. Truncation techniques have been applied to the classical setting in [18] and [4] (refer to [19] for a survey), whereas we propose a new unbiased truncated estimator in the finite-time setting.

The rest of the note is organized as follows. In Section II, we construct the framework for estimating rare event probabilities in finite time via importance splitting, and provide properties of the splitting estimator and conditions for optimal splitting curves. In Section III, we propose a new truncated estimator and characterize its properties. In Section IV, we propose a pilot algorithm to estimate the optimal splitting and truncation curves. In Section V, numerical examples illustrate the optimality of the splitting curves and the effectiveness of the new estimator. Section VI concludes.

II. Importance Splitting

Let $\{X_t, t \geq 0\}$ be a continuous-time continuous-state Markov process, and f be a real-valued measurable continuous function, with $Z_t := f(X_t)$ the performance of underlying process X_t. Let x_0 be the initial value of $\{X_t, t \geq 0\}$. Let l denote the rare event level, and the goal is to estimate the probability $\Pr\{\tau < \xi\}$, where $\tau = \inf\{t \geq 0 | Z_t \geq l\}$, and in the classical settings ([4], [6], [7]), ξ is the first time that the path $\{Z_t\}$ re-enters a given set, e.g., $\xi = \inf\{t > 0 | Z_t \leq l_0\}$ for some fixed $l_0 \ll l$.

However, in practice, another important setting is where the evolution of the underlying process is terminated by some constant time $\xi = T$, and the objective is to estimate the rare event probability $\gamma = \Pr\{\tau < T\}$. For example, $Z_t = f(X_t)$ could be a financial portfolio, and we are interested in estimating the probability that the portfolio defaults before some maturity date. Although [5] mentions the stopping rule $\tau < T$ when proving the unbiasedness of the splitting estimator, they provide no analysis of this setting nor provide any examples. In this note, we analyze properties of the splitting estimator for the finite-time setting, showing how to determine the splitting curves, and then consider a new estimator with truncated sample paths. Unlike the estimator in [18], the estimator proposed here is unbiased.

Let $l_0 < l_1 < \cdots < l_N = l$ be a sequence of levels for $\{Z_t\}$, and define the inter-level hitting time $\beta_1 = \inf\{t \geq 0 | Z_t \geq l_1, Z_0 = f(x_0)\}$, and $\beta_k = \inf\{t \geq 0 | Z_{t+\tau_{k-1}} \geq l_k, Z_{\tau_{k-1}} = l_{k-1}(\tau_{k-1})\}$ for $k = 2, \ldots, N$, where $\tau_0 = \beta_0 = 0$ and $\tau_k = \sum_{i=1}^{k} \beta_i, k = 1, \ldots, N$. Then, we define the event $D_k := \{\tau_k = \sum_{i=1}^{k} \beta_i < T\}$, which satisfies $D_k \subset D_{k-1}$ for $k = 2, \ldots, N$, and D_N is the rare event. Let $D_0 = \{0 < T\}$, which is a deterministic event with $\Pr\{D_0\} = 1$. Denote $p_i = \Pr\{D_i | D_{i-1}\}$ for $i = 1, \ldots, n$. Since $\Pr\{D_i\} = \Pr\{D_i \cap D_{i-1}\} = \Pr\{D_i | D_{i-1}\} \Pr\{D_{i-1}\}$, we have

$$\gamma = \Pr\{D_N\} = \prod_{i=1}^{N} \Pr\{D_i | D_{i-1}\} = \prod_{i=1}^{N} p_i.$$

Thus, estimating rare event probability $\Pr\{D_N\}$ is decomposed into estimating N "less rare" event probabilities $\Pr\{D_i|D_{i-1}\}, i = 1, \dots, N$.

Given the splitting levels, we say a path is in *Stage* i after it first hits Level $i-1$ and prior to hitting Level i. If a path in Stage i hits Level i, we call this path a *successful path* for Level i, and the corresponding value of X_t is called an *entrance state* for Level i, denoted by S_i. Let n_i be the number of splittings for each successful path for Level i, and R_i be the number of successful paths starting at Level $i-1$ and hitting Level i with $R_0 = 1$. For an illustration, see Fig. 1. The basic splitting procedure is provided in Algorithm 1.

Algorithm 1 Splitting algorithm

Input: Rare event level l, number of levels N, and splitting levels $l_0 < l_1 < \cdots < l_N = l$; number of initial simulation paths n_0, number of splittings n_i when a successful path hits Level $i, i = 1, \dots, N$. Let $R_0 = 1$.

for $i = 1$ to N **do**
  Let $R_i = 0$.
  for $j = 1$ to $R_{i-1} n_{i-1}$ **do**
    if the path hits l_i, it splits n_i times at next stage, and let $R_i = R_i + 1$, else go to next path.
  end for j
end for i

Output: Return the rare event estimator

$$\hat{\gamma} = \prod_{i=1}^{N} \frac{R_i}{R_{i-1} n_{i-1}} = \frac{R_N}{\prod_{i=1}^{N} n_{i-1}}. \tag{1}$$

Remark 1. [7] provides two ways to implement Algorithm 1. The first way is called *global stepping*, where we simulate a path by finishing all the consecutive stages before starting a new path. The second way is called *single stepping*, where we simulate all the paths simultaneously for one stage, and record all the entrance states for the stage, then go to the next stage. In this paper, we use the single stepping implementation, which is especially amenable to parallel computing implementation.

Remark 2. As pointed out by [5], the original proof of the unbiasedness of the estimator given in [7] is not correct in general, and [5] provided a new proof based on constructing a new strong Markov process. In this note, we fix

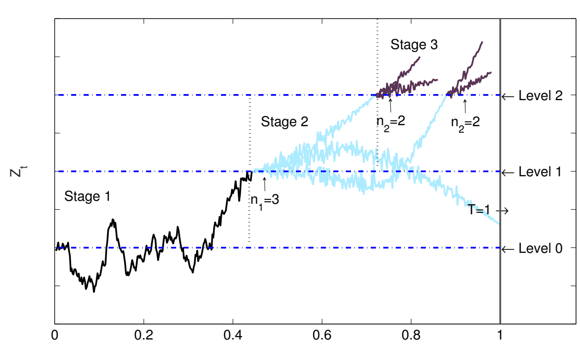

Fig. 1. Illustration of stages, splitting levels, and splitting paths.

the original proof in [7], which is much simpler than the proofs in [5].

A. *Fixed splitting*

Depending on how the $\{n_i\}$ are chosen, importance splitting can be basically divided into either fixed splitting or fixed effort. For fixed splitting, $\{n_i, i = 1, \ldots\}$ are constants (not necessarily the same constant), whereas for fixed effort, $\{n_i, i = 1, \ldots\}$ are chosen to make $n_i R_i$ approximately the same for all i. In the following, we analyze properties of the fixed splitting estimator.

