

Regenerative-Simulation-Based Estimators of Risk Measures for Hitting Times to Rarely Visited Sets

Peter W Glynn, Marvin K Nakayama, Bruno Tuffin

▶ To cite this version:

Peter W Glynn, Marvin K Nakayama, Bruno Tuffin. Regenerative-Simulation-Based Estimators of Risk Measures for Hitting Times to Rarely Visited Sets. RESIM 2021 13th International Workshop on Rare Event Simulation, May 2021, Paris, France. hal-03499024

HAL Id: hal-03499024 https://inria.hal.science/hal-03499024

Submitted on 21 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés. Regenerative-Simulation-Based Estimators of Risk Measures for Hitting Times to Rarely Visited Sets

Peter W. Glynn, Marvin K. Nakayama, Bruno Tuffin

Stanford University, NJIT, Inria

RESIM 2021



Context and goals

- Common in the literature to estimate $\mu = \mathbb{E}[T]$ with $T = \inf\{t \ge 0 : X(t) \in A\}$, hitting time of A
- But what about the distribution *F* of *T*, in a rare event context?
 - May be required to estimate q-quantiles (0 < q < 1) of hitting times:

 $\xi = F^{-1}(q) \equiv \inf\{t : F(t) \ge q\}$

and the conditional tail expectation (CTE)

 $\gamma = E[T \mid T > \xi].$

- We designed two (*exponential and convolution*) estimators in a regenerative context
 Glynn, Nakayama & T., WSC 2018
- And even 3 variations not particularly producing improvements Glynn, Nakayama & T., WSC 2020

Our Goals:

イロト 不得 トイヨト イヨト 二日

Context and goals

- Common in the literature to estimate $\mu = \mathbb{E}[T]$ with $T = \inf\{t \ge 0 : X(t) \in A\}$, hitting time of A
- But what about the distribution *F* of *T*, in a rare event context?
 - May be required to estimate q-quantiles (0 < q < 1) of hitting times:

 $\xi = F^{-1}(q) \equiv \inf\{t : F(t) \ge q\}$

and the conditional tail expectation (CTE)

 $\gamma = E[T \mid T > \xi].$

- We designed two (*exponential and convolution*) estimators in a regenerative context
 Glynn, Nakayama & T., WSC 2018
- And even 3 variations not particularly producing improvements

Glynn, Nakayama & T., WSC 2020

イロト 不得 トイヨト イヨト 三日

Our Goals:

1/ Recall the exponential and convolution estimators and their efficiency 2/ Illustrate their respective power on a simple example.

Outline



Estimators

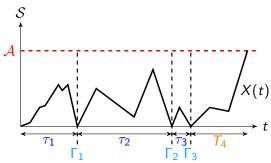
- Exponential Approximation Estimator
- Convolution Estimator
- 3 Numerical efficiency
- 4 Analysis on a toy example

Conclusions

< □ > < 同 > < 回 > < 回 > < 回

Regenerative system

• Regeneration times $0 = \Gamma_0 < \Gamma_1 < \cdots$, with iid cycles $((\tau_k, (X(\Gamma_{k-1} + s) : 0 \le s < \tau_k) : k \ge 1))$

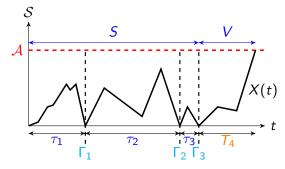


• $\tau_k = \Gamma_k - \Gamma_{k-1}$, length of the *k*th regenerative cycle

- $T_k = \inf\{t \ge 0 : X(\Gamma_{k-1} + t) \in A\}$ first hitting to A after regeneration Γ_{k-1}
- $M = \sup\{i > 0 : T_i > \tau_i\}$ (# cycles before first hitting A)

Regenerative system

• Regeneration times $0 = \Gamma_0 < \Gamma_1 < \cdots$, with iid cycles $((\tau_k, (X(\Gamma_{k-1} + s) : 0 \le s < \tau_k) : k \ge 1)$



• We can express

$$T = \mathbf{S} + \mathbf{V} \equiv \sum_{i=1}^{M} \tau_i + T_{M+1},$$

where the geometric sum S is independent of V.

