Confidence intervals

Let $X_1, X_2, \ldots, X_n$ be independent *realizations* of a random variable X with unknown mean μ and unknown variance σ^2.

Sample mean

$$\bar{X}(n) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Sample variance

$$S^2(n) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}(n))^2$$

An approximate $100(1-\delta)\%$ confidence interval for the unknown mean μ is given by

$$\bar{X}(n) \pm z_{1-\delta/2}\frac{S(n)}{\sqrt{n}}$$

Remark:

The width of a confidence interval can be reduced by decreasing the value of $S(n)$. The reduction obtained by halving $S(n)$ is the same as the one obtained by producing four times as much observations.

We will discuss the following **variance reduction techniques**:

- common random numbers;
- antithetic variables;
- control variates;
- conditioning;

Variance reduction technique:
Common random numbers

If we compare the performance of two systems with random components it is in general better to evaluate both systems with the *same realizations* of the random components.

If X and Y are estimators for the performance of the two systems, then

$$\mathrm{var}(X - Y) = \mathrm{var}(X) + \mathrm{var}(Y) - 2\,\mathrm{cov}(X, Y).$$

In general, use of common random numbers leads to *positively correlated* X and Y:

$$\mathrm{cov}(X, Y) > 0$$

Hence, the variance of $X - Y$ (and thus the corresponding confidence interval) will be smaller than in case X and Y are independent (when generated with independent random numbers).

Example: Job scheduling

Suppose that N jobs have to be processed on M identical machines. The processing times are independent and exponentially distributed with mean 1.

We want to compare the completion time of the last job, $C_{\max}$, under two different strategies:

- Longest Processing Time First (LPTF)
- Shortest Processing Time First (SPTF)

Results for $M = 2$, $N = 10$ *without* using common jobs for the SPTF and LPTF strategy (n is the number of experiments):

n	$C_{\max}^{SPTF} - C_{\max}^{LPTF}$		
	mean	st.dev.	half width 95% CI
10^3	0.860	2.43	0.151
10^4	0.867	2.49	0.049
10^5	0.849	2.50	0.016

and the results using common jobs:

n	$C_{\max}^{SPTF} - C_{\max}^{LPTF}$		
	mean	st.dev.	half width 95% CI
10^3	0.840	0.497	0.031
10^4	0.833	0.512	0.010
10^5	0.838	0.518	0.003

Hence, using common jobs reduces the (sample) standard deviation of $C_{\max}^{SPTF} - C_{\max}^{LPTF}$, and thus the width of the confidence interval for its mean with a factor 5!

Example: Single-stage production system

We want to determine the reduction in the mean waiting time when we have an extra machine. To compare the two situations we want to use the same realizations of arrival times and processing times in the simulation.

Assume that processing times of jobs are generated when they enter production. Then, the order in which arrival and processing times are generated depends on the number of machines: *synchronization problem*.

Solution:

- Use separate random number streams for different random variables (arrival and proccessing times);
- Design the simulation model such that it guarantees that exactly the same realizations of random variables are generated.

In this problem, the second aproach can be realized by assigning to each job a processing time immediately upon arrival.

Results for $\lambda = 4$, $\mu = 1$ and the number of machines M is 5, resp. 6. In each run $N = 10^5$ waiting times are generated;

i	$\bar{W}_N^{(i)}$		$\Delta^{(i)}$
	$M = 5$	$M = 6$	
1	0.607	0.150	0.457
2	0.545	0.138	0.407
3	0.527	0.139	0.388
4	0.526	0.135	0.391
5	0.595	0.157	0.438
6	0.569	0.144	0.425
7	0.587	0.150	0.437
8	0.577	0.149	0.428
9	0.553	0.145	0.408
10	0.554	0.138	0.416

The standard deviation of the $\Delta^{(i)}$ is equal to 0.022. If both systems use independent realizations of arrival and processing times the standard dev. of $\Delta^{(i)}$ is 0.029; so common random numbers yields a reduction of 25%.

Variance reduction technique:
Antithetic variables

If X_1 and X_2 are the outcomes of two successive simulation runs, then

$$\mathrm{var}\left(\frac{X_1+X_2}{2}\right) = \frac{1}{4}\mathrm{var}(X_1) + \frac{1}{4}\mathrm{var}(X_2) + \frac{1}{2}\mathrm{cov}(X_1, X_2).$$

Hence, when X_1 and X_2 are *negatively correlated*, the variance of $(X_1+X_2)/2$ will be smaller than when X_1 and X_2 are independent.

Observation:
If U is uniform on $(0, 1)$, then so is $1-U$ and U and $1-U$ are negatively correlated.

Hence, if we use the sequence $U_1, \ldots, U_N$ to compute the outcome X_1 of the first run, and after that the sequence $1-U_1, \ldots, 1-U_N$ to compute X_2 in the second run, then we expect that X_1 and X_2 are negatively correlated.

