IERG 3050 Week 8
Output Data Analysis and Variance Reduction Techniques

Bolei Zhou
Department of Information Engineering
The Chinese University of Hong Kong
Announcement

• Homework 3 is due tomorrow
• Homework 4 will be out by the end of today. No need to hand in, but very relevant to Quiz 2
• Quiz 2 will be on Nov. 6
• No lectures next week (Time for your course project)
• End of Oct. 2019: a 4-page check-point report for course project
Outline

• Output Data Analysis
• Stochastic Processes
• Ensemble Averages vs. Time Averages
• Transient State and Steady State Behavior
• Method of Replication
• Transient Removal Methods
• Terminating and Non-Terminating Simulations
• Obtaining Specified Precision
• Choosing Initial Conditions
• Variance-Reduction Techniques: Common Random Numbers, Antithetic Variates, Control Variates, Indirect Estimation, Conditioning

Reading: Chapters 9 and 11
Output Data Analysis

We want to find the answers from simulation runs.
However:
- Simulation runs are stochastic (random) by nature.
- We cannot get absolute “answers” from simulation runs.

Two different runs of the same model can produce different numerical results:
- Input values are random.
- Simulation model is fixed (same facilities, same server configuration).
- Output is random.
Output Data Analysis

The observed performance measures based on simulation runs (i.e., the output of the simulation runs) are samples from the probability distribution of the performance measures.

In order to interpret the simulation output, we perform statistical analysis of the output data.

E.g., to find out the answer of the average waiting time in queue $\overline{\text{WIQ}}$, we ask questions such as the following:

- Mean (expected value): $E\left[\overline{\text{WIQ}}\right]$.
- Variance: $\text{Var}\left(\overline{\text{WIQ}}\right)$.
- Probabilities: $\Pr\left(\overline{\text{WIQ}} > 250\ \text{minutes}\right)$.
- Quantiles: for what value of $x$ do we have $\Pr\left(\overline{\text{WIQ}} > x\right) \leq 0.02$?
Output Data Analysis

Simulation output data are usually

- non-normal;
- autocorrelated;
- non-stationary;
- ...

However, many statistical methods assume one or several of the following:

- normally distributed samples;
- i.i.d. samples;
- stationarity;
- ...

Therefore, we have to be careful to use appropriate statistical methods.
Review: Stochastic processes
Review: Random Variables

Let $X$ be a random variable. Recall that a random variable is defined to be a mapping from $\Omega$ to $\mathbb{R}$, i.e.,

$$X : \Omega \rightarrow \mathbb{R},$$

$$\omega \mapsto X(\omega).$$
Review: Random Variables

Let $X$ and $Y$ be two random variables, i.e.,

$X : \Omega \rightarrow \mathbb{R},$

$\omega \mapsto X(\omega);$

$Y : \Omega \rightarrow \mathbb{R},$

$\omega \mapsto Y(\omega).$
Review: Random Variables

Let $X$ and $Y$ be two random variables, i.e.,

$$X : \Omega \rightarrow \mathbb{R},$$

$$\omega \mapsto X(\omega);$$

$$Y : \Omega \rightarrow \mathbb{R},$$

$$\omega \mapsto Y(\omega).$$
Review: Stochastic Processes

Consider now a setup where we have random variables $X_0, X_1, X_2, X_3, \ldots$.

For a given $\omega \in \Omega$,

$$(X_t(\omega) : t \in \{0, 1, 2, 3, \ldots \})$$

represents a sample path / sample sequence.

As always, the only randomness is in which $\omega \in \Omega$ is picked. Once $\omega$ is picked,

$X_0(\omega), X_1(\omega), X_2(\omega), X_3(\omega), \ldots$

are determined.
Review: Stochastic Processes

Other stochastic processes:

- \( \{ X_t : t \in \{1, 2, 3, \ldots \} \} \)
- \( \{ X_t : t \in \mathbb{Z} \} \)
- \( \{ X_t : t \in \mathbb{R} \} \)
- etc.

In general, a stochastic process is a collection of random variables

\[ \{ X_t : t \in \mathcal{T} \}, \]

where \( \mathcal{T} \) is a totally ordered set.
Some Simple Stochastic Processes

In the following, we discuss the following processes:

• Example 1: i.i.d. process
• Example 2: “fixed-value” process
• Example 3: autoregressive process
• Example 4: hidden Markov model
• Example 5: M/M/1/1 queueing system
Example 1: I.I.D. Process

Let $X$ be some random variable. Define the stochastic process

$$\{X_t : t \in \{0, 1, 2, 3, \ldots \}\},$$

where

the $X_t$'s are i.i.d. random variables with the same distribution as $X$. 

\[\begin{array}{c}
\Omega \\
\omega \\
X_0(\omega) \\
X_1(\omega) \\
X_2(\omega) \\
X_3(\omega) \\
\vdots
\end{array}\]
Example 2: “Fixed-Value” Process

Let $X$ be some random variable. Define the stochastic process

$$\{X_t : t \in \{0, 1, 2, 3, \ldots \}\},$$

where

$$X_t = X, \quad t \in \{0, 1, 2, 3, \ldots \}.$$
Example 3: Autoregressive Process

Define the stochastic process

\[ X_t \triangleq a \cdot X_{t-1} + b \cdot Z_t, \quad t \in \{1, 2, 3, \ldots \}, \]

where

- \(X_0\) can be either fixed to some value or drawn according to some distribution,
- \(Z_t\) is drawn according to some (time-indep. or time-dep.) distribution,
- for all \(t \in \{1, 2, 3, \ldots \}\), the random variable \(Z_t\) is independent of \(X_0, X_1, \ldots, X_{t-1}\) and of \(Z_1, Z_2, \ldots, Z_{t-1}\).

