A regenerative modification of the multilevel splitting

Alexandra Borodina and Evsey Morozov

Institute of Applied Mathematical Research, Russian Academy of Sciences

and Petrozavodsk University, Russia

Supported by Russian Foundation for Basic Research, Grant 07-07-00088.

1 Introduction

The multilevel splitting (MS):

huge simulation time reduction vs crude Monte Carlo (CMC);

statistical properties of the estimators are not studied in detail.

Known:

– Markov queues: the overflow probability estimate is unbiased [Garvels];

– consistency and asymptotic normality in adaptive MS [Cerou, Guyader, 2005].

For a regenerative queue: splitted processes as a family of (dependent) regenerative cycles $\Rightarrow$ *regenerative* MS.

We estimate the probability of overflow within a cycle for:

I.1) workload process in M/G/1 (and GI/G/1);

I.2) *non-Markov* queue using an embedded Markov chain (MC);

II. A stationary probability to exceed a level.

As a result: a randomization of the thresholds and a change of algorithm.

2 Preliminaries

I. A Markov queueing process $X = \{X_n, n \geq 1\}$: state space E;

MS is (typically) used to estimate the probability $\gamma_c = \mathsf{P}(X\ hits\ set\ A\ before\ 0)$

1. Given thresholds $0 < L_1 < \cdots < L_M \equiv L$;

2. Partition of $E:\ A = C_M \subset \cdots \subset C_1$: $C_i = \{x \in E : x \geq L_i\}$;

3. R_i copies upon reaching L_i; termination at L.

Then $\gamma_c = p_1 p_2 \cdots p_M$ with unbiased estimate

$$\gamma(m) = \prod_{i=1}^{M} \hat{p}_i = \frac{N}{mR_1 \cdot \cdots \cdot R_M}; \tag{1}$$

estimate $\hat{p}_i$ of $p_i = P(C_i | C_{i-1})$ is evaluated by CMC;

$N = \#\{$ reaching $A = [L, \infty)\}$;

$m =$ the number of i.i.d. runs starting at 0. (More details: [5, 6, 7, 16].)

II. A process X_n, $n \geq 1$, with *regeneration instants*:

$$T(0) = 0,\ T(n+1) = \min(k > T(n) : X_k = 0),\ n \geq 0; \tag{2}$$

the same distribution of $X_{n+T(k)}$, $n \geq 0$, for each $k \geq 1$ and independent of the

X_n, $n < T(k)$;

i.i.d. regeneration cycles $G_n = (X_k,\, T(n-1) \leq k < T(n))$;

i.i.d. cycle periods $\beta_n = T(n) - T(n-1)$.

We estimate the *cycle overflow probability*

$$\gamma_c = \mathsf{P}(\max_{1 \leq n < \beta} X_n \in A), \tag{3}$$

and, if weak limit $X_n \Rightarrow X$ exists, the *stationary probability*

$$\gamma_s = \mathsf{P}(X \in A), \tag{4}$$

for queue size and for workload in M/G/1 queue. (A generic element with no index.)

I. **Estimating** γ_c (under stability condition). Consistency: i.i.d indicators $I_k =$

$I(\max_{T(k)\le n<T(k+1)} X_n \in A)$;

$$\gamma_c(n) = \frac{1}{n}\sum_{k=1}^{n} I_k \to \mathsf{E}I = \gamma_c; \tag{5}$$

asymptotical normality: by regenerative CLT.

II. **Estimating** γ_s: positive recurrence, $\mathsf{E}\beta < \infty$; (dependent) indicators $I_k = I(X_k \in A)$ and i.i.d. variables

$$Y_i = \sum_{k=T(i-1)}^{T(i)-1} I_k,\ i \geq 1. \tag{6}$$

Since $Y \leq \beta$, by regenerative theory,

$$\gamma_s(n) = \frac{1}{n}\sum_{k=1}^{n} I_k \to \frac{\mathsf{E}Y}{\mathsf{E}\beta} = \gamma_s,\ n \to \infty, \tag{7}$$

the time average $\mathsf{E}Y/\mathsf{E}\beta = \mathsf{P}(X \in A)$ if weak limit X exists (e.g. if β is aperiodic).

Confidence $100(1-\delta)\%$ interval for γ_s (by a regenerative CLT):

$$\left[\gamma_n - \frac{z_\delta \sqrt{v_s(n)}}{\sqrt{n}},\ \gamma_n + \frac{z_\delta \sqrt{v_s(n)}}{\sqrt{n}}\right], \tag{8}$$

$P(N(0,1) \leq z_\delta) = 1 - \delta/2$; where empirical variance

$$v_s(n) = \frac{1}{n}\frac{\sum_{i=1}^{n}(Y_i - \gamma_s(n)\beta_i)^2}{\overline{\beta}_n^2} \Rightarrow \sigma_s^2 \equiv Var(Y - \beta\gamma), \tag{9}$$

($\overline{\beta}_n =$ sample mean cycle length.) A minimal condition for the convergence:

$$\mathsf{E}(Y - \gamma\beta)^2 < \infty.$$

(Under E$Y^2 < \infty$, E$\beta^2 < \infty$, v_n is strongly consistent [8].)

Applicability to dependent cycles [1, 12, 15]: below D-groups of dependent cycles with $D = R_1 \times \cdots \times R_M$.

III. M/G/1 queue: input rate λ, service time S, distribution F, stability: $\rho = \lambda$E$S < 1$. The queue size at the departure instants: a positive recurrent (aperiodic) Markov chain (MC):

$$\nu_{n+1} = (\nu_n - 1)^+ + \Delta_n,\ n \geq 0,\ (\nu_0 = 0), \tag{10}$$

$\Delta_n = \#\{arrivals\}$ during service of customer $n+1$.

