Rare event simulation

Agnes Lagnoux

Probability in the Engineering and Informational Sciences, **20**, 2006, 45–66. Printed in the U.S.A.

RARE EVENT SIMULATION

AGNÈS LAGNOUX
Laboratoire de Statistique et Probabilités
Université Paul Sabatier
31062 Toulouse Cedex 4, France
E-mail: lagnoux@cict.fr

This article deals with estimations of probabilities of rare events using fast simulation based on the splitting method. In this technique, the sample paths are split into multiple copies at various stages in the simulation. Our aim is to optimize the algorithm and to obtain a precise confidence interval of the estimator using branching processes. The numerical results presented suggest that the method is reasonably efficient.

1. INTRODUCTION

The analysis of rare events is of great importance in many fields because of the risk associated with the event. Their probabilities are often about 10^{-9} to 10^{-12}. One can use many ways to study them: The first is statistical analysis, based on the standard extreme value distributions, but this needs a long observation period (see Aldous [1]); the second is modeling, which leads to estimating the rare event probability either by an analytical approach (see Sadowsky [10]) or by simulation.

In this article we focus on the simulation approach based on the Monte Carlo method. Nevertheless, a crude simulation is impractical for estimating such small probabilities: To estimate probabilities of order 10^{-10} with acceptable confidence would require the simulation of at least 10^{12} events (which corresponds to the occurrence of 100 rare events).

To overcome these limits, fast simulation techniques are applied. In particular, importance sampling (IS) is a refinement of Monte Carlo methods. The main idea of IS is to make the occurrence of the rare event more frequent. More precisely, IS consists of selecting a change of measure that minimizes the variance of the estimator. Using another method based on particles systems, Cerou, Del Moral, Legland,

 0269-9648/06 $16.00

and Lezaud [3] gave theoretical results on the convergence of this kind of algorithm. In this article, we deal with the RESTART (REpetitive Simulation Trials After Reaching Thresholds) algorithm presented by Villen-Altamirano and Villen-Altamirano [11] and based on splitting. The basic idea of splitting is to partition the space state of the system into a series of nested subsets and to consider the rare event as the intersection of a nested sequence of events. When a given subset is entered by a sample trajectory, random retrials are generated from the initial state corresponding to the state of the system at the entry point. Thus, the system trajectory has been split into a number of new subtrajectories. However, the analysis of the RESTART model presents numerous difficulties because of the lack of hypothesis and the complexity of formulas.

In this article we build a simple model of splitting for which we are able to derive precise conclusions. It is based on the same idea: Before entering the rare event A, there exists intermediate states visited more often than A by the trajectory: $A = B_{M+1} \subset B_M \subset \cdots \subset B_1$. Let $P_i = \mathbb{P}(B_i | B_{i-1})$, $i = 2, \ldots, M+1$, and $P_1 = \mathbb{P}(B_1)$. The fact that a sample trajectory enters B_i is represented by a Bernoulli trial. Every time a sample trajectory enters a subset B_i, $i = 1, \ldots, M$, it is divided in a number R_i of subtrajectories starting from level i. More precisely, we generate N random variables with common law Bernoulli $\mathrm{Ber}(P_1)$ and check whether the subset B_1 is reached. If so, we duplicate the trials in R_1 retrials of $\mathrm{Ber}(P_2)$ and check whether the subset B_2 is reached. Thus,

$$P = \mathbb{P}(A) = P_1 \cdots P_{M+1} \tag{1}$$

and an unbiased estimator of P is

$$\hat{P} := \frac{1}{N} \sum_{i=1}^{N} \hat{P}_i = \frac{N_A}{N R_1 \cdots R_M}, \tag{2}$$

where $\hat{P}_i$ are independent and identically distributed (i.i.d.), N_A is the number of trials that reach A during the simulation, and N is the number of particles initially generated. An optimal algorithm is chosen via the minimization of the variance of $\hat{P}$ for a given budget. For this, we have to describe the cost of a given simulation: Each time a particle is launched, it generates an average cost that is assumed here to be a function h of the transition probability. Therefore, the (average) cost is

$$C = N \sum_{i=0}^{M} r_i h(P_{i+1}) P_{i|0}, \tag{3}$$

where $r_i = R_1 \cdots R_i$, $i = 1, \ldots, M$, $r_0 = 1$, and $P_{i|0} = P_1 \cdots P_i$, $i = 1, \ldots, M+1$, and $P_{0|0} = 1$. Then the optimal algorithm is described by

$$P_i = P^{1/(M+1)}, \qquad i = 1, \ldots, M+1,$$

$$R_i = \frac{1}{P_i}, \qquad i = 1, \ldots, M,$$

$$N = \frac{C}{(M+1)h(P^{1/M+1})}, \tag{4}$$

and M is given by $M = [\ln P/y_0] - 1$ or $M = [\ln P/y_0]$, where y_0 is the solution of Eq. (30). The optimal sampling number is independent of the budget and this former only determines the optimal number of independent particles first generated. In the special case of $h = 1$,

$$M = [-0.6275 \ln P] - 1 \quad \text{or} \quad M = [-0.6275 \ln P],$$

$$R_i \approx 5, \quad \text{and} \quad P_i \approx \frac{1}{5}. \tag{5}$$

Thus, the optimal sampling number and the optimal transition probabilities are independent of the rare event probability. For example, if $P = 10^{-12}$ and $C = 10^3$, $M = 16$, $P_i \approx 0.2$, $R_i = 5$, and $N = 59$.

Example 1.1: To analyze the behavior of the different implementations described earlier, we perform a simulation experiment using these methods. We consider a queuing network and we want to estimate the occupancy of the finite buffer queuing system $M/M/1/C_0$. The results are presented in Figure 1. As expected and since we proceed for a given cost C ($C = 10^4$), the crude simulation stops after a few iterations, the number of samples run at the beginning being not sufficient. However, note that splitting simulation and theoretical analysis give very close results.

Example 1.2: This model can be applied to approximate counting (see Jerrum and Sinclair [7] and Diaconis and Holmes [5]). Given a positive real vector $\mathbf{a} = (a_i)_{i=1}^n$ and a real number b, we want to estimate the number of 0–1 vectors $\mathbf{x} = (x_i)_{i=1}^n$ s.t.

$$\mathbf{a} \cdot \mathbf{x} := \sum_{i=1}^{n} a_i x_i \leq b \tag{6}$$

For more details, see Section 3.2.

Remark 1.1: Hereafter we will take all the R_i equal to R and all the P_i equal to $P_0 = 1/R$. Thus, $RP_0 = 1$.

