Ch4. Variance Reduction Techniques

Zhang Jin-Ting

Department of Statistics and Applied Probability

July 17, 2012
Outline

Introduction

The Basic Problem

Variance Reduction Techniques
This chapter aims to improve the Monte Carlo Integration estimator via reducing its variance using some useful techniques.

- Stratified Sampling
- Importance Sampling
- Control Variates Method
- Antithetic Variates Method
This chapter aims to improve the Monte Carlo Integration estimator via reducing its variance using some useful techniques.

- Stratified Sampling
- Importance Sampling
- Control Variates Method
- Antithetic Variates Method
This chapter aims to improve the Monte Carlo Integration estimator via reducing its variance using some useful techniques.

- Stratified Sampling
- Importance Sampling
- Control Variates Method
- Antithetic Variates Method
This chapter aims to improve the Monte Carlo Integration estimator via reducing its variance using some useful techniques.

- Stratified Sampling
- Importance Sampling
- Control Variates Method
- Antithetic Variates Method
This chapter aims to improve the Monte Carlo Integration estimator via reducing its variance using some useful techniques.

- Stratified Sampling
- Importance Sampling
- Control Variates Method
- Antithetic Variates Method
The Integration Problem

- Suppose we want to estimate an integral over some region, such as

\[ I_A = \int_S k(x) \, dx \]

where \( S \) is subset of \( \mathbb{R}^d \), \( x \) denotes a generic point of \( \mathbb{R}^d \), and \( k \) is a given real-valued function on \( S \); or

\[ I_B = \int_{\mathbb{R}^d} h(x) f(x) \, dx \]

where \( h \) is a real-valued function on \( \mathbb{R}^d \) and \( f \) is a given pdf on \( \mathbb{R}^d \).
The Transformed Problem: Monte Carlo Integration

- It is clear that $I_B$ can be written as an expectation:

$$I_B = E(h(X)) \quad \text{where} \ X \sim f.$$ 

- Also, extend the definition of $k$ to all of $\mathbb{R}^d$ by saying that $k(x) = 0$ for every $x$ that is not in $S$, then

$$I_A = \int_{\mathbb{R}^d} k(x) = \int_{\mathbb{R}^d} \frac{k(x)}{f(x)} f(x) dx = E[k(x)/f(x)]. \quad (1)$$
The Transformed Problem: Monte Carlo Integration

- It is clear that $I_B$ can be written as an expectation:

$$I_B = E(h(X)) \text{ where } X \sim f.$$

- Also, extend the definition of $k$ to all of $\mathbb{R}^d$ by saying that $k(x) = 0$ for every $x$ that is not in $S$, then

$$I_A = \int_{\mathbb{R}^d} k(x) = \int_{\mathbb{R}^d} \frac{k(x)}{f(x)} f(x) dx = E\left[\frac{k(x)}{f(x)}\right]. \quad (1)$$
Notice that \( \frac{k(x)}{f(x)} \) is well-defined except where \( f \) equals 0, which is a set of probability 0.

This is a simple trick that will be especially useful in the method known as Importance Sampling.
Notice that \( \frac{k(x)}{f(x)} \) is well-defined except where \( f \) equals 0, which is a set of probability 0.

This is a simple trick that will be especially useful in the method known as Importance Sampling.
Simple Sampling

- This leads to a natural Monte Carlo strategy for estimating the value of \( I_B \), say.

- If we can generate iid random variables \( X_1, X_2, \ldots \) whose common pdf is \( f \), then for every \( n \),

\[
\hat{I}_n = \frac{1}{n} \sum_{i=1}^{n} h(X_i)
\]

is an unbiased estimator of \( I_B \).
Simple Sampling

- This leads to a natural Monte Carlo strategy for estimating the value of $I_B$, say.
- If we can generate iid random variables $X_1, X_2, \ldots$ whose common pdf is $f$, then for every $n$,

\[ \hat{I}_n = \frac{1}{n} \sum_{i=1}^{n} h(X_i) \]

is an unbiased estimator of $I_B$. 
Moreover, the strong law of large numbers implies that $\hat{I}_n$ converges to $I_B$ with probability 1 as $n \to \infty$.

This method for estimating $I_B$ is called simple sampling.
Moreover, the strong law of large numbers implies that $\hat{I}_n$ converges to $I_B$ with probability 1 as $n \to \infty$.