Waiting time (workload) Markov chain: W_n, $n \geq 1$; $W_n \Rightarrow W$ exists.

If distribution of W is analytically available [Ross,Seshadri] [13]:

$$\gamma_c = \mathsf{P}(\max_{1 \leq n < \beta} W_n \geq L) = 1 - \frac{\mathsf{P}(W + S \leq L)}{\mathsf{P}(W \leq L)}. \tag{11}$$

M/M/1: $\rho = \lambda/\mu < 1$, $\mathsf{P}(W \leq x) = 1 - \rho e^{-(\lambda-\mu)x}$:

$$\gamma_c = \frac{(1-\rho)e^{-(\mu-\lambda)L}}{1-\rho e^{-(\mu-\lambda)L}}, \tag{12}$$

while (applying technique from the birth-death processes)

$$\mathsf{P}(\max_{1\le n<\beta} \nu_n \ge L) = \frac{\rho^{L-1}-\rho^L}{1-\rho^L}. \tag{13}$$

NOTE 1. M/M/1 has been used to justify the quality of MS [5, 6, 7, 9, 10, 14].

If service time has subexponential integrated-tail distribution

$$\overline{F_e}(x) = \frac{1}{\mathsf{E}S}\int_x^\infty \overline{F}(y)dy, \quad (\overline{F} = 1 - F)$$

and

$$\lim_{x\to\infty} \frac{\overline{F}(xe^{y/\sqrt{x}})}{\overline{F}(x)} = 1, \text{ locally uniformly in } y, \in \mathsf{R}, \tag{14}$$

(holds for Pareto distribution) then stationary queue ν satisfies [2]

$$\gamma_s = P(\nu \geq L) \sim \frac{\rho}{1-\rho}\overline{F_e}(\frac{L-1}{\lambda}), \;\; L \to \infty. \tag{15}$$

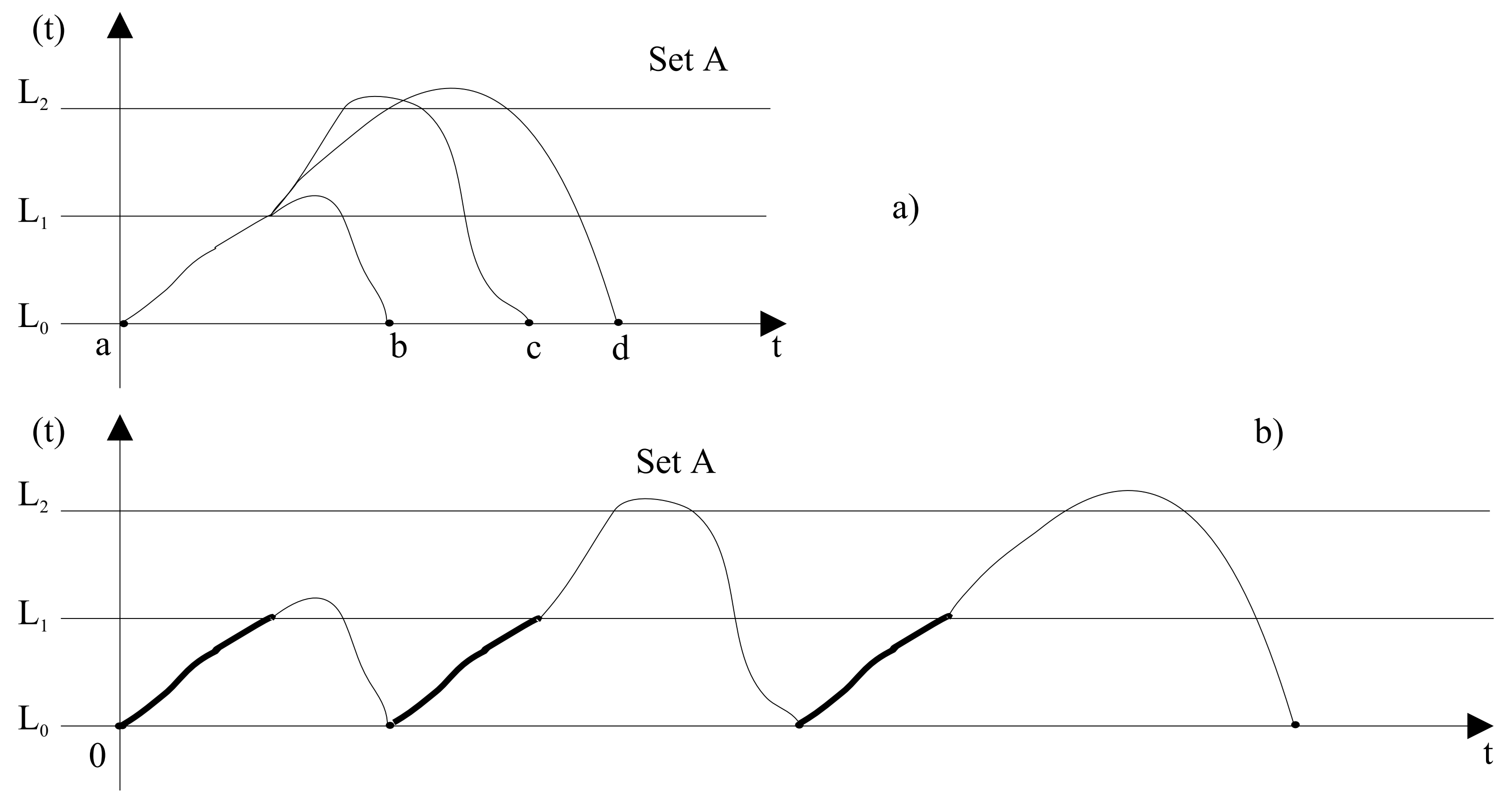
3. Splitting and regenerations in $M/G/1$

by splitting: dependent regeneration (sub)cycles:

$$\nu^{(i)} = \{\nu_n^{(i)},\, \tau_{i-1} \leq n < \tau_i\},\, \tau_0 = 0,\, i = 1, \ldots, mD \;\; (D = R_1 \times \cdots \times R_M)$$

with periods $\alpha_i = \tau_i - \tau_{i-1}$. D-group: $(\nu^{(i)} : kD < i \leq (k+1)D)$, $k = 0, \ldots, m-1$.

