Algorithmic and Theoretical Foundations of RL

Monte Carlo Methods

Ke Wei
School of Data Science
Fudan University

A Typical Example

Monte Carlo (MC) methods refer to a collection of computational methods where random simulation is used. A typical example is the computation of π.

Example 1 (Approximation of π)

▶ Ratio of the area of circle to the area of square: $\frac{A_c}{A_S} = \frac{\pi}{4}$

▶ Simulation steps:

- Randomly generate *n* points inside the square;
- Count the number of points inside the circle: n_c;
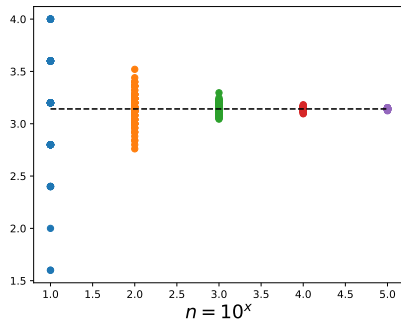- $\pi \approx 4 \cdot \frac{n_c}{n}$.
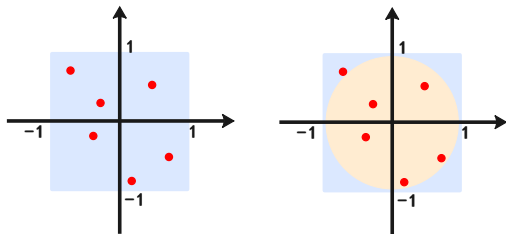
Simulation Result

$n = 10^x$

Table of Contents

Computing the expectation or integration

$$\mu = \mathbb{E}\left[f(X)\right] = \int f(x)p(x)dx$$

is a ubiquitous task in science and engineering.

► Numerical Integration: Trapezium rule, Simpson's rule, ...
► Monte Carlo Integration: Statistical methods based on **random** samples.

Simple Monte Carlo Integration

Let $X_1, \cdots, X_n \sim p(x)$ be i.i.d random samples. Then approximate μ by

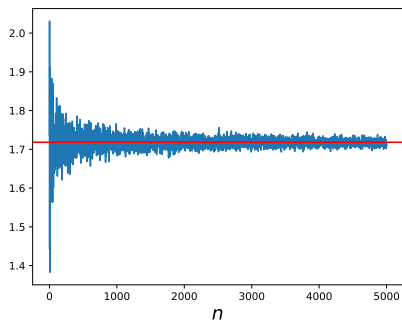$$\bar{\mu}_n = \frac{1}{n} \sum_{k=1}^{n} f(X_k).$$

▶ Un-bias and strong consistency

$$\mathbb{E}\left[\frac{1}{n} \sum_{k=1}^{n} f(X_k)\right] = \mu \quad \text{and} \quad \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} f(X_k) \xrightarrow{a.s.} \mu.$$

▶ Finite sample variance

$$\text{Var}\left[\frac{1}{n} \sum_{k=1}^{n} f(X_k)\right] = \frac{\text{Var}\left[f(X)\right]}{n}.$$

Illustrative Example

Estimate the integral $\int_0^1 e^x dx$.

Variance reduction is an important component in statistical computing.

$$数学上等价，计算上未必等效. － 冯康$$

Consider equivalent problems with more information, structure or control.

▶ Control variates

$$\mathbb{E}\left[f(X)\right] = \mathbb{E}\left[h(X)\right], \quad \text{where } h(X) = f(X) + c \cdot \left(g(X) - \mathbb{E}\left[g(X)\right]\right).$$

If $g(X)$ is highly correlated with $f(X)$, deducting $g(X)$ from $f(X)$ could reduce variance. Widely used in variance reduction of stochastic optimization.

Variance Reduction

Variance reduction is an important component in statistical computing.

$$数学上等价，计算上未必等效. - 冯康$$

Consider equivalent problems with more information, structure or control.

► Importance sampling

$$\mathbb{E}\left[f(X)\right] = \int f(x)p(x)dx = \int \frac{f(y)p(y)}{q(y)}q(y)dy = \mathbb{E}\left[\frac{f(Y)p(Y)}{q(Y)}\right].$$

If $q(Y)$ is similar to $f(Y)p(Y)$, then variance could be reduced (using quotient in contrast to control variate where addition is used).