The aim of the article is to give a precise confidence interval of $\hat{P}$. The bound involving the variance of $\hat{P}$ and given by the Markov inequality is not precise enough. Therefore, as done in the theory of large deviations, we introduce the Laplace transform of $\hat{P}_1$, which can be rewritten as $\mathbb{E}(e^{\lambda \hat{P}_1}) = P_0 f_M(e^{\lambda/R^M}) + 1 - P_0$, where f_M

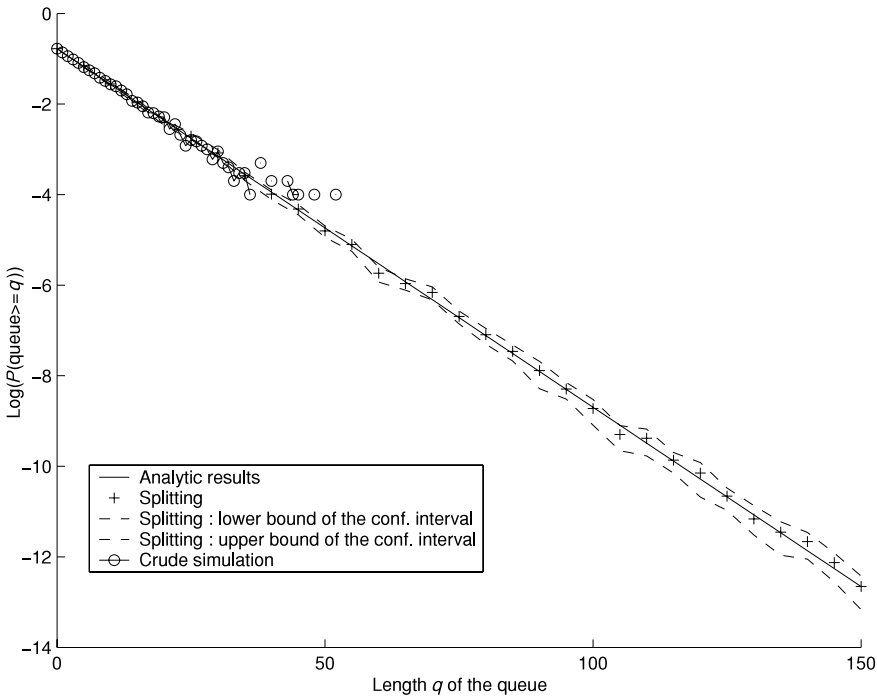


FIGURE 1. Comparison between the different methods: queuing theory model. Level of confidence 95/100.

is the Mth functional iterate of a $\mathrm{Bin}(R, P_0)$ generating function (g.f.). The elementary theory of branching processes leads to precise bounds of f_M and to a precise confidence interval that we can compare to the confidence interval if we only use the variance. For example, for $P = 10^{-9}$, $C = 10^8$, and $\alpha = 0.02$, the variance gives a bound about 10^{-2} and the Laplace transform gives a bound approximating 10^{-12}.

The article is organized as follows. Section 2 describes the importance splitting model, presents our model and goals (the analysis of the behavior of the probability P of a rare event), and introduces an estimator $\hat{P}$ of P. Section 3 is dedicated to the optimization of the algorithm. In Section 4 we obtain a precise confidence interval of the estimator via branching processes. Finally, in Section 5 we conclude and discuss the merits of this approach and potential directions for further researches.

2. IMPORTANCE SPLITTING MODEL

Our goal is to estimate the probability of a rare event A corresponding, for example, to the hit of a certain level L by a process $X(t)$. The main hypothesis is to suppose that before entering the target event, there exists intermediate states visited more

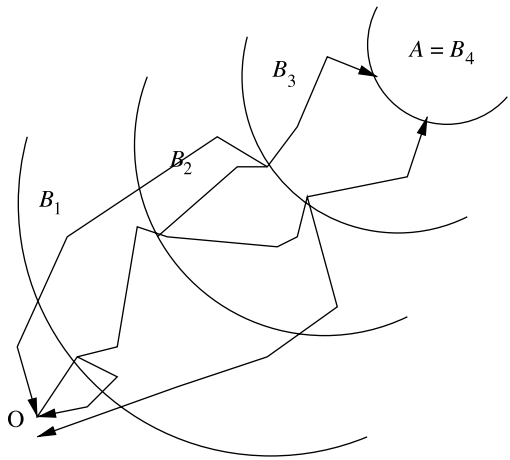


FIGURE 2. Splitting model.

frequently than A by the trajectory. Thus, define a sequence of sets of states B_i such as $A = B_{M+1} \subset B_M \subset \cdots \subset B_1$, which determines a partition of the state space into regions B_i–B_{i+1} called the *importance regions* (as represented in Figure 2). In general, these sets are defined through a function Φ called the *importance function* from the state space to $\mathbb{R}$ such that for all i, $B_i = \{\Phi \leq T_i\}$ for some value T_i called *thresholds*, with $T_1 \leq T_2 \leq \cdots \leq T_M \leq L$.

In this model a more frequent occurrence of the rare event is achieved by performing a number of simulation retrials when the process enters regions where the chance of occurrence of the rare event is higher. The fundamental idea consists of generating N Bernoulli $\mathrm{Ber}(P_1)$ and check whether the subset B_1 is reached. If so, we duplicate the trials in R_1 retrials of Bernoulli $\mathrm{Ber}(P_2)$ and check whether the subset B_2 is reached. If none of the higher levels is reached, the simulation stops.

Thus, by the Bayes formula,

$$\mathbb{P}(A) = \mathbb{P}(A|B_M)\mathbb{P}(B_M|B_{M-1})\cdots\mathbb{P}(B_2|B_1)\mathbb{P}(B_1) \tag{7}$$

$$:= P_{M+1}P_M\cdots P_2P_1. \tag{8}$$

Then P is the product of $M+1$ quantities (conditional probabilities) that are easier to estimate and with more accuracy than the probability P of the rare event itself, for a given simulation effort.

The estimator $\hat{P}$ of P defined in (2) can be rewritten as

$$\hat{P} = \frac{1}{NR_1\cdots R_M}\sum_{i_0=1}^{N}\sum_{i_1=1}^{R_1}\cdots\sum_{i_M=1}^{R_M}\mathbf{1}_{i_0}\mathbf{1}_{i_0 i_1}\cdots\mathbf{1}_{i_0 i_1\cdots i_M}, \tag{9}$$

where $\mathbf{1}_{i_0 i_1 \cdots i_j}$ represents the result of the jth trial. In that case,

$$\hat{P}_{i_0} = \frac{1}{R_1 \cdots R_M} \sum_{i_1=1}^{R_1} \cdots \sum_{i_M=1}^{R_M} \mathbf{1}_{i_0} \mathbf{1}_{i_0 i_1} \cdots \mathbf{1}_{i_0 i_1 \cdots i_M}. \tag{10}$$

Moreover, we define $\mathbb{P}(A)$ as the probability of reaching A and we suppose that the process forgets the past after reaching a level; this happens as soon as the process is Markov.

3. STUDY OF THE VARIANCE AND OPTIMIZATION

3.1. Variance of the Estimator

First, note that $\hat{P}$ is unbiased since

$$\begin{aligned} E(\hat{P}) &= \mathbb{E}\left(\frac{N_A}{NR_1 \cdots R_M}\right) \\ &= \frac{1}{NR_1 \cdots R_M} \sum_{i_0=1}^{N} \sum_{i_1=1}^{R_1} \cdots \sum_{i_M=1}^{R_M} \mathbb{E}(\mathbf{1}_{i_0} \mathbf{1}_{i_0 i_1} \cdots \mathbf{1}_{i_0 i_1 \cdots i_M}) = P. \end{aligned} \tag{11}$$

As done in [11], the variance of the estimator $\hat{P}$ is derived by induction and the variance for k thresholds is given by

$$\operatorname{var}(\hat{P}^{(k)}) = \frac{(P_1 \cdots P_{k+1})^2}{N} \left[\sum_{i=0}^{k} \frac{1}{r_i} \left(\frac{1}{P_{i+1|0}} - \frac{1}{P_{i|0}} \right) \right], \tag{12}$$

where $\hat{P}^{(k)}$ represents the estimator of P in a simulation with k thresholds.

Clearly, the formula holds in straightforward simulation (i.e., when $k = 0$), since $\hat{P}$ is a renormalized sum of i.i.d. Bernoulli variables with parameter P.