Let $P_{j:i}(x, t)$ be the probability that a path starting at state x at time t at Level i hits Level j before T, and let I_i denote a Bernoulli random variable with success probability $P_{i+1:i}(S_i, \tau_i)$. Thus, given the ith entrance state S_i and τ_i, $\mathbb{E}[I_i] = \tilde{P}_{i+1:i}$, where we introduce the notation $\tilde{P}_{j:i} \triangleq P_{j:i}(S_i, \tau_i)$, since by definition $X_{\tau_i} = S_i$. Let "$\overset{d}{=}$" denote equal in distribution. Then given $R_i = r_i$ and all the i.i.d. entrance states $(S_i^{(j)}, \tau_i^{(j)}) \overset{d}{=} (S_i, \tau_i), j = 1, \ldots, r_i$, and i.i.d. Bernoulli random variables $I_i^{(k,j)} \overset{d}{=} I_i^{(j)}, k = 1, \ldots, n_i$, for $j = 1, \ldots, r_i$, on Level i, where $I_i^{(j)}$ is a Bernoulli random variable with success probability $\tilde{P}_{i+1:i}^{(j)} \triangleq P_{i+1:i}(S_i^{(j)}, \tau_i^{(j)})$,

$$R_{i+1} = \sum_{j=1}^{r_i} \sum_{k=1}^{n_i} I_i^{(k,j)}. \tag{2}$$

Let $\mathbf{S}_i = \left(S_i^{(1)}, \ldots, S_i^{(r_i)}\right)$ and $\boldsymbol{\tau}_i = \left(\tau_i^{(1)}, \ldots, \tau_i^{(r_i)}\right)$, and consider the conditional expectation of R_{i+1}:

$$\begin{aligned} \mathbb{E}[R_{i+1}|R_i, \mathbf{S}_i, \boldsymbol{\tau}_i] &= \mathbb{E}\left[\sum_{j=1}^{R_i} \sum_{k=1}^{n_i} I_i^{(k,j)} \middle| R_i, \mathbf{S}_i, \boldsymbol{\tau}_i \right] \\ &= \sum_{j=1}^{R_i} \sum_{k=1}^{n_i} \mathbb{E}\left[I_i^{(k,j)} \middle| R_i, \mathbf{S}_i, \boldsymbol{\tau}_i \right] = \sum_{j=1}^{R_i} n_i \tilde{P}_{i+1:i}^{(j)}. \end{aligned} \tag{3}$$

When $i = 0$, $\mathbf{S}_0 = x_0$ is initial value, $\boldsymbol{\tau}_0 = \tau_0 = 0$ and $R_0 = 1$. Now we can prove the unbiasedness of the splitting estimator.

Theorem 1. *The estimator given by Equation (1) is an unbiased estimator of γ.*

Proof. The estimator is $\hat{\gamma} = \prod_{i=1}^{N} \hat{p}_i$ with $\hat{p}_i = R_i/(R_{i-1}n_{i-1})$. By Equation (3),

$$\mathbb{E}[R_N|R_{N-1}, \mathbf{S}_{N-1}, \boldsymbol{\tau}_{N-1}] = \sum_{j=1}^{R_{N-1}} n_{N-1} \tilde{P}_{N:N-1}^{(j)},$$

so that the conditional expectation of $\hat{p}_N$ is

$$\begin{aligned} &\mathbb{E}[\hat{p}_N|R_{N-1}, \mathbf{S}_{N-1}, \boldsymbol{\tau}_{N-1}] \\ &= \mathbb{E}\left[\frac{R_N}{R_{N-1}n_{N-1}} \middle| R_{N-1}, \mathbf{S}_{N-1}, \boldsymbol{\tau}_{N-1} \right] \\ &= \frac{1}{R_{N-1}n_{N-1}} \sum_{j=1}^{R_{N-1}} n_{N-1} \tilde{P}_{N:N-1}^{(j)} \\ &= \frac{1}{R_{N-1}} \sum_{j=1}^{R_{N-1}} \tilde{P}_{N:N-1}^{(j)}. \end{aligned} \tag{4}$$

Notice that

$$\begin{aligned}
\tilde{P}^{(j)}_{N:N-2} &= P_{N:N-2}\left(S^{(j)}_{N-2}, \tau^{(j)}_{N-2}\right) \\
&= \mathbb{E}\Big[I^{(j)}_{N-2} I_{N-1}\Big|\mathbf{S}_{N-2}, \boldsymbol{\tau}_{N-2}\Big] \\
&= \mathbb{E}\Big[\mathbb{E}\Big[I^{(j)}_{N-2} I_{N-1}\Big|S_{N-1}, \tau_{N-1}\Big]\Big|\mathbf{S}_{N-2}, \boldsymbol{\tau}_{N-2}\Big] \\
&= \mathbb{E}\Big[I^{(j)}_{N-2}\mathbb{E}\big[I_{N-1}\big|S_{N-1}, \tau_{N-1}\big]\Big|\mathbf{S}_{N-2}, \boldsymbol{\tau}_{N-2}\Big] \\
&= \mathbb{E}\Big[I^{(j)}_{N-2}\tilde{P}_{N:N-1}\Big|\mathbf{S}_{N-2}, \boldsymbol{\tau}_{N-2}\Big]. \qquad (5)
\end{aligned}$$

Next consider

$$\begin{aligned}
&\mathbb{E}\left[\hat{p}_N \hat{p}_{N-1} \right| R_{N-2}, \mathbf{S}_{N-2}, \boldsymbol{\tau}_{N-2}] \\
&= \mathbb{E}\Big[\ \mathbb{E}\Big[\hat{p}_N \hat{p}_{N-1}\Big|R_{N-1}, \mathbf{S}_{N-1}, \boldsymbol{\tau}_{N-1}\Big]\Big|\ R_{N-2}, \mathbf{S}_{N-2}, \boldsymbol{\tau}_{N-2}\Big] \\
&= \mathbb{E}\Big[\ \mathbb{E}\Big[\hat{p}_N\Big|R_{N-1}, \mathbf{S}_{N-1}, \boldsymbol{\tau}_{N-1}\Big]\hat{p}_{N-1}\Big|\ R_{N-2}, \mathbf{S}_{N-2}, \boldsymbol{\tau}_{N-2}\Big] \\
&= \mathbb{E}\left[\frac{1}{R_{N-1}}\sum_{j=1}^{R_{N-1}} \tilde{P}^{(j)}_{N:N-1}\frac{R_{N-1}}{R_{N-2} n_{N-2}}\right|R_{N-2}, \mathbf{S}_{N-2}, \boldsymbol{\tau}_{N-2}\Bigg] \\
&= \frac{1}{n_{N-2} R_{N-2}}\mathbb{E}\Big[\tilde{P}_{N:N-1} R_{N-1}\Big|R_{N-2}, \mathbf{S}_{N-2}, \boldsymbol{\tau}_{N-2}\Big] \\
&\quad \text{since } (S^{(j)}_{N-1}, \tau^{(j)}_{N-1}) \text{ are i.i.d., and independent of } R_{N-1} \\
&= \frac{1}{n_{N-2} R_{N-2}}\mathbb{E}\left[\sum_{j=1}^{R_{N-2}}\sum_{k=1}^{n_{N-2}} I^{(k,j)}_{N-2}\tilde{P}_{N:N-1}\right|R_{N-2}, \mathbf{S}_{N-2}, \boldsymbol{\tau}_{N-2}\Bigg] \\
&= \frac{1}{R_{N-2}}\sum_{j=1}^{R_{N-2}}\mathbb{E}\Big[I^{(j)}_{N-2}\tilde{P}_{N:N-1}\Big|\mathbf{S}_{N-2}, \boldsymbol{\tau}_{N-2}\Big]. \qquad (6)
\end{aligned}$$

