IEOR E4703: Monte-Carlo Simulation

Output Analysis for Monte-Carlo

Martin Haugh

Department of Industrial Engineering and Operations Research
Columbia University
Email: martin.b.haugh@gmail.com

Output Analysis

Recall our simulation framework for estimating $\theta := \mathbb{E}[h(\mathbf{X})]$ where $\mathbf{X} \in \mathbb{R}^n$.

We first simulate $\mathbf{X}_1, \ldots, \mathbf{X}_n$ IID and then set

$$\widehat{\theta}_n = \frac{h(\mathbf{X}_1) + \ldots + h(\mathbf{X}_n)}{n}.$$

The Strong Law of Large Numbers (SLLN) implies

$$\widehat{\theta}_n \rightarrow \theta \text{ as } n \rightarrow \infty \text{ w.p. } 1.$$

But how large n should be so that we can have confidence in our estimator, $\hat{\theta}_n$?

We can figure this out through the use of confidence intervals.

Output Analysis

Suppose then that we wish to estimate θ and we have a random vector $\mathbf{Y} = (Y_1, \ldots, Y_n)$ whose distribution depends on θ.

We seek $L(\mathbf{Y})$ and $U(\mathbf{Y})$ such that

$$P\left(L(\mathbf{Y}) \leq \theta \leq U(\mathbf{Y})\right) = 1 - \alpha$$

where $0 \leq \alpha \leq 1$ is a pre-specified number.

We then say that $[L(\mathbf{Y}), U(\mathbf{Y})]$ is a $100(1-\alpha)\%$ confidence interval for θ.

Note that $[L(\mathbf{Y}), U(\mathbf{Y})]$ is a **random interval**.

However, once we replace $\mathbf{Y}$ with a sample vector, $\mathbf{y}$, then $[L(\mathbf{y}), U(\mathbf{y})]$ becomes a real interval.

Question: How can we find $L(\mathbf{Y})$ and $U(\mathbf{Y})$?

The Chebyshev Inequality

Y_i's are assumed IID so $\text{Var}(\widehat{\theta}_n) = \sigma^2/n$ where $\sigma^2 := \text{Var}(Y)$

- clearly a small value of $\text{Var}(\widehat{\theta}_n)$ implies a more accurate estimate of θ.

Indeed Chebyshev's Inequality states that for any $k > 0$ we have

$$P\left(|\widehat{\theta}_n - \theta| \geq k\right) \leq \frac{\text{Var}(\widehat{\theta}_n)}{k^2}. \tag{1}$$

Could easily use Chebyshev's Inequality to construct (how?) confidence intervals for θ but it is generally very conservative.

Question: Why does Chebyshev's Inequality generally lead to conservative confidence intervals?

Instead, will use the Central Limit Theorem (CLT) to obtain better estimates of $P\left(|\widehat{\theta}_n - \theta| \geq k\right)$

- and therefore narrower confidence intervals for θ.

The Central Limit Theorem

Theorem. (Central Limit Theorem)
Suppose $Y_1, \ldots, Y_n$ are IID and $\mathbb{E}[Y_i^2] < \infty$. Then

$$\frac{\widehat{\theta}_n - \theta}{\sigma/\sqrt{n}} \stackrel{d}{\rightarrow} \mathsf{N}(0,1) \quad \text{as} \quad n \rightarrow \infty$$

where $\widehat{\theta}_n = \sum_{i=1}^n Y_i/n$, $\theta := \mathbb{E}[Y_i]$ and $\sigma^2 := \mathsf{Var}(Y_i)$. $\square$

Using the CLT to Construct Confidence Intervals

Let $z_{1-\alpha/2}$ be the the $(1-\alpha/2)$ percentile point of the $\mathsf{N}(0,1)$ distribution so that

$$P(-z_{1-\alpha/2} \leq Z \leq z_{1-\alpha/2}) = 1-\alpha$$

when $Z \sim \mathsf{N}(0,1)$.

Suppose that we have simulated IID samples, Y_i, for $i = 1, \ldots, n$.

Now want to construct a $100(1-\alpha)\%$ CI for $\theta = \mathbb{E}[Y]$. That is, we want $L(\mathbf{Y})$ and $U(\mathbf{Y})$ such that

$$P\left(L(\mathbf{Y}) \leq \theta \leq U(\mathbf{Y})\right) = 1-\alpha.$$

The CLT implies $\sqrt{n}\left(\widehat{\theta}_n - \theta\right)/\sigma$ is approximately $\mathsf{N}(0,1)$ for large n.