This method for estimating $I_B$ is called simple sampling.
The Variance Reduction Problem

- The variance of simple sampling estimator \( \hat{I}_n \) of \( I_B \) is

\[
\text{var}(\hat{I}_n) = \frac{\text{var}(h(X))}{n} = \frac{\left( \int_S h(x)^2 f(x) dx - I_B^2 \right)}{n}.
\] (2)

- The variance of the estimator determines the size of the confidence interval.
- The \( n \) in the denominator is hard to avoid in Monte Carlo, but there are various ways to reduce the numerator.
The Variance Reduction Problem

- The variance of simple sampling estimator $\hat{I}_n$ of $I_B$ is

$$\text{var}(\hat{I}_n) = \frac{\text{var}(h(X))}{n} = \frac{(\int_S h(x)^2 f(x) dx - I_B^2)}{n}.$$  \hspace{1cm} (2)

- The variance of the estimator determines the size of the confidence interval.

- The $n$ in the denominator is hard to avoid in Monte Carlo, but there are various ways to reduce the numerator.
The Variance Reduction Problem

The variance of simple sampling estimator $\hat{l}_n$ of $l_B$ is

$$\text{var}(\hat{l}_n) = \frac{\text{var}(h(X))}{n} = \frac{\left( \int_S h(x)^2 f(x) dx - l_B^2 \right)}{n}. \quad (2)$$

- The variance of the estimator determines the size of the confidence interval.
- The $n$ in the denominator is hard to avoid in Monte Carlo, but there are various ways to reduce the numerator.
The goal of this chapter is to explore alternative sampling schemes which can achieve smaller variance for the same amount of computational efforts.
Stratified Sampling

Step 1: Range Partition

- Stratified sampling is a powerful and commonly used technique in population survey and is also very useful in Monte Carlo computations.

- To evaluate $I_B$, the stratified sampling is to partition $S$ into several disjoint sets $S^{(1)}, \ldots, S^{(M)}$ (so that $S = \bigcup_{i=1}^{M} S^{(i)}$).
Stratified Sampling

**Step 1: Range Partition**

- Stratified sampling is a powerful and commonly used technique in population survey and is also very useful in Monte Carlo computations.
- To evaluate $I_B$, the stratified sampling is to partition $S$ into several disjoint sets $S^{(1)}, \ldots, S^{(M)}$ (so that $S = \bigcup_{i=1}^{M} S^{(i)}$).
For $i = 1, \ldots, M$, let

$$a_i = \int_{S^{(i)}} f(x)dx = P(X \in S^{(i)}).$$

Observe that $a_1 + \cdots + a_M = 1$. Fix integers $n_1, \ldots, n_M$ such that $n_1 + \cdots + n_M = n$. 
For $i = 1, \ldots, M$, let

$$a_i = \int_{S^{(i)}} f(x) \, dx = P(X \in S^{(i)}).$$

Observe that $a_1 + \cdots + a_M = 1$. Fix integers $n_1, \ldots, n_M$ such that $n_1 + \cdots + n_M = n$. 
Step 2: Sub-sampling

- For each $i$, generate $n_i$ samples $X_1^{(i)}, \ldots, X_{n_i}^{(i)}$ from $S^{(i)}$ having the conditional pdf

$$g(x) = \begin{cases} \frac{f(x)}{a_i} & \text{if } x \in S^{(i)} \\ 0 & \text{otherwise} \end{cases}$$
Let $T_i = n_i^{-1} \sum_{j=1}^{n_i} h(X_j^{(i)})$. Then

$$E(T_i) = \int_{S^{(i)}} h(x) \frac{f(x)}{a_i} \, dx = \frac{1}{a_i} \int_{S^{(i)}} h(x)f(x) \, dx = l_i/a_i,$$

by defining $l_i = \int_{S^{(i)}} h(x)f(x) \, dx$. 
Step 3: The Stratified Estimator

- Observe that $I_1 + \cdots + I_M = I_B$. The stratified estimator is

$$T = \sum_{i=1}^{M} a_i T_i.$$ 

- It is unbiased because of

$$E(T) = \sum_{i=1}^{M} a_i E(T_i) = \sum_{i=1}^{M} a_i l_i/a_i = I_B.$$
Step 3: The Stratified Estimator