Then by Equations (5) and (6),

$$\mathbb{E}\left[\hat{p}_{N-1} \hat{p}_N \right| R_{N-2}, \mathbf{S}_{N-2}, \boldsymbol{\tau}_{N-2}] = \frac{1}{R_{N-2}}\sum_{j=1}^{R_{N-2}} \tilde{P}^{(j)}_{N:N-2} \qquad (7)$$

Continuing by induction, following the analysis used to establish Equations (4) and (7),

$$\begin{aligned}
\mathbb{E}[\hat{\gamma}] &= \mathbb{E}[\hat{p}_N \hat{p}_{N-1} \cdots \hat{p}_1 | R_0 = 1, \mathbf{S}_0 = x_0, \boldsymbol{\tau}_0 = 0] \\
&= \frac{1}{R_0}\sum_{j=1}^{R_0} \tilde{P}^{(j)}_{N:0} = P_{N:0}(x_0, 0) = \gamma,
\end{aligned}$$

and the theorem is proved. □

[7] claimed that the unconditional expectation $\mathbb{E}[\hat{p}_i] = \Pr\{D_i | D_{i-1}\} = p_i$, whereas our proof makes clear (see Equation (4)) that probability depends on the entrance states and the success paths of previous level, i.e., the unconditional expectation of Equation (4) does not necessarily equal p_i. Thus, our proof uses induction on the

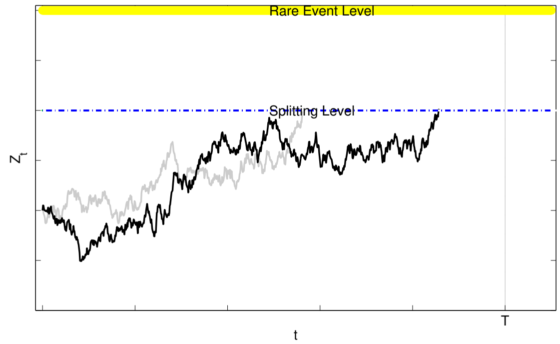

Fig. 2. The importance of the paths.

conditional expectation to establish the unbiasedness, which does not depend on the choice of the splitting levels. However, good splitting levels can make the estimator more efficient.

B. Splitting curves for finite time

When to split in the finite-time setting differs from the classical splitting setting, because in the finite-time setting, a path loses its importance not only when it strays far away from the rare event set but also when it approaches the termination time. For example, in Fig. 2, both paths hit the splitting level, but the black path is less likely to hit the rare event level than the gray one, because it hits the splitting level much closer to the termination time. By Equation (1), minimizing the variance of estimator $\hat{\gamma}$ is equivalent to minimizing the variance of R_N, the number of final-level successful paths. Based on those observations, we derive the properties that the splitting curve should satisfy to minimize the variance of the estimator. To distinguish from the classical constant splitting level l_i, we use $\mathcal{L}_i = \{l_i(t)\}$ to denote the ith splitting *curve*, which is a function of t, and Algorithm 1 can be modified for the finite time setting accordingly by replacing l_i with $\mathcal{L}_i$.

Theorem 2. *The set of optimal splitting curves $\mathcal{L}_i \equiv \{l_i(t)\}, i = 1, \ldots, N-1$, that minimizes the variance of estimator $\hat{\gamma}$, must satisfy $P_{N:i}\left(l_i(t), t\right) = c_i, 0 \le t \le T$, for constants $c_i > 0$, where $P_{N:i}\left(x, t\right)$ is the probability that a path starting from state x at time t at Level i hits the rare event set.*

Proof. We first analyze how the entrance state affects the variance of R_{i+1}. Given $(R_i, \mathbf{S}_i, \boldsymbol{\tau}_i)$, R_{i+1} follows a binomial distribution $Bin(n_i, P_{i+1:i}(S_i, \tau_i))$, and by the conditional variance decomposition,

$$\begin{aligned}&\text{Var}[R_{i+1}|R_i = r_i] = \\ &\mathbb{E}[\text{Var}(R_{i+1}|R_i = r_i, \mathbf{S}_i, \boldsymbol{\tau}_i)] + \text{Var}(\mathbb{E}[R_{i+1}|R_i = r_i, \mathbf{S}_i, \boldsymbol{\tau}_i]).\end{aligned} \tag{8}$$

For the first term,

$$\begin{aligned}&\mathbb{E}[\text{Var}(R_{i+1}|R_i = r_i, \mathbf{S}_i, \boldsymbol{\tau}_i)] \\ &= \mathbb{E}\left[\text{Var}\left(\sum_{j=1}^{r_i}\sum_{k=1}^{n_i} I_i^{(k,j)}\right)\right] = \mathbb{E}\left[\sum_{j=1}^{r_i}\text{Var}\left(\sum_{k=1}^{n_i} I_i^{(k,j)}\right)\right] \\ &= \mathbb{E}\left[\sum_{j=1}^{r_i} n_i \tilde{P}_{i+1:i}^{(j)}\left(1 - \tilde{P}_{i+1:i}^{(j)}\right)\right]\end{aligned}$$

$$= r_i n_i \mathbb{E}[\tilde{P}_{i+1:i}] - r_i n_i \mathbb{E}\left[\left(\tilde{P}_{i+1:i}\right)^2\right]$$
$$= r_i n_i q_{i+1} - r_i n_i \left(\text{Var}\left(\tilde{P}_{i+1:i}\right) + q_{i+1}^2\right), \tag{9}$$

where $q_{i+1} \triangleq \mathbb{E}[\tilde{P}_{i+1:i}]$. For the second term, by Equations (2) and (3),

$$\text{Var}\left(\mathbb{E}[R_{i+1}|R_i = r_i, \mathbf{S}_i, \boldsymbol{\tau}_i]\right) = \text{Var}\left(\sum_{j=1}^{r_i} n_i \tilde{P}_{i+1:i}^{(j)}\right)$$
$$= r_i n_i^2 \text{Var}\left(\tilde{P}_{i+1:i}\right). \tag{10}$$

Substituting (9) and (10) into (8):

$$\text{Var}[R_{i+1}|R_i = r_i]$$
$$= r_i n_i q_{i+1} - r_i n_i q_{i+1}^2 + r_i n_i (n_i - 1)\text{Var}(\tilde{P}_{i+1:i}),$$

so

$$\mathbb{E}\left[R_{i+1}^2|R_i = r_i\right]$$
$$= \text{Var}[R_{i+1}|R_i = r_i] + \left(\mathbb{E}[R_{i+1}|R_i = r_i]\right)^2$$
$$= r_i\left(n_i q_{i+1} - n_i q_{i+1}^2 + n_i(n_i - 1)\text{Var}(\tilde{P}_{i+1:i})\right)$$
$$+ r_i^2 n_i^2 q_{i+1}^2. \tag{11}$$

For notational convenience, denote $n_i q_{i+1} - n_i q_{i+1}^2 + n_i(n_i - 1)\text{Var}(\tilde{P}_{i+1:i}) \triangleq A_i$ and $n_i^2 q_{i+1}^2 \triangleq B_i$, which are both deterministic values. Then (11) becomes

$$\mathbb{E}\left[R_N^2|R_{N-1}\right] = R_{N-1}A_{N-1} + R_{N-1}^2 B_{N-1}.$$

Consider

$$\mathbb{E}\left[R_N^2|R_{N-2} = r_{N-2}\right] = \mathbb{E}\left[\mathbb{E}\left[R_N^2|R_{N-1}, R_{N-2} = r_{N-2}\right]\right]$$
$$= \mathbb{E}\left[R_{N-1}|R_{N-2} = r_{N-2}\right] A_{N-1}$$
$$+ \mathbb{E}\left[R_{N-1}^2|R_{N-2} = r_{N-2}\right] B_{N-1}$$