Instead of generating n independent outcomes

$$X_1, X_2, \ldots, X_n,$$

we now generate $n/2$ outcomes

$$(X_{2i-1} + X_{2i})/2, \qquad i = 1, \ldots, n/2.$$

For each pair X_{2i-1} and X_{2i}, the second one uses *antithetic variables*.
Then we expect that the variance of the average $(X_{2i-1} + X_{2i})/2$ will be less than one half of the variance of an individual outcome.

Remark:
The method is easy to implement if all random variables are generated using the *inverse transform method*.

Example: Job scheduling

We want to estimate $E(C_{\max}^{LPTF})$, the mean completion time of the last job under the LPTF strategy.

The random variables Y_i, $i = 1, \ldots, N$, used to generate an ordered list of jobs are exponential with mean $1/i$; they are sampled according to (the inverse transform method)

$$Y_i = -\ln(U_i)/i$$

where U_i is uniform on $(0, 1)$.

To apply the method of antithetic variables, we generate pairs of experiments; if $U_1, \ldots, U_N$ are used to generate $Y_1, \ldots, Y_N$ in the first experiment, then we use $1 - U_1, \ldots, 1 - U_N$ in the second one.

The idea is that when the random variables Y_i are small (U_i are close to one), then they will be large (U_i are close to zero) in the second experiment.

Results for $M = 2$, $N = 10$ *without* using antithetic variables:

n	$C_{\max}^{LPTF}$		
	mean	st.dev.	half width 95% CI
10^3	4.998	1.51	0.094
10^4	5.036	1.59	0.031
10^5	5.053	1.60	0.010

Results for $M = 2$, $N = 10$ using antithetic variables; n is the number of *pairs of experiments*, where in each pair the second experiment uses antithetic variables; the table shows the mean and st.dev. of the *average completion time of a pair of experiments*:

$2n$	mean	st.dev.	half width 95% CI
10^3	5.032	0.70	0.061
10^4	5.070	0.71	0.020
10^5	5.054	0.69	0.006

We conclude that use of antithetic variables reduces the width of the confidence interval for $E(C_{\max}^{LPTF})$ with a factor 1.5.

Example: Single-stage production system

The exponential processing times B (with mean $1/\mu$) in the single-stage production system are sampled as

$$B = -\ln(U)/\mu$$

Thus a small U gives a large processing time.

So, if we produce in the first run $U_1, \ldots, U_N$ yielding large processing times and thus large waiting times, and then, by using in the second run the random variables $1 - U_1, \ldots, 1 - U_N$, we get small processing times and thus small waiting times.

Note that we also have to take care of synchronization.

Results for $\lambda = 4$, $\mu = 1$ and $M = 5$ *without using antithetic variables.* In each run $N = 10^5$ waiting times are generated;

i	$\bar{W}_N^{(i)}$	i	$\bar{W}_N^{(i)}$
1	0.607	11	0.510
2	0.545	12	0.612
3	0.527	13	0.535
4	0.526	14	0.547
5	0.595	15	0.586
6	0.569	16	0.526
7	0.587	17	0.580
8	0.577	18	0.629
9	0.553	19	0.516
10	0.554	20	0.572

Hence $E(W) = 0.563 \pm 0.015$ (95% confidence interval)

Results for $\lambda = 4$, $\mu = 1$ and $M = 5$ *using antithetic variables.* In each run $N = 10^5$ waiting times are generated; in each pair the second experiment uses antithetic variables.

i	$\bar{W}_N^{(2i-1)}$	$\bar{W}_N^{(2i)}$
1	0.607	0.533
2	0.545	0.564
3	0.527	0.537
4	0.526	0.559
5	0.595	0.530
6	0.569	0.526
7	0.587	0.529
8	0.577	0.499
9	0.553	0.567
10	0.554	0.547

Hence $E(W) = 0.552 \pm 0.007$ (95% confidence interval)
So use of antithetic variables reduces the width of the confidence interval for $E(W)$ with a factor 2.

Variance reduction technique:
Control variates

Let $X_1, \ldots, X_n$ be realizations of the random variable X with *unknown mean* $E(X)$; then the sample mean $\bar{X}(n)$ is an *unbiased* estimator of $E(X)$:

$$E(\bar{X}(n)) = E(X).$$

Let $Y_1, \ldots, Y_n$ be realizations of Y with *known mean* $E(Y)$, which can be generated at the same time as the sequence $X_1, \ldots, X_n$.

Then also $\bar{X}(n) + c(\bar{Y}(n) - E(Y))$ is an unbiased estimator for $E(X)$; its variance is equal to

$$\mathrm{var}(\bar{X}(n)) + c^2 \mathrm{var}(\bar{Y}(n)) + 2\,c\,\mathrm{cov}(\bar{X}(n), \bar{Y}(n)),$$

which is minimized for

$$c = -\frac{\mathrm{cov}(\bar{X}(n), \bar{Y}(n))}{\mathrm{var}(\bar{Y}(n))} = -\frac{\mathrm{cov}(X, Y)}{\mathrm{var}(Y)}$$

The quantities $\mathrm{var}(Y)$ and $\mathrm{cov}(X, Y)$ are usually unknown, and may be estimated from the simulated data, by their sample estimates

$$S^2(n) = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y}(n))^2$$

$$C(n) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}(n))(Y_i - \bar{Y}(n))$$

Why does this method work?