B. Tuffin (Inria)

Rare event and exponential limit

- Context: $p = \mathbb{P}(T < \tau)$ is small (rare event)
 - ▶ Model indexed by ϵ , rarity parameter, such that $p \equiv p_{\epsilon} \rightarrow 0$
- Ex: GI/G/1 queue
 - buffer size $b \equiv b_{\epsilon} = \lceil 1/\epsilon \rceil$
 - $\mathcal{A} = \mathcal{A}_{\epsilon} = \{b_{\epsilon}, b_{\epsilon} + 1, \dots\}$
- Ex: Highly Reliable System (HRS; HRMS in the Markovian case)
 - Multicomponent system with component j failure rate λ_j = c_j ϵ^{d_j} (d_j > 0)
 - Repair distributions independent of ϵ
 - Set \mathcal{A} : states with combinations of components down.

Theorem (Renyi's)

In the above contexts, if $p_\epsilon \to 0$ as $\epsilon \to 0$, T_ϵ/μ_ϵ converges weakly to an exponential

$$\lim_{\epsilon o 0} \mathbb{P}_\epsilon(T_\epsilon/\mu_\epsilon \leq t) = 1 - e^{-t}, \quad \forall t \geq 0.$$

(a)

Outline

1 Model: regenerative process

2 Estimators

- Exponential Approximation Estimator
- Convolution Estimator

3 Numerical efficiency

4 Analysis on a toy example

Conclusions

< □ > < 同 > < 回 > < 回 > < 回

Exponential Approximation Estimator

Glynn, Nakayama and T., WSC'18

•
$$F(t) = \mathbb{P}_{\epsilon}(T_{\epsilon} < t) = \mathbb{P}_{\epsilon}(T_{\epsilon}/\mu_{\epsilon} < t/\mu_{\epsilon}) \approx 1 - e^{-t/\mu_{\epsilon}}$$

- "Just" estimate the mean μ_ϵ
- Use the expression

$$\mu_{\epsilon} = \frac{\mathbb{E}_{\epsilon}[T_{\epsilon} \wedge \tau_{\epsilon}]}{\mathbb{P}_{\epsilon}(T_{\epsilon} < \tau_{\epsilon})} \equiv \frac{\zeta_{\epsilon}}{p_{\epsilon}}$$

- Measure-specific IS (MSIS): simulate *n* cycles Shahabuddin et al. (1988)
 - $n_{CS} \equiv \gamma n$ cycles to estimate by crude simulation (CS): $\hat{\zeta}_n \approx \zeta_{\epsilon}$
 - ▶ $n_{IS} \equiv (1 \gamma)n$ cycles to estimate by importance sampling (IS): $\hat{p}_n \approx p_\epsilon$

• Resulting estimator
$$\widehat{\mu}_n = rac{\zeta_n}{\widehat{p}_n}$$

Estimator

The exponential estimator of the cdf F(t) of T is

$$\widehat{F}_{exp,n}(t) = 1 - e^{-t/\widehat{\mu}_n}$$

Exponential estimators

• Remarkably, estimating the distribution is reduced to estimating its mean.

From

$$F(t) = \mathbb{P}(T \leq t) = \mathbb{P}(T/\mu \leq t/\mu) pprox 1 - e^{-t/\mu} \equiv \widetilde{F}_{\mathrm{exp}}(t),$$

we get

$$\widetilde{\xi}_{exp} = \widetilde{F}_{exp}^{-1}(q) = -\mu \ln(1-q)$$

 $\widetilde{\gamma}_{exp} = \widetilde{\xi}_{exp} + \mu = \mu [1 - \ln(1-q)].$

Using an efficient estimator $\widehat{\mu}$ of μ from the literature

L'Ecuyer & T., Annals of OR 2011

< ロト < 同ト < ヨト < ヨト

$$egin{aligned} \widehat{\xi}_{\mathrm{exp}} &= \widehat{F}_{\mathrm{exp}}^{-1}(q) = -\widehat{\mu}\ln(1-q) \ \widehat{\gamma}_{\mathrm{exp}} &= \widehat{\xi}_{\mathrm{exp}} + \widehat{\mu} = \widehat{\mu}[1-\ln(1-q)] \end{aligned}$$

Glynn, Nakayama and T., WSC'18

• From T = S + V, the cdf F can be expressed as the convolution

 $F = G \star H$ with $S \sim G$ and $V \sim H$.