Substituting (11) by letting $i = N - 2$,

$$= \mathbb{E}\left[R_{N-1}|R_{N-2} = r_{N-2}\right] A_{N-1}$$
$$+ r_{N-2}A_{N-2}B_{N-1} + r_{N-2}^2 B_{N-1}B_{N-2}$$

Similarly,

$$\mathbb{E}\left[R_N^2|R_{N-3} = r_{N-3}\right] = \mathbb{E}\left[R_{N-1}|R_{N-3} = r_{N-3}\right] A_{N-1}$$
$$+ \mathbb{E}\left[R_{N-2}|R_{N-3} = r_{N-3}\right] A_{N-2}B_{N-1}$$
$$+ r_{N-3}A_{N-3}B_{N-1}B_{N-2} + r_{N-3}^2 B_{N-1}B_{N-2}B_{N-3}.$$

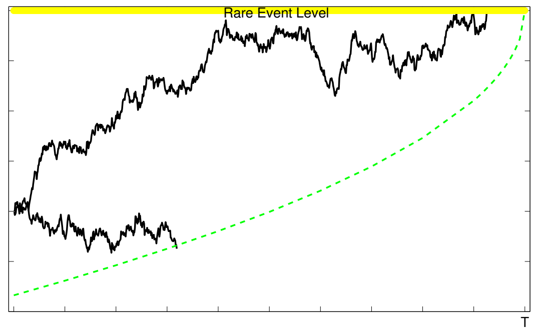

Fig. 3. Truncated path strategy.

Continuing by iteration,

$$\begin{aligned}
&\mathbb{E}\left[R_N^2|R_0=1\right]=\mathbb{E}\left[R_{N-1}|R_0=1\right]A_{N-1}\\
&+\sum_{i=1}^{N-2}\left(\mathbb{E}[R_i|R_0=1]A_i\prod_{j=i}^{N-2}B_{j+1}\right)\\
&+\mathbb{E}[R_1^2|R_0=1]\prod_{j=0}^{N-2}B_{j+1}\\
&=\mathbb{E}[R_{N-1}|R_0=1]\Big(n_{N-1}q_N-n_{N-1}q_N^2\\
&+n_{N-1}(n_{N-1}-1)\text{Var}\left(\tilde{P}_{N:N-1}\right)\Big)\\
&+\sum_{i=1}^{N-2}\Big\{\mathbb{E}[R_i|R_0=1]\Big(n_iq_{i+1}-n_iq_{i+1}^2+n_i(n_i-1)\times\\
&\text{Var}\left(\tilde{P}_{i+1:i}\right)\Big)\prod_{j=i}^{N-2}n_{j+1}^2q_{j+2}^2\Big\}+q_1\prod_{j=0}^{N-2}n_{j+1}^2q_{j+2}^2.
\end{aligned}$$

Minimizing the variance is equivalent to minimizing $\mathbb{E}(R_N^2)$. Since $n_i>1$ it is optimal to let $\text{Var}(\tilde{P}_{i+1:i})=\text{Var}(P_{i:i-1}\left(S_{i-1},\tau_{i-1}\right))=0$, i.e., $P_{i:i-1}\left(l_{i-1}(t),t\right)$ is a constant, which is equivalent to $P_{N:i}\left(l_i(t),t\right)=c_i, 0\leq t\leq T$, where c_i is a constant, for all $i=1,\ldots,N-1$. □

The finite-time case can also be modeled by augmenting the state of the original process, i.e., for Markov process $\{X_t\}$, the finite-time case could redefine the state as $\tilde{X}_t=(X_t,t)$, so [7] or [20] could be applied to find the optimal splitting levels, but the increased dimensionality of the process makes it computationally more challenging.

III. The Truncation Estimator

In our setting, with a terminal time T, it may be computationally advantageous to stop early if it becomes clear that a path is highly unlikely to hit the target set, complementing the splitting of paths that look promising. Let B be the truncation boundary, i.e., when a path hits B, the path is terminated, e.g., the dashed green line in Fig. 3. We assume that each point on B has the same probability γ_K of hitting the rare event set (prior to T) and B is chosen such that $\gamma_K\ll\gamma$. Under the truncation setting, let R_i' be the number of successful paths, starting at Level $i-1$ and hitting Level i, and K_i be the number of killed paths between Level $i-1$ and Level i.

The next theorem shows that the truncation estimator (12) is also unbiased.

Algorithm 2 Truncated splitting algorithm

Input: Rare event level l, truncation boundary B, correction probability γ_K, number of splitting curves N, and splitting curves $l_0 < \mathcal{L}_1 < \cdots < \mathcal{L}_N = l$; number of initial simulation paths n_0, number of splittings n_i when a successful path hits splitting curve $\mathcal{L}_i, i = 1, \ldots, N-1$.

for $i = 1$ to N **do**
  Let $R'_i = 0$ and $K_i = 0$.
  for $j = 1$ to $R'_{i-1} n_{i-1}$ **do**
    if the path hits l_i, it splits n_i times at next stage, and let $R'_i = R'_i + 1$.
    if the path hits B, terminate this path, and let $K_i = K_i + 1$.
  end for j
end for i

Output: Return the estimator

$$\hat{\gamma}_{tr} = \frac{R'_N}{\prod_{i=0}^{N-1} n_i} + \gamma_K \sum_{i=1}^{N} \frac{K_i}{\prod_{j=0}^{i-1} n_j}. \tag{12}$$

Theorem 3. *Under the setting in Theorem 2, the estimator $\hat{\gamma}_{tr}$ is an unbiased estimator of γ.*

Proof. Under the setting of Theorem 2, the probability of any point starting from Level $i-1$ hitting Level i is a constant (denoted by p_i). For $i = 1, \ldots, N-1$, consider the reduction in the number of Stage i successful paths after truncation, which can be divided into two sets: paths killed in Stage i and paths killed prior to Stage i. Thus, conditioning on the number of paths reaching Stage i:

$$\begin{aligned} &\mathbb{E}\left[R_i - R'_i \middle| R_{i-1} = r_{i-1}, R'_{i-1} = r'_{i-1}\right] \\ &= \mathbb{E}\left[K_i \middle| R_{i-1} = r_{i-1}, R'_{i-1} = r'_{i-1}\right] \frac{\gamma_K}{\prod_{j=i+1}^{N} p_j} \\ &+ \left(r_{i-1} - r'_{i-1}\right) n_{i-1} p_i, \end{aligned} \tag{13}$$

where $\gamma_K / \prod_{j=i+1}^{N} p_j$ is the probability that a path starting from truncation boundary B hits Level i. Similarly, for Level N,

$$\begin{aligned} &\mathbb{E}\left[R_N - R'_N \middle| R_{N-1} = r_{N-1}, R'_{N-1} = r'_{N-1}\right] \\ &= \mathbb{E}\left[K_N \middle| R_{N-1} = r_{N-1}, R'_{N-1} = r'_{N-1}\right] \gamma_K \\ &+ \left(r_{N-1} - r'_{N-1}\right) n_{N-1} p_N. \end{aligned}$$