Note: the above is a simple variant of an autoregressive process.
Example 4: Hidden Markov Process

Define the stochastic process

\[ S_t \triangleq g_t(S_{t-1}), \quad X_t \triangleq h_t(S_t), \quad t \in \{1, 2, 3, \ldots \}, \]

where

- \( S_0 \) can be either fixed to some value or drawn according to some distribution,
- \( g_t \) is a stochastic mapping,
- \( h_t \) is a stochastic mapping.
Example 5: M/M/1/1 Queueing System

Fix some arrival rate $\lambda$ and some service rate $\mu$. Let $\Delta t > 0$. Define

$$S_t = \begin{cases} 0 & \text{(if } S_{t-\Delta t} = 0, Z_t = 0 \text{ or } S_{t-\Delta t} = 1, X_t = 1) \\ 1 & \text{(if } S_{t-\Delta t} = 0, Z_t = 1 \text{ or } S_{t-\Delta t} = 1, X_t = 0) \end{cases}, \quad t \in \{\Delta t, 2\Delta t, 3\Delta t, \ldots \},$$

where

- $S_0$ can be either fixed to 0 or 1, or be drawn according to some distribution;
- $Z_t \sim \text{Bernoulli}(\lambda \cdot \Delta t)$; \quad \text{[}Z_t \text{ indicates an arrival]}\]
- $X_t = 0$ if $S_t = 0$ and $X_t \sim \text{Bernoulli}(\mu \cdot \Delta t)$ if $S_t = 1$; \quad \text{[}X_t \text{ indicates a departure]}\]
- for all $t \in \{\Delta t, 2\Delta t, 3\Delta t, \ldots \}$, the random variable $Z_t$ is independent of all earlier random variables; for all $t \in \{\Delta t, 2\Delta t, 3\Delta t, \ldots \}$, conditioned on $S_{t-\Delta t}$, the random variable $X_t$ is independent of earlier random variables and of $Z_t$.

In the limit $\Delta t \to 0$, we obtain an M/M/1/1 queueing system.
Notations and Terms of Queueing Systems

Kendall notation: A/S/m/B/K/SD
- A: interarrival time distribution
- S: service time distribution
- m: number of servers
- B: system capacity: max number of users accommodated in the system
Interesting Quantities

In the following, we will consider

ensemble averages

and

time averages.

It is very important not to confuse these two types of averages!!!
Interesting Quantities

\[ X_t(\omega_1) \]
\[ \Omega = \omega_1 \]

\[ X_t(\omega_2) \]
\[ \Omega = \omega_2 \]

\[ X_t(\omega_3) \]
\[ \Omega = \omega_3 \]

\[ X_t(\omega_4) \]
\[ \Omega = \omega_4 \]
Interesting Quantities

\[ X_t(\omega_1) \]
\[ \bar{X}^{(m)}(\omega_1) = \frac{1}{m} \sum_{t=1}^{m} X_t(\omega_1) \]  
(time average based on \( \Omega = \omega_1 \))

\[ X_t(\omega_2) \]
\[ \bar{X}^{(m)}(\omega_2) = \frac{1}{m} \sum_{t=1}^{m} X_t(\omega_2) \]  
(time average based on \( \Omega = \omega_2 \))

\[ X_t(\omega_3) \]
\[ \bar{X}^{(m)}(\omega_3) = \frac{1}{m} \sum_{t=1}^{m} X_t(\omega_3) \]  
(time average based on \( \Omega = \omega_3 \))

\[ X_t(\omega_4) \]
\[ \bar{X}^{(m)}(\omega_4) = \frac{1}{m} \sum_{t=1}^{m} X_t(\omega_4) \]  
(time average based on \( \Omega = \omega_4 \))

\[ E[X_T(\omega)] = \sum_\omega X_T(\omega) \rho_{\Omega}(\omega) \]  
(ensemble average)
Ensemble Averages

Given some stochastic process \( \{ X_t : t \in T \} \), here are some ensemble averages that we might be interested in obtaining:

- \( \text{E}[X_t] \) for any \( t \),
- \( \text{Var}(X_t) = \text{E}\left[ (X_t - \text{E}[X_t])^2 \right] \) for any \( t \),
- \( \text{Cov}(X_t, X_{t+\Delta t}) = \text{E}\left[ (X_t - \text{E}[X_t]) \cdot (X_{t+\Delta t} - \text{E}[X_{t+\Delta t}]) \right] \) for any \( t \) and \( \Delta t \),
- \( \ldots \)

Recall the definitions of these expressions. E.g., \( \text{E}[X_t] \) and \( \text{Var}(X_t) \) are given by

\[
\text{E}[X_t] = \sum_{x_t} x_t \cdot p_{X_t}(x_t) = \sum_{\omega} X_t(\omega) \cdot p_{\Omega}(\omega),
\]
\[
\text{Var}[X_t] = \sum_{x_t} (x_t - \text{E}[X_t])^2 \cdot p_{X_t}(x_t) = \sum_{\omega} (X_t(\omega) - \text{E}[X_t])^2 \cdot p_{\Omega}(\omega).
\]

(Here we have assumed that \( \Omega \) is a countable set. If \( \Omega \) is not a countable set, the notation has to be suitably adapted.)
Example 1: I.I.D. Process

Let $X$ be some random variable. Define the stochastic process

$$\{X_t : t \in \{0, 1, 2, 3, \ldots \}\},$$

where

the $X_t$'s are i.i.d. random variables with the same distribution as $X$.