Using the CLT to Construct Confidence Intervals

Therefore have

$$
\begin{aligned}
P\left(-z_{1-\alpha/2} \;\leq\; \frac{\sqrt{n}(\widehat{\theta}_n - \theta)}{\sigma} \;\leq\; z_{1-\alpha/2}\right) &\;\approx\; 1-\alpha \\
\Rightarrow\; P\left(-z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \;\leq\; \widehat{\theta}_n - \theta \;\leq\; z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right) &\;\approx\; 1-\alpha \\
\Rightarrow\; P\left(\widehat{\theta}_n - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \;\leq\; \theta \;\leq\; \widehat{\theta}_n + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right) &\;\approx\; 1-\alpha.
\end{aligned}
$$

Our approximate $100(1-\alpha)\%$ CI for θ is therefore given by

$$
[L(\mathbf{Y}),\; U(\mathbf{Y})] \;=\; \left[\widehat{\theta}_n - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}},\;\; \widehat{\theta}_n + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]. \tag{2}
$$

Recall that $\widehat{\theta}_n = (Y_1 + \ldots + Y_n)/n$, so L and U are indeed functions of $\mathbf{Y}$.

Using the CLT to Construct Confidence Intervals

There is still a problem, however, as we do not usually know σ^2. We resolve this issue by estimating σ^2 with

$$\hat{\sigma}_n^2 = \frac{\sum_{i=1}^n (Y_i - \hat{\theta}_n)^2}{n-1}.$$

Easy to show that $\hat{\sigma}_n^2$ is an unbiased estimator of σ^2 and that $\hat{\sigma}_n^2 \to \sigma^2$ w.p. 1 as $n \to \infty$.

So now replace σ with $\hat{\sigma}_n$ in (2) to obtain

$$[L(\mathbf{Y}),\ U(\mathbf{Y})] = \left[\widehat{\theta}_n - z_{1-\alpha/2}\frac{\hat{\sigma}_n}{\sqrt{n}},\ \ \widehat{\theta}_n + z_{1-\alpha/2}\frac{\hat{\sigma}_n}{\sqrt{n}}\right] \qquad (3)$$

as our **approximate** $100(1-\alpha)\%$ CI for θ when n is "large".

Using the CLT to Construct Confidence Intervals

When we obtain sample values $\mathbf{y} = (y_1, \ldots, y_n)$, then $[L(\mathbf{y}),\ U(\mathbf{y})]$ becomes a **real interval**.

Then can no longer say (why not?) that $P\left(\theta \in [L(\mathbf{y}),\ U(\mathbf{y})]\right) = 1 - \alpha$.

Instead, say that we are $100(1-\alpha)\%$ **confident** that $[L(\mathbf{y}),\ U(\mathbf{y})]$ contains θ.

The width of the confidence interval is given by

$$U - L = \frac{2\hat{\sigma}_n z_{1-\alpha/2}}{\sqrt{n}}$$

and so the half-width is $(U - L)/2$.

For a fixed α, must increase n if we are to decrease the width of the CI.

$U - L \propto 1/\sqrt{n}$ so (for example) would need to increase n by a factor of four in order to decrease the width of CI by only a factor of two.

Run-Length Control

Up to this point we have selected n and then computed the approximate CI. The width of the CI is then a measure of the **error** in our estimator.

Now will do the reverse by first choosing some error criterion that we want our estimator to satisfy, and then choosing n so that this criterion is satisfied.

There are two types of error that we will consider:

1. Absolute error which is given by $E_a := |\hat{\theta}_n - \theta|$
2. Relative error which is given by $E_r := \left|\frac{\hat{\theta}_n - \theta}{\theta}\right|$.

We know that $\widehat{\theta}_n \rightarrow \theta$ w.p. 1 as $n \rightarrow \infty$ so that E_a and E_r both $\rightarrow 0$ w.p.1 as $n \rightarrow \infty$.

However, in practice $n \neq \infty$ and so the errors will be non-zero. We specify the following error criterion:

Error Criterion: Given $0 \leq \alpha \leq 1$ and $\epsilon \geq 0$, we want $P(E \leq \epsilon) = 1 - \alpha$ where $E = E_a$ or $E = E_r$.

Run-Length Control

Goal then is to choose n so that error criterion is (approximately satisfied) and this is easily done.