Next, we consider

$$\begin{aligned} &\mathbb{E}\left[R_N - R'_N \middle| R_{N-2} = r_{N-2}, R'_{N-2} = r'_{N-2}\right] \\ &= \mathbb{E}\Big[\mathbb{E}\big[R_N - R'_N \\ &\big| R_{N-1}, R'_{N-1}, R_{N-2} = r_{N-2}, R'_{N-2} = r'_{N-2}\big]\Big] \end{aligned}$$

$$= \mathbb{E}[K_N|R_{N-2}=r_{N-2},R'_{N-2}=r'_{N-2}]\gamma_K + n_{N-1}p_N\times$$
$$\mathbb{E}\left[R_{N-1}-R'_{N-1}\right| R_{N-2}=r_{N-2},R'_{N-2}=r'_{N-2}\Big]$$
$$= \mathbb{E}[K_N|R_{N-2}=r_{N-2},R'_{N-2}=r'_{N-2}]\gamma_K$$
$$+\Big(\mathbb{E}[K_{N-1}|R_{N-2}=r_{N-2},R'_{N-2}=r'_{N-2}]\frac{\gamma_K}{p_N}$$
$$+\left(r_{N-2}-r'_{N-2}\right)n_{N-2}p_{N-1}\Big)n_{N-1}p_N$$

by applying (13) with $i=N-1$,

$$= \mathbb{E}[K_N|R_{N-2}=r_{N-2},R'_{N-2}=r'_{N-2}]\gamma_K$$
$$+\mathbb{E}[K_{N-1}|R_{N-2}=r_{N-2},R'_{N-2}=r'_{N-2}]n_{N-1}\gamma_K$$
$$+\left(r_{N-2}-r'_{N-2}\right)n_{N-1}n_{N-2}p_Np_{N-1}.$$

Denoting $k_i=\mathbb{E}[K_i|R_0=1,R'_0=1]=\mathbb{E}[K_i]$, then by iteration,

$$\mathbb{E}[R_N-R'_N]=\mathbb{E}[R_N-R'_N|R_0=1,R'_0=1]=\gamma_K\times$$
$$\left(k_N+k_{N-1}n_{N-1}+k_{N-2}n_{N-1}n_{N-2}+\cdots+k_1\prod_{q=1}^{N-1}n_q\right)$$
$$=\gamma_K\left(k_N+\sum_{l=1}^{N-1}\left(k_l\prod_{q=l}^{N-1}n_q\right)\right).$$

Let $\hat{\gamma}'=R'_N/\prod_{i=0}^{N-1}n_i$. Then

$$\mathbb{E}[\hat{\gamma}-\hat{\gamma}']=\mathbb{E}\left[\frac{R_N}{\prod_{i=0}^{N-1}n_i}-\frac{R'_N}{\prod_{i=0}^{N-1}n_i}\right]$$
$$=\frac{1}{\prod_{i=0}^{N-1}n_i}\mathbb{E}[R_N-R'_N]=\gamma_K\sum_{i=1}^{N}\frac{k_i}{\prod_{j=0}^{i-1}n_j}.$$

So

$$\mathbb{E}[\hat{\gamma}_{tr}]=\mathbb{E}\left[\hat{\gamma}'+\gamma_K\sum_{i=1}^{N}\frac{K_i}{\prod_{j=0}^{i-1}n_j}\right]$$
$$=\mathbb{E}\left[\hat{\gamma}'\right]+\gamma_K\sum_{i=1}^{N}\frac{k_i}{\prod_{j=0}^{i-1}n_j}=\mathbb{E}[\hat{\gamma}]=\gamma.$$

The last equality holds by Theorem 1. □

IV. Pilot Algorithm

[7] and [4] suggest using a pilot algorithm to determine the splitting parameters such as the splitting levels, the number of splitting levels, and the number of splittings for each successful point n_i. Here, we propose a new pilot algorithm given p_i and n_i to determine the splitting curves $\mathcal{L}_i$, which differs from previous work, because the curves are time dependent. We first make the following assumption.

Assumption 1. *$\{Z_t\}$ is a Markov process with stationary increments, i.e., $Z_t-Z_s\overset{d}{=}Z_{t-s},\forall s,t>0$.*

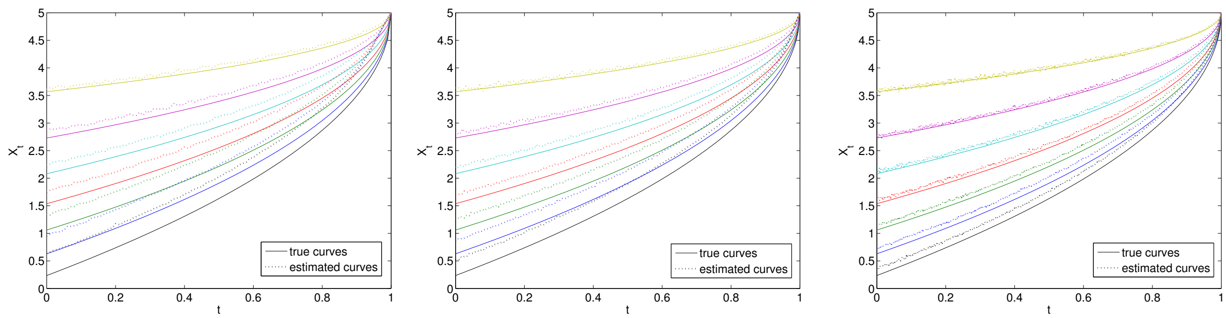

Fig. 4. Pilot run curves; $m = 100$ for left panel; $m = 200$ for middle panel; $m = 800$ for right panel.

By Theorem 2, the splitting curves should satisfy the condition that all paths starting from any point on the Level i splitting curve share a constant probability c_i of hitting the rare event set, which implies that all the paths starting from any point on $\mathcal{L}_{i-1}$ have a constant probability $c_i/c_{i-1} = p_i$ of hitting the next level $\mathcal{L}_i$. Therefore, we can determine each splitting curve backwards step by step, i.e., given $\mathcal{L}_i$, we use a pilot run to estimate $\mathcal{L}_{i-1}$

Suppose that a path hits Level $i-1$ at s, i.e., the entrance state is $x_s := l_{i-1}(s)$. Let $\tau = \inf\{t \geq s | Z_t \geq l_i(t)\}$ be the first time that the path hits Level i. Then, by Theorem 2, for any $t \in [s, T]$,

$$\Pr\{\tau \leq T | Z_s = x_s\} = p_i,$$

which is equivalent to

$$\begin{aligned}
& \Pr\left\{ \max_{t\in(s,T]} \{Z_t - l_i(t)\} \geq 0 \middle| Z_s = x_s \right\} \\
= \quad & \Pr\left\{ \max_{t\in(s,T]} \{x_s + Z_{t-s} - l_i(t)\} \geq 0 \right\} \\
= \quad & \Pr\left\{ \min_{t\in(s,T]} \{l_i(t) - Z_{t-s}\} \leq x_s \right\} = p_i.
\end{aligned}$$

The first equality holds because of the stationary increments property, and the second equality holds since, for fixed s, x_s is a constant value. Let $B_s = \min_{t\in(s,T]}\{l_i(t) - Z_{t-s}\}$, and then we need to estimate the p_i-quantile of B_s. The pilot run algorithm is presented in Algorithm 3.