We have

• $E[X_t] = E[X]$ for any $t$,
• $\text{Var}(X_t) = \text{Var}(X)$ for any $t$,
• $\text{Cov}(X_t, X_{t+\Delta t}) = \text{Var}(X)$ for any $t$ and $\Delta t = 0$,
• $\text{Cov}(X_t, X_{t+\Delta t}) = 0$ for any $t$ and $\Delta t \neq 0$,
• $\ldots$
Example 2: “Fixed-Value” Process

Let $X$ be some random variable. Define the stochastic process

$$\{X_t : t \in \{0, 1, 2, 3, \ldots \}\},$$

where

$$X_t = X, \quad t \in \{0, 1, 2, 3, \ldots \}.$$  

We have

- $E[X_t] = E[X]$ for any $t$,
- $\text{Var}(X_t) = \text{Var}(X)$ for any $t$,
- $\text{Cov}(X_t, X_{t+\Delta t}) = \text{Var}(X)$ for any $t$ and $\Delta t$.
- ...
Example 3: Autoregressive Process

Define the stochastic process

$$X_t = a \cdot X_{t-1} + b \cdot Z_t, \quad t \in \{1, 2, 3, \ldots \},$$

where

- $X_0$ can be either fixed to some value or drawn according to some distribution,
- $Z_t$ is drawn according to some (time-indep. or time-dep.) distribution,
- for all $t \in \{1, 2, 3, \ldots \}$, the random variable $Z_t$ is independent of $X_0, X_1, \ldots, X_{t-1}$ and of $Z_1, Z_2, \ldots, Z_{t-1}$.

We have

- $E[X_t] = a \cdot E[X_{t-1}] + b \cdot E[Z_t], \quad t \in \{1, 2, 3, \ldots \},$
- $\text{Var}(X_t) = a^2 \cdot \text{Var}(X_{t-1}) + b^2 \cdot \text{Var}(Z_t), \quad t \in \{1, 2, 3, \ldots \},$
- $\text{Cov}(X_t, X_{t+\Delta t}) = a^{\Delta t} \cdot \text{Var}(X_t)$ for any $t$ and any $\Delta t \in \{0, 1, 2, 3, \ldots \}$,
- ...
Estimating Ensemble Averages

Fix some $t \in T$. Define $\mu_{X_t} \triangleq \mathbb{E}[X_t]$.

For estimating a quantity like $\mu_{X_t}$, we can use the following procedure:

- Randomly generate $n$ independent samples $\omega_1, \omega_2, \ldots, \omega_n$.
- Compute
  \[
  \hat{\mu}_{X_t}(n) \triangleq \frac{1}{n} \cdot \sum_{i=1}^{n} X_t(\omega_i).
  \]
- Compute
  \[
  s^2_{X_t}(n) \triangleq \frac{1}{n-1} \cdot \sum_{i=1}^{n} \left( X_t(\omega_i) - \hat{\mu}_{X_t}(n) \right)^2.
  \]
- Based on $\hat{\mu}_{X_t}(n)$ and $s^2_{X_t}(n)$, compute a confidence interval for $\mu_{X_t}$ with confidence level $1 - \alpha$. (See earlier slides on how to do that.)
Estimating Time Averages

Assume $\mathcal{T} = \{1, 2, 3, \ldots \}$.

Given some stochastic process $\{X_t : t \in \mathcal{T}\}$ and some $\omega$, here are some time averages that we are interested in obtaining:

- $\overline{X}(m)(\omega) \triangleq \frac{1}{m} \cdot \sum_{t=1}^{m} X_t(\omega),$
- $\overline{X^2}(m)(\omega) \triangleq \frac{1}{m} \cdot \sum_{t=1}^{m} X_t^2(\omega),$
- $\ldots$

Note that for given a $\omega$, the quantities $\overline{X}(m)(\omega), \overline{X^2}(m)(\omega), \ldots$ are constants.
Time Averages

Question:
Can we use time averages based on a single sample path to estimate ensemble averages?

Answer:
No, not in general!!!
Example 1: I.I.D. Process

Let $X$ be some random variable. Define the stochastic process

$$\{X_t : t \in \{0, 1, 2, 3, \ldots \}\},$$

where

the $X_t$'s are i.i.d. random variables with the same distribution as $X$.

Using the strong law of large numbers, we have for a randomly chosen $\omega$ that

$$\lim_{m \to \infty} \frac{1}{m} \cdot \sum_{t' = 1}^{m} X_{t'}(\omega) = E[X] \quad \text{with probability 1.}$$

Combining this with the earlier result stating that $E[X_t] = E[X]$ for any $t$, we obtain for any $t$ and a randomly chosen $\omega$ that

$$E[X_t] = \lim_{m \to \infty} \frac{1}{m} \cdot \sum_{t' = 1}^{m} X_{t'}(\omega) \quad \text{with probability 1.}$$

Conclusion: For this process, we can use the time average of a single sample path to estimate the ensemble average $E[X_t]$ for any $t$. 
Example 2: “Fixed-Value” Process

Let $X$ be some random variable. Define the stochastic process

$$\left\{ X_t : t \in \{1, 2, 3, \ldots \} \right\},$$

where

$$X_t = X, \quad t \in \{1, 2, 3, \ldots \}.$$

For any $\omega$ and any finite integer $m$ we have

$$\lim_{m \to \infty} \frac{1}{m} \cdot \sum_{t'=1}^{m} X_{t'}(\omega) = \frac{1}{m} \cdot (X(\omega) + \cdots + X(\omega)) = X(\omega).$$

Recall the earlier result $E[X_t] = E[X]$ for any $t$. Therefore, unless $X$ takes a constant value with probability 1, we have for any $t$ and a randomly chosen $\omega$ that

$$E[X_t] \neq \lim_{m \to \infty} \frac{1}{m} \cdot \sum_{t'=1}^{m} X_{t'}(\omega)$$

with probability strictly larger than 0.

Conclusion: For this process, despite $E[X_t]$ being independent of $t$, we cannot use the time average of a single sample path to estimate the ensemble average $E[X_t]$. 
Example 3: Autoregressive Process

Define the stochastic process

$$X_t = a \cdot X_{t-1} + b \cdot Z_t, \quad t \in \{1, 2, 3, \ldots \},$$

where

- $X_0$ can be either fixed to some value or drawn according to some distribution,
- $Z_t$ is drawn according to some (time-indep. or time-dep.) distribution,
- for all $t \in \{1, 2, 3, \ldots \}$, the random variable $Z_t$ is independent of $X_0, X_1, \ldots, X_{t-1}$ and of $Z_1, Z_2, \ldots, Z_{t-1}$.