- Observe that \( I_1 + \cdots + I_M = I_B \). The stratified estimator is

\[
T = \sum_{i=1}^{M} a_i T_i.
\]

- It is unbiased because of

\[
E(T) = \sum_{i=1}^{M} a_i E(T_i) = \sum_{i=1}^{M} a_i I_i / a_i = I_B.
\]
The variance of $T$ is

$$\text{var}(T) = \sum_{i=1}^{M} a_i^2 \text{var}(T_i),$$

where

$$\text{var}(T_i) = \frac{\int_{S(i)} h(x)^2 \frac{f(x)}{a_i} \, dx}{n_i} - \left( \frac{l_i}{a_i} \right)^2.$$ 

following from (2).
Theorem

The Foundation Theory of the Stratified Sampling

If $n_i = na_i$ for $i = 1, \ldots, M$. then the stratified estimator has smaller variance than the simple estimator $\hat{I}_n$. In fact,

$$\text{var}(\hat{I}_n) = \text{var}(T) + \frac{1}{n} \sum_{i=1}^{M} a_i (\frac{l_i}{a_i} - l_B)^2.$$ 

- The choice $n_i = na_i$, called “proportional allocation”, give a stratified estimator which has smaller variance than the simple estimator.
Theorem

The Foundation Theory of the Stratified Sampling

If \( n_i = na_i \) for \( i = 1, \ldots, M \). then the stratified estimator has smaller variance than the simple estimator \( \hat{I}_n \). In fact,

\[
\text{var}(\hat{I}_n) = \text{var}(T) + \frac{1}{n} \sum_{i=1}^{M} a_i \left( \frac{l_i}{a_i} - l_B \right)^2.
\]

The choice \( n_i = na_i \), called “proportional allocation”, give a stratified estimator which has smaller variance than the simple estimator.
Importance Sampling

Property of the Important Sampling

- Importance sampling is a very powerful method that can improve Monte Carlo efficiency by orders of magnitude in some problems.

- But it requires Caution: an inappropriate implementation can reduce efficiency by orders of magnitude!
Importance Sampling

Property of the Important Sampling

- Importance sampling is a very powerful method that can improve Monte Carlo efficiency by orders of magnitude in some problems.

- But it requires Caution: an inappropriate implementation can reduce efficiency by orders of magnitude!
The Basic Idea

- The method works by sampling from an artificial probability distribution that is chosen by the user, and then reweighting the observations to get an unbiased estimate.
- The idea is based on the identity \[ I_A = \int_{\mathbb{R}^d} k(x) = \int_{\mathbb{R}^d} \frac{k(x)}{f(x)} f(x) dx = E\left[ \frac{k(x)}{f(x)} \right]. \]
The Basic Idea

- The method works by sampling from an artificial probability distribution that is chosen by the user, and then reweighting the observations to get an unbiased estimate.
- The idea is based on the identity (1)

\[
I_A = \int_{\mathbb{R}^d} k(x) = \int_{\mathbb{R}^d} \frac{k(x)}{f(x)} f(x) dx = E\left[\frac{k(x)}{f(x)}\right].
\]
It implies that $I_A$ can be estimated by

$$\hat{J}_n = \frac{1}{n} \sum_{i=1}^{n} \frac{k(X_i)}{f(X_i)},$$

where $X_i$’s are iid from $f$.

We call $\hat{J}_n$ the importance sampling estimator based on $f$.

The identity (1) implies that $\hat{J}_n$ is unbiased.
It implies that $I_A$ can be estimated by

$$\hat{J}_n = \frac{1}{n} \sum_{i=1}^{n} \frac{k(X_i)}{f(X_i)},$$

where $X_i$’s are iid from $f$.

We call $\hat{J}_n$ the importance sampling estimator based on $f$.

The identity (1) implies that $\hat{J}_n$ is unbiased.
It implies that $I_A$ can be estimated by

$$\hat{J}_n = \frac{1}{n} \sum_{i=1}^{n} \frac{k(X_i)}{f(X_i)},$$

where $X_i$'s are iid from $f$.

We call $\hat{J}_n$ the importance sampling estimator based on $f$.

