Sequential Stopping for Parallel Monte Carlo

Peter W. Glynn
Stanford University

Joint work with Jing Dong

ICTS, Bangalore, August 14, 2019

Topics in Applied Probability

- Today: Sequential Stopping for Parallel Monte Carlo
- Tomorrow: Non-stationary Markov Processes: Approximations and Numerical Methods
- Friday: Experimentation with Temporal Interference: Poisson's Equation and Adaptive Markov Chain Sampling

Outline:

1. Error Assessment for Monte Carlo
2. Sequential Stopping
3. Our New Sequential Stopping Rule

The Monte Carlo Method:

Goal: Compute $\alpha = EX$

Method: Generate iid copies $X_1, \ldots, X_n$ of X and form

$$\alpha_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Consistency (via the LLN):

$$\alpha_n \to \alpha \text{ a.s.}$$

In the presence of $\sigma^2 = \text{var} X < \infty$,

$$n^{1/2}(\alpha_n - \alpha) \Rightarrow \sigma N(0,1)$$

as $n \to \infty$. So,

$$\alpha_n \overset{\mathcal{D}}{\approx} \alpha + \frac{\sigma}{\sqrt{n}} N(0,1)$$

Consequence:

- Slowly converging
- Error assessment is important
- Error is asymptotically normally distributed

Fixed Sample Size Procedure:

- Choose "sample size" n large
- Estimate σ^2 via
$$s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \alpha_n)^2$$
- For $\delta > 0$, choose z so that
$$P(-z \leq N(0,1) \leq z) = 1 - \delta$$
- Then
$$I_n \triangleq \left[\alpha_n - \frac{z s_n}{\sqrt{n}}, \alpha_n + \frac{z s_n}{\sqrt{n}}\right]$$
is an approximate $100(1-\delta)$% confidence interval for α

Confidence interval asymptotically covers $100(1-\delta)$% of the time:

$$P(\alpha \in I_n) \to 1 - \delta$$

as $n \to \infty$

Remark:

- For finite and fixed n, coverage can be arbitrarily bad
- Algorithms focused on uniform error bounds are not used in practice (because they would be based on boundedness assumptions that cannot be validated from data alone)
- Above algorithm is asymptotically valid “instance-by-instance” (fundamental difference between statistics algorithmic perspective and CS algorithmic perspective)

Problem: No control over accuracy of the computation

$$\text{Half-width } \approx \frac{\sigma z}{\sqrt{n}}$$

What is typically preferred:

Run computation until accuracy ϵ is achieved

The Easy Fix: Two-Stage Procedures

- First generate n_0 "trial runs"
- Estimate σ^2 via sample variance s^2 computed from trial runs
- Compute
$$n \approx \sigma^2 z^2 / \epsilon^2$$
- Run n "production runs" and compute interval I_n (as earlier)

Problems:

- To achieve accuracy asymptotically, we need to let $n_0 = n_0(\epsilon) \nearrow \infty$ as $\epsilon \downarrow 0$
- We need a consistent estimator s^2 for σ^2
- Then,

  $$|I_\epsilon| \sim 2\epsilon$$

  as $\epsilon \downarrow 0$, and $P(\alpha \in I_\epsilon) \to 1 - \delta$ as $\epsilon \downarrow 0$

A Fully Sequential Alternative (due to Chow and Robbins)

- Compute "standard interval" I_n until

$$|I_n| \leq 2\epsilon$$

- Provided that one avoids "early termination", one has

$$P(\alpha \in [\alpha_n - \epsilon, \alpha_n + \epsilon]) \to 1 - \delta$$

as $\epsilon \downarrow 0$

- Early termination avoided by inflating s_n^2 a bit:

$$s_n^2 \leftarrow s_n^2 + a_n \quad (a_n \downarrow 0)$$

Obstacle: There are many Monte Carlo problems for which estimation is cumbersome or challenging