Except for special cases of the above autoregressive process, $E[X_t]$ will depend on $t$ and so estimating $E[X_t]$ for any $t$ based on

$$\lim_{m \to \infty} \frac{1}{m} \cdot \sum_{t'=1}^{m} X_{t'}(\omega)$$

is rather meaningless.

**Conclusion:** For this process, we **cannot** use the time average of a single sample path to estimate the ensemble average $E[X_t]$. 
Time Averages

Question:

Can we use time averages based on a single sample path to estimate ensemble averages?

Answer:

No, not in general!!

However: if a process is stationary and ergodic then
time averages based on a single sample path can be used to estimate ensemble averages.

Example 1: stationary and ergodic.
Example 2: stationary, but not ergodic.
Example 3: neither stationary nor ergodic (except for special cases).
Transient state and steady-state behavior
Example 3: Autoregressive Process

Define the stochastic process

\[ X_t = a \cdot X_{t-1} + b \cdot Z_t, \quad t \in \{1, 2, 3, \ldots \}, \]

where

- \( X_0 \) can be either fixed to some value or drawn according to some distribution,
- \( Z_t \) is drawn according to some (time-indep. or time-dep.) distribution,
- for all \( t \in \{1, 2, 3, \ldots \} \), the random variable \( Z_t \) is independent of \( X_0, X_1, \ldots, X_{t-1} \) and of \( Z_1, Z_2, \ldots, Z_{t-1} \).

Consider the following setup:

- \( X_0 = 0; \)
- \( Z_t \sim \mathcal{N}(\mu_Z, \sigma_Z^2) \), \( t \in \{1, 2, 3, \ldots \} \), for some fixed \( \mu_Z \) and \( \sigma_Z^2; \)
- \( 0 \leq a < 1 \) and \( b = 1 - a. \)
Example 3: Autoregressive Process

Define the stochastic process

\[ X_t = a \cdot X_{t-1} + b \cdot Z_t, \quad t \in \{1, 2, 3, \ldots \}, \]

where

- \( X_0 \) can be either fixed to some value or drawn according to some distribution,
- \( Z_t \) is drawn according to some (time-indep. or time-dep.) distribution,
- for all \( t \in \{1, 2, 3, \ldots \} \), the random variable \( Z_t \) is independent of \( X_0, X_1, \ldots, X_{t-1} \) and of \( Z_1, Z_2, \ldots, Z_{t-1} \).

It follows that

\[ E(X_t) = a \cdot E(X_{t-1}) + b \cdot E(Z_t), \quad t \in \{1, 2, 3, \ldots \}, \]

\[ \text{Var}(X_t) = a^2 \cdot \text{Var}(X_{t-1}) + b^2 \cdot \text{Var}(Z_t), \quad t \in \{1, 2, 3, \ldots \}. \]

Plugging in the given values, we get

\[ E(X_t) = a \cdot E(X_{t-1}) + (1 - a) \cdot \mu_Z, \quad t \in \{1, 2, 3, \ldots \}, \]

\[ \text{Var}(X_t) = a^2 \cdot \text{Var}(X_{t-1}) + (1 - a)^2 \cdot \sigma_Z^2, \quad t \in \{1, 2, 3, \ldots \}. \]
Example 3: Autoregressive Process

Solving these iterative expressions, we obtain
\[ E[X_t] = \mu_Z \cdot (1 - a^t), \quad t \in \{0, 1, 2, 3, \ldots \}, \]
\[ \text{Var}(X_t) = \frac{1 - a}{1 + a} \cdot \sigma_Z^2 \cdot (1 - a^{2t}), \quad t \in \{0, 1, 2, 3, \ldots \}. \]

Plots: \( a = 0.50 \) (black), \( a = 0.75 \) (blue), \( a = 0.95 \) (magenta), \( a = 0.99 \) (green); \( \mu_Z = 10, \sigma_Z^2 = 1 \).
Example 3: Autoregressive Process

Solving these iterative expressions, we obtain

\[ E[X_t] = \mu_Z \cdot (1 - a^t), \quad t \in \{0, 1, 2, 3, \ldots \} \]

\[ \text{Var}(X_t) = \frac{1 - a}{1 + a} \cdot \sigma_Z^2 \cdot (1 - a^{2t}), \quad t \in \{0, 1, 2, 3, \ldots \} . \]

In fact, for the given setup one can show that \( X_t \sim \mathcal{N}(E[X_t], \text{Var}(X_t)) \) for all \( t \).

**Initial transient**: for \( t \lesssim \frac{5}{\log(1/a)} \), the distribution of \( X_t \) is changing considerably from time index to time index.

**(Approximate) steady state**: for \( t \gtrsim \frac{5}{\log(1/a)} \), the distribution of \( X_t \) is hardly changing from time index to time index. We obtain the approximations

\[ E[X_t] \approx E[X_\infty] = \mu_{X_\infty} = \mu_Z , \]

\[ \text{Var}(X_t) \approx \text{Var}(X_\infty) = \sigma_{X_\infty}^2 = \frac{1 - a}{1 + a} \cdot \sigma_Z^2 , \]

\[ f_{X_t}(x_t) \approx f_{X_\infty}(x_t) = \frac{1}{\sqrt{2\pi \cdot \sigma_{X_\infty}^2}} \cdot \exp \left( -\frac{(x_t - \mu_{X_\infty})^2}{2 \cdot \sigma_{X_\infty}^2} \right) . \]
Example 3: Autoregressive Process

Because \( \frac{5}{\log(1/a)} \) is finite, we have the following results.

For \( t \gtrsim \frac{5}{\log(1/a)} \), i.e., \( t \) in the (approximate) steady state, we have for a randomly chosen \( \omega \) that

\[
E[X_t] \approx \lim_{m \to \infty} \frac{1}{m} \cdot \sum_{t'=1}^{m} X_{t'}(\omega) \quad \text{with probability 1.}
\]

Note that we were able to formulate such a result despite the fact that the process \( \{X_t\} \) is neither stationary nor ergodic.