Notice that

$$\text{Var}\left[f(X) + c \cdot (g(X) - \mathbb{E}\left[g(X)\right])\right]$$
$$= \text{Var}\left[f(X)\right] + c^2 \cdot \text{Var}\left[g(X)\right] + 2c \cdot \text{Cov}\left(f(X), g(X)\right).$$

The best c^* is given by

$$c^* = -\frac{\text{Cov}\left(f(X), g(X)\right)}{\text{Var}\left[g(X)\right]}.$$

Then variance is reduced by

$$\frac{\left[\text{Cov}\left(f(X), g(X)\right)\right]^2}{\text{Var}\left[g(X)\right]}.$$

Illustrative Example

Estimate the integral $\int_0^1 e^x dx$. Let $n = 1000$. Take $g(x) = x$ and $c = -1.5$.

```python
import numpy as np
```

```python
n = 1000
x = np.random.rand(n)
y = np.exp(x)
mu = np.mean(y)
var = np.var(y)
print('Simple MC Integration: mu = %.5f,' % mu, 'var = %.5f' % var)
```

```
  Simple MC Integration: mu = 1.70772, var = 0.24851
```

```python
c = - 1.5
y = np.exp(x)+c*(x-1/2)
mu = np.mean(y)
var = np.var(y)
print('Control variate: mu = %.5f,' % mu, 'var = %.5f' % var)
```

```
  Control variate: mu = 1.71893, var = 0.00715
```

Monte Carlo Integration via Importance Sampling

$$\underbrace{\mathbb{E}\left[f(X)\right]}_{\mu} = \mathbb{E}\left[\frac{f(Y)p(Y)}{q(Y)}\right] \approx \underbrace{\frac{1}{n}\sum_{k=1}^{n}f(Y_k)\frac{p(Y_k)}{q(Y_k)}}_{\bar{\mu}_{IS}},$$

where $Y_1, \cdots, Y_n \sim q(y)$.

▶ If a region is important for computing $\mathbb{E}\left[f(X)\right]$, more data should be obtained from this region. This can be achieved by importance sampling distribution, followed by correction of bias caused by distinction between target distribution and importance sampling distribution via $w(y) = p(y)/q(y)$.

▶ The expression for importance sampling indeed indicates that evaluation of $\mathbb{E}\left[f(X)\right]$ is not intrinsically associated with target distribution. This is particularly useful when only data from a different distribution are available, **which is often the case in RL**.

In order to reduce variance, $q(y)$ should be selected such that $f(y)p(y)/q(y)$ varies not much over the integration region.

Theorem 1

The minimum variance of $\bar{\mu}_{IS}$ under q is achieved at

$$q^*(y) = \frac{|f(y)|p(y)}{\mathbb{E}_p\left[|f(X)|\right]},$$

with the minimum variance given by

$$\mathsf{Var}_{q^*}\left[\bar{\mu}_{IS}^*\right] = \frac{1}{n}\left(\left(\mathbb{E}_p\left[|f(Y)|\right]\right)^2 - \mu^2\right).$$

▶ Though $q^*(y)$ is not available due to $\mathbb{E}_p\left[|f(y)|\right]$, the theorem does provide a principle for selecting $q(y)$: $q(y)$ should be "similar" to $|f(y)|p(y)$ in shape.

Proof of Theorem 1

Note that $\forall q$,

$$\text{Var}\left[\frac{f(Y)p(Y)}{q(Y)}\right] = \int \frac{f(y)^2 p(y)^2}{q(y)} dy - \mu^2$$

$$\geq \left(\int |f(y)|p(y)dy\right)^2 - \mu^2.$$

It is easy to see that the equality holds if

$$q(y) = \frac{|f(y)|p(y)}{\int |f(x)|p(x)dx}.$$

Illustrative Example

Estimate the integral $\int_0^1 1_{[3/4,1]}(x)dx$. Let $n = 1000$. Take

$$p(x) = 1 \text{ for } 0 \leq x \leq 1 \quad \text{and} \quad q(x) = \begin{cases} 0 & \text{if } 0 \leq x < 0.5 \\ 2 & \text{if } 0.5 \leq x < 1. \end{cases}$$

```python
import numpy as np

n = 1000
x = np.random.rand(1,n)
fx = (x>=0.75)
mu = np.mean(fx)
var = np.var(fx)
print('Simple MC Integration: mu = %.5f,' % mu, 'var = %.5f' % var)
```

```
 Simple MC Integration: mu = 0.23900, var = 0.18188
```

```python
y = 0.5*np.random.rand(1,n)+0.5
fy = (y>=0.75)/2
mu = np.mean(fy)
var = np.var(fy)
print('Importance Sampling: mu = %.5f,' % mu, 'var = %.5f' % var)
```

```
 Importance Sampling: mu = 0.25450, var = 0.06248
```