Fig. 1: splitted processes with termination instants b, c, d (case a)) and cycles with regenerations τ_n *pasted together* with common pre-history, with no idle-time, case b).



1. **Estimating** $\gamma_c = \mathsf{P}(\max_{1\le n<\beta} \nu_n \in A)$:

1.a) Markov queue: standard MS, termination upon reaching A with (at most D-dependent) indicators

$$I_i = I(\max_{\tau_{i-1}\le n<\tau_i} \nu_n^{(i)} \in A), \;\; i = 1, \ldots, mD. \tag{16}$$

To compensate overestimation if a process starting at L_i reaches 0 before L_{i+1}: take extra $R_{i+1}R_{i+2}\cdots R_M$ (virtual) cycles, $i = 1, \ldots, M-1$.

1.b) Non-Markov M/G/1 queue, to keep independence after splitting: use MC

(10): makes transitions between regions $G_i = [L_i,\, L_{i+1}) \Rightarrow$:

A *randomization of the thresholds* and modification of the algorithm:

Condition A: Under transition $y \in G_i \to x \in G_k,\ k > i$ (crossing the thresholds $L_{i+1}, \ldots, L_k$), generate $R_{i+1} \cdots R_k$ processes at state x, $i = 0, \ldots, M-1;\ k = i+1, \ldots, M\ (R_M = 1)$.

2. **Estimating** $\gamma_s = \mathsf{P}(\nu \geq L)$ (weak limit $\nu_n \Rightarrow \nu$ exists): evaluate the time (=number of arrivals) the process is in A using indicators $I_{i,n} = I(\nu_n^{(i)} \in A)$.

Apply condition **A** evaluating the *number of arrivals* in all processes.

For W_n, $n \geq 1$ with jumps $X_n(\neq \pm 1)$: the same randomization.

NOTE 2. *By PASTA: estimating overflow probability for continuous time.*

Randomization is widely used, [3, 11].

4. Consistency and asymptotic normality

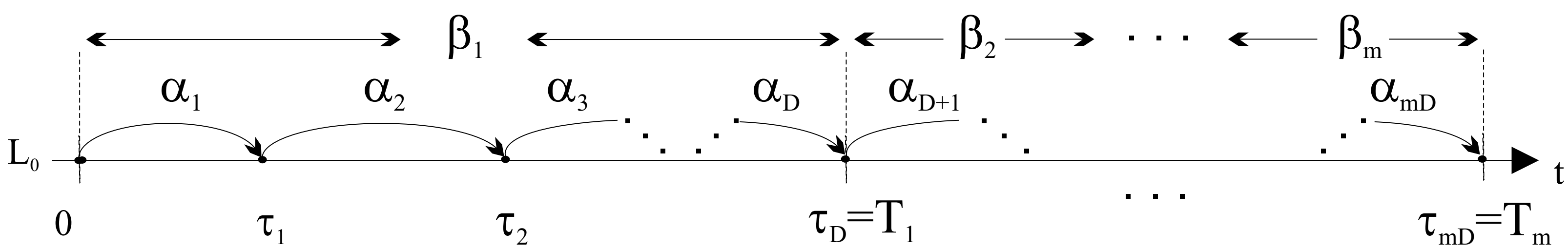
1. Consistency: estimate $\gamma(m)=$ regenerative estimate (5) based on $n = mD$

dependent indicators (16), so:

$$\frac{1}{n}\sum_{i=1}^{n} I_i \to \frac{\mathsf{E}\sum_{i=1}^{D} I_i}{D} = \gamma_c. \tag{17}$$

NOTE 2. *No need $n = mD$ for the convergence.*

2. Estimating $\gamma_s = \mathsf{P}(\nu \geq L)$: cycle structure, Fig 2.



Variables

$$Z_i = \sum_{k=\tau_{i-1}}^{\tau_i - 1} I(\nu_k^{(i)} \geq L), \tag{18}$$

belonging to the same D-group are dependent; i.i.d variables

$$Y_i = \sum_{j=(i-1)D+1}^{iD} Z_j, \quad i = 1, \ldots, m; \tag{19}$$

the length of the ith D-group is

$$\beta_i = \sum_{j=(i-1)D+1}^{iD} \alpha_j,$$

so $\beta_i = T(i) - T(i-1)$, $i = 1, \ldots, m$.

Since $Y \leq \beta =_{st} \alpha_1 + \cdots + \alpha_D$ (see Fig. 2) and E$\alpha < \infty$ by $\rho < 1$, then E$Y < \infty$, and (by regenerative theory) the estimate

$$\gamma_s(n) = \frac{\sum_{i=1}^n Z_i}{\sum_{i=1}^n \alpha_i} \to \frac{\mathsf{E}Y}{\mathsf{E}\beta} = \mathsf{P}(\nu \geq L) \equiv \gamma_s,\ n \to \infty \tag{20}$$

is consistent. The limit $\nu_n \Rightarrow \nu$ exists in the system M/G/1.