To go from k to $k + 1$, assume (12); thus, we have to prove that this formula holds for $k + 1$ thresholds. First, note that for all X and Y random variables, which are independent given the set B and X $\sigma(B)$-measurable, we have

$$\operatorname{var}(XY) = \operatorname{var}(X)\operatorname{var}(Y) + \operatorname{var}(X)\mathbb{E}(Y)^2 + \operatorname{var}(Y)\mathbb{E}(X)^2. \tag{13}$$

Now let

$$X_{i_0} = \mathbf{1}_{i_0}, \qquad Z_{i_0} = \frac{1}{R_1 \cdots R_{k+1}} \sum_{i_1=1}^{R_1} \cdots \sum_{i_{k+1}=1}^{R_{k+1}} \mathbf{1}_{i_0 i_1} \cdots \mathbf{1}_{i_0 i_1 \cdots i_{k+1}}. \tag{14}$$

The random variables X_{i_0} are i.i.d. with common law $\operatorname{Ber}(P_1)$, and conditionally at the event B_1, X_{i_0} and Z_{i_0} are independent. Note that each Z_{i_0} is the estimator of P in

a model with k thresholds, T_2 to T_{k+1} for the trajectory issued from the success of X_{i_0}. Thus,

$$\mathbb{E}(Z) = P_2 \cdots P_{k+2}, \tag{15}$$

and by the induction hypothesis,

$$\operatorname{var}(Z) = (P_2 \cdots P_{k+2})^2 \left[\sum_{i=1}^{k+1} \frac{1}{R_1 \cdots R_i} \left(\frac{1}{P_{i+1|1}} - \frac{1}{P_{i|1}} \right) \right]. \tag{16}$$

So applying (13) with $X \sim \operatorname{Ber}(P_1)$ and $Z \sim Z_{i_0}$, we have

$$\operatorname{var}(\hat{P}^{(k+1)}) := \frac{1}{N^2} \operatorname{var}\left(\sum_{i_0=1}^{N} X_{i_0} Z_{i_0} \right) \tag{17}$$

$$= \frac{P_1}{N} \left[\operatorname{var}(Z) + (1 - P_1)\mathbb{E}(Z)^2 \right] \tag{18}$$

$$= \frac{(P_1 P_2 \cdots P_{k+2})^2}{N} \left[\sum_{i=0}^{k+1} \frac{1}{r_i} \left(\frac{1}{P_{i+1|0}} - \frac{1}{P_{i|0}} \right) \right]. \tag{19}$$

Thus, for M thresholds,

$$\operatorname{var}(\hat{P}) = \frac{P^2}{N} \left[\sum_{i=0}^{M} \frac{1}{r_i} \left(\frac{1}{P_{i+1|0}} - \frac{1}{P_{i|0}} \right) \right]. \tag{20}$$

Remark 3.1: The induction principle has a concrete interpretation: If in a simulation with M steps, the retrials generated in the first level are not taken into account except one that we call the *main trial*, we have a simulation with $M - 1$ steps.

3.2. Optimization of the Parameters

As stated in Section 1, our aim is to minimize the variance for a fixed budget, giving optimal values for $N, R_1, \ldots, R_M$, $P_1, \ldots, P_{M+1}$, and M. Therefore, we have to describe the cost of a given simulation: Each time a particle is launched, it generates an average cost function h. We assume the following:

- The cost h for a particle to reach B_i starting from B_{i-1} depends only on P_i (not on the starting level).
- h is decreasing in x (which means that the smaller the transition probability is, the harder the transition is and the higher the cost is).
- h is nonnegative.
- h converges to a constant (in general, small) when x converges to 1.

The (average) cost is then

$$C = \mathbb{E}(Nh(P_1) + R_1 N_1 h(P_2) + R_2 N_2 h(P_3) + \cdots + R_M N_M h(P_{M+1})), \tag{21}$$

where N_i is the number of trials that have reached threshold i. Finally,

$$C = N \sum_{i=0}^{M} r_i h(P_{i+1}) P_{i|0}. \tag{22}$$

Example 3.1: We want to study the model of the simple random walk on $\mathbb{Z}$ starting from zero that we kill as soon as it reaches the level -1 or k (success if we reach k, failure otherwise).

So let X_n be such that $X_0 = 0$ and $X_n = \sum_{i=1}^{n} Y_n$, where $\{Y_n\}$ is a sequence of random variables valued in $\{-1,1\}$ with $\mathbb{P}(Y_n = 1) = \mathbb{P}(Y_n = -1) = \frac{1}{2}$ and define $T_k = \inf\{n \geq 0 : X_n = -1 \ or \ k\}$.

One can easily check that X_n and $X_n^2 - n$ are martingales. By Doob's stopping theorem, $\mathbb{E}(X_{T_k}) = 0$ and $\mathbb{E}(X_{T_k}^2) = \mathbb{E}(T_k)$, which yields

$$p := \mathbb{P}(X_{T_k} = k) = \frac{1}{k+1} \quad \text{and} \quad \mathbb{E}(T_k) = k = \frac{1}{p} - 1 \tag{23}$$

(i.e., the cost needed to reach the next level is $(1/p) - 1$ if p is the success probability).

To minimize the variance of $\hat{P}$, the optimal values are derived in three steps:

1. The optimal values of $N, R_1, \ldots, R_M$ are derived when we consider that $P_1, \ldots, P_{M+1}$ are constant (i.e., the thresholds B_i are fixed).
2. Replacing these optimal values in the variance, we derive the optimal transition probabilities: $P_1, \ldots, P_{M+1}$.
3. Replacing these optimal values in the variance, we derive M, the optimal number of thresholds.

Optimal values for $N, R_1, \ldots, R_M$. Using the method of Lagrange multipliers, we get

$$R_i = \frac{r_i}{r_{i-1}} = \sqrt{\frac{h(P_i)}{h(P_{i+1})}} \sqrt{\frac{1}{P_i P_{i+1}}} \sqrt{\frac{1 - P_{i+1}}{1 - P_i}}, \qquad i = 1, \ldots, M, \tag{24}$$

$$N = \frac{1}{\sqrt{h(P_1)}} \frac{C\sqrt{1/P_1 - 1}}{\sum_{i=1}^{M+1} \sqrt{h(P_i)} \sqrt{\frac{1}{P_i} - 1}}. \tag{25}$$

Optimal values for $P_1, \ldots, P_{M+1}$. Thus, the variance becomes

$$\operatorname{var}(\hat{P}) = \frac{P^2}{C} \left[\sum_{i=1}^{M+1} \sqrt{h(P_i)} \sqrt{\frac{1}{P_i} - 1} \right]^2. \tag{26}$$

Proceeding as previously under the constraint $P = P_1 \cdots P_{M+1}$, we obtain that all of the P_i's satisfy $2\sqrt{C}\lambda\sqrt{h(x)((1/x) - 1)} = h'(x)(1 - x) - (h(x)/x)$. If we assume

that there exists a unique solution to this equation, we have $P_i = g(\lambda)$; hence, $P = g(\lambda)^{M+1}$ and $g(\lambda) = P^{1/(M+1)}$. Finally,

$$P_i = P^{1/(M+1)}, \qquad i = 1, \ldots, M+1. \tag{27}$$

Optimal value for M. The optimal values for $P_1, \ldots, P_{M+1}$ imply that the optimal R_i becomes $1/P_i$, $i = 1, \ldots, M$; thus,

$$\operatorname{var}(\hat{P}) = \frac{P^2}{C}(M+1)^2 h(P^{1/M+1})(P^{-1/M+1} - 1), \tag{28}$$

which we want to minimize in M. Note that $R_i P_i = 1$. Let

$$f(M) = \frac{P^2}{C}(M+1)^2 h(P^{1/M+1})(P^{-1/M+1} - 1), \tag{29}$$

whose derivative cancels in

$$F(y) := (2(1 - e^y) + y)h(e^y) - y(1 - e^y)e^y h'(e^y) = 0 \quad \text{with } y = \frac{\ln P}{M+1}. \tag{30}$$

In general, this does not give an integer. We have $y_0 = \ln P/(M+1)$ (i.e., $M + 1 = [\ln P/y_0]$ or $[\ln P/y_0] + 1$). Let $\ln P/y_0 = n + x$ with $0 < x < 1$. Then the following hold:

- If we take $M + 1 = n$, $y = \ln P/n$.
- If we take $M + 1 = n + 1$, $y = \ln P/(n+1)$.