A Word about Bias-Variance Tradeoff

Let $\bar{\mu} \in \mathbb{R}^d$ be an estimator of parameter $\mu \in \mathbb{R}^d$ (not necessarily $\mathbb{E}[\bar{\mu}] = \mu$). Then the MSE is given by

$$\mathbb{E}\left[\|\bar{\mu} - \mu\|_2^2\right] = \mathbb{E}\left[\|\bar{\mu} - \mathbb{E}[\bar{\mu}] + \mathbb{E}[\bar{\mu}] - \mu\|_2^2\right]$$
$$= \underbrace{\|\mathbb{E}[\bar{\mu}] - \mu\|_2^2}_{\text{bias}^2} + \underbrace{\mathbb{E}\left[\|\bar{\mu} - \mathbb{E}[\bar{\mu}]\|_2^2\right]}_{\text{variance}}.$$

▶ Unbiased estimator is not always the one the achieves the smallest MSE.

A Word about Bias-Variance Tradeoff: Example

Consider the problem of estimating a vector $x \in \mathbb{R}^d$ from the noisy observation

$$y = x + e,$$

where $e \sim \mathcal{N}(0, \sigma^2 I_d)$. Suppose x is k-sparse, i.e., it only has k-nonzero entries.

▶ Least squares (LS) estimator: $\bar{x}_1 = y$. Unbiased; MSE is variance, given by

$$\mathbb{E}\left[\|\bar{x}_1 - x\|_2^2\right] = d \cdot \sigma^2.$$

▶ Zero estimator $\bar{x}_2 = 0$: $\mathbb{E}\left[(\bar{x}_2)_j\right] \neq x_j$ if $x_j \neq 0$ and $\text{Var}\left[\bar{x}_2\right] = 0$. Moreover,

$$\mathbb{E}\left[\|\bar{x}_2 - x\|_2^2\right] = \|x\|_2^2 \leq k\|x\|_\infty^2.$$

Assume $\|x\|_\infty < \sigma$. Then it can be seen that the MSE of $\bar{x}_2$ can be much smaller than that of $\bar{x}_1$.

Without assumption $\|x\|_\infty < \sigma$, one can use the soft-thresholding denoiser $\bar{x} = \text{sgn}(y)(|y| - \lambda)_+$ for suitable λ. See "Ideal spatial adaptation by wavelet shrinkage" by Donoho and Johnstone, 1994.

Table of Contents

Incremental Update for Sample Mean

Calculate sample mean $\bar{\mu}_k = \frac{1}{k} \sum_{j=1}^{k} X_j$ in an incremental manner rather than wait until all the samples are collected:

$$\bar{\mu}_{k+1} = \bar{\mu}_k + \frac{1}{k+1} \left(X_{k+1} - \bar{\mu}_k \right).$$

A more general form

$$\bar{\mu}_{k+1} = \bar{\mu}_k + \alpha_k \left(X_{k+1} - \bar{\mu}_k \right), \quad \alpha_k > 0.$$

▶ This update can be viewed as a special case of Robbins-Monro Algorithm.

Robbins-Monro (RM) Algorithm

Consider the fixed point problem

$$x = h(x),$$

where $h : \mathbb{R}^d \to \mathbb{R}^d$ is a function. The fixed point iteration after introducing stepsize is given by

$$x_{k+1} = x_k + \alpha_k(h(x_k) - x_k).$$

RM considers the case where only a perturbation or noisy version of $h(x)$, denoted $\tilde{h}(x) = h(x) + e$, is available for every input x.

$$x_{k+1} = x_k + \alpha_k \left(\tilde{h}(x_k) - x_k \right) = x_k + \alpha_k \left(h(x_k) - x_k + e_k \right).$$

When $h(x_k)$ cannot be accurately known, using $\alpha_k = 1$ may be too aggressive. Then adjusting the stepsize becomes necessary, which is also one reason that it is introduced.

Special Cases

▶ Incremental update for sample mean: $h(x) = \mu$. For each $\bar{\mu}$ only the X ($\mathbb{E}[X] = \mu$) can be obtained. Thus RM reduces to

$$\bar{\mu}_{k+1} = \bar{\mu}_k + \alpha_k \cdot (X_{k+1} - \mu_k).$$

▶ Stochastic gradient: $\min_x \mathbb{E}_\xi [f(x; \xi)]$ is equivalent to $x - \mathbb{E}_\xi [\nabla f(x; \xi)] = x$. For each x only $\nabla f(x; \xi)$ can be returned. Then RM reduces to