NOTE 4. *To estimate* γ_s *in discrete-time setting one can choose any discrete-event scale because, for each cycle, the number of arrivals equals the number of departures.*

An asymptotic normality: by regenerative interpretation.

Applicability of CLT for estimating $\gamma_s = \mathsf{P}(\nu \geq L)$:

assumption $\mathsf{E}S^2 < \infty$ implies $\mathsf{E}\alpha^2 < \infty$, Wolff [20].

Since $\beta =_{st} \alpha_1 + \cdots + \alpha_D$ then $\mathsf{E}\beta^2 < \infty$.

Similarly, consistency and asymptotic normality of the estimates for the workload.

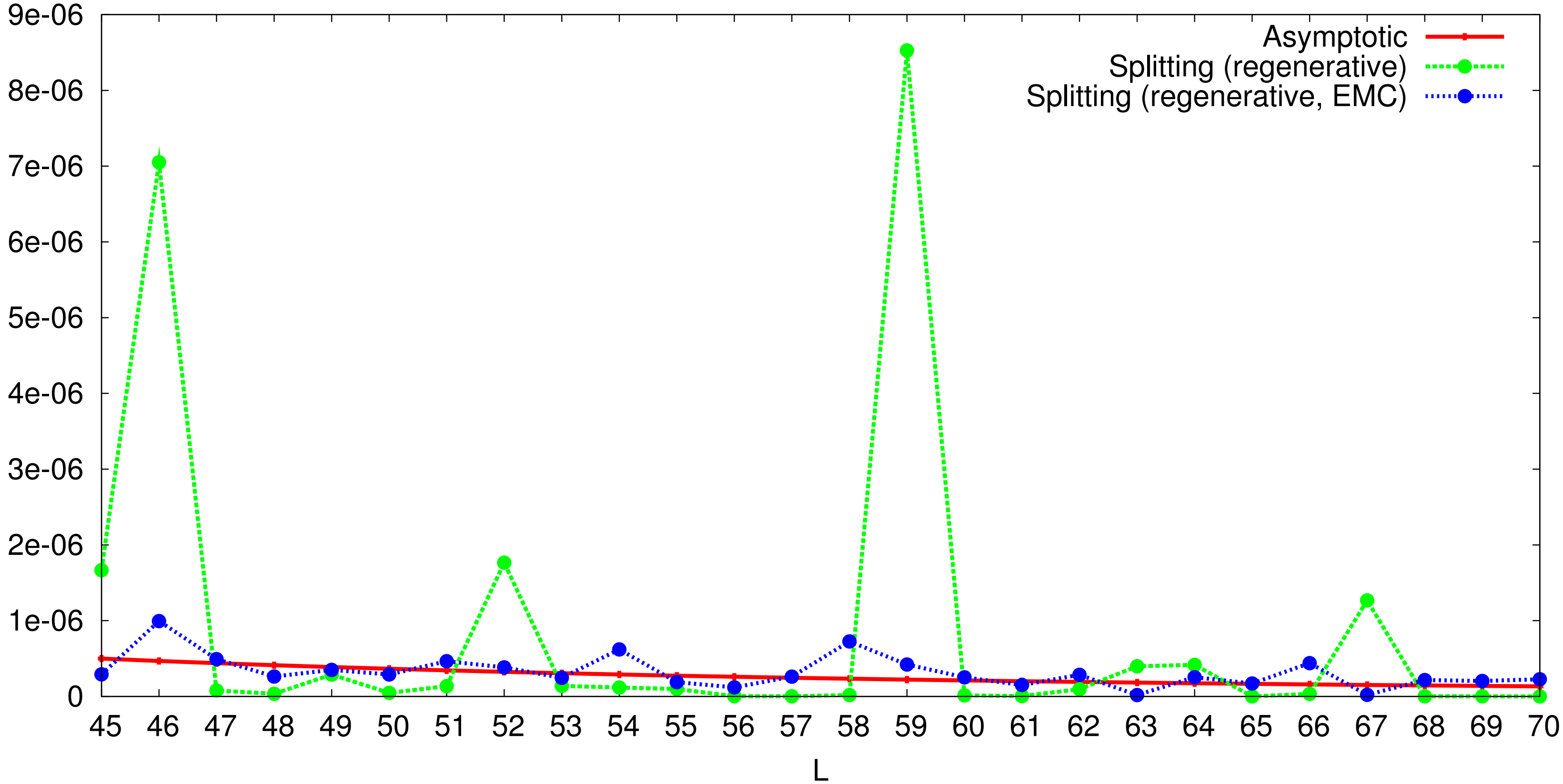
In confidence interval (8): variance includes Y_i depending on target probability.

5. Simulation results

Fig. 3 estimate of $\gamma_s = \mathsf{P}(\nu \geq L)$ in $M/Pareto/1$ with service time distribution

$$\overline{F}(x) = 1/x^4,\ x \geq 1 (= 1,\ x \leq 1), \tag{21}$$

input rate $\lambda = 0.45$ ($\rho = 0.6$), $N_{M+1} = 10^6$.


9e-06
8e-06
7e-06
6e-06
5e-06
4e-06
3e-06
2e-06
1e-06
0
45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70
L
Asymptotic
Splitting (regenerative)
Splitting (regenerative, EMC)

The simulation stopping rule: a given number N of the target events to be achieved.

For instance, for $L = 50$ we take $L_i = \{7, 14, 21, 28, 35, 42\}$ (that is $M = 6$), $N = 10000$ and $R_i \equiv 10$.

More details in Table 1: S=standard MS, EMC= embedded MC.

The estimate $\gamma_s(n)$ is (20), and "asymp" = (15).

(In all experiments the number of replications $m = nD$ is widely varied in $[10^3, 10^4]$.)

