On Initial Transient Detection in the Setting of Steady-State Simulation

Peter W. Glynn

Stanford University

Efficient Monte Carlo, Sandbjerg , Denmark, July 18, 2008
A Conference in Honor of Reuven Rubinstein

Based on joint work with Hernan Awad and Jose Blanchet

Last talk of meeting

↓

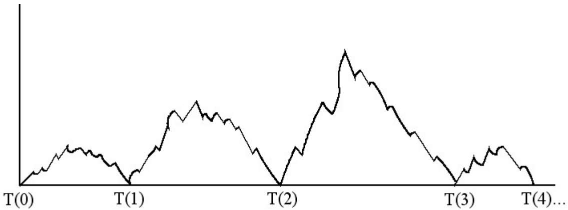
Reuven's early career

↓

"Simulation and the Monte Carlo Method"
by Reuven Rubinstein, John Wiley (1981)

↓

Early and careful treatment of
regenerative steady-state simulation



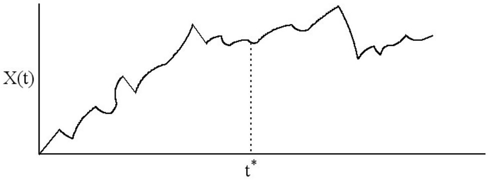
Key Insight:
On regenerative time-scale, iid cycles

Apparently solves the problem of the "initial transient".

What is the Initial Transient Problem?

- $X = (X(t) : t \geq 0)$ S-valued Markov process
- $X(t) \Rightarrow X(\infty)$ as $t \to \infty$
- Because $X(0)$ has a distribution atypical of equilibrium behavior,
$$P(X(t) \in \cdot) \neq P(X(\infty) \in \cdot).$$
- How large must t be in order that
$$P(X(t) \in \cdot) \approx P(X(\infty) \in \cdot)?$$

Three different methodologies



- Perfect Simulation:
  Sample $X(0)$ from equilibrium distribution
- Initial transient detection:
  Find t^* so that

$$P((X(t^* + u) : u \geq 0) \in \cdot) \approx P((X^*(u) : u \geq 0) \in \cdot)$$

(continued)

- Low-bias estimators: Modify estimator $\widehat{\alpha}(t)$ of $\mathbb{E}f(X(\infty))$ so that

$$\mathbb{E}\widehat{\alpha}(t) \approx \mathbb{E}f(X(\infty))$$

i.e.

$$|\mathbb{E}\widehat{\alpha}_2(t) - \mathbb{E}f(X(\infty))| \ll |\mathbb{E}\widehat{\alpha}_1(t) - \mathbb{E}f(X(\infty))|$$

This is the approach that was introduced in the regenerative setting in the late 70s/early 80s.

Two different user communities

- Discrete-event simulation / performance engineering community
  - User input: Transition kernel of Markov chain
  - Equilibrium distribution unknown
- Markov chain Monte Carlo community:
  - User input: Equilibrium distribution of Markov chain (known up to a normalization constant)
  - Have freedom to simulate any chain with prescribed equilibrium distribution.
  - Can enforce reversibility (if one wishes)

Time to Equilibrium

Let

$$|||B|||_w \triangleq \sup_x \int_S \frac{|B(x,dy)|w(y)}{w(x)}$$

Set

$$\begin{aligned} P^t &= (P_x(X(t) \in dy) : x, y \in S) \\ \Pi &= (P_x(X(\infty) \in dy) : x, y \in S) \end{aligned}$$

Then,

$$|||P^{t+s} - \Pi|||_w \leq |||P^t - \Pi\,|||_w\;|||P^s - \Pi|||_w$$

so

$$\left(\log |||P^t - \Pi|||_w : t \geq 0\right)$$

is subadditive.

(Continued)

Hence, if $|||P^t - \Pi|||_w \to 0$,

$$\frac{1}{t}\log |||P^t - \Pi|||_w \to -\lambda \leq 0$$

as $t \to \infty$.
Roughly speaking,

$$|||P^t - \Pi|||_w \approx \exp(-\lambda t)$$

$$\frac{1}{\lambda} \approx \text{"time to equilibrium"}$$

De-correlation Time

$$\frac{1}{t} \sup_{|f|=w} \log |\mathrm{cov}(f(X^*(0)), f(X^*(t)))| \to -\gamma$$

as $t \to \infty$, suggesting that

$$|\mathrm{cov}(f(X^*(0)), f(X^*(t)))| = O(e^{-\gamma t})$$

as $t \to \infty$.

$$\frac{1}{\gamma} = \text{“de-correlation time”}$$

Note that

$$\overline{\lim_{n\to\infty}} \frac{\text{var}\left(\sum_{i=1}^{n} f(X^*(ih))\right)}{n \cdot \text{var} f(X^*(0))} = O\left(\frac{1}{\gamma}\right)$$

so γ is the rate at which "fresh independent samples" are produced.