The value of the ratio $\rho := f(n-1)/f(n)$ gives the best choice for M as follows:

- If $\rho < 1$, $M = n - 1$.
- If $\rho > 1$, $M = n$.

Thus, the optimal number of thresholds is given by $M = [\ln P/y_0] - 1$ or $M = [\ln P/y_0]$, where y_0 solves $F(y) = 0$. Then M minimizes

$$\operatorname{var}(\hat{P}) = \frac{P^2}{C}(\ln P)^2 y^{-2} h(e^y)(e^{-y} - 1). \tag{31}$$

Example 3.2: For $h = 1$, we have to solve $y = 2(e^y - 1)$. We get $y_1 = 0$ and $y_2 \approx -1.5936$. y_2 is a minimum and the optimal value of M is

$$M = [-0.6275 \ln P] - 1 \quad \text{or} \quad [-0.6275 \ln P]. \tag{32}$$

With $P = 10^{-k}$, we have

k	n	ratio(ρ) >1,<1	M	k	n	ratio(ρ) >1,<1	M
1	1	>	1	6	8	>	8
2	2	>	2	9	13	<	12
3	4	<	3	12	17	<	16
4	5	>	5	15	21	>	21
5	7	<	6	18	26	<	25

Note that M increases while P decreases, and with this value of M, each R_i and P_i become

$$R_i \approx 5 \quad \text{and} \quad P_i \approx \frac{1}{5}. \tag{33}$$

Thus, the optimal sampling number and the optimal transition probabilities are independent of the rare event probability.

Moreover, asymptotically, $M = n = [\ln P/y_0] - 1$; thus,

$$P_i = P^{1/(M+1)} = e^{\ln P/(M+1)} = e^{y_0} \quad \text{and} \quad P = e^{-(n+1)|y_0|}. \tag{34}$$

Application 3.1: In approximate counting, remember that the goal is to estimate the number of Knapsack solutions (i.e., the cardinal of Ω defined by

$$\Omega := \left\{ x \in \{0,1\}^n : \mathbf{a}\cdot\mathbf{x} := \sum_{i=1}^{n} a_i x_i \le b \right\}$$

for a given positive real vector $\mathbf{a} = (a_i)_{i=1}^n$ and real number b). We might try to apply the Markov chain Monte Carlo method (MCMC) [9]: Construct a Markov chain $\mathcal{M}_{\text{Knap}}$ with state space $\Omega = \{x \in \{0,1\}^n : \mathbf{a}\cdot\mathbf{x} \le b\}$ and transitions from each state $x = (x_1, \ldots, x_n) \in \Omega$ defined by the following:

- With probability $\frac{1}{2}$, let $y = x$; otherwise
- select i uniformly at random in $\{1, \ldots, n\}$ and let $y' = (x_1, \ldots, x_{i-1}, 1 - x_i, x_{i+1}, \ldots, x_n)$
- If $ay' \le b$, then let $y = y'$, else let $y = x$.

The new state is y. This random walk on the hypercube truncated by the hyperplane $\mathbf{a}\cdot\mathbf{x} = b$ converges to the uniform distribution over Ω. This suggests a procedure for selecting Knapsack solutions almost uniformly at random. Starting in state $(0,..,0)$, simulate $\mathcal{M}_{\text{Knap}}$ for sufficiently many steps that the distribution over states is "close"[1]

[1]The problem is to bound the number of steps necessary to make the Markov chain $\mathcal{M}_{\text{Knap}}(b)$ "close" to stationarity. More precisely, we need a bound of the *mixing time*:

$$\tau_{\text{mix}}(\nu) := \min\{t : \Delta_x(t') \le \nu \quad \text{for all} \quad t' \ge t\},$$

where $\Delta_x(t) = \max_{S\subset\Omega} |P^t(x,S) - \Pi(S)|$ and Π is the stationary distribution. In [7], it is shown that $\mathcal{O}(n^{9/2+\nu})$ steps suffice.

to uniform; then return to the current state. Of course, sampling over Ω is not the same as estimating the size of Ω. However, the first task leads to the second.

Keep the vector **a** fixed but allow b to vary. Use $\Omega(b)$ and $\mathcal{M}_{\text{Knap}}(b)$ instead of Ω and $\mathcal{M}_{\text{Knap}}$ to emphasize the dependence on b. Assume without loss of generality that $a_1 \leq \cdots \leq a_n$ and define $b_1 = 0$ and $b_i = \min\{b, \sum_{i=1}^{i-1} a_j\}$. One can check that

$$|\Omega(b_{i-1})| \leq |\Omega(b_i)| \leq (n+1)|\Omega(b_{i-1})|. \tag{35}$$

Now write

$$\begin{aligned} |\Omega(b)| &= |\Omega(b_{n+1})| \\ &= \frac{|\Omega(b_{n+1})|}{|\Omega(b_n)|} \frac{|\Omega(b_n)|}{|\Omega(b_{n-1})|} \cdots \frac{|\Omega(b_2)|}{|\Omega(b_1)|} |\Omega(b_1)| \\ &:= \rho_n^{-1} \cdots \rho_1^{-1}. \end{aligned} \tag{36}$$

The ratio $\rho_i = |\Omega(b_i)|/|\Omega(b_{i+1})|$ may be estimated by sampling almost uniformly from $\Omega(b_{i+1})$ using the Markov chain $\mathcal{M}_{\text{Knap}}(b_{i+1})$ and computing the fraction of the samples that lie within $\Omega(b_i)$.

Now take $a = [1,2,3,4]$, $b = 3$, $h = 1$, $R = 5$, and $C = 2600$. We chose the levels as follows: First, define $b_1 = 0$, $b_2 = 1$, $b_3 = 3$, $b_4 = 3$, and $b_5 = b$; second, define $B_0 = \Omega$, $B_1 = \Omega(b_4)$, $B_2 = \Omega(b_3)$, $B_3 = \Omega(b_2)$, and $B_4 = \Omega(b_1)$. Thus, here, $M = n - 1$, $N = C/n$, and $n_{step} = 1020$. Obviously, $\text{Card}(\Omega) = 5$. We run three different simulations: The first, suggested in [7], consists of estimating the n ratios independently, the crude and splitting ones. We obtain different estimations for $\text{Card}(\Omega)$:

- Estimation by crude simulation = 4.088
- Estimation by the n ratios independently = 5.44
- Estimation by splitting = 5.019

Even though the levels are not optimal, splitting provides an improvement.