Example 1: Smooth Function of Expectations

- Goal: Compute $\alpha = g(EX)$, where $g : \mathbb{R}^d \to \mathbb{R}$
- Method: $\alpha_n = g(\overline{X}_n)$
- CLT:

$$n^{1/2}(\alpha_n - \alpha) \Rightarrow \sigma N(0,1),$$

$$\text{where } \sigma^2 = \text{var}(\nabla g(\alpha) X)$$

- Need $\nabla g(\alpha)$

Example 2: Quantiles

- Goal: Compute q such that $P(X \leq q) = p$
- Estimator: Sample quantile Q_n defined via

$$\frac{1}{n} \sum_{i=1}^{n} I(X_i \leq Q_n) \approx p$$

- CLT:

$$n^{1/2}(Q_n - q) \Rightarrow \sigma N(0,1),$$

$$\text{where } \sigma^2 = p(1-p)/f_X(q)$$

Example 3: Equilibrium Calculations

- Goal: Compute $\alpha = Ef(X_\infty)$, where X_∞ has equilibrium distribution of Markov chain $X = (X_n : n \geq 0)$
- Estimator:
$$\alpha_n = \frac{1}{n}\sum_{i=0}^{n-1} f(X_i)$$
- CLT:
$$n^{1/2}(\alpha_n - \alpha) \Rightarrow \sigma N(0,1),$$
where $\sigma^2 = \mathrm{var} f(X_0^*) + 2\sum_{i=1}^{\infty} \mathrm{cov}(f(X_0^*), f(X_i^*))$

Example 4: Optimization

Goal: Compute $\alpha = \max_{\theta \in \Lambda} EX(\theta)$

Estimator: $\max_{\theta \in \Lambda} \alpha_n(\theta)$

When variance estimation is difficult, the *method of replication* is a useful alternative for constructing confidence intervals when a *fixed sample size approach* is used

- Assume that
$$\alpha_n \overset{\mathcal{D}}{\approx} \alpha + \eta N(0,1)$$
- Replicate α_n independently on m different processors
$$\alpha_n^i \approx \alpha + \eta N(0,1), \quad 1 \le i \le m$$
- Set
$$\overline{\alpha}_n = \frac{1}{m}\sum_{i=1}^{m} \alpha_n^i$$
$$s_n^2 = \frac{1}{m-1}\sum_{i=1}^{m} (\alpha_n^i - \overline{\alpha}_n)^2$$

- Then,
$$m^{1/2}\left(\frac{\overline{\alpha}_n - \alpha}{s_n}\right) \Rightarrow t_{m-1},$$
where t_{m-1} is a Student-t rv with $m-1$ degrees of freedom

- So,
$$P(\alpha \in I_n) \to 1 - \delta$$
as $n \to \infty$, if
$$I_n = \left[\overline{\alpha}_n - \frac{z_{m-1} s_n}{\sqrt{m}}, \overline{\alpha}_n + \frac{z_{m-1} s_n}{\sqrt{m}}\right],$$
where z_{m-1} is chosen so that $P(-z_{m-1} \leq t_{m-1} \leq z_{m-1}) = 1 - \delta$

Method of replication:

- ideal for parallel implementation (linear speed-up!)
- can also be implemented in conventional computing environment

Does method of replication extend to the sequential setting?

If we run fixed sample size procedure to time $N(\epsilon)$ for which

$$|I_{N(\epsilon)}| \leq 2\epsilon,$$

is it the case that

$$P(\alpha \in [\overline{\alpha}_n - \epsilon, \overline{\alpha}_n + \epsilon]) \to 1 - \delta$$

as $\epsilon \downarrow 0$?

Answer: No!

Fundamental Problem:

- When variance is consistently estimated, $N(\epsilon) \approx \sigma^2 z^2/\epsilon^2$ when ϵ is small, so looks asymptotically deterministic
- When variance is not consistently estimated,

$$\epsilon^2 N(\epsilon) \Rightarrow W,$$

  where W is random
- So, joint distribution between $N(\epsilon)$ and $(\overline{\alpha}_n : n \geq 0)$ must be taken into account

$$\overline{\alpha}_{N(\epsilon)} \overset{\mathcal{D}}{\approx} \alpha + \frac{\sigma B(N(\epsilon))}{N(\epsilon)}$$

Our solution assumes the existence of an estimator $\alpha(t)$ satisfying a strong approximation principle:

> There exists a probability space supporting $(\alpha(t) : t \geq 0)$ and a standard Brownian motion for which
>
> $$\alpha(t) = \alpha + \frac{\sigma B(t)}{t} + o(t^{-1/2-\delta}) \quad \text{a.s.}$$
>
> as $t \to \infty$, for some $\delta > 0$ and $\sigma \in \mathbb{R}$.

