Simulation Methodology: An Overview

Peter W. Glynn
Stanford University

Theory of RL Bootcamp, Simons Institute
September 4, 2020

Outline:

I. Efficiency Improvement Techniques

- Suppose that we have two different simulation algorithms for computing α:

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

and

$$\beta_n \stackrel{a.s.}{\to} \alpha$$

- We want to use the algorithm that is computationally more efficient
- Suppose

$$n^{1/2}(\alpha_n - \alpha) \Rightarrow \sigma_1 N(0,1)$$

and

$$n^{1/2}(\beta_n - \alpha) \Rightarrow \sigma_2 N(0,1)$$

- Then:

$$\alpha_n \overset{D}{\approx} N(\alpha, \sigma_1^2/n)$$
$$\beta_n \overset{D}{\approx} N(\alpha, \sigma_2^2/n)$$

- Choose α_n over β_n if $\sigma_1^2 \leq \sigma_2^2$
- Constructing estimators with such a smaller variance is called a *variance reduction technique*

- But each iteration of α_n may be more costly than an iteration of β_n:

$$T_1(n) = \text{total computer time expended to compute } \alpha_n$$

$$T_2(n) = \text{total computer time expended to compute } \beta_n$$

- Then, the estimators available after c units of computer time have been expended are

$$\alpha(c) = \alpha_{N_1(c)}, \quad \beta(c) = \beta_{N_2(c)},$$

where

$$N_i(c) = \max\{n : T_i(n) \leq c\}$$

- If $N_i(c)/c \to \lambda_i$ as $c \to \infty$, then (typically)

$$c^{1/2}(\alpha(c) - \alpha) \Rightarrow \lambda_1^{-1/2}\sigma_1 N(0,1)$$

  and

$$c^{1/2}(\beta(c) - \alpha) \Rightarrow \lambda_2^{-1/2}\sigma_2 N(0,1)$$

- Choose $\alpha(c)$ over $\beta(c)$ if $\lambda_1^{-1}\sigma_1^2 \leq \lambda_2^{-1}\sigma_2^2$
- Constructing estimators with such a smaller *work-normalized variance* is called an *efficiency improvement technique*

A Philosophical Distinction

Statistics and simulation/*Monte Carlo* may seem very clearly related

BUT

In statistics, one is sampling because one does not know P

In simulation/*Monte Carlo*, one samples as a computational vehicle for computing

$$\int_\Omega W(\omega)P(d\omega)\,(= E[W])$$

One knows the associated P, at least implicitly

We can hope to use available problem structure to obtain efficiency improvements

II. Control Variates

Goal: Compute $\alpha = E[W]$

Given: A rv Z with known expectation

- Put $C = Z - E[Z]$ and $W(\lambda) = W - \lambda C$
- Then, $E[W(\lambda)] = \alpha$ for all $\lambda \in \mathbb{R}$
- $\text{Var}(W(\lambda)) = \text{Var}(W) - 2\lambda \text{Cov}(W, C) + \lambda^2 \text{Var}(C)$
- Minimizing λ:

$$\lambda^* = \text{Cov}(W, C)/\text{Var}(C)$$

- Minimum variance:

$$\text{Var}(W(\lambda^*)) = \text{Var}(W) \cdot (1 - \rho^2)$$

  $\rho =$ coefficient of correlation between W and C

- $$\widehat{\lambda}_n = \widehat{\text{Cov}}(W, C)/\widehat{\text{Var}}(C)$$

- No asymptotic loss of efficiency

Markov Chains and Martingale Controls

Goal: Compute $\alpha = E_x \left[\sum_{j=0}^{\infty} e^{-\alpha j} r(X_j) \right] \left(\triangleq u^*(x) \right)$

- It is known that u^* satisfies

$$u = r + e^{-\alpha} P u$$

- Also,

$$M_n = \sum_{j=0}^{n-1} e^{-\alpha j} r(X_j) + e^{-\alpha n} u^*(X_n)$$

  is a martingale adapted to $(X_n : n \geq 0)$, i.e.,

$$E \left[M_{n+1} \mid X_0, \ldots, X_n \right] \overset{a.s.}{=} M_n$$

- So, $C_n = M_n - M_0$ has mean zero

- Put $\lambda = 1$. Then,

$$W - \lambda C_\infty = u^*(x)$$

  So,

$$\text{Var}(W(\lambda)) = 0$$

- We don't know u^* ... but if $\widetilde{u}$ is a good approximation to u^*, use

$$\widetilde{M}_n = \sum_{j=0}^{n-1} e^{-\alpha j}\widetilde{r}(X_j) + e^{-\alpha n}\widetilde{u}(X_n),$$