Suppose (for example) we want to control absolute error, E_a. Then

$$P\left(\hat{\theta}_n - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \ \leq \ \theta \ \leq \ \hat{\theta}_n + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \approx 1-\alpha.$$

Therefore $P\left(|\widehat{\theta}_n - \theta| \leq z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \approx 1-\alpha$, i.e.

$$P\left(E_a \leq z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \approx 1-\alpha.$$

If we then want $P\left(E_a \leq \epsilon\right) \approx 1-\alpha$, then clearly suffices to choose n such that

$$n = \sigma^2 z_{1-\alpha/2}^2/\epsilon^2.$$

Similar argument implies that $P\left(E_r \leq \epsilon\right) \approx 1-\alpha$ if

$$n = \sigma^2 z_{1-\alpha/2}^2/\theta^2\epsilon^2.$$

Run-Length Control

There are still some problems, however:

1. When we are controlling E_r, we need to know σ and θ in advance.
2. When we are controlling E_a, we need to know σ in advance.

Of course we do not usually know σ or θ in advance. In fact, θ is what we are trying to estimate!

There are two methods we can use to overcome this problem: the two-stage method and the sequential method.

The Bootstrap Approach to Output Analysis

Can view output analysis problem as one of estimating

$$\mathsf{MSE}(F) := \mathbb{E}_F\left[(g(Y_1,\ldots,Y_n) - \theta(F))^2\right] \tag{4}$$

where $\theta(F) = \mathbb{E}_F[X]$, $g(Y_1,\ldots,Y_n) := \bar{Y}$ and F denotes the CDF of Y.

Saw earlier how we could use the CLT to construct approximate CI's for θ.

But there are situations where the CLT cannot be easily used to obtain a CI.
e.g. If $\theta(F) = \mathsf{Var}(Y)$ or $\theta(F) = \mathbb{E}[Y \mid Y \geq \alpha]$.

Need an alternative method of constructing a CI for θ in such situations.

The bootstrap method provides such an alternative.

To begin, recall the empirical distribution, F_e, is the CDF that places a weight of $1/n$ on each of the simulated values $Y_1,\ldots,Y_n$.

The Bootstrap Approach to Output Analysis

The empirical CDF therefore satisfies

$$F_e(y) = \frac{\sum_{i=1}^{n} 1_{\{Y_i \leq y\}}}{n}.$$

For large n can be shown (and should be intuitively clear) that F_e should be a good approximation to F.

Therefore, if θ sufficiently **well-behaved** function of F, then for sufficiently large n should have

$$\mathrm{MSE}(F) \approx \mathrm{MSE}(F_e) = \mathbb{E}_{F_e}\left[\left(g(Y_1, \ldots, Y_n) - \theta(F_e)\right)^2\right]. \tag{5}$$

$\mathrm{MSE}(F_e)$ is known as the bootstrap approximation to $\mathrm{MSE}(F)$

- easy to estimate via simulation.

But first will consider an example where $\mathrm{MSE}(F_e)$ can be computed exactly.

Indeed the bootstrap is not required in this case but nonetheless instructive to see the explicit calculations.

E.G. Applying the Bootstrap to the Sample Mean

We wish to estimate $\theta(F) = \mathbb{E}_F[Y]$ via the estimator $\hat{\theta} = g(Y_1, \ldots, Y_n) := \bar{Y}$.

As noted above, the bootstrap is not necessary in this case as we can apply the CLT directly to obtain CI's for $\hat{\theta}$.

Equivalently, we can estimate the MSE, $\mathbb{E}\left[\left(\bar{Y} - \theta\right)^2\right] = \sigma^2/n$, with $\hat{\sigma}_n^2/n = \sum_{i=1}^n (y_i - \bar{y})^2/(n(n-1))$.

Let $\bar{y}$ denote mean of simulated data-points $y_1, \ldots, y_n$. Bootstrap estimator then given by

$$
\begin{aligned}
\text{MSE}(F_e) &= \mathbb{E}_{F_e}\left[\left(\frac{\sum_{i=1}^n Y_i}{n} - \bar{y}\right)^2\right] \\
&= \text{Var}_{F_e}\left(\frac{\sum_{i=1}^n Y_i}{n}\right) && (6) \\
&= \frac{\text{Var}_{F_e}(Y)}{n} && (7) \\
&= \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n^2}.
\end{aligned}
$$

The Bootstrap Algorithm

In general, however, cannot usually compute $\text{MSE}(F_e)$ explicitly. But it's an expectation so can use Monte-Carlo!