Table 1

L	asymp	$\gamma_s(n)[S]$	$\gamma_s(n)[EMC]$	t[S]	t[EMC]	Var[S]	Var[EMC]
31	$1.53 \cdot 10^{-6}$	$2.98 \cdot 10^{-7}$	$1.53 \cdot 10^{-6}$	18.5	24.2	$1.01 \cdot 10^{-8}$	$2.40 \cdot 10^{-10}$
36	$9.77 \cdot 10^{-7}$	$3.49 \cdot 10^{-8}$	$1.78 \cdot 10^{-7}$	18.7	34.3	$7.20 \cdot 10^{-9}$	$7.33 \cdot 10^{-11}$
40	$7.12 \cdot 10^{-7}$	$9.21 \cdot 10^{-9}$	$2.35 \cdot 10^{-7}$	17.2	45.3	$1.49 \cdot 10^{-10}$	$6.15 \cdot 10^{-11}$
45	$5.0 \cdot 10^{-7}$	$4.46 \cdot 10^{-8}$	$1.24 \cdot 10^{-7}$	19.7	38.6	$2.56 \cdot 10^{-8}$	$5.77 \cdot 10^{-10}$
50	$3.65 \cdot 10^{-7}$	$2.24 \cdot 10^{-8}$	$3.59 \cdot 10^{-7}$	51.4	91.1	$4.20 \cdot 10^{-11}$	$3.54 \cdot 10^{-13}$

EMC: an increasing of simulation time t and

HUGE variance reduction.

NOTE: a consistency between EMC simulation and asymptotic formula (15).

Fig. 4: estimation of P$(\max_{1 \leq n < \beta} W_n \geq L)$ in $M/M/1$ (with $\rho = 0.6$).

$$\gamma_c(EMC) = 3.12 \cdot 10^{-11},\ 4.09 \cdot 10^{-12},\ 4.49 \cdot 10^{-13},$$

(12): $\gamma_c = 1.51 \cdot 10^{-11},\ 2.04 \cdot 10^{-12},\ 2.77 \cdot 10^{-13}.$

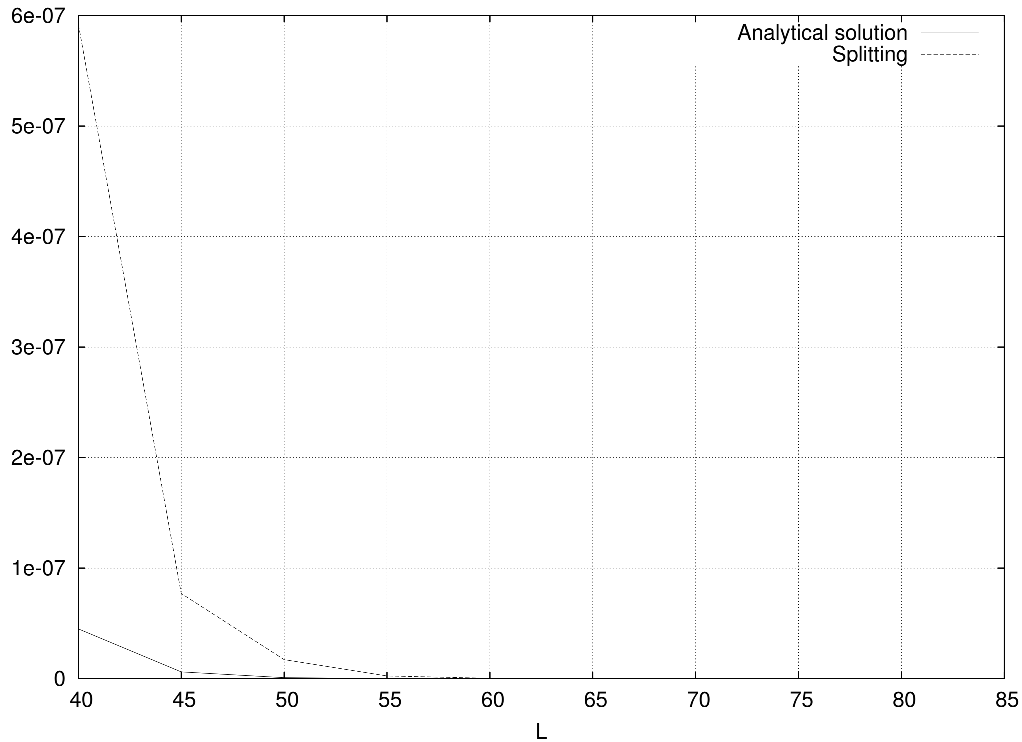


Fig. 5: estimate of $\gamma_c = \mathsf{P}(\max_{1 \le n < \beta} W_n \ge L)$ in $M/Pareto/1$, service time distribution (21).