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k; \xi_k).$$

▶ Finding root of $f(x) = 0$ is equivalent to fixed point problem $x - f(x) = x$. Suppose only $f(x) + e$ can be obtained for each x. Then, RM reduces to

$$x_{k+1} = x_k - \alpha_k(f(x_k) + e_k).$$

▶ **RL is stochastic optimization problem due to existence of expectation and Bellman optimality equation is in the form of $x = h(x)$ and TD algorithms in RL are in the form of RM.**

Illustrative Example

Find the root of $f(x) = \tanh(x - 1)$ with starting point $x_1 = 3$.
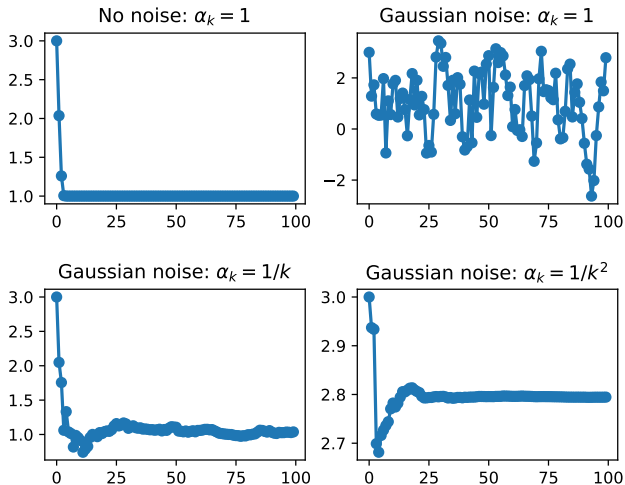
Motivation

Consider the mean estimation problem via $\bar{\mu}_k = \frac{1}{k}\sum_{j=1}^{k} X_k$, where X_k are i.i.d samples with μ and variance σ^2.

▶ By strong law of large numbers, one has

$$\bar{\mu}_k \to \mu \quad \text{almost surely.}$$

Can we establish similar asymptotic almost convergence for RM?

▶ Moreover, it is easy to show that

$$\mathbb{E}\left[|\bar{\mu}_k - \mu|^2\right] = \frac{\sigma^2}{k}.$$

Can we establish similar finite iteration convergence rate for RM?

Theorem 2 (Convergence of RM)

Assume the following conditions hold:

- $|h'(x)| \leq \gamma < 1$ *for all x;*
- $\sum_{k=1}^{\infty} \alpha_k = \infty$ *and* $\sum_{k=1}^{\infty} \alpha_k^2 < \infty$*;*
- $\mathbb{E}[e_k | \mathcal{F}_k] = 0$ *and* $\mathbb{E}[e_k^2 | \mathcal{F}_k] < \infty$ *with probability 1, where* $\mathcal{F}_k = \{x_k, x_{k-1}, \cdots\}$.

Then x_k converges to the fixed point solution x^* with probability 1.

The proof of this theorem relies on a powerful tool: Dvoretzky's Theorem.

We only consider the scalar case for asymptotic almost sure convergence.

Theorem 3 (Dvoretzky's Theorem)

Consider the stochastic process

$$w_{k+1} = (1 - \alpha_k)w_k + \beta_k e_k,$$

where $\{\alpha_k\}$, $\{\beta_k\}$, $\{\eta_k\}$ *are stochastic sequences and* $\alpha_k \geq 0$, $\beta_k \geq 0$. *Suppose the following conditions hold:*

► $\sum_{k=1}^{\infty} \alpha_k = \infty$, $\sum_{k=1}^{\infty} \alpha_k^2 < \infty$, $\sum_{k=1}^{\infty} \beta_k^2 < \infty$ *uniformly with probability* 1;
► $\mathbb{E}\left[e_k | \mathcal{F}_k\right] = 0$ *and* $\mathbb{E}\left[e_k^2 | \mathcal{F}_k\right] < \infty$, *where*

$$\mathcal{F}_k = \{w_k, w_{k-1}, \cdots, e_{k-1}, \cdots, \alpha_{k-1}, \cdots, \beta_{k-1}, \cdots\}.$$

Then w_k *converges to* 0 *with probability* 1.

Proof of Dvoretzky's Theorem requires the knowledge of martingales, which will not be discussed; see "On the Convergence of Stochastic Iterative Dynamic Programming Algorithms" by Jaakkola et al., 1994.