  where

$$\widetilde{r} \triangleq \widetilde{u} - e^{-\alpha} P\widetilde{u}$$

III. Common Random Numbers

Suppose we have two policies we wish to compare:

$$\kappa_1 = E[W_1] \quad \text{vs} \quad \kappa_2 = E[W_2]$$

Goal: Compute $\alpha = \kappa_1 - \kappa_2$

- EIT 1: Estimate α via

$$\widehat{\alpha} = \overline{W}_1(n_1) - \overline{W}_2(n_2)$$

"stratified sampling"

$$n_i \propto \lambda_i^{-1/2}\sigma_i, \quad i = 1, 2$$

- EIT 2: "Couple" W_1 and W_2 with a well-chosen joint distribution (not independent)

$$
\begin{aligned}
W &= W_1 - W_2 \\
\mathsf{Var}(W) &= \mathsf{Var}(W_1) - 2\mathsf{Cov}(W_1, W_2) + \mathsf{Var}(W_2)
\end{aligned}
$$

- Want $\mathsf{Cov}(W_1, W_2)$ to be as large as possible

Suppose

$$W_1 = \widetilde{f}_1(\xi_1, \ldots, \xi_d)$$
$$W_2 = \widetilde{f}_2(\xi_1, \ldots, \xi_d)$$

Guaranteed efficiency improvement if $\widetilde{f}_i \nearrow$, $i = 1, 2$

“common random numbers”

IV. Importance Sampling

Goal: Compute $\alpha = E[W] = E_P[W]$

- Note that

$$
\begin{aligned}
E_P[W] = \int_\Omega W(\omega)P(d\omega) &= \int_\Omega W(\omega)\frac{P(d\omega)}{Q(d\omega)}Q(d\omega) \\
&\triangleq \int_\Omega W(\omega)L(\omega)Q(d\omega) \\
&= E_Q[WL]
\end{aligned}
$$

- Put $Q^*(d\omega) = |W(\omega)|P(d\omega)/E_P[|W|]$
- If $W \geq 0$, $WL^* = \alpha$
- Of course, we do not know Q^*. Instead, we hope to use a $\widetilde{Q}$ that approximates Q^*

For example, $\alpha = E_P[r(X_n)]$

- Then,

  $$\alpha = E_Q[r(X_n)L_n]$$

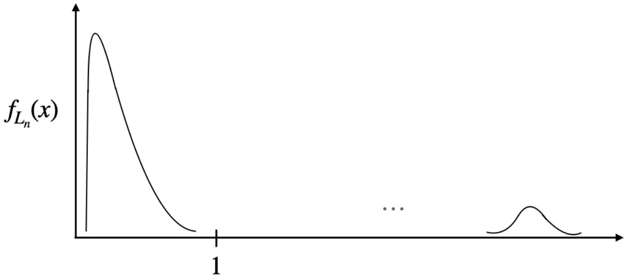
  where

  $$L_n = \prod_{i=0}^{n-1} \frac{P(X_i, X_{i+1})}{Q(X_i, X_{i+1})}$$

- $\mathrm{Var}_Q(L_n) \sim a\beta^n$, $\beta > 1$
- On the other hand,

  $$\frac{1}{n} \log L_n \to \sum_{x,y} \log\left(\frac{P(x,y)}{Q(x,y)}\right) Q(x,y)\pi_Q(x) < 0$$

  so $L_n \to 0$, Q a.s.



- $\widehat{\text{Var}}(L_n)$ is highly misleading in many settings
- If
$$Q - P = O\left(\frac{1}{\sqrt{n}}\right),$$
then,
$$\text{Var}_Q(L_n) = O(1)$$

V. Gradient Estimation

- Suppose that θ is a decision variable:

$$\alpha(\theta) = \int_\Omega W(\theta, \omega) P(d\omega)$$

or

$$\alpha(\theta) = \int_\Omega W(\omega) P_\theta(d\omega)$$

- How to efficiently compute $\nabla\alpha(\theta)$?
- Why it is of interest:
  - Stochastic gradient descent algorithm
  - Statistical analysis:

$\widehat{\theta}$: statistical estimator for "true" parameter θ_0

$$\begin{aligned}\alpha(\widehat{\theta}) - \alpha(\theta_0) &\approx \nabla\alpha(\theta_0)\left(\widehat{\theta} - \theta_0\right) \\ &\overset{D}{\approx} \nabla\alpha(\theta_0) N(0, C)\end{aligned}$$

One can often move parametric dependence from $W(\theta)$ to P_θ and vice versa...