Let us describe briefly the possible solutions of (30). Remember that we want to solve (30); that is, if $z = e^y$ and $z \neq 0,1$,

$$H(z) := \frac{h'(z)}{h(z)} = \frac{1}{z}\left(\frac{2}{\ln z} + \frac{1}{1-z}\right) =: \frac{l'(z)}{l(z)} =: L(z). \tag{37}$$

First, let z_0 be the solution of $2(z-1) = \ln z$. Since $h' \leq 0$, H is negative and a quick survey shows that L is positive on $]0, z_0[$ and negative on $]z_0, 1[$. As a consequence, the solutions of (37) lie in $]z_0, 1[$, if they exist. Thus, solving (30) is equivalent to studying the intersections between H and L. A quick survey of these functions shows that we have two cases (see Fig. 3):

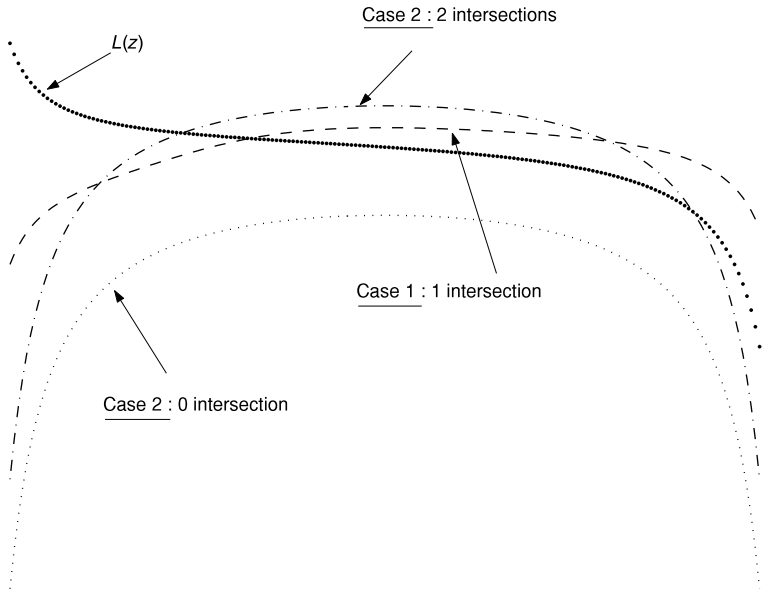


FIGURE 3. Behavior of H and L.

Case 1: An odd number of intersections between L and H

$$\Leftrightarrow H(z) > L(z) \quad \text{near} \quad 1, \tag{38}$$

$$\Leftrightarrow h''(1) < 0 \tag{39}$$

Case 2: An even number of intersections or 0 between L and H

Note that $y = 0$ is a solution of (30). In case 1, it corresponds to a maximum, and in case 2, it corresponds to a minimum. The second case is excluded since we made the assumption $h(1) > 0$.

Remark 3.2: The solution $y = 0$ corresponds to the following optimal values:

$$M = \infty, \qquad P_i = 1, \qquad R_i = 1, \qquad N \sim_{M\to\infty} \frac{C}{(M+1)h(1) + \ln(P)h'(1)}. \tag{40}$$

However $P_i = 1$ implies that $P = 1$ and $R_i = 1$ means that we just perform a crude simulation.

Example 3.3: Here, $P = 10^{-12}$ and $C = 10^4$.

1. In Example 1, $h(1) = 0$ and we are in the second case: The unique solution $y = 0$ is the minimum.

2. Let $h(x) = 1/x - 8x^2 + 12x - 5$. $h(1) = 0$ and we are in the first case: $y = 0$ and $y \approx -0.9919$ are the solutions. $y = 0$ is the maximum and the other solution is the minimum. Taking $y \approx -0.9919$, we obtain

$$M = 26, \qquad P_0 \approx 0.3594, \qquad R \approx 2.7826, \quad \text{and} \quad N \approx 22.9 \tag{41}$$

and we can take $R = 3$ and $N = 23$.
3. Let $h(x) = (1/x - 1)^2 e^{6x}$. $h(1) = 0$ and we are in the second case: $y = 0$, $y_1 \approx -0.4612$, and $y_2 \approx -0.5645$ are the solutions. $y = 0$ is the minimum and the second solution is the maximum.
4. Let $h(x) = 1/x$. Here, $h(1) = 1$. We want to solve (30), whose solutions are $y = 0$ and $y \approx -0.6438$. Taking $y \approx -0.6438$, we obtain

$$M = 41, \qquad P_0 \approx 0.5179, \qquad R \approx 1.9307, \quad \text{and} \quad N \approx 34.5 \tag{42}$$

and we can take $R = 2$ and $N = 34$.

Thus, the control of the variance of $\hat{P}$ gives a crude confidence interval for P. Indeed, we get

$$\mathbb{P}\left(\frac{|\hat{P} - P|}{P} \geq \alpha\right) \leq \frac{1}{P^2\alpha^2}\,\mathbb{E}((\hat{P} - \mathbb{E}(\hat{P}))^2) \tag{43}$$

$$\leq \frac{1}{\alpha^2 C}\,[(M+1)^2(P^{-1/M+1} - 1)h(P^{1/(M+1)})] \tag{44}$$

$$\approx \frac{4(M+1)}{\alpha^2 N}\,h(P^{1/(M+1)}). \tag{45}$$

This estimation is, in general, useless. For example, for $h = 1$, $M = 12$, and $\alpha = 10^{-2}$, the upper bound becomes $\approx 5 \times 10^5/N$. To obtain a bound lower than 1, we need $N \geq 5 \times 10^5$. To improve it, we will use Chernoff's bounding method instead of the Markov inequality: For all $\lambda > 0$,

$$\mathbb{P}(\hat{P} \geq P(1+\alpha)) = \mathbb{P}\left(\frac{1}{N}\sum_{i=1}^{N}\hat{P}_i \geq P(1+\alpha)\right) \tag{46}$$

$$= \mathbb{P}(e^{\lambda \sum_{i=1}^{N}\hat{P}_i} \geq e^{\lambda NP(1+\alpha)}) \tag{47}$$

$$\leq e^{-\lambda NP(1+\alpha)}\,\mathbb{E}(e^{\lambda\hat{P}_1})^N \tag{48}$$

$$\leq e^{-N[\lambda P(1+\alpha) - \psi(\lambda)]}, \tag{49}$$

where $\psi(\lambda) = \mathbb{E}(e^{\lambda\hat{P}_1})$ is the log-Laplace of $\hat{P}_1$. Optimization on $\lambda > 0$ provides

$$\mathbb{P}(\hat{P} \geq P(1+\alpha)) \leq e^{-N\sup_{\lambda>0}[\lambda P(1+\alpha) - \psi(\lambda)]}. \tag{50}$$

Similarly,

$$\mathbb{P}(\hat{P} \leq P(1+\alpha)) \leq e^{-N \sup_{\lambda<0}[\lambda P(1-\alpha)-\psi(\lambda)]}.$$

Let ψ^* be the Crämer transform of ψ: $\psi^*(\tau) = \sup_\lambda[\lambda\tau - \psi(\lambda)]$. Thus,

$$\mathbb{P}\left(\frac{|\hat{P}-P|}{P} \geq \alpha\right) \leq e^{-N\psi^*(P(1-\alpha))} + e^{-N\psi^*(P(1+\alpha))} \tag{51}$$

$$\leq 2e^{-N\min(\psi^*(P(1-\alpha)),\psi^*(P(1-\alpha)))}. \tag{52}$$

So we want to obtain an accurate lower bound of ψ^*.

Remark 3.3: Although we would therefore like to take R_i so that $R_i P_i = 1$, we are constrained to choose R_i to be a positive integer. Hereafter, we suppose that we are in the optimal case, where $R_i = 1/P_i$ is an integer.