Remark 3. p_i can be estimated by $p_i = \hat{\gamma}^{1/N}, i = 1, \ldots, N$, where $\hat{\gamma}$ is obtained by another pilot run with some other levels, e.g., linear levels with ending points at L. Alternatively, we can artificially set $p_i = c$, and input N is determined by this probability c or artificially setting N to a constant.

To estimate the truncation boundary, we carry out an additional pilot run after $i = 1$, and set $p_i = c_\alpha$ such that $\gamma_K = \prod_{i=1}^{N} p_i c_\alpha < \gamma$ (or $\gamma_K = c^N c_\alpha < \gamma$). Specifically, when obtaining $l_1(t)$, we do another loop from Step 3 to Step 6 of Algorithm 3 with $p_i = c_\alpha$, and obtain $l_0(t)$, which is the truncation boundary. Notice that if the truncation boundary is estimated, the truncated estimator $\hat{\gamma}_{tr}$ given by Equation (12) is no longer unbiased, because the bias may be introduced by the estimated truncation boundary.

Algorithm 3 Pilot algorithm for determining splitting curves

Input: rare event level l; number of splitting curves N; hitting probabilities $p_i, i = 1, \dots, N$, ; terminal time T, discretization points $\{0 = t_0 < t_1 < \cdots < t_m = T\}$, and number of pilot runs M.

Initialization: initial value Z_0; set $i = N$ and $l_N = l$.

for $i = N$ to 1 **do**
    for $j = 0$ to m **do**
        Generate $Z_{t_k - t_j}$ from $k = j + 1$ to m.
        Let $B_{t_k} = \min_{k=j+1\dots,m}\{l_i(t_k) - Z_{t_k - t_j}\}$.
        Repeat M times to obtain $B^s_{t_k}, s = 1, \dots, M$, and the order statistics $\{B^{(1)}_{t_k} \le B^{(2)}_{t_k} \le \cdots \le B^{(M)}_{t_k}\}$.
        Set $l_{i-1}(t_j) = B^{(\lceil p_i M \rceil)}_{t_j}$.
    end for j
end for i

Output: Discrete splitting curves $l_i(t_j), i = 1, \dots, N, j = 0, \dots, m$, where $l_N(t_j) = l$ for all j.

V. NUMERICAL EXAMPLES

In this section, we first test our pilot algorithm, then verify the optimality of the splitting curves derived in Section II and test the effectiveness of the truncation estimator proposed in Section III. Throughout, we take $p_i = \gamma^{1/N}, i = 1, \dots, N-1$ and $n_i = \lceil 1/p_i \rceil, i = 1, \dots, N-1$.

A. *Testing the pilot algorithm*

Let $Z_t = X_t$ be a standard Brownian motion, and consider the probability that Z_t hits $l = 5$ before time $T = 1$. Let $N = 8$, $M = 2000$. By Theorem 2 and [21], we can calculate γ and the splitting curves analytically ($\gamma = 5.73 \cdot 10^{-7}$, $p_i = 0.169$), and Fig. 4 shows the results from the pilot runs for $m = 100, 200, 800$, indicating how finer discretizations lead to more accurate estimated curves.

B. *Optimality of the splitting curves*

We compare the optimal splitting curves (denoted SC) derived in this note with the optimal constant levels (denoted CL) for classical splitting setting. For $l = 5$ and $T = 1$ again (so $\gamma = 5.73 \cdot 10^{-7}$ again), $N = 5$ (so $p_i = 0.056$ and $n_i = 18$), $m = 100$, $n_0 = 2000$, the results are graphed in Fig. 5. For the estimated splitting curves (denoted EC) via Algorithm 3, we take the fixed inter-level probability $c = 0.04$.

To balance the variance and computational effort when comparing estimator performance, we consider the *work-normalized variance* as in [6], which is defined as $\mathrm{Var}(\hat{\gamma})w$, where w is the average computational effort per replication. More specifically, for a replication, if the path does not split, then $w = T$; otherwise $w = t_1 + \sum_{i=2}^{N}(t_i - t_{i-1})n_{i-1}$, where t_i is the epoch at which it splits on hitting Level i. Fig. 6 shows the variance, average computational effort w, and the work-normalized variance with respect to the number of initial simulations n_0 based on 1000 independent replications. Both the EC and SC estimators have smaller variance than the CL estimator, and the superiority is more significant for work-normalized variance. Since the average cost is almost linearly increasing

TABLE I

COMPARISON OF ESTIMATORS BASED ON $n_0 = 2000$ AND 100 INDEPENDENT MACROREPLICATIONS. C1: $m = 400$, $l = 3$, $N = 5$, $n_i = 4$. C2: $m = 100$, $l = 5$, $N = 5$, $n_i = 18$. C3: $m = 400$, $l = 5$, $N = 5$, $n_i = 18$. C4: $m = 1000$, $l = 5$, $N = 5$, $n_i = 18$. C5: $m = 100$, $l = 7$, $N = 8$, $n_i = 29$ (STANDARD ERRORS IN PARENTHESES). DEFINE $w^* = w/n_0$.

		SC			EC			T-SC			T-EC		
	P_{true}	P_e (std err)	w^*	RE (%)	P_e (std err)	w^*	RE (%)	P_e (std err)	w^*	RE (%)	P_e (std err)	w^*	RE (%)
C1	$2.70 \cdot 10^{-3}$	$2.50 \cdot 10^{-3}$ $(.017 \cdot 10^{-3})$	4.3	7.4	$2.52 \cdot 10^{-3}$ $(.019 \cdot 10^{-3})$	3.6	6.6	$2.78 \cdot 10^{-3}$ $(.015 \cdot 10^{-3})$	2.2	2.9	$2.77 \cdot 10^{-3}$ $(.015 \cdot 10^{-3})$	1.9	2.6
C2	$5.73 \cdot 10^{-7}$	$4.47 \cdot 10^{-7}$ $(.081 \cdot 10^{-7})$	4.3	22.0	$4.39 \cdot 10^{-7}$ $(.065 \cdot 10^{-7})$	5.5	23.3	$4.41 \cdot 10^{-7}$ $(.074 \cdot 10^{-7})$	2.5	23.1	$4.06 \cdot 10^{-7}$ $(.063 \cdot 10^{-7})$	2.5	29.4
C3	$5.73 \cdot 10^{-7}$	$4.90 \cdot 10^{-7}$ $(.079 \cdot 10^{-7})$	5.0	14.5	$4.94 \cdot 10^{-7}$ $(.057 \cdot 10^{-7})$	8.5	13.9	$5.07 \cdot 10^{-7}$ $(.078 \cdot 10^{-7})$	2.8	11.6	$4.85 \cdot 10^{-7}$ $(.060 \cdot 10^{-7})$	3.9	15.4
C4	$5.73 \cdot 10^{-7}$	$5.28 \cdot 10^{-7}$ $(.091 \cdot 10^{-7})$	5.2	7.9	$5.23 \cdot 10^{-7}$ $(.061 \cdot 10^{-7})$	12	8.8	$5.32 \cdot 10^{-7}$ $(.086 \cdot 10^{-7})$	2.9	7.1	$5.20 \cdot 10^{-7}$ $(.053 \cdot 10^{-7})$	5.7	9.3
C5	$2.56 \cdot 10^{-12}$	$1.73 \cdot 10^{-12}$ $(.058 \cdot 10^{-12})$	9.1	32.4	$1.69 \cdot 10^{-12}$ $(.056 \cdot 10^{-12})$	8.2	34.1	$1.85 \cdot 10^{-12}$ $(.049 \cdot 10^{-12})$	5.9	27.8	$1.73 \cdot 10^{-12}$ $(.050 \cdot 10^{-12})$	4.5	32.3

with respect to n_0, the work-normalized variance is almost a constant. The EC estimator costs more than the SC estimator on average due to the fluctuations in the estimated splitting curves.