- When $W(\theta)$ depends smoothly on θ:

$$\nabla\alpha(\theta_0) = E_P\left[\nabla W(\theta_0)\right]$$

"infinitesimal perturbation analysis"

- When P_θ depends smoothly on θ:

$$\alpha(\theta) = E_{\theta_0}\left[W L(\theta)\right]$$

so

$$\nabla\alpha(\theta) = E_{\theta_0}\left[W \nabla L(\theta_0)\right]$$

where

$$L(\theta, \omega) = \frac{P_\theta(d\omega)}{P_{\theta_0}(d\omega)}$$

"likelihood ratio gradient estimation"

Application to Markov Chains

- Compute $\nabla\alpha(\theta_0)$ where $\alpha(\theta) = E_\theta\left[r(X_\infty)\right]$
- Here, $W = \frac{1}{n}\sum_{j=1}^{n} r(X_j)$
- Then,

$$\nabla\alpha(\theta_0) \approx E_{\theta_0}\left[W\nabla L_n(\theta_0)\right]$$

  where

$$\nabla L_n(\theta_0) = \sum_{j=1}^{n} \frac{\nabla p(\theta_0, X_{j-1}, X_j)}{p(\theta_0, X_{j-1}, X_j)}$$

Remark: $(\nabla L_n(\theta_0) : n \geq 1)$ is a zero-mean martingale adapted to $(X_n : n \geq 0)$

IPA versus Likelihood Ratio Gradient Estimation

IPA:

$$\frac{1}{n}\sum_{j=1}^{n} \nabla r(\theta_0, X_j) \approx \nabla \alpha(\theta_0) + \frac{1}{\sqrt{n}} N(0, C)$$

Likelihood ratio:

$$\begin{aligned}
\frac{1}{n}\sum_{j=1}^{n} r(X_j)\nabla L_n(\theta) &= \frac{1}{n}\sum_{j=1}^{n} r(X_j) \sum_{i=1}^{n} D_i \\
&= \frac{1}{n}\sum_{j=1}^{n} r_c(X_j) \sum_{i=1}^{n} D_i + E_{\theta_0}\left[r(X_\infty)\right] \sum_{i=1}^{n} D_i \\
&\qquad\qquad (r_c(x) = r(x) - E_{\theta_0}\left[r(X_\infty)\right]) \\
&\overset{D}{\approx} \nabla \alpha(\theta_0) + N_1(0, \sigma^2) N_2(0, C_2) + \sqrt{n} E_{\theta_0}\left[r(X_\infty)\right] N_2(0, C_2)
\end{aligned}$$

Since the D_j's are martingale differences,

$$E\left[r(X_j)D_i\right] = 0, \quad i > j$$

Modify estimator:

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n} D_i \sum_{j=i}^{n} r(X_j) \\ &\overset{D}{\approx} \sqrt{n} E_{\theta_0}\left[r(X_\infty)\right] \int_0^1 (1-s)dB(s) \end{aligned}$$

- If $E_{\theta_0}\left[r(X_\infty)\right] = 0$, then

$$\frac{1}{n}\sum_{i=1}^{n} D_i \sum_{j=i}^{n} r(X_j)$$

$$\overset{D}{\approx} \sigma_1 C^{1/2} \int_0^1 B_2(s) d\vec{B}_1(s)$$ Olvera-Cravioto + G (2018)

- So, work with $r_c(x) = r(x) - E_{\theta_0}\left[r(X_\infty)\right]$
- Effectively equivalent to using $\sum_{j=1}^{n} D_j$ as a control variate

Finite Difference Estimators

- Central differences:

$$\frac{\overline{W}_n(\theta_0 + h) - \overline{W}_n(\theta_0 - h)}{2h} \stackrel{D}{\approx} \alpha'(\theta_0) + \frac{h^2}{3}\alpha^{(3)}(\theta_0) + \frac{\sigma}{\sqrt{n}h}N(0,1)$$

- To balance bias and variance, put $h \approx cn^{-1/6}$
- Convergence rate: $n^{-1/3}$
- If we use common random numbers, convergence rate $\approx n^{-2/5}$

VI. Stochastic Optimization

- r policies
- Which policy maximizes reward?

  "Selection of best system"

  Connections to multi-armed bandit literature

$$\min_{\theta} \alpha(\theta)$$

- $\theta_{n+1} = \theta_n - C_n \widehat{\nabla\alpha}(\theta_n)$

  "stochastic gradient descent"
- Optimal choice of C_n depends on Hessian of $\alpha(\cdot)$, covariance structure of $\widehat{\nabla\alpha}(\theta_\infty)$
- Polyak averaging can be effective in implicitly finding C_n

- Large literature that intersects with many different applications domains
- Many areas not covered in today's lectures
  - *Stochastic Simulation: Algorithms and Analysis*, Asmussen + G (2007)
  - Winter Simulation Conference
  - ACM Transactions on Modeling and Computer Simulation