< ロト < 同ト < ヨト < ヨト

Glynn, Nakayama and T., WSC'18

• From T = S + V, the cdf F can be expressed as the convolution

$$F = G \star H$$
 with $S \sim G$ and $V \sim H$.

• Exponential approximation for G when for $p \approx 0$: For each $t \ge 0$, $G(t) \approx \widetilde{G}_{exp}(t) = 1 - e^{-t/\eta}$ where $\eta = \mathbb{E}[S] = \mathbb{E}[M] \cdot \mathbb{E}[\tau \mid \tau < T].$ Kalashnikov (1997)

< ロト < 同ト < ヨト < ヨト

Glynn, Nakayama and T., WSC'18

n . .

< ロト < 同ト < ヨト < ヨト

• From T = S + V, the cdf F can be expressed as the convolution

$$F = G \star H$$
 with $S \sim G$ and $V \sim H$.

• Exponential approximation for G when for $p \approx 0$: For each $t \ge 0$, $G(t) \approx \widetilde{G}_{exp}(t) = 1 - e^{-t/\eta}$ where $\eta = \mathbb{E}[S] = \mathbb{E}[M] \cdot \mathbb{E}[\tau \mid \tau < T].$ Kalashnikov (1997)

• Estimator (CS):
$$\widehat{G}_{\exp,n}(t) = 1 - e^{-t/\widehat{\eta}_n}$$
 with $\widehat{\eta}_n = \frac{1}{\widehat{p}_n n_{CS}} \sum_{i=1}^{n_{CS}} \tau_i \mathcal{I}(\tau_i < T_i).$

Glynn, Nakayama and T., WSC'18

n . .

< ロト < 同ト < ヨト < ヨト

• From T = S + V, the cdf F can be expressed as the convolution

$$F = G \star H$$
 with $S \sim G$ and $V \sim H$.

• Exponential approximation for G when for $p \approx 0$: For each $t \ge 0$, $G(t) \approx \widetilde{G}_{exp}(t) = 1 - e^{-t/\eta}$ where $\eta = \mathbb{E}[S] = \mathbb{E}[M] \cdot \mathbb{E}[\tau \mid \tau < T].$ Kalashnikov (1997)

• Estimator (CS):
$$\widehat{G}_{\exp,n}(t) = 1 - e^{-t/\widehat{\eta}_n}$$
 with $\widehat{\eta}_n = \frac{1}{\widehat{p}_n n_{CS}} \sum_{i=1}^{n_{CS}} \tau_i \mathcal{I}(\tau_i < T_i).$

• Estimator (IS) of H:
$$\widehat{H}_n(x) = \frac{1}{\widehat{p}_n n_{lS}} \sum_{i=1}^{n_{lS}} \mathcal{I}(T'_i \wedge \tau'_i \leq x, T'_i < \tau'_i) L'_i$$

Glynn, Nakayama and T., WSC'18

• From T = S + V, the cdf F can be expressed as the convolution

$$F = G \star H$$
 with $S \sim G$ and $V \sim H$.

• Exponential approximation for G when for $p \approx 0$: For each $t \ge 0$, $G(t) \approx \widetilde{G}_{exp}(t) = 1 - e^{-t/\eta}$ where $\eta = \mathbb{E}[S] = \mathbb{E}[M] \cdot \mathbb{E}[\tau \mid \tau < T].$ Kalashnikov (1997)

• Estimator (CS):
$$\widehat{G}_{\exp,n}(t) = 1 - e^{-t/\widehat{\eta}_n}$$
 with $\widehat{\eta}_n = \frac{1}{\widehat{p}_n n_{CS}} \sum_{i=1}^{n_{CS}} \tau_i \mathcal{I}(\tau_i < T_i).$

• Estimator (IS) of H:
$$\widehat{H}_n(x) = \frac{1}{\widehat{\rho}_n n_{IS}} \sum_{i=1}^{n_{IS}} \mathcal{I}(T'_i \wedge \tau'_i \leq x, T'_i < \tau'_i) L'_i$$