Proposition:

$$\gamma = \lambda$$

Upper Bounds on Time to Equilibrium

- In MCMC setting, theoretical bounds are sometimes possible (e.g. Cheeger, etc).
- Much harder in performance engineering setting
- Can be deceptive:
  For queue fed by fBM,

$$|||P^t - \Pi||| = O\left(\exp(-\lambda t^{\alpha})\right), \quad 0 < \alpha < 1$$

$$\mathrm{cov}\left(W^*(0), W^*(t)\right) \approx t^{-p}, \quad 0 < p < 1$$

(conjecture)

Our Approach

- We observe $(X(s) : 0 \leq s \leq t)$
- Construct a rv T_t from $(X(s) : 0 \leq s \leq t)$
- The goal is that

$$(X(T_t + u) : u \geq 0) \approx (X^*(u) : u \geq 0).$$

Note: One possibility is $T_t = \log t$.

- Would like

$$(T_t : t \geq 0)$$

to be stochastically bounded.

- We do not demand that $(T_t : t \geq 0)$ work uniformly well over all problem instances.
- Even the CLT does not work uniformly well over all P's under which the X_i's are iid with finite variance.
- For any given P, we minimally would like:

$$\text{dist}\,(P((X(T_t + u) : u \geq 0) \in \cdot), P((X^*(u) : u \geq 0) \in \cdot)) \to 0$$

and

$$(T_t : t \geq 0)$$

to be stochastically bounded.

Proposal 1

- Glynn and Iglehart (1987)
- Analyzed by Awad and Glynn (2007, 2008)

Sample Z from $\pi_t(\cdot) = \frac{1}{t}\int_0^t I(X(s) \in \cdot)\, ds$,
$T_t = \inf\{0 \le s \le t : X(s) = Z\}$.

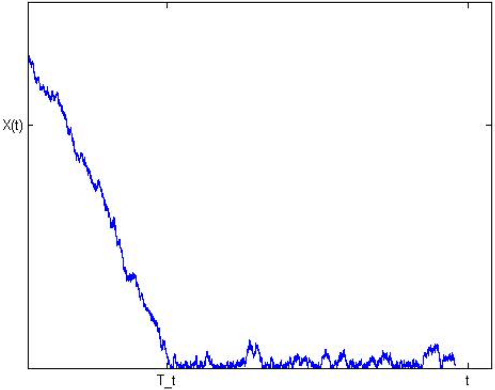


Figure: Single-server queue with unusual initialization

Theorem (Awad and Glynn)

$$\mathbb{E}\frac{1}{t - T_t}\int_0^{t-T_t} I(X(s) \in \cdot)\, ds = P(X^*(0) \in \cdot) + O\left(\frac{1}{t^2}\right)$$

Note: Compare with

$$\mathbb{E}\frac{1}{t}\int_0^{t} I(X(s) \in \cdot)\, ds = P(X^*(0) \in \cdot) + \frac{b}{t} + o\left(\frac{1}{t}\right)$$

Example: Steady-state mean of M/M/1 model

x_0	t	# of trials	avg over $[0, t]$	avg over $[T_t, t]$	$\mathbb{E}T_t/t$
0	100	400	8.7431	9.0771	4.10%
50	100	400	20.5136	14.5742	22.59%
50	400	400	11.8578	9.2529	9.38%
50	400	100	13.1981	10.7683	8.28%
100	400	100	22.5237	11.2899	20.75%
100	600	100	16.6735	9.2193	13.52%

Table: Arrival rate $\lambda = 9$, service rate $\mu = 10$ and initial state x_0. We take time t of the order $(1-\rho)^{-2}$, where $\rho = \lambda/\mu = 0.9$ is the traffic intensity. The steady-state mean is $\rho/(1-\rho) = 9$.

x_0	t	# of trials	avg over $[0, t]$	avg over $[T_t, t]$	$\mathbb{E}T_t/t$
0	400	100	17.4229	19.1983	8.41%
0	800	100	17.4416	18.0553	6.15%
50	800	100	21.6572	19.7996	11.04%
50	1200	100	20.7071	19.7541	7.55%
100	800	100	29.5073	20.6968	15.84%
100	1200	100	27.5538	19.5757	12.66%

Table: Arrival rate $\lambda = 9.5$, service rate $\mu = 10$, $\rho = 0.95$ We take time t of the order $(1-\rho)^{-2} = 400$. The steady-state mean is $\rho/(1-\rho) = 19$.

x_0	t	# of trials	avg over $[0, t]$	avg over $[T_t, t]$	$\mathbb{E}T_t/t$
0	1111	100	27.9184	30.8362	8.52%
0	2222	100	30.6561	31.7969	10.26%
50	1111	100	32.5568	32.0916	11.99%
50	2222	100	34.0016	33.7515	6.74%
100	1111	100	44.1968	35.8125	14.81%
100	2222	100	36.8861	32.5032	8.00%

Table: Arrival rate $\lambda = 9.7$, service rate $\mu = 10$, $\rho = 0.97$. We take time t of the order $(1-\rho)^{-2} = 1111$. The steady-state mean is $\rho/(1-\rho) = 32.33$.