4. LAPLACE TRANSFORM OF $\hat{P}_1$

To study the Laplace transform of $\hat{P}_1$, we turn to the theory of branching processes (see Harris [6], Lyons [8], and Athreya and Ney [2]). More precisely, we consider our splitting model as a Galton–Watson process, the thresholds representing the different generations.

4.1. Description of the Model and First Results

We consider a Galton–Watson model (Z_n), where the size of the nth-generation Z_n is the number of particles that have reached the level B_n, with one particle run at the beginning. Then $Z_0 = 1$ and Z_n satisfies the following recurrence relation:

$$Z_{n+1} = \sum_{i=1}^{Z_n} X_i^n, \tag{53}$$

where X_i^n is the number of particles among R_i that have reached the $(n+1)$-st level. The $(X_i^n)_{n\geq 1}$ are i.i.d. with common law Binomial, with parameters (R_n, P_{n+1}) and $X_i^0 \sim \mathrm{Ber}(P_1)$. Take the optimal values of Section 3.2:

$$R_i = R, \qquad i = 1,\ldots,M, \qquad P_i = P_0, \qquad i = 1,\ldots,M+1. \tag{54}$$

Let $f(s) = \mathbb{E}(s^{Z_1})$, the g.f. of Z_1. Then the g.f. of Z_n is the nth iterate of f. Since $\hat{P}_1 = (1/R^M)Z_{M+1}$, we get

$$\mathbb{E}(e^{\lambda\hat{P}_1}) = \mathbb{E}(e^{(\lambda/R^M)Z_{M+1}}) = g(f_M(e^{\lambda/R^M})) = g(f^{oM}(e^{\lambda/R^M})), \tag{55}$$

where g is the g.f. of a $\mathrm{Ber}(P_0)$ and f the g.f. of a $\mathrm{Bin}(R, P_0)$. Thus, we are interested in the expression of f_M, the Mth functional iterate of f.

Here, $m = \mathbb{E}(Z_1) = RP_0 = 1$, so we are in the critical case of the branching process that ensures that the algorithm of the simulation stops with probability 1

when $M \to \infty$ (see [6]), since if $f^{(3)}(1) < \infty$,

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{2Z_n}{nf''(1)} > u \,|\, Z_n \neq 0\right) = e^{-u}, \qquad u \geq 0. \tag{56}$$

This emphasizes the rarity character when the number M of thresholds increases and the probabilities between the levels decrease.

In our case,

$$f(s) = [P_0 s + (1 - P_0)]^R = [P_0(s-1) + 1]^R. \tag{57}$$

The iterated function f_M has no explicit tractable form and we will derive bounds for $f_M(s)$ around $s = 1$. To do this, we state a general result on the Laplace transform in critical Galton–Watson models, which we could not find in the literature.

4.2. Bounds of $f_n(s)$ for $0 \leq s < 1$ and $m = 1$

Remark 4.1: Remember that f_n and its derivatives are convex. Furthermore, for all $0 \leq s \leq 1$, $s \leq f(s) \leq f(1) = 1$, and by induction, $f(s) \leq f_2(s) \leq \cdots \leq 1$. Finally, we obtain $f_n(s) \to 1$ since $f_n(s) \geq f_n(0)$.

PROPOSITION 4.1: *Let* $\alpha_1 = f''(1)/2$, $C = (\max_{s\in[0,1]} f'''(s))/6\alpha_1$, *and* $\gamma_n = n\alpha_1[1 - (C/n)(\log n + 1)] - \alpha_1$. *Then, for s close to* 1 *and large n,*

$$1 - \frac{1-s}{1+\gamma_n(1-s)} \leq f_n(s) \leq 1 - \frac{(1-s)[1-\alpha_1(1-s)]}{1+\alpha_1(1-s)\left(n - 1 - \dfrac{\alpha_1^2(1-s)^2}{2}\right)}. \tag{58}$$

PROOF: **Upper bound**: Using Taylor's expansion, with $f_n(s) \leq \theta_n \leq f_n(1) = 1$,

$$f_{n+1}(s) = f(f_n(s)) = f(1) + (f_n(s) - 1)f'(1) + \frac{(f_n(s)-1)^2}{2} f''(\theta_n) \tag{59}$$

$$= f_n(s) + \frac{(f_n(s)-1)^2}{2} f''(\theta_n), \tag{60}$$

since $f'(1) = 1$. Let $r_n = 1 - f_n(s)$; r_n satisfies

$$r_{n+1} = r_n - r_n^2 \frac{f''(\theta_n)}{2}. \tag{61}$$

Now let $\alpha_0 = f''(0)/2$. Define the decreasing sequences (a_n) and (b_n) satisfying

$$a_{n+1} = a_n - a_n^2\alpha_1, \qquad b_{n+1} = b_n - b_n^2\alpha_0, \qquad a_0 = b_0 = 1 - s. \tag{62}$$

Then

$$a_n \le r_n \le b_n. \tag{63}$$

1. b_n's upper bound: Since $0 \le b_j \le 1$, we have

$$\frac{1}{b_n} = \frac{1}{b_{n-1}} + \alpha_0 \frac{1}{1-\alpha_0 b_{n-1}} = \frac{1}{b_0} + \alpha_0 \sum_{j=0}^{n-1} \frac{1}{1-\alpha_0 b_j} \ge \frac{1}{b_0} + n\alpha_0. \tag{64}$$

   Thus,

$$b_n \le \frac{1-s}{1+\alpha_0 n(1-s)}. \tag{65}$$

2. a_n's lower bound: Apply this upper bound to a_n (α_0 becoming α_1):

$$a_n \le \frac{1-s}{1+n\alpha_1(1-s)}. \tag{66}$$

   By substituting (66) in $1/a_n = (1/a_0) + \alpha_1 \sum_{j=0}^{n-1}(1/(1-\alpha_1 a_j))$, we get

$$a_n \ge \frac{(1-s)[1-\alpha_1(1-s)]}{1+\alpha_1(1-s)\left(n-1-\dfrac{\alpha_1^2(1-s)^2}{2}\right)}. \tag{67}$$

   Finally, (63) and (67) lead to the upper bound of f_n in (58).

Lower bound: In fact, we prove by induction that

$$h_{\hat{\gamma}_n}(s) := 1 - \frac{1-s}{1+\hat{\gamma}_n(1-s)} \le f_n(s) \qquad \text{with} \quad \begin{cases} \hat{\gamma}_{n+1} = c_n + \hat{\gamma}_n \\ \hat{\gamma}_1 = 0 \\ c_n = \alpha_1\left(1-\dfrac{C}{n}\right). \end{cases} \tag{68}$$

For $n = 1$, the left-hand side of (68) is given by Remark 4.1. Then note that $h_{\hat{\gamma}_n}(s) \to_{n\to\infty} 1$; thus, for n large enough, $1 - h_{\hat{\gamma}_n}(s) \le 1/n$. For all $1 - (1/n) \le s \le 1$,

$$h_{c_n}(s) = 1 + (s-1) + c_n(s-1)^2 + \frac{(s-1)^3}{6} h'''_{c_n}(\theta_n^1) \tag{69}$$

$$\le 1 + (s-1) + c_n(s-1)^2 \tag{70}$$

$$= f(s) - (s-1)^2\left[\frac{s-1}{6} f'''(\theta_n^2) + \frac{C\alpha_1}{n}\right] \tag{71}$$

$$\le f(s) \qquad \text{by definition of } C. \tag{72}$$

However, by induction, we have $h_{\hat{\gamma}_n}(s) \le f_n(s)$, and so, since f is increasing, taking $s = h_{\hat{\gamma}_n}(t)$,

$$h_{c_n}(h_{\hat{\gamma}_n}(t)) = h_{c_n+\hat{\gamma}_n}(t) \le f(h_{\hat{\gamma}_n}(t)) \le f(f_n(t)) = f_{n+1}(t) \tag{73}$$