Estimator (For cdf F(t); estimators of quantile and CTE deduced)

$$\widehat{F}_{conv,n}(t) = (\widehat{G}_{exp,n} \star \widehat{H}_n)(t) = 1 - \frac{1}{\widehat{p}_n \cdot n_{IS}} \sum_{i=1}^{n_{IS}} \mathcal{I}(T'_i < \tau'_i) L'_i e^{-(t - (T'_i \wedge \tau'_i))^+ / \widehat{\eta}_n}.$$

Outline

1 Model: regenerative process

2 Estimators

- Exponential Approximation Estimator
- Convolution Estimator

On the second second

4 Analysis on a toy example

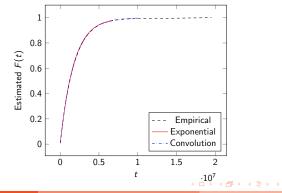
Conclusions

• • • • • • • • • • • •

Numerical example

- Highly reliable Markovian system with three component types
- five components of each type
- 15 repairmen
- system up whenever at least two components of each type work
- Each component has failure rate ϵ and repair rate 1.

With $\epsilon = 10^{-2}$



Numerical results

 ε	q	Empirical 95% CI	CPU	Exp. Est.	Exp. 95% CI	CPU
 0.01	0.1	(1.701e+05, 1.971e+05)	890	1.830e+05	(1.764e+05, 1.896e+05)	0.3
0.01	0.5	(1.206e+06, 1.271e+06)	890	1.204e+06	(1.161e+06, 1.247e+06)	0.3
0.01	0.9	(3.958e+06, 4.135e+06)	890	4.000e+06	(3.856e+06, 4.143e+06)	0.3
10^{-4}	0.1	N/A	N/A	1.757e+13	(1.756e+13, 1.758e+13)	0.3 sec
10-4	0.5	N/A	N/A	1.155e+14	(1.154e+14, 1.157e+14)	0.3 sec
10 ⁻⁴	0.9	N/A	N/A	3.840e+14	(3.838e+14, 3.842e+14)	0.3 sec

• Quantile estimators (CPU in sec.)

• CTE estimators (CPU in sec.)

ε	q	Empirical 95% Cl	CPÚ	Exp. Est.	Exp. 95% CI	CPU	Convol. Est.	CPU
0.01	0.1	(1.701e+05, 1.971e+05)	890	1.830e+05	(1.764e+05, 1.896e+05)	0.3	1.865e+05	0.4
0.01	0.5	(1.206e+06, 1.271e+06)	890	1.204e+06	(1.161e+06, 1.247e+06)	0.3	1.227e+06	0.4
0.01	0.9	(3.958e+06, 4.135e+06)	890	4.000e+06	(3.856e+06, 4.143e+06)	0.3	4.075e+06	0.4
10^{-4}	0.1	N/A	N/A	1.757e+13	(1.756e+13, 1.758e+13)	0.3	1.762e+13	0.4
10^{-4}	0.5	N/A	N/A	1.155e+14	(1.154e+14, 1.157e+14)	0.3	1.159e+14	0.4
10^{-4}	0.9	N/A	N/A	3.840e+14	(3.838e+14, 3.842e+14)	0.3	3.850e+14	0.4

Very efficient

• But biased.... for small ϵ , it does not seem a problem in practice.

Image: A math a math

Outline

1 Model: regenerative process

2 Estimators

- Exponential Approximation Estimator
- Convolution Estimator

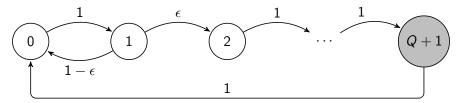
3 Numerical efficiency

Analysis on a toy example

Conclusions

• • • • • • • • • • • •

Simple (Q + 2)-states example (discrete time)



