CONTROL VARIATES

• Suppose there is another function $g(u)$ that resembles $f(u)$ but simpler for which I can compute the integral. Use it to improve the Monte Carlo integration of $f$.

$$
\int_{0}^{1} f(u)du = \int_{0}^{1} g(u)du + \int_{0}^{1} [f(u) - g(u)]du
$$

Use the known value for the first term $\int_{0}^{1} g(u)du$

and estimate the second term, the difference, using Monte Carlo.

Estimator is

$$
\int_{0}^{1} g(u)du + \frac{1}{n} \sum_{i=1}^{n} (f(U_i) - g(U_i))
$$
Control variates

Estimator is

\[
\int_0^1 g(u)du + \frac{1}{n} \sum_{i=1}^n (f(U_i) - g(U_i))
\]

VARIANCE

\[
\frac{1}{n} \text{var}(f(U_1) - g(U_1))
\]

This is better than crude if \( \text{var}(f(U_1) - g(U_1)) < \text{var}(f(U_1)) \)

When \( g(U_1) \) is close to \( f(U_1) \) variance of difference is small.
Control Variate

I tried a function of the form

\[ g(u) = (u - 0.47) + 6(u - 0.47)^2 \quad \text{if } u > 0.47 \]
\[ g(u) = 0 \quad \text{if } u \leq 0.47. \]

This is easy to integrate since it is quadratic.

- function g=GG(u)
  % control variate for callopt2.
  % this function integrates to \(2 \times (0.53)^3 + 0.53^2/2\)
  \(u=\max(0,u-.47)\);
  \(g=6*u.^2+u;\)
Function $GG(u)$ and $fn(u)$

- Only need to estimate the area between the curves by Monte Carlo
Performance of Control Variate

- \(u = \text{rand}(1,500000)\);
- \(F = \text{fn}(u)\);
- \(G = \text{GG}(u)\);
- \(\text{est} = 2 \times (0.53)^3 + 0.53^2/2 + (F - G)\);
- \(\text{mean}({\text{est}})\)
- \(\text{var}({\text{est}})/\text{length}({\text{est}})\)
- \(\text{mean} = 0.4616 \quad \text{var} = 2.93 \times 10^{-08}\)
- Efficiency compared with crude \(8.7 \times 10^{-07}\) ratio = 30 (equivalent to 15 million crude!)
The "best" approximation to \( \Gamma \) using \( g \) is not necessarily \( g \) itself. As an alternative, we can approximate \( f(U) \) using a linear regression on \( g(U) \)

\[
f(U) = \theta + \beta (g(U) - Eg(U)) + \varepsilon,
\]

where \( \varepsilon \) is a mean zero "error" random variable.

\[
\theta = E(f(U)) = \int_0^1 f(u)du
\]

Slope parameter given by \( \beta = \frac{\text{cov}(f(U), g(U))}{\text{var}(g(U))} \).

We wish to estimate the parameter \( \theta \) and since

\[
\theta = E[f(U) - \beta (g(U) - Eg(U))]
\]

we can use a sample mean of terms of this form, i.e.

\[
\frac{1}{n} \sum_{i=1}^{n} [f(U_i) - \beta (g(U_i) - Eg(U_i))].
\]

\[
= \beta E(g(U)) + \frac{1}{n} \sum_{i=1}^{n} [f(U_i) - \beta (g(U_i))] .
\]
This is identical to the earlier Control Variate if \( \beta = 1 \). However since \( \beta \) is usually unknown, we replace it by the regression estimator of slope, equivalent to replacing variance and covariance by sample variance and covariance.

\[
\hat{\beta} = \frac{\sum_{i=1}^{n} [f(U_i) - f(\bar{U})][g(U_i) - g(\bar{U})]}{\sum_{i=1}^{n} [g(U_i) - g(\bar{U})]^2}
\]

where \( g(U) = \frac{1}{n} \sum_{i=1}^{n} g(U_i) \).

It is easy to show that the variance of this control estimator

\[
\text{var}(\beta E(g(U)) + \frac{1}{n} \sum_{i=1}^{n} [f(U_i) - \beta (g(U_i))] ) = (1 - r^2) \text{var}(\frac{1}{n} \sum_{i=1}^{n} f(U_i))
\]

so the efficiency is given by \( \frac{1}{1 - r^2} \) where

\( r = \text{correlation coefficient between } f(U_i) \text{ and } g(U_i) \).
Example of Control Variate II

Find \( \int_{0}^{10} \frac{x}{2 + x^{1/4}} \, dx = E(f(X)) \) where \( X \) is \( U[0,10] \) and

\[ f(x) = \frac{10x}{2 + x^{1/4}}. \]

Consider as control variate \( g(x) = x \).