(i.e., $h_{\hat{\gamma}_{n+1}}(t) \le f_{n+1}(t)$, where $\hat{\gamma}_{n+1} = c_n + \hat{\gamma}_n$). Note that $\gamma_n \sim \hat{\gamma}_n$; more precisely, we have $\gamma_n \le \hat{\gamma}_n$ and we finally obtain the left-hand side of (58) since $\gamma \to h_\gamma$ is increasing. ■

In the particular case of $f(s) = (P_0 s + 1 - P_0)^R$, we can derive a more precise lower bound:

PROPOSITION 4.2: *For s close to 1,*

$$1 - \frac{1-s}{1 + n\alpha_1(1-s)} \le f_n(s). \tag{74}$$

Observe that this is precise at $s = 1$.

PROOF: Let $h(s) = 1 - ((1-s)/(1+\alpha_1(1-s)))$. Since $f(1) = h(1) = 1$, $f'(1) = h'(1) = 1$, and $f''(1) = h''(1) = 2\alpha_1$, the sign of $f - h$ trivially depends on the sign of the third derivative of $f - h$, which is obviously negative here. Then $h \le f$. Since f is increasing, we deduce (74) by induction. ■

We plot in Figure 4a the upper bound and the two lower bounds for $P = 10^{-12}$ and s near 1.

4.3. Bounds of $f_n(s)$ for $1 \le s$ and $m = 1$

Remark 4.2: First, let us note that, by convexity, for all $s \ge 1$,

$$(s-1)f'(1) \le f(s) - f(1) = f(s) - 1; \tag{75}$$

hence, $f(s) \ge s$, and by induction on n,

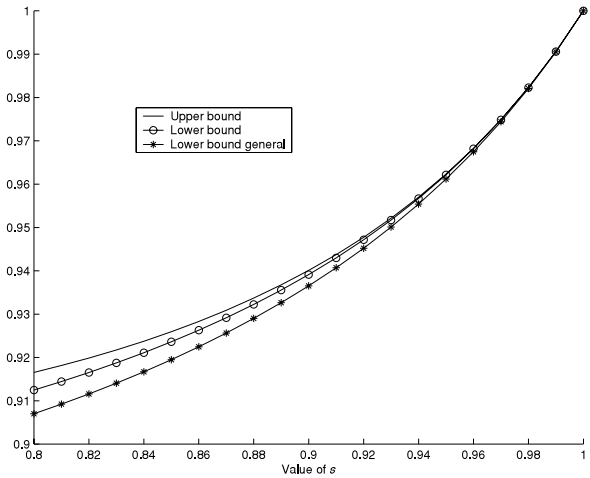
$$f_{n+1}(s) \ge f_n(s) \ge \cdots \ge f(s) \ge s \ge 1. \tag{76}$$

We remark that for $s > 1$, the iterated function increases rapidly to infinity.

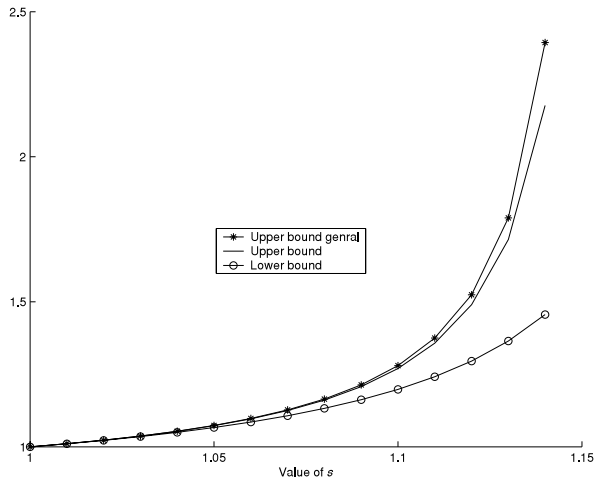
PROPOSITION 4.3: *Let* $\gamma'_n = n\alpha_1[1 + (C/n)(\log n + 1)] - \alpha_1$. *Then, for s close to 1 and large n,*

$$1 + \frac{(s-1)}{1 - n\alpha_1 s_n^{P_0-2}(s-1)} \le f_n(s) \le 1 + \frac{s-1}{1 - n\gamma'_n(s-1)}. \tag{77}$$

PROOF: Proceeding as in Proposition 4.1 leads to the upper bound. Here, $f_n \to_{n\to+\infty} \infty$, which prevents us from making a Taylor expansion around 1. To overcome this difficulty, consider k_n, the inverse function of f_n; it is the nth functional iterate of the g.f. k (inverse function of f) that takes the value 1 in 1, whose



(a) for $0 \leq s \leq 1$



(b) for $1 \leq s$

FIGURE 4. Bounds of $f_M(s)$.

derivative is 1 in 1 and second derivative is negative, and $k_n \to_{n\to+\infty} 1$. Thus, making a Taylor development and using the same tools as previously, we get

$$1 + \frac{(s-1)(1-\alpha_1(s-1))}{1+(n-1)\alpha_1(s-1)} \leq k_n(s) \leq 1 + \frac{s-1}{1+n\alpha_1 s_n^{P_0-2}(s-1)}, \tag{78}$$

where $\beta_2 = k''(s)/2$ and $s_n := 1 + (1/n\alpha_1)$. Using the link between k_n and f_n and the upper bound of k_n,

$$1 + \frac{(s-1)}{1 - n\alpha_1 s_n^{P_0-2}(s-1)} \leq f_n(s). \tag{79}$$

The lower bound of k_n leads to an upper bound of f_n. However, it provides no improvement. ■

As done earlier, we can derive a more precise upper bound in the particular case of $f(s) = (P_0 s + 1 - P_0)^R$:

PROPOSITION 4.4: *For s close to* 1,

$$f_n(s) \leq 1 + \frac{s-1}{1 - n\alpha_1(s-1)}. \tag{80}$$

We plot in Figure 4b these three bounds for $P = 10^{-12}$ and s near 1.

About the geometric distribution. If the law of X is such that the probabilities p_k are in a geometric proportion ($p_k = \mathbb{P}(X = k) = bc^{k-1}$ for $k = 1, 2 \ldots$. and $p_0 = 1 - p_1 - p_2 \ldots$ with $b, c > 0$ and $b \leq 1 - c$), then the associated g.f. is a rational function:

$$h(s) = 1 - \frac{b}{1-c} + \frac{bs}{1-cs}. \tag{81}$$

Taking $b = (1-c)^2$ and $c = \alpha_1/(1+\alpha_1)$ leads to

$$h(s) = 1 + \frac{s-1}{1 - \alpha_1(s-1)}. \tag{82}$$

So we have compared the nth functional iterate of a Binomial g.f. to the one of a geometric g.f. It suggests comparing the importance splitting models with Binomial and with geometric laws. The geometric laws model is set in the following way: We run particles one after the other. As long as the next level is not reached, we keep on generating particles; then we start again from the level the particle is at (the geometric distribution is the law of the first success).

This link is also stressed by Cosnard and Demongeot in [4]: for $m = 1$ and $\sigma^2 = f''(1) = 2\alpha_1$, the asymptotic behavior of f^{2^n} is the same as the geometric distribution with the same variance (i.e., h).