- With $Q = \lfloor \epsilon^{-w} \rfloor$ $(w \ge 0)$ increasing as $\epsilon \to 0$
- Mixes both asymptotic regimes; what about the convergence to an exponential?
- Only two possibilities for paths of $T \wedge \tau$.
- We have $p = \mathbb{P}(T < \tau) = \epsilon$ and

$$T=2M+Q+1,$$

with S = 2M and V = Q + 1 with M number of cycles between 0 and 1,

$$\mathbb{P}(M=m) = (1-\epsilon)^m \epsilon \quad (m=0,1,\dots)$$

On the weak convergence of $S_{\epsilon}/\mathbb{E}_{\epsilon}[S_{\epsilon}]$ to an exponential

M geometric with starting value 0; $\mathbb{E}[S_{\epsilon}] = 2(1-\epsilon)/\epsilon$

 $S_{\epsilon}/\mathbb{E}_{\epsilon}[S_{\epsilon}]$ converges to an exponential whatever $w \geq 0$.

$$P\left(\frac{S}{\mathbb{E}[S]} \le y\right) = P\left(M \le \frac{y(1-\epsilon)}{\epsilon}\right)$$
$$= P\left(M+1 \le \frac{y(1-\epsilon)}{\epsilon}+1\right)$$
$$= 1-P\left(M+1 > \frac{y(1-\epsilon)}{\epsilon}+1\right)$$
$$= 1-(1-\epsilon)^{\left\lceil\frac{y(1-\epsilon)}{\epsilon}\right\rceil+1}$$
$$= 1-e^{\log(1-\epsilon)\left(\left\lceil\frac{y(1-\epsilon)}{\epsilon}\right\rceil+1\right)}$$
$$= 1-e^{-(\epsilon+o(\epsilon))\left(\left\lceil\frac{y(1-\epsilon)}{\epsilon}\right\rceil+1\right)}$$
$$= 1-e^{-y+o(1)}.$$

イロト イポト イヨト イヨト

On the weak convergence of $T_{\epsilon}/\mathbb{E}_{\epsilon}[T_{\epsilon}]$ to an exponential

Recall that T = 2M + Q + 1 and $\mathbb{E}[T] = 2(1 - \epsilon)/\epsilon + 1/\epsilon^{w} + 1$.

$$P\left(\frac{T}{\mathbb{E}[T]} \le x\right) = P\left(M+1 \le x\left(\frac{(1-\epsilon)}{\epsilon} + \frac{Q+1}{2}\right) - \frac{Q+1}{2} + 1\right)$$
$$= 1 - (1-\epsilon)^{\lceil x\left(\frac{(1-\epsilon)}{\epsilon} + \frac{Q+1}{2}\right) - \frac{Q+1}{2} + 1\rceil}$$
$$= 1 - e^{\log(1-\epsilon)\lceil x\left(\frac{(1-\epsilon)}{\epsilon} + \frac{\epsilon^{-w}+1}{2}\right) - \frac{\epsilon^{-w}+1}{2} + 1\rceil}.$$

 $T_{\epsilon}/\mathbb{E}_{\epsilon}[T_{\epsilon}]$ converges to an exponential if and only is $0 \leq w < 1$.

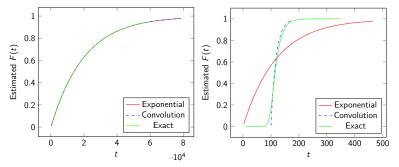
Basically, the exponential approximation to T = S + V is valid when V is negligible compared to S.

イロト イポト イヨト イヨト 二日

Plots of CDFs

IS: replacing transition probability ϵ from state 0 to state 1 by $\alpha = 0.5$.

With $\epsilon = 0.0001$ and w = 0.5 (left) and $\epsilon = 0.1$ and w = 2 (right)



The convolution estimator is the only one always matching the true distribution.

Outline

1 Model: regenerative process

2 Estimators

- Exponential Approximation Estimator
- Convolution Estimator

3 Numerical efficiency

4 Analysis on a toy example

5 Conclusions

イロト イポト イヨト イヨ

Conclusions

- Rare-event estimation often focuses on determining a mean
- But other measures are of interest: quantiles, conditional tail expectation, etc.
- For rare events, not so many existing techniques
- We have described two estimators making use of an exponential limit for regenerative systems.
- The convolution estimator is
 - more robust to asymptotic settings
 - and expected to be less biased.

Thank you!