C. *Effectiveness of the truncation estimator*

Now we compare the SC estimator with the truncated SC estimator (T-SC) and truncated EC estimator (T-EC). For T-SC, the truncated level is calculated analytically, and for T-EC, the truncated level is also calculated by the pilot algorithm (see Remark 3) with $c_\alpha = 0.1$. As illustrated in Fig. 7 (based on 1000 independent replications), both T-SC and T-EC estimators require less computation than SC, and have superior work-normalized variances. Table I provides results comparing SC, EC, T-SC, and T-EC for other parameter settings (keeping $T = 1$) based on 100 replications. Since the process is discretized, between the discrete time points there is a possibility that the rare event is reached but undetected, so the estimator is biased, and we use the relative error (RE) to measure the bias of the estimators, and the standard error (std err) to measure the variance in the table. Comparing the results in C2, C3

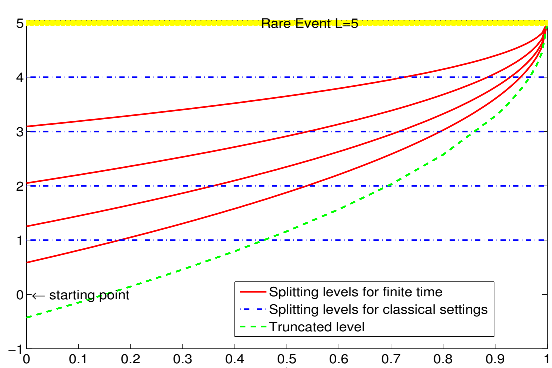

Fig. 5. Splitting curves for finite time and classical setting.

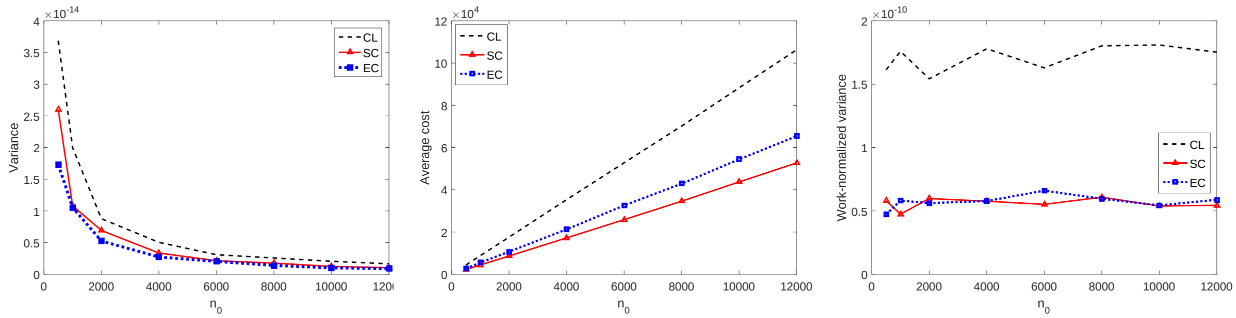

Fig. 6. Variance, average cost, and work-normalized variance for three different splitting curves based on 1000 independent replications.

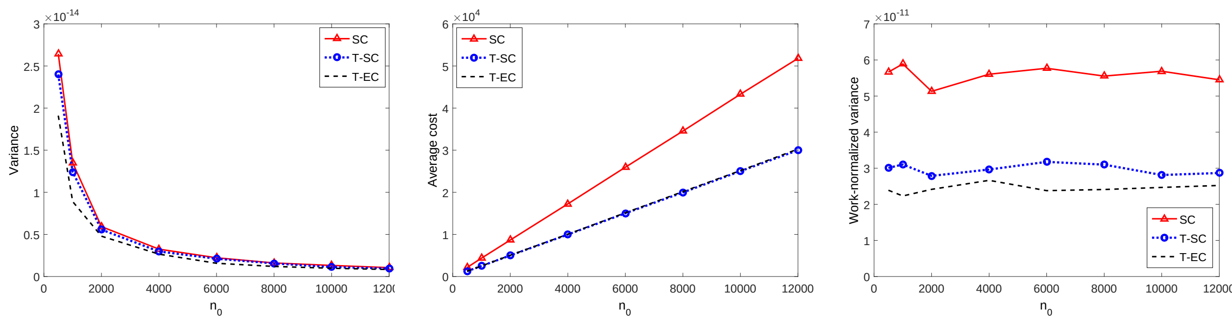

Fig. 7. Variance, average cost, and work-normalized variance estimators with and without truncation based on 1000 independent replications.

and C4 illustrates how increasing m reduces bias. The truncation estimators perform as well as the non-truncation estimators, but with less computation.

D. *Risk management example*

In this subsection, we provide an application in risk management. The bankruptcy probability or default probability is an important indictor in risk management and insurance. For example, the US bank default probability is used to determine FDIC insurance premiums that banks are charged. In reduced-form models of credit risk (see [22] for more details), if the assets level of a company sinks below the debt level of the company, default occurs. Let Z_t be the assets of the company and d be its debt, and $\tau = \inf\{t \geq 0 | Z_t < d\}$ be the default time. Investors are usually interested in the default probability before a given time T, i.e., $\Pr\{\tau \leq T\}$, where T may be the bond maturity or the annual date at which a portfolio loss risk needs to be reported to shareholders.

In this example, we assume the assets of the company contain three factors, which are modeled by the variance-gamma process (VG) (see [23], [24]). Let $Z_t = \mathbf{w}^\top \mathbf{S}_t$, where $\mathbf{S}_t = (S_t^{(1)}, S_t^{(2)}, S_t^{(3)})^\top$ is a three-dimensional VG process with initial value $\mathbf{S}_0 = (1,1,1)^\top$ and parameter values $\theta = (0, 0.1, -0.1)^\top$, $\nu = (0, 2, 0.3, 0.4)^\top$ and $\sigma = (0.2, 0.2, 0.3)^\top$. Let $\mathbf{w} = (1,1,1)^\top$, i.e., each factor has equal weight, and $d = 1$. In the splitting algorithms, let $N = 6$, $m = 100$, $M = 1000$, $p_i = c = 0.1$, $n_i = \lceil 1/p_i \rceil = 10, i = 1, \dots, N-1$, $c_\alpha = 0.05$. Then we compare the classical Monte Carlo (CMC) with EC and T-EC based on 100 replications. Since we do not have an analytical formula for the true default probability, we use the estimated probability of CMC from 10^8 paths as the true probability. In Fig. 8, CMC has significantly larger variance than EC and T-EC. When $n_0 = 5000$, most of the estimated values are 0, whereas EC and T-EC provide reasonable estimates. Further, we show the work-normalized

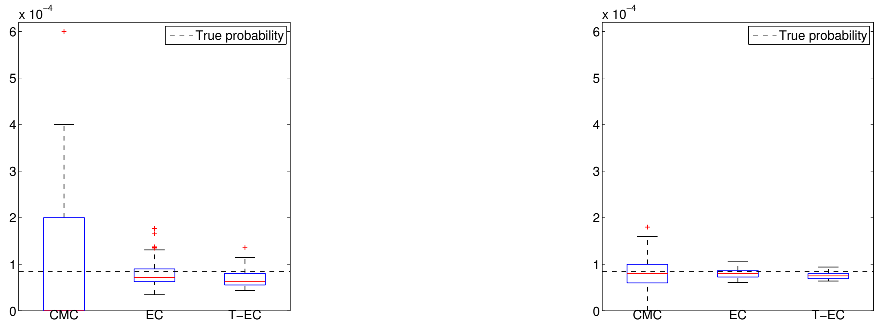

Fig. 8. Boxplots for default probability: $n_0 = 5000$ (left), 50000 (right).