The identity (1) implies that $\hat{J}_n$ is unbiased.
The Important Sampling Procedure

Suppose now one is interested in evaluating

\[ I_B = \int_{\mathbb{R}^d} h(x)f(x) \, dx, \]

the procedure of the importance sampling is as follows:

(a) Draw \( X_1, \ldots, X_n \) from a trial density \( g \).
(b) Calculate the importance weight

\[ w_j = \frac{f(X_j)}{g(X_j)}, \quad \text{for } j = 1, \ldots, n. \]

(c) Approximate \( I_B \) by

\[ \hat{I}_{g,n} = \frac{\sum_{j=1}^{n} w_j h(X_j)}{\sum_{j=1}^{n} w_j}. \]
The Important Sampling Procedure

Suppose now one is interested in evaluating

\[ I_B = \int_{\mathbb{R}^d} h(x)f(x)dx, \]

the procedure of the importance sampling is as follows:

(a) Draw \( X_1, \ldots, X_n \) from a trial density \( g \).

(b) Calculate the importance weight

\[ w_j = f(X_j)/g(X_j), \quad \text{for } j = 1, \ldots, n. \]

(c) Approximate \( I_B \) by

\[ \hat{J}_{g,n} = \frac{\sum_{j=1}^n w_j h(X_j)}{\sum_{j=1}^n w_j}. \]
The Important Sampling Procedure

- Suppose now one is interested in evaluating

\[ I_B = \int_{\mathbb{R}^d} h(x) f(x) \, dx, \]

the procedure of the importance sampling is as follows:

(a) Draw \( X_1, \ldots, X_n \) from a trial density \( g \).

(b) Calculate the importance weight

\[ w_j = \frac{f(X_j)}{g(X_j)}, \quad \text{for } j = 1, \ldots, n. \]

(c) Approximate \( I_B \) by

\[ \hat{J}_{g,n} = \frac{\sum_{j=1}^{n} w_j h(X_j)}{\sum_{j=1}^{n} w_j}. \]

(3)
The Important Sampling Procedure

Suppose now one is interested in evaluating

\[ I_B = \int_{\mathbb{R}^d} h(x)f(x)dx, \]

the procedure of the importance sampling is as follows:

(a) Draw \( X_1, \ldots, X_n \) from a trial density \( g \).
(b) Calculate the importance weight

\[ w_j = \frac{f(X_j)}{g(X_j)}, \quad \text{for } j = 1, \ldots, n. \]

(c) Approximate \( I_B \) by

\[ \hat{I}_{g,n} = \frac{\sum_{j=1}^{n} w_j h(X_j)}{\sum_{j=1}^{n} w_j}. \]
Thus, in order to make the estimation error small, one wants to choose $g$ as “close” in shape to $h(x)f(x)$ as possible.
An Alternative Important Sampling Procedure

- A major advantage of using (3) instead of the unbiased estimate, $\tilde{I}_B = \frac{1}{n} \sum_{j=1}^{n} w_j h(X_j)$, is that
  - in using the former, we need only to know the ratio $f(X)/g(X)$ up to a multiplicative constant; whereas in the latter, the ratio needs to be known exactly.
  - Although introducing a small bias, (3) often has a smaller mean squared error than the unbiased one $\tilde{I}_B$. 
An Alternative Important Sampling Procedure

- A major advantage of using (3) instead of the unbiased estimate, 
  \[ \tilde{I}_B = \frac{1}{n} \sum_{j=1}^{n} w_j h(X_j) \]

  is that

  - in using the former, we need only to know the ratio \( f(X)/g(X) \) up to a multiplicative constant; whereas in the latter, the ratio needs to be known exactly.
  - Although introducing a small bias, (3) often has a smaller mean squared error than the unbiased one \( \tilde{I}_B \).
An Alternative Important Sampling Procedure

- A major advantage of using (3) instead of the unbiased estimate,

\[
\tilde{I}_B = \frac{1}{n} \sum_{j=1}^{n} w_j h(X_j)
\]

is that

- in using the former, we need only to know the ratio \( f(X)/g(X) \) up to a multiplicative constant; whereas in the latter, the ratio needs to be known exactly.
- Although introducing a small bias, (3) often has a smaller mean squared error than the unbiased one \( \tilde{I}_B \).
Example 1

Let $h(x) = 4\sqrt{1 - x^2}$, $x \in [0, 1]$. Let us imagine that we do not know how to evaluate $I = \int_0^1 h(x)dx$ (which is $\pi$, of course).
Use Simple Sampling

- The simple sampling estimate is

\[ \hat{I}_n = \frac{1}{n} \sum_{i=1}^{n} 4 \sqrt{1 - U_i^2}, \]

where \( U_i \) are iid U[0,1] random variables.