CMC vs EMC splitting , $\rho = 0.6$, $N[EMC] = 1000$, $N[CMC] = 100$

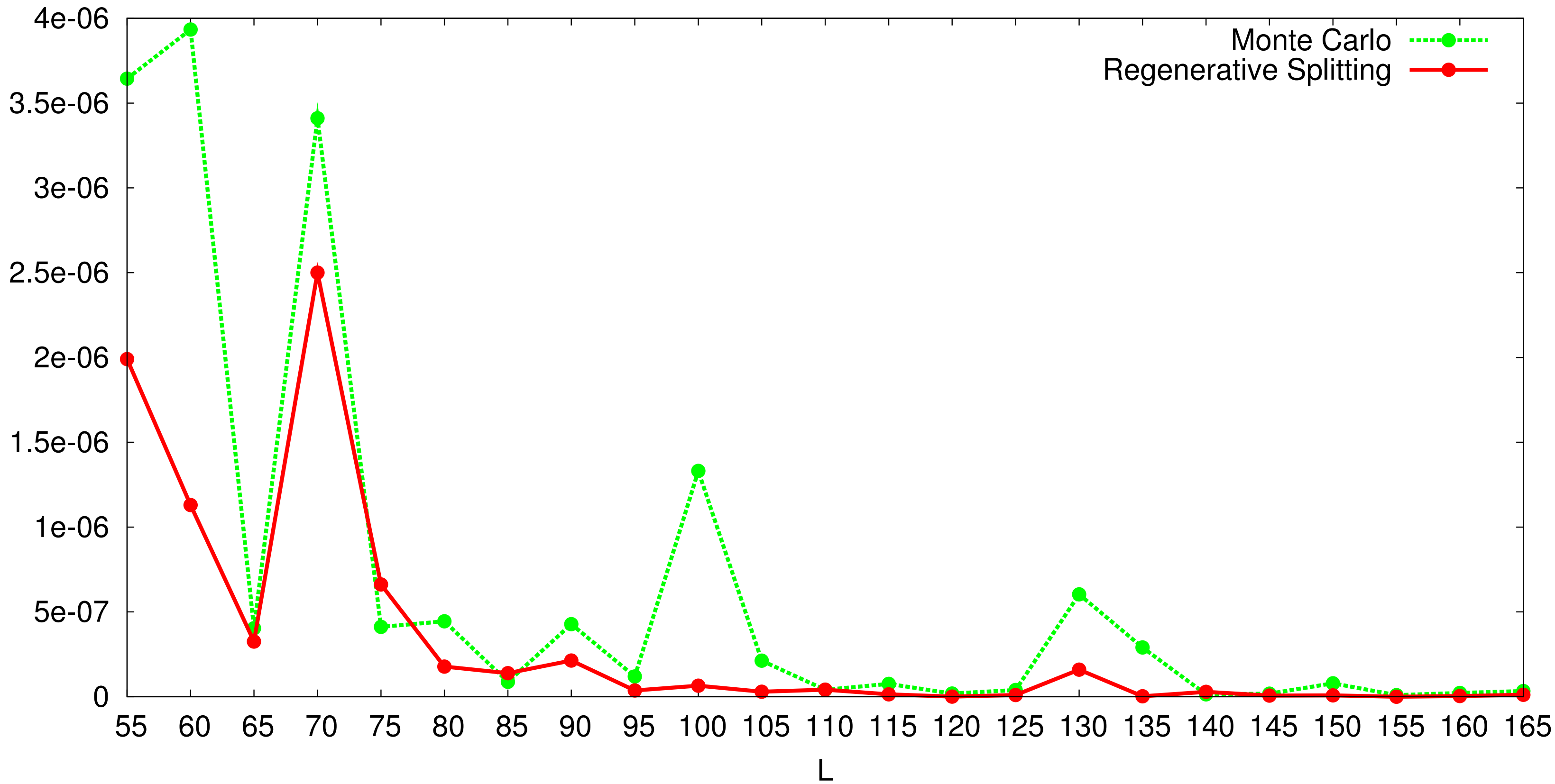
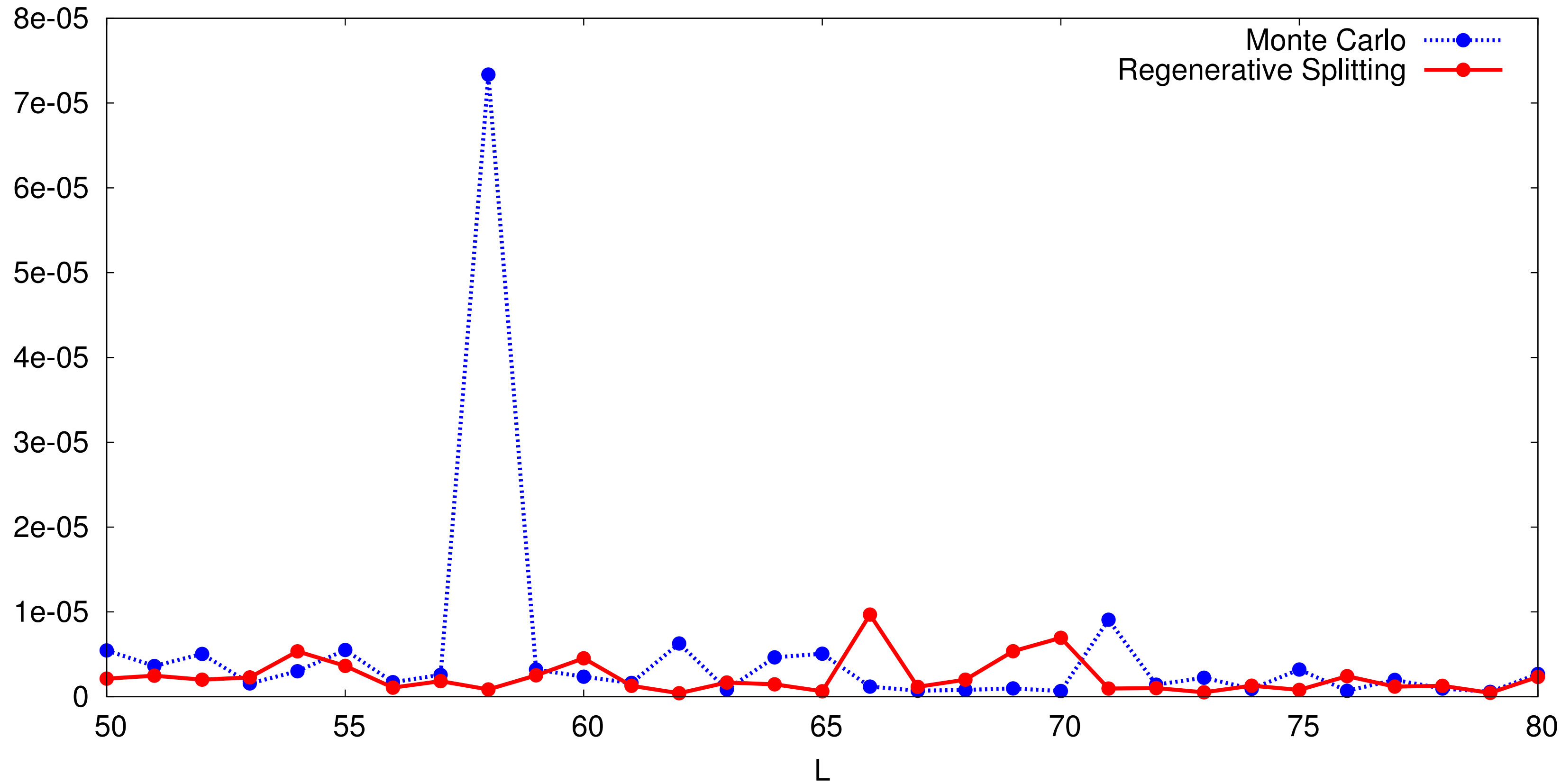