4.4. Optimization of the Crämer Transform

Remember that

$$\psi^*(P(1+\alpha)) = \sup_{\lambda > 0}\{\lambda P(1+\alpha) - \ln(P_0 f_M(e^{\lambda/R^M}) + 1 - P_0)\}, \tag{83}$$

$$\psi^*(P(1-\alpha)) = \sup_{\lambda < 0}\{\lambda P(1-\alpha) - \ln(P_0 f_M(e^{\lambda/R^M}) + 1 - P_0)\}. \tag{84}$$

Considering the gradient of the functions, we prove that the supremum for $\lambda \geq 0$ is reached near zero. So we can use the upper bounds for f_M obtained in the previous subsection, which leads to lower bounds for ψ^*:

$$\psi^*(P(1+\alpha)) \geq F(P(1+\alpha)) \quad \text{and} \quad \psi^*(P(1-\alpha)) \geq G(P(1-\alpha)), \tag{85}$$

where

$$F(x) = \sup_{\lambda>0}\left[\lambda x - \ln\left(1 + P_0 \frac{(e^{\lambda/R^M} - 1)}{1 - M\alpha_1(e^{\lambda/R^M} - 1)}\right)\right]$$

and

$$G(x) = \sup_{\lambda<0}\left[\lambda x - \ln\left(1 - P_0 \frac{(1 - e^{\lambda/R^M})[1 - \alpha_1(1 - e^{\lambda/R^M})]}{u_0}\right)\right].$$

Finally,

$$\mathbb{P}\left(\frac{|\hat{P} - P|}{P} \geq \alpha\right) \leq 2e^{-N\min(F(P(1+\alpha)),G(P(1-\alpha)))}. \tag{86}$$

One can easily obtain explicit but complex expressions for $F(x)$ and $G(x)$. We plot in Figure 5 the upper bounds obtained by the variance and by the Laplace transform, for different values of α, the prescribed error of the confidence interval. We take $P = 10^{-9}$ and the optimal values obtained above for the parameters. Note that the upper bound given by the Laplace transform is better than the bound given by Chebychev's inequality, with the variance. We obtain $\mathbb{P}(|(\hat{P} - P)/P| \geq \alpha) \leq L$. In the preceding example where $P = 10^{-9}$, if we fix $\alpha = 0.05$ and L close to 0.01, then the corresponding costs needed are 3×10^7 for the variance and 3×10^6 for the Laplace transform.

5. CONCLUSION

The simplified model described here has two main faults. First, we cannot choose in general the optimal level P_i. In practice, we just have an empirical estimation of the P_i, and we can adjust the levels according to them. A more precise analysis is then needed to get confidence intervals of the estimation. Second, the optimal sampling number at each level is not an integer in general. Therefore, in practice, the number of particles generated at each step should be chosen at random, either such that $\mathbb{E}(R) = 1/P_0$ or $\mathbb{E}(1/R) = P_0$. Thus, we finally need to work in a random environment. This requires a precise asymptotic of random iterates of the Laplace transform, where analysis is more delicate than the one presented here and will be the subject of a forthcoming paper.

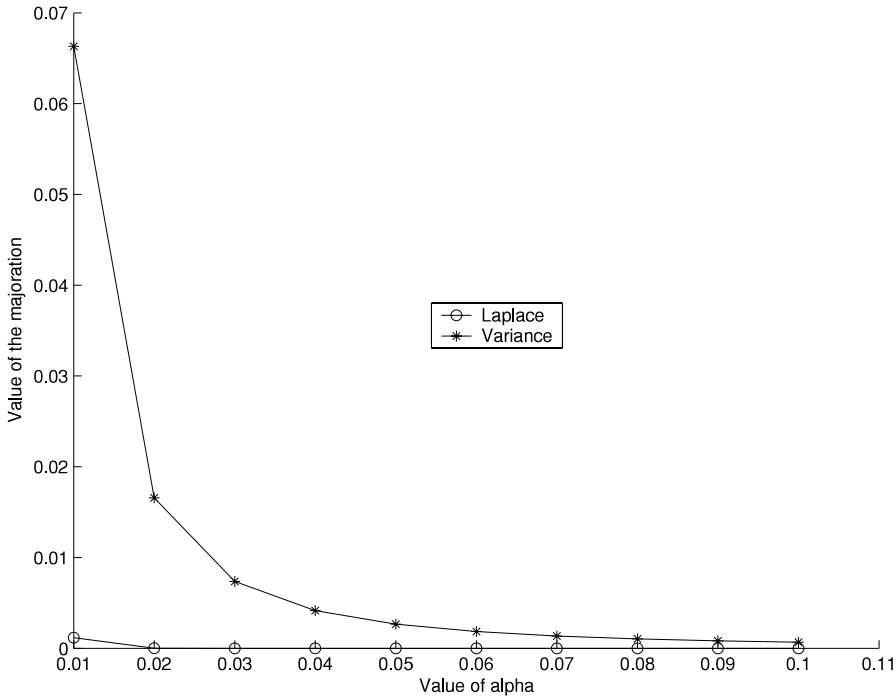


FIGURE 5. Upper bounds obtained by the variance and the Laplace transform.

Acknowledgments

I gratefully acknowledge my supervisors Dominique Bakry and Pascal Lezaud for their helpful suggestions and comments. I would also like to thank Persi Diaconis for telling me about approximate counting and Pieter-Tjerk de Boer for making a number of valuable remarks that improved the precision of the figures.

References

1. Aldous, D. (1989). *Probability approximations via the Poisson clumping heuristic*. Applied Mathematical Sciences. Vol. 77. New York: Springer-Verlag.
2. Athreya, K.B. & Ney, P.E. (1972). *Branching processes*. Die Grundlehren der mathematischen Wissenschaften Band 196. New York: Springer-Verlag.
3. Cerou, F., Del Moral, P., Legland, F., & Lezaud, P. (2002). Genetic genealogical models in rare event analysis. Preprint. Available from http://www.lsp.ups-tlse.fr/delmoral/preprints.html.
4. Cosnard, M. & Demongeot, J. (1984). Théorèmes de point fixe et processus de Galton–Watson. *Annales des sciences mathématiques du Québec* 8(1): 5–21.
5. Diaconis, P. & Holmes, S. (1995). Three examples of Monte-Carlo Markov chains: At the interface between statistical computing, computer science, and statistical mechanics. In D. Aldous, P. Diaconis, J. Spencer, & J.M. Steele (eds.), *Discrete probability and algorithms*. IMA Volumes in Mathematics and its Applications, Vol. 72. New York: Springer-Verlag, pp. 43–56.
6. Harris, T.E. (2002). *The theory of branching processes*. Dover Phoenix Editions. Mineola, NY: Dover.

7. Jerrum, M. & Sinclair, A. (1996). The Markov chain Monte Carlo method: An approach to approximate counting and integration. In D. Hochbaum (ed.), *Approximation algorithms for NP-hard problems*. Boston: PWS Publishing, pp. 482–520.
8. Lyons, R. (2002). Probability on trees and networks. Preprint. Available from http://mypage.iu.edu/rdlyons/.
9. Norris, J.R. (1998). *Markov chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press.
10. Sadowsky, J.S. (1996). On Monte Carlo estimation of large deviations probabilities. *Annals of Applied Probability* 6(2): 399–422.
11. Villen-Altamirano, J. & Villen-Altamirano, M. (1994). *RESTART: A method for accelerating rare event simulations*. Amsterdam: North-Holland, pp. 71–76.