- This is unbiased, with variance

\[ \text{var}(\hat{I}_n) = \frac{1}{n} \left( \int_0^1 h(x)^2 \, dx - I^2 \right) = \frac{1}{n} \left( \int_0^1 16(1-x^2) \, dx - \pi^2 \right) = \frac{0.797}{n}. \]
Use Simple Sampling

- The simple sampling estimate is

\[ \hat{I}_n = \frac{1}{n} \sum_{i=1}^{n} 4 \sqrt{1 - U_i^2}, \]

where \( U_i \) are iid U[0,1] random variables.

- This is unbiased, with variance

\[ \text{var}(\hat{I}_n) = \frac{1}{n} \left( \int_0^1 h(x)^2 \, dx - I^2 \right) = \frac{1}{n} \left( \int_0^1 16(1-x^2) \, dx - \pi^2 \right) = \frac{0.797}{n}. \]
Use Inappropriate Important Sampling

Consider the importance sampling estimate based on the pdf \( g_b(x) = 2x, \ x \in [0, 1] \).

It is easy to generate \( Y_i \sim g_b \) (the cdf is \( F(t) = t^2 \), so we can set \( Y_i \leftarrow F^{-1}(U_i) = \sqrt{U_i} \), where \( U_i \sim U[0, 1] \)).

The importance sampling estimator is

\[
\hat{J}_n^{(b)} = \frac{1}{n} \sum_{i=1}^{n} \frac{h(Y_i)}{g_b(Y_i)} = \frac{1}{n} \sum_{i=1}^{n} \frac{4\sqrt{1 - Y_i^2}}{2Y_i}.
\]
Use Inappropriate Important Sampling

Consider the importance sampling estimate based on the pdf $g_b(x) = 2x, \ x \in [0, 1]$.

It is easy to generate $Y_i \sim g_b$ (the cdf is $F(t) = t^2$, so we can set $Y_i \leftarrow F^{-1}(U_i) = \sqrt{U_i}$, where $U_i \sim U[0, 1]$).

The importance sampling estimator is

$$\hat{J}^{(b)}_n = \frac{1}{n} \sum_{i=1}^{n} h(Y_i) / g_b(Y_i) = \frac{1}{n} \sum_{i=1}^{n} \frac{4\sqrt{1 - Y_i^2}}{2Y_i}.$$
Use Inappropriate Important Sampling

- Consider the importance sampling estimate based on the pdf $g_b(x) = 2x$, $x \in [0, 1]$.
- It is easy to generate $Y_i \sim g_b$ (the cdf is $F(t) = t^2$, so we can set $Y_i \leftarrow F^{-1}(U_i) = \sqrt{U_i}$, where $U_i \sim U[0, 1]$).
- The importance sampling estimator is

$$
\hat{J}_n^{(b)} = \frac{1}{n} \sum_{i=1}^{n} \frac{h(Y_i)}{g_b(Y_i)} = \frac{1}{n} \sum_{i=1}^{n} \frac{4\sqrt{1 - Y_i^2}}{2Y_i}.
$$
The $\hat{J}_n^{(b)}$ has mean $I$ and variance

$$\text{var}(\hat{J}_n^{(b)}) = \frac{1}{n} \text{var}\left( \frac{h(Y)}{g_b(Y)} \right) = \frac{1}{n} \int_0^1 (\frac{h(x)}{g_b(x)} - I)^2 dx = +\infty.$$ 

Hence, the trial density $g(x) = 2x$ is very bad, and we need try a different one.
The $\hat{J}_n^{(b)}$ has mean $I$ and variance

$$\text{var}(\hat{J}_n^{(b)}) = \frac{1}{n} \text{var}\left( \frac{h(Y)}{g_b(Y)} \right) = \frac{1}{n} \int_0^1 \left( \frac{h(x)}{g_b(x)} - I \right)^2 dx = +\infty.$$ 

Hence, the trial density $g(x) = 2x$ is very bad, and we need try a different one.
Use Appropriate Important Sampling