Fig. 6: estimate of $\gamma_c = \mathsf{P}(\max_{1 \le n < \beta} W_n \ge L)$ in $Pareto/Pareto/1$, input distribution $P(u \ge x) = 1/x^3,\ x \ge 1 (= 1\ x \le 1)$, service time distribution (21) ($\rho = 8/9$) EMC vs CMC.



NOTE:

i) a consistency between EMC and CMC;

ii) more unstable CMC estimate;

iii) LRD stationary workload: $\mathsf{E}S^3 < \infty$, $\mathsf{E}S^4 = \infty$ [4, 12].

In all experiments, the cycles from the same D-group are highly dependable.

3 Concluding remarks

1. Extension of experiments;
2. Source of variance reduction (Markovity)?
3. Selection of the thresholds;
4. Combination randomization with given thresholds;
5. More...

References

[1] S. Asmussen. *Applied Probability and Queues*, 2nd ed., Springer, NY, 2003.

[2] S. Asmussen, C. Klüppelberg, K. Sigman. Sampling at subexponential times, with queueing applications. Stochastic Process. Appl. 79 ,1999, 265-286.

[3] F.Cerou, A. Guyader. Adaptive multilevel splitting for rare event analysis. INRIA, Research report No 5710, Oct. 2005.

[4] D. J. Daley. The serial correlation coefficients of waiting times in a stationary single server queue. J.Aust.Math.Soc. 8, 1968, 683-699.

[5] M. Garvels. PhD Thesis *The splitting method in rare event simulation*. The University of Twente, The Nether-

lands May, 2000.

[6] P. Glasserman, P. Heidelberger, P. Shahabuddin, and T. Zajic. A look at multilevel splitting. In H. Niederreiter (ed.): Monte Carlo and Quasi Monte Carlo Methods, Lecture Notes in Statistics, vol. 127, 1996, 99-108, Springer-Verlag.

[7] P. Glasserman, P. Heidelberger, P. Shahabuddin, and T. Zajic. Splitting for rare event simulation: analysis of simple cases. Proceedings of the 1996 Winter Simulation Conference, 1996, 302-308.

[8] P. W. Glynn, D. L. Iglehart. Conditions for the applicability of the regenerative method. Management Science 39, 1993, 1108-1111.

[9] P. E. Heegaard. A survey of Speedup simulation techniques. Workshop tutorial on Rare Event Simulation, Aachen, Germany, 1997.

[10] P. Heidelberger. Fast simulation of rare events in queueing and relaibility models. Performance Evaluation of Computers and Communications Systems Springer-Verlag, LN in Computer Sci., 729, 1993, 165-202.

[11] S.Ermakov, V.Melas *Mathematical experiment with complex stochstic models*, S.-Petersburg University, 1993. (In Russian)

[12] E. Morozov. Weak regeneration in modeling of queueing processes. Queueing Systems, 46, 2004, 295-315.

[13] S. M. Ross, S. Seshadri. Hitting Time in an $M/G/1$ Queue. J. Appl. Prob., 36, 1999, 934-940.

[14] P. Shahabuddin. Rare event simulation in stochastic models. Proceedings of the WSC 1995, IEEE Press., 178-185 1995.

[15] K. Sigman, R. Wolff. A review of regenerative processes. SIAM Review, 35, No. 2, 1993, 269-288.

[16] M. Villen-Altamirano, J. Villen-Altamirano. RESTART: A Method for Accelerating Rare Event Simulations. Proceedings of the 13-th International Teletraffic Congress, Queueing, Perform-ance and Control in ATM, 1991, 71-76.

[17] M. Villen-Altamirano, J. Villen-Altamirano. RESTART: A Straightforward Method for Fast Simulation of Rare Events. Proceedings of the 1994 Winter Simulation Conference, 1994, 282-289.

[18] M. Villen-Altamirano, J. Villen-Altamirano. About the Efficiency of RESTART. Proceedings of the RESIM'99 Workshop, University of Twente, The Netherlands, 1999, 99-128.

[19] M. Villen-Altamirano, J. Villen-Altamirano. Analysis of RESTART Simulation: Theoretical Basis and Sensitivity Study. European Transactions on Telecommunications vol. 13, No. 4, 2002, 373-385.

[20] R.W. Wolff. *Stochastic Modeling and the Theory of Queues*, Prentice-Hall, 1989.