If we run m independent replications $\alpha_1(\cdot), \ldots, \alpha_m(\cdot)$ of $\alpha(\cdot)$, then

$$\overline{\alpha}_m(t) = \alpha + \frac{\sigma \overline{B}_m(t)}{t} + o(t^{-1/2-\delta}) \quad \text{a.s.}$$

and

$$s_m(t) = \sigma \frac{1}{\sqrt{m-1}} \frac{R_m(t)}{t} + o(t^{-1/2-\delta}) \quad \text{a.s.,}$$

where

$$R_m(t) = \sqrt{\sum_{i=1}^{m} (B_i(t) - \overline{B}_m(t))^2}$$

Key Facts:

- $\overline{B}_m = (\overline{B}_m(t) : t \geq 0)$ and $R_m = (R_m(t) : t \geq 0)$ are independent as processes

- $\sqrt{m}\overline{B}_m \stackrel{\mathcal{D}}{=} B$

- $R_m \stackrel{\mathcal{D}}{=} Z_{m-1}$, where $Z_d = (Z_d(t) : t \geq 0)$ is a Bessel process of dimension d starting from the origin, so that

$$(Z_d(t) : t \geq 0) \stackrel{\mathcal{D}}{=} \left(\sqrt{\sum_{i=1}^{d} B_i^2(t)} : t \geq 0 \right)$$

- For $r > 0$, let

$$\kappa_m(\epsilon, r) = \inf\left\{t : \frac{s_m(t)}{\sqrt{m}} < \frac{\epsilon}{r}\right\}$$

- Using the fact that $(tB(1/t) : t \geq 0) \stackrel{\mathcal{D}}{=} (B(t) : t \geq 0)$, we can show that

$$\epsilon^2 \kappa_m(\epsilon, r) \Rightarrow \frac{r^2\sigma^2}{m(m-1)}\chi^2_{m-3}$$

where χ^2_{m-3} is a chi-square rv with $m-3$ degrees of freedom

Key to calculation: Reducing computation to level crossing calculation for the Bessel process

- The pivotal quantity

$$\frac{\sqrt{m}(\overline{\alpha}_m(\kappa_m(\epsilon,r))-\alpha)}{s_m(\kappa_m(\epsilon,r))} \overset{\mathcal{D}}{\approx} \frac{r}{\epsilon}\frac{\sigma\overline{B}_m\left(\frac{r^2\sigma^2}{m(m-1)}\chi^2_{m-3}\right)}{\frac{r^2\sigma^2}{m(m-1)}\chi^2_{m-3}} \overset{\mathcal{D}}{=} \sqrt{\frac{m-1}{m-3}}t_{m-3}$$

- If we choose $r=\sqrt{(m-1)/(m-3)}z_{m-3}$, we conclude that

$$P(\alpha\in[\overline{\alpha}_m(\kappa_m(\epsilon,r))-\epsilon,\overline{\alpha}_m(\kappa_m(\epsilon,r))+\epsilon])\to 1-\delta$$

as $\epsilon\downarrow 0$

- This is our desired sequential procedure!

- Easily implemented in parallel setting
- Can also be implemented easily in standard sequential computing environment
- Note that the procedure involves t_{m-3}, so $m \geq 4$ is required
- For $m \leq 3$, procedure terminates early because Z_d's recurrence in $d = 1, 2$ causes early termination (in scale of $o(1/\epsilon^2)$)

Conclusions:

- We have surveyed error assessment and sequential stopping in the Monte Carlo setting
- We have provided a new sequential procedure for settings in which consistent estimation of the variance is difficult
- We are studying settings in which $\alpha(t)$ cannot be approximated by $\alpha + \sigma B(t)/t \ldots$ future work

Thank you!