- Let \( g_c(x) = (4 - 2x)/3, \ x \in [0, 1] \).
- The importance sampling estimator is

\[
\hat{J}_n^{(c)} = \frac{1}{n} \sum_{i=1}^{n} \frac{4 \sqrt{1 - Y_i^2}}{(4 - 2Y_i)/3},
\]

whose variance is

\[
\text{var}(\hat{J}_n^{(c)}) = \frac{1}{n} \text{var} \left( \frac{h(Y)}{g_c(Y)} \right) = \frac{1}{n} \int_0^1 \left( \frac{h(x)}{g_c(x)} - 1 \right)^2 dx
\]

\[
= \frac{1}{n} \left\{ \int_0^1 \frac{16(1 - x^2)}{(4 - 2x)/3} dx - \pi^2 \right\} = 0.224/n.
\]
Use Appropriate Important Sampling

- Let $g_c(x) = (4 - 2x)/3$, $x \in [0, 1]$.
- The importance sampling estimator is

$$\hat{J}_n^{(c)} = \frac{1}{n} \sum_{i=1}^{n} \frac{4 \sqrt{1 - Y_i^2}}{(4 - 2Y_i)/3},$$

whose variance is

$$\text{var}(\hat{J}_n^{(c)}) = \frac{1}{n} \text{var}\left(\frac{h(Y)}{g_c(Y)}\right) = \frac{1}{n} \int_{0}^{1} \left(\frac{h(x)}{g_c(x)} - 1\right)^2 dx$$

$$= \frac{1}{n} \left[ \int_{0}^{1} \frac{16(1 - x^2)}{(4 - 2x)/3} dx - \pi^2 \right] = 0.224/n.$$
Thus, the importance sampling estimate of (c) can achieve the same size confidence interval as the simple sampling estimate of (a) while using only one third as many generated random variables.
Control Variates Method

The Main Idea

▶ In this method, one uses a control variate $C$, which is Correlated with the sample $X$, to produce a better estimate.

The Procedure

▶ Suppose the estimation of $\mu = E(X)$ is of interest and $\mu_C = E(C)$ is known.
▶ Then we can construct Monte Carlo samples of the form

$$X(b) = X + b(C - \mu_C),$$

which have the same mean as $X$, but a new variance

$$\text{var}(X(b)) = \text{var}(X) - 2b\text{Cov}(X, C) + b^2\text{var}(C).$$
Control Variates Method

The Main Idea

- In this method, one uses a control variate $C$, which is correlated with the sample $X$, to produce a better estimate.

The Procedure

- Suppose the estimation of $\mu = E(X)$ is of interest and $\mu_C = E(C)$ is known.
- Then we can construct Monte Carlo samples of the form

$$X(b) = X + b(C - \mu_C),$$

which have the same mean as $X$, but a new variance

$$\text{var}(X(b)) = \text{var}(X) - 2b\text{Cov}(X, C) + b^2\text{var}(C).$$
Control Variates Method

The Main Idea

- In this method, one uses a control variate $C$, which is correlated with the sample $X$, to produce a better estimate.

The Procedure

- Suppose the estimation of $\mu = E(X)$ is of interest and $\mu_C = E(C)$ is known.
- Then we can construct Monte Carlo samples of the form

$$X(b) = X + b(C - \mu_C),$$

which have the same mean as $X$, but a new variance

$$\text{var}(X(b)) = \text{var}(X) - 2b\text{Cov}(X, C) + b^2\text{var}(C).$$
If the computation of $\text{Cov}(X, C)$ and $\text{var}(C)$ is easy, then we can let $b = \text{Cov}(X, C)/\text{Var}(C)$, in which case

$$\text{var}(X(b)) = (1 - \rho_{XC}^2)\text{var}(X) < \text{var}(X).$$

A Special Case

- Another situation is when we know only that $E(C)$ is equal to $\mu$. Then, we can form $X(b) = bX + (1 - b)C$.
- It is easy to show that if $C$ is Correlated with $X$, we can always choose a proper $b$ so that $X(b)$ has a smaller variance than $X$. 

If the computation of $\text{Cov}(X, C)$ and $\text{var}(C)$ is easy, then we can let $b = \frac{\text{Cov}(X, C)}{\text{Var}(C)}$, in which case

$$\text{var}(X(b)) = (1 - \rho_{XC}^2)\text{var}(X) < \text{var}(X).$$

A Special Case

Another situation is when we know only that $E(C)$ is equal to $\mu$. Then, we can form $X(b) = bX + (1 - b)C$.