Suppose $\bar{X}(n)$ and $\bar{Y}(n)$ are positively correlated; if it happens that $\bar{X}(n) > E(X)$, then probably also $\bar{Y}(n) > E(Y)$, and hence, we correct the estimate for $E(X)$ by lowering its value.

A similar argument applies when $\bar{X}(n)$ and $\bar{Y}(n)$ are negatively correlated.

Example: Fluid flow model for the two-machine production line

A natural control variate is the throughput TH for buffer size $K = 0$, which is given by

$$TH = r_2 \cdot \frac{\mu}{\lambda + \mu},$$

where $1/\lambda$ and $1/\mu$ are the mean up and down time of the first machine.

Results for $\lambda = 1/9$, $\mu = 1$, $r_1 = 5$, $r_2 = 4$ and $K = 1$; in each run, at $t = 0$ the buffer is empty and machine 1 is up, and the runlength is 10^5 time units. The variables X_i and Y_i denote the throughput for $K = 1$ and $K = 0$ in run i. In this example $E(Y) = 3.6$.

i	X_i	Y_i	$X_i + c(Y_i - E(Y))$
1	3.6848	3.6006	3.6842
2	3.6832	3.5983	3.6847
3	3.6841	3.5998	3.6843
4	3.6857	3.6008	3.6849
5	3.6879	3.6035	3.6848
6	3.6904	3.6057	3.6853
7	3.6814	3.5958	3.6852
8	3.6895	3.6056	3.6845
9	3.6846	3.6001	3.6845
10	3.6874	3.6033	3.6844

The estimated value for c is given by

$$c = -0.88991$$

In case the control variate Y is not used, the resulting 95% confidence interval is

$$E(X) = 3.6859 \pm 0.0018$$

and when Y is used,

$$E(X) = 3.6847 \pm 0.0002$$

Variance reduction technique:
Conditioning

If X and Y are two random variables, then

$$E(E(X|Y)) = E(X)$$

and

$$\text{var}(X) = E(\text{var}(X|Y)) + \text{var}(E(X|Y))$$

So

$$\text{var}(X) \geq \text{var}(E(X|Y))$$

How can we use this property?

Suppose we are interested in some performance characteristic X and we want to estimate its mean $E(X)$. If Y is some random variable for which $E(X|Y)$ is known, then instead of simulating X, we can better simulate Y and thus $E(X|Y)$, and estimate the mean of $E(X|Y)$.

Example: single-stage production system with a finite buffer

A single machine processes jobs in order of arrival. The interarrival times and processing times are exponential with parameters λ and μ. In front of the machine there is a small buffer with a capacity of $N-1$ jobs. Jobs who find upon arrival $N-1$ jobs in the buffer are lost (i.e., they will be processed somewhere else).

We are interested in $E(X_t)$, which is the mean total number of jobs that is lost upto time t, given that the system is empty at time $t=0$.
Let $X_t^{(i)}$ be the total number of lost jobs in the i-th simulation run, $i = 1,\ldots,n$. Then the sample mean

$$\bar{X}_{nt} = \frac{1}{n}\sum_{i=1}^{n} X_t^{(i)}$$

is an unbiased estimator of $E(X_t)$.
But an estimator with a smaller variance can be obtained as follows.

Let Y_t be the total amount of time in $(0, t)$ that the buffer is full. Since jobs arrive according to a Poisson stream, we have

$$E(X_t|Y_t) = \lambda Y_t$$

Hence, if $Y_t^{(i)}$ is the total amount of time the buffer is full in the i-th simulation run, then

$$\lambda \bar{Y}_{nt}$$

is an improved estimator for $E(X_t)$, where $\bar{Y}_{nt}$ is the sample mean

$$\bar{Y}_{nt} = \frac{1}{n} \sum_{i=1}^{n} Y_t^{(i)}$$

Results for $\lambda = 0.5$, $\mu = 1$, $N = 3$ and $t = 1000$ without using conditioning.

n	$E(X_t)$	$\sigma(X_t)$	half width 95% CI
10^1	33.1	10	6.2
10^2	34.1	9.2	1.8
10^3	33.6	8.9	0.6

Results for $\lambda = 0.5$, $\mu = 1$, $N = 3$ and $t = 1000$ using conditioning.

n	$E(X_t) = \lambda E(Y_t)$	$\lambda\sigma(Y_t)$	half width 95% CI
10^1	33.6	7.2	4.4
10^2	33.8	6.9	1.4
10^3	33.4	6.6	0.4

Hence, use of conditionig reduces the standard deviation of the estimator for $E(X_t)$ and thus the width of the confidence interval with a factor 1.3.