In order to improve the convergence speed of \( \frac{1}{m} \cdot \sum_{t'=1}^{m} X_{t'}(\omega) \) as \( m \to \infty \), we can omit the first \( k \) terms. With this, we have for a randomly chosen \( \omega \) that

\[
E[X_t] \approx \lim_{m \to \infty} \frac{1}{m-k} \cdot \sum_{t'=k+1}^{m} X_{t'}(\omega) \quad \text{with probability 1.}
\]

(Here, \( k \) can be chosen to be roughly the length of the initial transient.)
Transient vs. Steady State

(Figure from book by Averill M. Law.)

Some points to notice:

- Even though the distribution of the $Y_i$’s after time $k$ is not appreciably changing, observations on the $Y_i$’s could still have large variance.
- Even in steady state, the $Y_i$’s are generally not independent, and could be heavily (auto)correlated.
- The nature and rate of convergence of the transient distributions can depend heavily on the initial conditions.
Method of Replication

Assume that we want to estimate some quantity like $\mu_Y \triangleq E[Y]$.

In the system simulation literature, the technique typically used to estimate an ensemble average like $\mu_Y$ goes also under the name of “method of replication.”

Wrong method:

estimate the mean $\mu_Y$ directly from a single simulation run.

Correct method:

make $n$ independent replications of the whole simulation and estimate $\mu_Y$ based on the results of these $n$ simulation runs.

(More details on the next slide.)

Note:

- One simulation run is one sample.
- $n$ simulation runs are $n$ samples.
Method of Replication

Assume that we want to estimate some quantity like \( \mu_Y \triangleq \mathbb{E}[Y] \).

For estimating a quantity like \( \mu_Y \), we can use the following procedure:

- Randomly generate \( n \) independent samples \( \omega_1, \omega_2, \ldots, \omega_n \).
- Compute
  \[
  \hat{\mu}_Y(n) \triangleq \frac{1}{n} \cdot \sum_{i=1}^{n} Y(\omega_i).
  \]
- Compute
  \[
  s^2_Y(n) \triangleq \frac{1}{n-1} \cdot \sum_{i=1}^{n} (Y(\omega_i) - \hat{\mu}_Y(n))^2.
  \]
- Based on \( \hat{\mu}_Y(n) \) and \( s^2_Y(n) \), compute a confidence interval for \( \mu_Y \) with confidence level \( 1 - \alpha \). (See earlier slides on how to do that.)
Method of Replication

Example:

- Assume that $X_t$ is the waiting time in the queue for the $t$-th customer.
- Let $\bar{X}^{(m)} \triangleq \frac{1}{m} \cdot \sum_{t=1}^{m} X_t$ be the average waiting time of the first $m$ customers.
- We would like to estimate $\mu_{\bar{X}^{(m)}} = E[\bar{X}^{(m)}] = E \left[ \frac{1}{m} \cdot \sum_{t=1}^{m} X_t \right]$.

For estimating a quantity like $\mu_{\bar{X}^{(m)}}$, we can use the following procedure:

- Randomly generate $n$ independent samples $\omega_1, \omega_2, \ldots, \omega_n$.
- Compute $\bar{X}^{(m)}(\omega_i) \triangleq \frac{1}{m} \cdot \sum_{t=1}^{m} X_t(\omega_i), \quad i = 1, 2, \ldots, n$.
- Compute $\hat{\mu}_{\bar{X}^{(m)}}(n) \triangleq \frac{1}{n} \cdot \sum_{i=1}^{n} \bar{X}^{(m)}(\omega_i)$.
- Compute $s^2_{\bar{X}^{(m)}}(n) \triangleq \frac{1}{n-1} \cdot \sum_{i=1}^{n} (\bar{X}^{(m)}(\omega_i) - \hat{\mu}_{\bar{X}^{(m)}}(n))^2$.
- Based on $\hat{\mu}_{\bar{X}^{(m)}}(n)$ and $s^2_{\bar{X}^{(m)}}(n)$, compute a confidence interval for $\mu_{\bar{X}^{(m)}}$ with confidence level $1 - \alpha$. (See earlier slides on how to do that.)
Method of Replication

**Numerical Example:**

- \( n = 10 \).
- \( \bar{X}^{(m)}(\omega_i) \) values (\( n \) samples):
  
  \[
  2.02, 0.73, 3.20, 6.23, 1.76, 0.47, 3.89, 5.45, 1.44, 1.23.
  \]
- Want 90% confidence interval, i.e., \( \alpha = 0.10 \).
- \( \hat{\mu}_{\bar{X}^{(m)}}(n) = 2.64 \).
- \( s^2_{\bar{X}^{(m)}}(n) = 3.96 \).
- \( t_{10-1,0.95} = 1.833 \) (based on Student’s \( t \)-distribution).

With this, the approximate 90% confidence interval is

\[
2.64 \pm 1.15, \text{ or } [1.49, 3.79].
\]
Transient Removal

In many cases we are only interested in the steady-state system performance.

Examples:

- We hope to estimate the average waiting time in an M/M/1 queue when the system is in a steady state.
- A company that is going to build a new manufacturing system would like to find out the long-run (steady-state) hourly throughput of the system.

However, the performance is biased by the initial conditions.

Therefore, we hope to remove the biasing effect due to the initial conditions.
Some Transient Removal Methods

**Long simulation run:** use very long runs so that the effect of the transient state becomes very small.

⇒ Wastes resources.
⇒ Cannot ensure whether the runs are long enough.

**Proper initialization:** use an initial setting that is similar to the steady state.
⇒ No guarantee that the selected initial condition is representative.

**Truncation / initial data deletion:**

- Warm up the model.
- Most widely used method for transient removal.
- Given observations $X_1(\omega), X_2(\omega), \ldots, X_m(\omega)$, for some suitably chosen $k$ discard the first $k$ observations when computing averages, i.e.,

$$\text{compute } \frac{1}{m-k} \sum_{t'=k+1}^{m} X_{t'}(\omega) \text{ instead of } \frac{1}{m} \sum_{t'=1}^{m} X_{t'}(\omega).$$
Terminating and non-terminating simulations
Terminating and Non-Terminating Simulations

**Terminating simulation:**
- Parameters to be estimated are defined relative to specific initial and stopping conditions that are part of the model.
- There is a “natural” and realistic way to model both the initial and stopping conditions.
- Output performance measures generally depend on both the initial and stopping conditions.