Bootstrap Simulation Algorithm for Estimating MSE(F)

> **for** $i = 1$ **to** B
>
> > **generate** $Y_1, \ldots, Y_n$ IID from F_e
> >
> > **set** $\hat{\theta}_i^b = g(Y_1, \ldots, Y_n)$
> >
> > **set** $Z_i^b = \left[\hat{\theta}_i^b - \theta(F_e)\right]^2$
>
> **end for**
>
> **set** $\widehat{\text{MSE}(F)} = \sum_{b=1}^{B} Z^{(b)}/B$

The Z_i^b's (or equivalently the $\hat{\theta}_i^b$'s) are the bootstrap samples and a value of $B = 100$ is often sufficient to obtain a sufficiently accurate estimate.

Next example applies bootstrap in a historical simulation context where we have **real data** observations as opposed to simulated data

- have no control over n in historical simulations – a disadvantage!

e.g. Estimating the Minimum Variance Portfolio

Wish to invest a fixed sum of money in two financial assets, X and Z say, that yield random returns of R_x and R_z, respectively.

Let θ and $1-\theta$ be fractions of our wealth invested in X and Z.

Goal is to choose θ to minimize the **total variance**

$$\text{Var}(\theta R_x + (1-\theta)R_z).$$

Easy to see the minimizing θ satisfies

$$\theta = \frac{\sigma_z^2 - \sigma_{xz}}{\sigma_x^2 + \sigma_z^2 - 2\sigma_{xz}}. \tag{8}$$

In practice have to estimate σ_z^2, σ_x^2 and σ_{xz} from historical return data $Y_1, \ldots, Y_n$ with $Y_i := \left(R_x^{(i)}, R_z^{(i)}\right)$ the joint return in period i.

Therefore obtain

$$\hat{\theta} = \frac{\hat{\sigma}_z^2 - \hat{\sigma}_{xz}}{\hat{\sigma}_x^2 + \hat{\sigma}_z^2 - 2\hat{\sigma}_{xz}}. \tag{9}$$

as our estimator.

e.g. Estimating the Minimum Variance Portfolio

Would like to know how good an estimator $\hat{\theta}$ is. More specifically, what is MSE $\left(\hat{\theta}\right)$?

Can answer this question using the bootstrap with $\theta(F) := \theta$ and $g(Y_1, \ldots, Y_n) = \hat{\theta}$.

Question: Provide pseudo-code for estimating $\text{MSE}(\hat{\theta}) := \text{MSE}(F)$.

Estimating Confidence Intervals Via the Bootstrap

Bootstrap also widely used to construct CI's. Here we consider the so-called basic bootstrap interval.

Consider the bootstrap samples $\hat{\theta}_1^b, \ldots, \hat{\theta}_B^b$ and suppose we want a $1 - \alpha$ CI for $\theta = \theta(F)$.

Let q_l and q_u be the $\alpha/2$ **lower-** and **upper-sample quantiles**, respectively, of the bootstrap samples.

Fraction of bootstrap samples satisfying

$$q_l \leq \hat{\theta}^b \leq q_u \tag{10}$$

is $1 - \alpha$. But (10) is equivalent to

$$\hat{\theta} - q_u \leq \hat{\theta} - \hat{\theta}^b \leq \hat{\theta} - q_l \tag{11}$$

where $\hat{\theta} = g(y_1, \ldots, y_n)$ is our estimate of θ computed using original data-set.

Estimating Confidence Intervals Via the Bootstrap

This implies $\hat{\theta} - q_u$ and $\hat{\theta} - q_l$ are the lower and upper quantiles for $\hat{\theta} - \hat{\theta}^b$.

The basic bootstrap assumes they are **also** the quantiles for $\theta - \hat{\theta}$

- makes sense intuitively – and can be justified mathematically as $n \to \infty$ and if θ is a **well-behaved** function of F.

Therefore follows that

$$\hat{\theta} - q_u \leq \theta - \hat{\theta} \leq \hat{\theta} - q_l \tag{12}$$

will occur in approximately in a fraction $1 - \alpha$ of samples.

Adding $\hat{\theta}$ across (12) yields an approximate $(1-\alpha)\%$ CI for θ of

$$(2\hat{\theta} - q_u,\, 2\hat{\theta} - q_l).$$