Since $|h'(x)| \leq \gamma < 1$, it is easy to see $h(x)$ is a contraction. Thus the fixed point solution x^* exists. Thus,

$$
\begin{aligned}
x_{k+1} - x^* &= x_k - x^* + \alpha_k(h(x_k) - h(x^*) + x^* - x_k) + \alpha_k e_k \\
&= x_k - x^* + \alpha_k(h(x_k')(x_k - x^*) + x^* - x_k) + \alpha_k e_k \\
&= x_k - x^* - \alpha_k(1 - h(x_k'))(x_k - x^*) + \alpha_k e_k.
\end{aligned}
$$

It is easy to verify that the conditions of Dvoretzky's Theorem holds for $w_k = x_k - x^*$. Thus $x_k \to x^*$ almost surely.

Theorem 4

Assume the following conditions hold:

▶ $\|h(x) - h(y)\|_2 \leq \gamma \|x - y\|_2$ *for* $0 < \gamma < 1$;

▶ $\mathbb{E}\left[e_k | \mathcal{F}_k\right] = 0$ *and* $\mathbb{E}\left[e_k^2 | \mathcal{F}_k\right] \leq B$ *with probability 1, where* $\mathcal{F}_k = \{x_k, x_{k-1}, \cdots\}$.

Set $\alpha_k = \frac{1}{k}$. *Then,*

$$\mathbb{E}\left[\|x_k - x^*\|_2^2\right] \lesssim \begin{cases} \frac{1}{k} & \text{if } 0 \leq \gamma < 1/2 \\ \frac{\log k}{k} & \text{if } \gamma = 1/2 \\ \frac{1}{k^{2(1-\gamma)}} & \text{if } 1/2 < \gamma < 1. \end{cases}$$

Proof of Theorem 4

Without loss of generality, assume that e_k are chosen independently in each iteration. First we have

$$
\begin{aligned}
\|x_{k+1} - x^*\|_2^2 &= \|x_k - x^* + x_{k+1} - x_k\|_2^2 \\
&= \|x_k - x^*\|_2^2 + 2\langle x_k - x^*, x_{k+1} - x_k\rangle + \|x_{k+1} - x_k\|_2^2 \\
&= \|x_k - x^*\|_2^2 + 2\alpha_k \langle x_k - x^*, h(x_k) - x_k + e_k\rangle + \alpha_k^2 \|h(x_k) - x_k + e_k\|_2^2 \\
&= \|x_k - x^*\|_2^2 + 2\alpha_k \langle x_k - x^*, h(x_k) - h(x^*)\rangle - 2\alpha_k \langle x_k - x^*, x_k - x^*\rangle \\
&\quad + 2\alpha_k \langle x_k - x^*, e_k\rangle + \alpha_k^2 \|h(x_k) - x_k + e_k\|_2^2 \\
&\leq \|x_k - x^*\|_2^2 - 2\alpha_k(1 - \gamma)\|x_k - x^*\|_2^2 + 2\alpha_k \langle x_k - x^*, e_k\rangle \\
&\quad + \alpha_k^2 \|h(x_k) - x_k + e_k\|_2^2 \\
&= \|x_k - x^*\|_2^2 - 2\alpha_k(1 - \gamma)\|x_k - x^*\|_2^2 + 2\alpha_k \langle x_k - x^*, e_k\rangle \\
&\quad + \alpha_k^2 \|h(x_k) - h(x^*) + x^* - x_k + e_k\|_2^2 \\
&\leq \|x_k - x^*\|_2^2 - 2\alpha_k(1 - \gamma)\|x_k - x^*\|_2^2 + 2\alpha_k \langle x_k - x^*, e_k\rangle \\
&\quad + \alpha_k^2 \left(2\|x_k - x^*\|_2 + \|e_k\|_2\right)^2 \\
&\leq \|x_k - x^*\|_2^2 - 2\alpha_k(1 - \gamma)\|x_k - x^*\|_2^2 + 2\alpha_k \langle x_k - x^*, e_k\rangle \\
&\quad + 8\alpha_k^2 \|x_k - x^*\|_2^2 + 2\alpha_k^2 \|e_k\|_2^2.
\end{aligned}
$$

Note that $\mathbb{E}\left[\langle x_k - x^*, e_k \rangle\right] = 0$. Taking an expectation on both sides yields that

$$\mathbb{E}\left[\|x_{k+1} - x^*\|_2^2\right] \le (1 - 2\alpha_k(1 - \gamma) + 8\alpha_k^2)\mathbb{E}\left[\|x_k - x^*\|_2^2\right] + 2\alpha_k^2 B.$$

Note that when k is sufficiently large $1 - 2\alpha_k(1 - \gamma) + 8\alpha_k^2 < 1$. Then the rate can be established by some algebraic recursive inequalities.

Questions?