**Nonterminating simulation:**
- There is no natural and realistic event that terminates the model.
- Interested in “long-run” behavior characteristic of “normal” operation.
- If the performance measure of interest is a characteristic of a steady-state distribution of the process, it is a steady-state parameter of the model.
- Must ensure that the run is long enough so that initial-condition effects have dissipated.
Terminating and Non-Terminating Simulations

**Note:** Not all non-terminating systems are steady-state: there could be a periodic “cycle” in the long run, giving rise to steady-state cycle parameters.

Examples: linear congruential generator, a bank operating at 9am to 5pm, etc.

---

**Examples of terminating vs. steady-state simulations:**

<table>
<thead>
<tr>
<th>physical model</th>
<th>terminating estimand</th>
<th>steady-state estimand</th>
</tr>
</thead>
<tbody>
<tr>
<td>single-server queue</td>
<td>expected average delay in queue of the first 25 customers, given empty-and-idle conditions</td>
<td>long-run expected queueing delay of a customer</td>
</tr>
<tr>
<td>manufacturing system</td>
<td>expected daily production, given some number of workpieces in process initially</td>
<td>expected long-run daily production</td>
</tr>
</tbody>
</table>
Obtaining a specified precision
Obtaining a Specified Precision

Throughout this slide, fix some confidence level.

Thanks to the central limit theorem, the confidence interval for some quantity of interest becomes smaller as \( n \) grows.

**Problem:** If \( n \) is chosen too small, the confidence interval might be too large.

\[ \Rightarrow \text{We need to choose } n \text{ large enough so that the confidence interval is not larger than some pre-specified length.} \]

Side note: a possibility to specify a confidence interval length is to require that the ratio \( \frac{\text{confidence interval length}}{\text{midpoint value}} \) is smaller than some value.

**Sequential procedure** to get a suitable confidence interval:

1. Make \( n \) replications of the simulation for some small \( n \).
2. Compute the confidence interval for the quantity of interest.
3. If the confidence interval meets the requirements, stop.
4. Otherwise, increase \( n \) by 1, make another replication of the simulation, and go to Step 2.
Choosing initial conditions
Choosing Initial Conditions

For terminating simulations, the initial conditions can affect the output performance measure, so the simulations should be initialized appropriately.

Example:
- Want to estimate the expected average delay in a queue of bank customers who arrive and complete their delay between noon and 1:00pm.
- Bank is likely to be crowded already at noon, so starting empty and idle at noon will probably bias the results low.

Two possible remedies:
- If bank actually opens at 9:00am, start the simulation empty and idle, let it run for 3 simulated hours, clear the statistical accumulators, and observe statistics for the next simulated hour.
- Take data in the field on number of customers present at noon, fit a (discrete) distribution to it, and draw from this distribution to initialize the simulation at time $t_0 = \text{noon}$. Draw independently from this distribution to initialize multiple replications.

Note: This could be difficult in simulation software, depending on the modeling constructs available.
Variance reduction techniques
Variance-Reduction Techniques

Main drawback of using simulation to study stochastic models:

- Results are uncertain — have variance associated with them.
- We would like to have as little variance as possible — so as to obtain a more precise results.

One sure way to decrease the variance:

- Run it some more (longer runs, additional repl.). However, this is not for free.

Sometimes we can “manipulate” the simulation to reduce the variance of the output at no additional cost — not just by having longer runs and/or additional replications.

⇒ Variance-reduction techniques:
  - Common random numbers
  - Antithetic variates
  - Control variates
  - Indirect estimation
  - Conditioning
Common Random Numbers

Common random numbers:

• Other names: correlated sampling, matched streams, matched pairs.
• Applies when goal is to compare two (or more) alternative systems.
• Main idea: use the same random number stream as the input to the two systems.

Basic relation used:

• Let $V$ and $W$ be some random variables.
• Let $a$ and $b$ be some real constants.
• It holds that

$$\text{Var}(aV + bW) = a^2 \cdot \text{Var}(V) + b^2 \cdot \text{Var}(W) + 2ab \cdot \text{Cov}(V, W).$$
Common Random Numbers

**Method A**

Simulation run $i$

Interarrival times

$u_{t,i}^{(1)}, \ldots, u_{t,i}^{(k)}$

System 1

$\bar{y}_i^{(1)}$

$z_i = \bar{y}_i^{(1)} - \bar{y}_i^{(2)}$

System 2

$\bar{y}_i^{(2)}$

**Method B**

Simulation run $i$

Interarrival times

$u_{1,i}, \ldots, u_{k,i}$

System 1

$\bar{y}_i^{(1)}$

$z_i = \bar{y}_i^{(1)} - \bar{y}_i^{(2)}$

System 2

$\bar{y}_i^{(2)}$

**Note:** $u_{t,i}^{(1)}$ and $u_{t,i}^{(2)}$ should be used for the same purpose in System 1 and System 2, respectively.
Common Random Numbers

**Method A:** feed different random numbers to Systems 1 and 2.

**Method B:** feed the same random numbers to Systems 1 and 2.

**Method B is better. Why?**

- If you feed different data to two systems, then the discrepancy between the outputs may be due to
  1. differences between the systems,
  2. differences between the input data.
- If you feed the same input data to two systems, you eliminate uncertainty due to factor (2).
  \[\Rightarrow\] Differences can only be due to the difference in systems.
Common Random Numbers

For \( i = 1, \ldots, n \), let

\[
Z_i \triangleq \bar{Y}_i^{(1)} - \bar{Y}_i^{(2)}.
\]

For the expectation value of \( Z_i \) we obtain:

\[
E(Z_i) = E(\bar{Y}_i^{(1)} - \bar{Y}_i^{(2)})
\]

\[
= E(\bar{Y}_i^{(1)}) - E(\bar{Y}_i^{(2)}).
\]

For the variance of \( Z_i \) we obtain:

\[
\text{Var}(Z_i) = \text{Var}(\bar{Y}_i^{(1)} - \bar{Y}_i^{(2)})
\]

\[
= \text{Var}(\bar{Y}_i^{(1)}) + \text{Var}(\bar{Y}_i^{(2)}) - 2 \cdot \text{Cov}(\bar{Y}_i^{(1)}, \bar{Y}_i^{(2)}).
\]

Assume \( \text{Cov}(\bar{Y}_i^{(1)}, \bar{Y}_i^{(2)}) > 0 \).