4 APPENDIX

Example A. $\gamma_s = \mathsf{P}(\nu \geq L)$ in $M/Pareto/1$, $\rho = 0.6$

Standard S vs EMC for L small;

EMC: consistency with asymptotic formula;

EMC: less variance.

Table A

L	asymp	$\hat{\gamma}_s[S]$	$\hat{\gamma}_s[EMC]$	t[S]	t[EMC]	Var[S]	Var[EMC]
10	$4.55 \cdot 10^{-5}$	$4.11 \cdot 10^{-4}$	$4.41 \cdot 10^{-4}$	2.1	1.6	$1.85 \cdot 10^{-9}$	$1.56 \cdot 10^{-9}$
15	$1.35 \cdot 10^{-5}$	$9.77 \cdot 10^{-5}$	$2.23 \cdot 10^{-5}$	4.2	3.2	$5.32 \cdot 10^{-9}$	$5.63 \cdot 10^{-10}$
20	$5.69 \cdot 10^{-6}$	$3.52 \cdot 10^{-6}$	$4.14 \cdot 10^{-6}$	3.2	3.6	$1.41 \cdot 10^{-7}$	$8.21 \cdot 10^{-9}$
25	$2.91 \cdot 10^{-6}$	$6.67 \cdot 10^{-6}$	$2.32 \cdot 10^{-6}$	3.2	5.3	$7.54 \cdot 10^{-9}$	$2.04 \cdot 10^{-10}$
30	$1.68 \cdot 10^{-6}$	$5.23 \cdot 10^{-7}$	$1.03 \cdot 10^{-6}$	11.4	10.9	$4.21 \cdot 10^{-9}$	$4.74 \cdot 10^{-11}$

Example A. (continued)

Table B: M (R_i) - number of thresholds (copies);

D$=$ number of dependent paths;

N$=$number of rare events to be achieved$=$stopping rule;

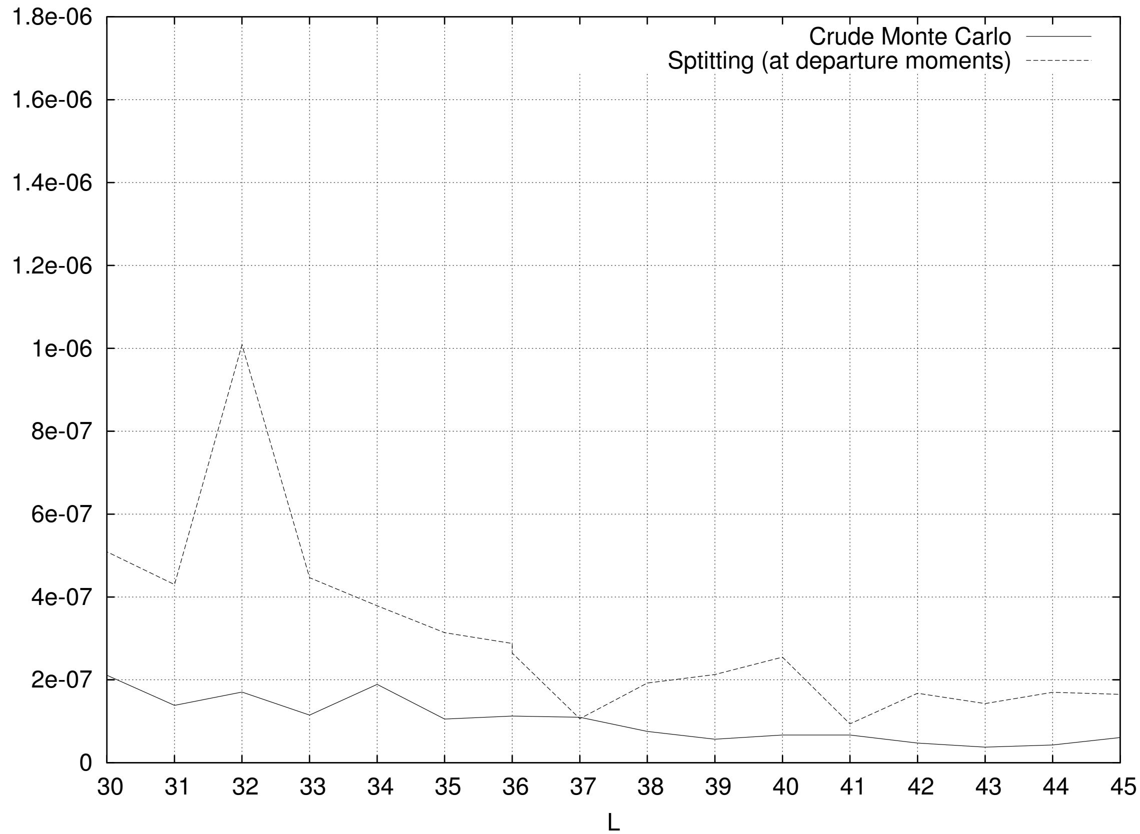
N[S],N[EMC] $=$actual number of reaching set A

$m$$=$ number of the i.i.d. trials.

Table B

L	M(R_i)	D	N	N[S]	N[EMC]	m[S]	m[EMC]
10	1(10)	10	10000	10051	10131	491749	520347
15	2(10)	100	10000	17882	11265	367591	1145046
20	3(10)	1000	10000	10028	10176	71631	954783
25	4(10)	10000	10000	13587	129396	117678	451050
30	5(10)	100000	100000	252883	188500	971845	1251389

Example B. CMC vs EMC: $\gamma_s = \mathsf{P}(\nu \geq L)$ in $M/Pareto/1$.



Example B(continued): simulation time (sec) S vs EMC