It is easy to show that if $C$ is correlated with $X$, we can always choose a proper $b$ so that $X(b)$ has a smaller variance than $X$. 

Zhang J.T. Ch4. Variance Reduction Techniques
If the computation of $\text{Cov}(X, C)$ and $\text{var}(C)$ is easy, then we can let $b = \text{Cov}(X, C)/\text{Var}(C)$, in which case

$$\text{var}(X(b)) = (1 - \rho_{XC}^2)\text{var}(X) < \text{var}(X).$$

A Special Case

Another situation is when we know only that $E(C)$ is equal to $\mu$. Then, we can form $X(b) = bX + (1 - b)C$.

It is easy to show that if $C$ is Correlated with $X$, we can always choose a proper $b$ so that $X(b)$ has a smaller variance than $X$. 
Antithetic Variates Method

The Main Idea

- Suppose $U$ is a random number used in the production of a sample $X$ that follows a distribution with cdf $F$, that is, $X = F^{-1}(U)$, then $X' = F^{-1}(1 - U)$ also follows distribution $F$.

- More generally, if $g$ is a monotone function, then

  $$[g(u_1) - g(u_2)][g(1 - u_1) - g(1 - u_2)] \leq 0$$

  for any $u_1, u_2 \in [0, 1]$. 

Zhang J.T. Ch4. Variance Reduction Techniques
Antithetic Variates Method

The Main Idea

- Suppose $U$ is a random number used in the production of a sample $X$ that follows a distribution with cdf $F$, that is, $X = F^{-1}(U)$, then $X' = F^{-1}(1 - U)$ also follows distribution $F$.

- More generally, if $g$ is a monotone function, then

\[
[g(u_1) - g(u_2)][g(1 - u_1) - g(1 - u_2)] \leq 0
\]

for any $u_1, u_2 \in [0, 1]$. 

Zhang J.T. Ch4. Variance Reduction Techniques
For two independent uniform random variable $U_1$ and $U_2$, we have

$$E\{[g(U_1) - g(U_2)][g(1-U_1) - g(1-U_2)]\} = \text{Cov}(X, X') \leq 0,$$

where $X = g(U)$ and $X' = g(1 - U)$.

Therefore, $\text{var}[(X + X')/2] \leq \text{var}(X)/2$, implying that using the pair $X$ and $X'$ is better than using two independent Monte Carlo draws for estimating $E(X)$. 
For two independent uniform random variable $U_1$ and $U_2$, we have

$$E\{[g(U_1) - g(U_2)][g(1 - U_1) - g(1 - U_2)]\} = \text{Cov}(X, X') \leq 0,$$

where $X = g(U)$ and $X' = g(1 - U)$.

Therefore, $\text{var}[(X + X')/2] \leq \text{var}(X)/2$, implying that using the pair $X$ and $X'$ is better than using two independent Monte Carlo draws for estimating $E(X)$. 
Example 2

- We return once more to the problem of estimating the integral $I = \int_0^1 4\sqrt{1 - x^2} \, dx$.

- Choose a large even value of $n$. As usual, our Simple Estimator and its Variance are

$$\hat{I}_n = \frac{1}{n} \sum_{i=1}^{n} h(U_i), \quad \text{var}(\hat{I}_n) = \frac{0.797}{n}.$$
Example 2

- We return once more to the problem of estimating the integral \( I = \int_0^1 4\sqrt{1 - x^2} \, dx \).

- Choose a large **even** value of \( n \). As usual, our Simple Estimator and its Variance are

\[
\hat{I}_n = \frac{1}{n} \sum_{i=1}^{n} h(U_i), \quad \text{var}(\hat{I}_n) = 0.797/n.
\]
Our corresponding Antithetic Estimator and its Variance are

\[ \hat{I}^{An}_n = \frac{1}{n} \sum_{i=1}^{n/2} (h(U_i) + h(1 - U_i)). \]

\[ \text{var}(\hat{I}^{An}_n) = \frac{1}{n^2} \left\{ \frac{n}{2} [\text{var}(h(U_1)) + 2\text{Cov}(h(U_1), h(1 - U_1)) + \text{var}(h(1 - U_1))] \right\} \]

\[ = \frac{1}{n} [\text{var}(h(U_1)) + \text{Cov}(h(U_1), h(1 - U_1))] \]

\[ = 0.219/n \]