Usually

\[
\text{Cov}(\bar{Y}_i^{(1)}, \bar{Y}_i^{(2)}) \text{ under Method A} < \text{Cov}(\bar{Y}_i^{(1)}, \bar{Y}_i^{(2)}) \text{ under Method B}
\]

and so

\[
\text{Var}(Z_i) \text{ under Method A} > \text{Var}(Z_i) \text{ under Method B}.
\]
Antithetic variates
Antithetic Variates

Example:

- Consider a random variable $X$ with mean $\mu_X$ and variance $\sigma_X^2$.
- Assume that $X = g(U)$ for some function $g$ and $U \sim U(0, 1)$.
- Assume we want to estimate $\mu_X$ based on sampling $X$.

First estimator:

- Generate i.i.d. random numbers $U_1, U_2 \sim U(0, 1)$.
- Compute random variates $X_1 = g(U_1)$ and $X_2 = g(U_2)$.
- Compute the estimate $\hat{\mu}_X \triangleq \frac{1}{2} \cdot (X_1 + X_2)$.

This estimator has expected value $E[\hat{\mu}_X] = \mu_X$, i.e., it is a bias-free estimator.

Moreover, because $X_1$ and $X_2$ are independent, its variance is $\text{Var}(\hat{\mu}_X) = \frac{1}{2} \sigma_X^2$. 
Antithetic Variates

Example:

- Consider a random variable $X$ with mean $\mu_X$ and variance $\sigma_X^2$.
- Assume that $X = g(U)$ for some function $g$ and $U \sim U(0, 1)$.
- Assume we want to estimate $\mu_X$ based on sampling $X$.

Second estimator:

- Generate random number $U \sim U(0, 1)$.
- Compute random variates $X_1 = g(U)$ and $X_2 = g(1 - U)$.
- Compute the estimate $\hat{\mu}_X \triangleq \frac{1}{2} \cdot (X_1 + X_2)$.

This estimator has expected value $E[\hat{\mu}_X] = \mu_X$, i.e., it is a bias-free estimator.

Moreover, its variance is

$$\text{Var}(\hat{\mu}_X) = \text{Var} \left( \frac{1}{2} \cdot (X_1 + X_2) \right) = \frac{1}{4} \cdot \text{Var}(X_1) + \frac{1}{4} \cdot \text{Var}(X_2) + \frac{1}{2} \cdot \text{Cov}(X_1, X_2).$$

For many setups we have $\text{Cov}(X_1, X_2) < 0$. For these setups, $\text{Var}(\hat{\mu}_X) < \frac{1}{2} \sigma_X^2$. 
Antithetic Variates

The above idea can be used as follows when simulating systems.

**Note:** can be used for analyzing a single system. Cannot be used for comparisons.

**Implementation:** For $i = 1, \ldots, n$ do

- If $i$ is odd: generate i.i.d. random numbers $U_1, U_2, U_3, \ldots \sim U(0,1)$ as input to the simulation (for generating other random numbers, etc.) Simulate the system.

- If $i$ is even: use $1 - U_1$, $1 - U_2$, $1 - U_3$ \ldots as input to the simulation (for generating other random numbers, etc.), where $U_1, U_2, U_3, \ldots$ are from the previous simulation run. Simulate the system.

**Note:**

- **Must re-use the same random numbers for the same purposes** in two adjacent simulation runs of the system. I.e., if $U_\ell$ was used for some purpose in the $i$-th simulation run (where $i$ is odd), then $1 - U_\ell$ must be used for the same purpose in the $(i+1)$-st simulation run.

- **Failure to maintain this synchronization** of random-number usage can get things mixed up and dilute the effect of antithetic variates, or even make it backfire (increase the variance).
Control Variates
Control Variates

Setup:
- $X$ is some random variable with mean $\mu_X$ and variance $\sigma_X^2$.
- $Y$ is some random variable with mean $\mu_Y$ and variance $\sigma_Y^2$.
- Suppose that we know $\mu_Y$ and that we want to estimate $\mu_X$.
- We assume that $X$ and $Y$ are (pos. or neg.) correlated, i.e., $\text{Corr}(X, Y) \neq 0$.

Control-Variate-Based Estimator:
- Generate random variates $X$ and $Y$.
- Fix some $a \in \mathbb{R}$ and compute the estimate $\hat{\mu}_{X,cv(a,Y)} \triangleq X - a \cdot (Y - \mu_Y)$.

Properties of the Control-Variate-Based Estimator:
- $\hat{\mu}_{X,cv(a,Y)}$ is bias free.
  **Proof:** $E[\hat{\mu}_{X,cv(a,Y)}] = E[X - a \cdot (Y - \mu_Y)] = \mu_X - a \cdot (\mu_Y - \mu_Y) = \mu_X$.
- The parameter $a$ can be chosen such that $\text{Var}(\hat{\mu}_{X,cv(a,Y)}) < \sigma_X^2$.
  (For details, see the next slide.)
Control Variates

The variance of $\text{Var}(\hat{\mu}_{X,cv(a,Y)})$ is given by

$$\text{Var}(\hat{\mu}_{X,cv(a,Y)}) = \text{Var}(X - a \cdot (Y - \mu_Y))$$

$$= \text{Var}(X) + a^2 \cdot \text{Var}(Y) - 2a \cdot \text{Cov}(X, Y).$$

This expression is minimized by the choice

$$a^* \triangleq \frac{\text{Cov}(X, Y)}{\text{Var}(Y)}.$$

For this special choice of $a$, the variance of $\hat{\mu}_{X,cv(a,Y)}$ becomes

$$\text{Var}(\hat{\mu}_{X,cv(a^*)}) = \left(1 - (\text{Corr}(X, Y))^2\right) \cdot \sigma_X^2,$$

i.e., $\text{Var}(\hat{\mu}_{X,cv(a^*)}) < \sigma_X^2$ as promised.
Control Variates

Comments:

• In order to determine \( a \), we need to know the ratio \( \frac{\text{Cov}(X,Y)}{\text{Var}(Y)} \), or at least have a reasonable estimate of it.