TABLE II

WORK-NORMALIZED VARIANCE FOR DEFAULT PROBABILITY

Unit:10^{-5}	CMC	EC	T-EC
$n_0 = 5000$	9.07	0.757	0.217
$n_0 = 50000$	7.60	0.863	0.236

variance in Table II, which indicates that T-EC has smaller work-normalized variance than EC, and both T-EC and EC have much smaller work-normalized variance than CMC.

VI. CONCLUSION

In this note, we study importance splitting for rare event simulation of Markov processes under a finite-time constraint, deriving an unbiased estimator and characterizing the properties of the new splitting curves. We also provide a new unbiased estimator that incorporates truncated paths to improve the computational efficiency. A pilot algorithm is given to estimate the splitting and truncation curves. Numerical examples illustrate the effectiveness of the new estimators.

ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation (NSF) under Grants CMMI-1362303 and CMMI-1434419, by the Air Force of Scientific Research (AFOSR) under Grant FA9550-15-10050, and by the National Natural Science Foundation of China (Project 11171256).

REFERENCES

[1] M. Garvels, "A combined splitting-cross entropy method for rare-event probability estimation of queueing networks," *Ann. Oper. Res.*, vol. 189, pp. 167–185, 2011.

[2] J. F. Shortle, C.-H. Chen, B. Crain, A. Brodsky, and D. Brod, "Optimal splitting for rare-event simulation," *IIE Trans.*, vol. 44, pp. 352–367, 2012.

[3] P. Glasserman, P. Heidelberger, and P. Shahabuddin, "Variance reduction techniques for estimating value-at-risk," *Manage. Sci.*, vol. 46, pp. 1349–1364, 2000.

[4] P. Glasserman, P. Heidelberger, P. Shahabuddin, and T. Zajic, "Multilevel splitting for estimating rare event probabilities," *Oper. Res.*, vol. 47, pp. 585–600, 1999.

[5] M. Amerind and H. R. Künsch, "A variant of importance splitting for rare event estimation: Fixed number of successes," *ACM Trans. Model. Comput. Simul.*, vol. 21, Article 13, 2011.

[6] P. Glasserman, P. Heidelberger, P. Shahabuddin, and T. Zajic, "A large deviations perspective on the efficiency of multilevel splitting," *IEEE Trans. Automat. Control*, vol. 43, pp. 1666–1679, 1998.

[7] M. Garvels, "The splitting method in rare event simulation," Ph.D. Thesis, University of Twente, The Netherlands, 2000.

[8] F. Cérou and A. Guyader, "Adaptive multilevel splitting for rare event analysis," *Stoch. Anal. Appl.*, vol. 25, no. 2, pp. 417–443, 2007.

[9] T. Dean and P. Dupuis, "Splitting for rare event simulation: A large deviation approach to design and analysis," *Stochastic Process. Appl.*, vol. 119, pp. 562–587, 2009.

[10] M. Villén-Altamirano and J. Villén-Altamirano, "RESTART: A straightforward method for fast simulation of rare events," in *Proceedings of the 1994 Winter Simulation Conference*, J. D. Tew, S. Manivannan, D. A. Sadowski, and A. F. Seila, Eds., 1994, pp. 282–289.

[11] M. Balesdent, J. Morio, and J. Marzat, "Recommendations for the tuning of rare event probability estimators," *Reliab. Eng. Syst. Safe*, vol. 133, pp. 68–78, 2015.

[12] M. Villén-Altamirano and J. Villén-Altamirano, "Rare event simulation of non-Markovian queueing networks using RESTART method," *Simul. Model. Pract. Th.*, vol. 37, pp. 70–78, 2013.

[13] J. Blanchet and Y. Shi, "Efficient splitting-based rare event simulation algorithm for heavy-tailed sums," in *Proceedings of the 2013 Winter Simulation Conference*, R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, Eds., 2013, pp. 724–754.

[14] J. Blanchet, K. Leder, and Y. Shi, "Analysis of a splitting estimator for rare event probabilities in Jackson networks," *Stoch. Syst.*, vol. 1, no. 2, pp. 306–339, 2011.

[15] W. S. Wadman, D. Crommelin, and B. Zwart, "A large-deviation-based splitting estimation of power flow reliability," *ACM Trans. Model. Comput. Simul.*, vol. 26, Article 23, 2016.

[16] D. Bhaumik, D. Crommelin, and B. Zwart, "A computational method for optimizing storage placement to maximize power network reliability," in *Proceedings of the 2016 Winter Simulation Conference*, T. M. K. Roeder, P. I. Frazier, R. Szechtman, and E. Zhou, Eds., 2016, pp. 883–894.

[17] W. S. Wadman, M. S. Squillante, and S. Ghosh, "Accelerating splitting algorithms for power grid reliability estimation," in *Proceedings of the 2016 Winter Simulation Conference*, T. M. K. Roeder, P. I. Frazier, R. Szechtman, and E. Zhou, Eds., 2016, pp. 1756–1768.

[18] P. Glasserman, P. Heidelberger, P. Shahabuddin, and T. Zajic, "Splitting for rare event simulation: analysis of simple cases," in *Proceedings of the 1996 Winter Simulation Conference*, J. M. Charnes, D. Morrice, D. T. Brunner, and J. J. Swain, Eds., 1996, pp. 302–308.

[19] P. L'Ecuyer, V. Demers, and B. Tuffin, "Rare events, splitting, and quasi-Monte Carlo," *ACM Trans. Model. Comput. Simul.*, vol. 17, Article 9, 2007.

[20] J. F. Shortle and C.-H. Chen, "Sensitivity analysis of rare-event splitting applied to cascading blackout models," in *Proceedings of the 2013 Winter Simulation Conference*, R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, Eds., 2013, pp. 745–735.

[21] N. Kahale, "Analytic crossing probabilities for certain barriers by Brownian motion," *Ann. Appl. Probab.*, vol. 18, pp. 1424–1440, 2008.

[22] R. A. Jarrow and S. M. Turnbull, "Pricing derivatives on financial securities subject to credit risk," *J. Finance*, vol. 50, pp. 53–85, 1995.

[23] M. C. Fu, "Variance-gamma and Monte Carlo," in *Advances in Mathematical Finance*, M. C. Fu, R. A. Jarrow, J. Y. Yen, and R. J. Elliott, Eds. Boston: Birkhäuser, 2007, pp. 21–34.

[24] J. Cariboni and W. Schoutens, "Pricing credit default swaps under lévy models," *J. Comput. Financ.*, vol. 10, pp. 71–91, 2007.