Generalizing to Continuous State Space

In continuous state space, T_t will often be uniform on $[0, t]$. So, partition the state space into a finite or countably infinite number of subsets, and apply the method to the "discretized process".

Philosophy: We are attempting to build a methodology that reliably mitigates most of the impact of the initial transient for those models where the problem is potentially serious. We want to know whether the initial transient has contaminated 1% or 50% of the simulation.

Pragmatically, we can tolerate some approximation error here.

Proposal 2 (joint work with Jose Blanchet)

Let $X = (X_n : n \geq 0)$ be an irreducible finite state Markov chain. How might we estimate the rate of convergence to equilibrium for X, based on observing $X_0, \ldots, X_t$?

- Non-parametric Maximum Likelihood
  - Compute the second eigenvalue of empirical transition matrix $\widehat{P}_t$.
  - This converges a.s. to true second eigenvalue.
  - How might one apply such ideas in the setting of a more general Markov process?

Resampling:

- Assume $X = X(t) : t \geq 0$ regenerative with regeneration times $0 = T(0) < T(1) < \cdots$
- Simulate $(X(s) : 0 \leq s \leq t)$.

$a(t) \triangleq \mathbb{E}f(X(t))$ satisfies

$$\begin{aligned} a(t) &= b(t) + (F * a)(t) \\ b(t) &= \mathbb{E}f(X(t))I(\tau_1 > t) \quad (\tau_1 = T(1) - T(0)) \\ F(dt) &= P(\tau_1 \in dt) \end{aligned}$$

Compute

$$\widehat{b}_t(\cdot) = \frac{1}{N(t)} \sum_{i=1}^{N(t)} f(X(T(i-1)+\cdot))I(\tau_i > \cdot)$$

$$\widehat{F}_t(\cdot) = \frac{1}{N(t)} \sum_{i=1}^{N(t)} I(\tau_i \leq \cdot)$$

$$\widehat{a}_t(s) \xrightarrow{\text{a.s.}} \widehat{a}_t(\infty) = \frac{\sum_{i=1}^{N(t)} \int_{T(i-1)}^{T(i)} f(X(s))\,ds}{\sum_{i=1}^{N(t)} \tau_i}$$

as $s \longrightarrow \infty$.

Given $\epsilon > 0$, choose T_t so that

$$|\widehat{a}_t(u) - \widehat{a}_t(\infty)| < \epsilon$$

for $u \geq T_t$.

How to compute T_t?

- Fourier / Laplace methods:

$$\int_{[0,\infty)} e^{su}\widehat{a}_t(u)\,du = \frac{\int_{[0,\infty)} e^{su}\widehat{b}_t(u)\,du}{1 - \int_{[0,\infty)} e^{su}\widehat{F}(du)}$$

- Simulate the cycle-quantities (not the full chain) to compute T_t.

Uniform Convergence of Empirical Estimator

Theorem Assume that τ_1 is discrete, aperiodic, and has finite p'th moment, $p > 1$. If f is bounded, then

$$\sup_{s \in (0,\infty)} |\widehat{a}_t(s) - a(s)| \xrightarrow{\text{a.s.}} 0$$

as $t \to \infty$.

Corollary $T_t \to t(\epsilon)$ as $t \to \infty$, where $t(\epsilon)$ is such that

$$|a(u) - a(\infty)| < \epsilon$$

for $u \geq t(\epsilon)$.

But more can be said...

$$t^{1/2}\left(\widehat{a}_t(\cdot) - a(\cdot)\right) \Rightarrow Z(\cdot)$$

where $Z = (Z(s) : s \geq 0)$ is a Gaussian process. This suggests that

$$t^{1/2}\left(\widehat{a}_t(s) - \widehat{a}(\infty)\right)$$

is of order $t^{-1/2}$. However:

The error $\widehat{a}_t(s) - \widehat{a}_t(\infty)$ tracks the true error $a(s) - a(\infty)$ much more closely than this analysis suggests.

Theorem Suppose that τ_1 is discrete, aperiodic, and satisfies $\mathbb{E}\exp(\eta^*\tau_1) < \infty$ for some $\eta > 0$. If f is bounded, then there exists $c > 0$ such that

$$\sup_{s\in(0,\infty)} \exp(cs)\left|(\widehat{a}_t(s) - \widehat{a}_t(\infty)) - (a(s) - a(\infty))\right| \xrightarrow{a.s.} 0$$

as $t \to \infty$.

Proof: Apply a uniform version of renewal theorem path-by-path.

In view of the previous result, one can reliably compute $t(\epsilon)$ for quite small values of epsilon given a sample of size t:

If $(\log(1/\epsilon))^2 = o(t)$, then $T_t - t(\epsilon)$ converges in probability to 0. As in previous methodology, one can consider applying this to "approximate regenerations".

Conclusions

- The use of sampling-based methods for determining the duration of the initial transient seems both pragmatically useful, and can sometimes be supported at a theoretical level.
- Particularly appealing for performance engineering applications where it will often be hopeless to compute useful theoretical estimates of rates of convergence.

Thank you to Reuven for your many years of inspiring contributions to our field and with the hope of many more such years to come!