(For details on how to estimate this ratio, see Section 11.4 in the book by Averill M. Law.)

• The above approach can easily be extended to the case when multiple samples of \((X, Y)\) are available.

• In the above setup, \( Y \) is called the “control variate.” Clearly, the above approach can easily be extended to multiple control variates.
Indirect Estimation
Indirection Estimation

Usually we want to estimate several performance measures.
Examples: expected time in queue(s) of parts, average queue length(s), ... 
For some classes of models, there are relationships among these measures.

Example: consider a queueing model with the following performance measures:

- $\lambda$: arrival rate
- $\ell$: expected time-average number in the system
- $q$: expected time-average number in the queue
- $w$: expected time a customer spends in the system
- $d$: expected time a customer spends in the queue
- $S$: random variable for service time

The following relationships hold for this queueing system:

\[
\ell = \lambda \cdot w, \\
q = \lambda \cdot d, \\
w = d + \mathbb{E}[S].
\]
Indirection Estimation

**Example (continued):** assume we want to estimate $w$.

- **Directly:** Collect times in system, average them.
- **Indirectly:** Estimate $d$, then add $E[S]$ (which would be known).

The *indirect method is better* since the direct method essentially estimates the known value of $E[S]$.

---

**Moral:** Do not estimate things that you know.

---

A more careful discussion is given in Ch. 11.5 of the book by A. M. Law.

(The analysis is not trivial because the waiting time of a customer in the queue and his/her service time are not independent random variables.)
Conditioning

Use knowledge of “parts” of the output, rather than estimating them (similar in spirit to indirect estimation).

Consider the following setup:

- Assume that we want to estimate \( \mu_X = E[X] \) for some random variable \( X \).
- Suppose there is another random variable \( Z \) such that if we knew the value of \( Z \), then we would know the expected value of \( X \) for sure, i.e., the conditional expectation \( E[X \mid Z = z] \) is known.

Recall that \( \mu_X = E[X] = E_Z \left[ E[X \mid Z] \right] \), where the latter stands for \( \int_{-\infty}^{+\infty} E[X \mid Z = z] \cdot f_Z(z) \, dz \).

Therefore, \( \mu_X \) can be estimated by the following procedure:

- Let \( z_1, \ldots, z_n \) be i.i.d. samples distributed according to \( F_Z \).
- Compute the estimate \( \hat{\mu}_X = \frac{1}{n} \cdot \sum_{i=1}^{n} E[X \mid Z = z_i] \).

Clearly, \( E[\hat{\mu}_X] = E_Z \left[ E[X \mid Z] \right] = E[X] = \mu_X \), i.e., \( \hat{\mu}_X \) is a bias-free estimate of \( \mu_X \).
Conditioning

Recall the following result (also known as "law of total variance"):

\[
\text{Var}(X) = \mathbb{E}_Z [\text{Var}(X \mid Z)] + \text{Var}_Z (\mathbb{E}[X \mid Z]),
\]

where (here for a continuous random variable \( Z \)):

\[
\mathbb{E}_Z [\text{Var}(X \mid Z)] = \int_{-\infty}^{+\infty} \text{Var}(X \mid Z = z) \cdot f_Z(z) \, dz,
\]

\[
\text{Var}_Z (\mathbb{E}[X \mid Z]) = \int_{-\infty}^{+\infty} \left( \mathbb{E}[X \mid Z = z] - \mathbb{E}[X] \right)^2 \cdot f_Z(z) \, dz.
\]

This result can be used to show the following variance advantage of the above estimator of \( \mu_X \). Namely,

\[
\text{Var}(\hat{\mu}_X) = \frac{1}{n} \cdot \text{Var}_Z (\mathbb{E}[X \mid Z]) = \frac{1}{n} \cdot \left( \text{Var}(X) - \mathbb{E}_Z [\text{Var}(X \mid Z)] \right)
\]

\[
\leq \frac{1}{n} \cdot \text{Var}(X).
\]
Conditioning

**Example:** Single-server queueing system.

- Let $X$ be the delay in queue of an arriving customer.
- Service times are exponential (and therefore memoryless) with mean $\text{E}[S]$.
- Let $Z$ be the number of customers already in the queue when customer arrives.
- Then $\text{E}[X \mid Z = z] = (z + 1) \cdot \text{E}[S]$.

**Trick:** Specify the “right” $Z$. — **Requirements for $Z$:**

- Ease of generating samples with distribution $F_Z$ (still have to simulate it).
- Ease of computing $\text{E}[X \mid Z = z]$ for any $z$.

**Issues:**

- Clearly need to understand the model, exploit its special properties.
- Knowing $\text{E}[X \mid Z = z]$ is a very strong assumption. Note:
  - If $X$ and $Z$ are independent then no benefit because estimating $\text{E}[X \mid Z = z]$ is the same as estimating $\text{E}[X]$.
  - If $X = Z$ then no benefit because sampling from $F_Z$ is like sampling from $F_X$.  

Thank you!

• Homework 3 is due tomorrow
• Homework 4 will be out by the end of today. No need to hand in, but very relevant to Quiz 2
• Quiz 2 will be on Nov. 6
• No lectures next week (Time for your course project)
• End of Oct. 2019: a 4-page check-point report for course project