Note that \( E(g(X)) = E(10U) = 5 \).

- \( x=10*\text{rand}(1,500000); \)
- \( f=10*x./(2+x.^{.25}); \quad g=x; \quad C=\text{cov}([f'\ g']); \)
- \( \beta=C(1,2)/C(2,2); \quad \text{EG}=5; \)
- \( \text{est}=\beta*\text{EG}+f-\beta*g; \)
- \( \text{mean(est); var(est)/length(est)} \)
- \( \text{(mean= 14.000011 var= 1.8e-007)} \)
- \( \text{var(f)/500000= 2.8160e-04} \)
- \( \text{Eff=622 (equivalent to 311 million crude)} \)
Stratified Sample

• Consider choosing one point in interval $[0,a]$ and another in the interval $[a,1]$

$aU_1$ is in $[0,a]$ and $a + (1-a)U_2$ is in $[a,1]$. Consider using a weighted average of the function at these two points to estimate $\int_0^1 f(u)du$. What should the weights be?

Let $\theta_{ST} = af(aU_1) + (1-a)f(a + (1-a)U_2)$

where $U_1, U_1$ are independent $U[0,1]$. Why weights $a,1-a$?

$E(\theta_{ST}) = E(af(aU_1) + (1-a)f(a + (1-a)U_2)) = aE(f(aU_1)) + (1-a)E(f(a + (1-a)U_2)

= a\int_0^1 f(au)du + (1-a)\int_0^1 f(a + (1-a)u)du = \int_0^1 f(u)du + \int_0^a f(u)du = \int_0^1 f(u)du$
Weights in a stratified sample

• The weights attached to a given stratum (interval, region) are proportional to the length of the interval (or volume of region).

• We may allow many observations in a given stratum and apply weight to the stratum mean.
General stratified sample

• Choose

\( n_i \) from stratum \( i, i = 1,2,...m \)

Evaluate the function at \( n_i \) random points, uniformly distributed over stratum \( i \) and average the result

\[
AV_i = \frac{1}{n_i} \sum_{j=1}^{n_i} f(V_{ij}) \quad \text{where } V_{ij} \text{ is uniformly distributed over stratum } i.
\]

Use WEIGHTED AVERAGE of the stratum mean, weight proportional to length, (area, or volume) of stratum.
Variance of Stratified Sample

Estimator:

$$\sum_i (x_{i+1} - x_i) AV_i$$

Variance of Stratified Sample estimator:

$$\sum_i (x_{i+1} - x_i)^2 \frac{\text{var}(f(V_{ij}))}{n_i}$$
Performance of stratified sample

- Example. We chose \( a = 0.7 \) and sampled 100,000 on \([0, 0.7]\) and another 100,000 on \([0.7, 1]\).

- \( a = 0.7; \)
  - \( F = a \cdot fn(a \cdot \text{rand}(1, 500000)) + (1-a) \cdot fn(a + (1-a) \cdot \text{rand}(1, 500000)) \);
  - mean(F) \( \% = 0.4608 \)
  - \( \text{var}(F)/\text{length}(F) \) \( \% = 9.18 \times 10^{-8} \)
  - Compare with Crude with \( n = 1000000 \). Variance = \( 4.3 \times 10^{-7} \); Efficiency around 5.
Optimal sample sizes for stratified sample

The optimal sample size in each stratum is proportional to
(a) the size of the stratum (e.g. interval length)
(b) the stratum standard deviation
So if stratum $i$ is the interval $(x_i, x_{i+1})$ then

optimal sample size $n_i$ proportional to

$$(x_{i+1} - x_i)\sqrt{\text{var}(f(x_i + U(x_{i+1} - x_i)))}$$

$$\text{var}(f(x_i + U(x_{i+1} - x_i))) \text{ estimated from a preliminary sample.}$$
The function “stratified”

- function [est,v,n]=stratified(x,nsample)
- % input x=vector of strata (e.g. x=[0 .25 .5 .75 1] and nsample=sample size
- est=0; n=[]; m=length(x);
- for i=1:m-1 %preliminary simulation of 10000 to optimize sample size
  - v= var(fn(unifrnd(x(i),x(i+1),1,1000)));  
  - n=[n (x(i+1)-x(i))*sqrt(v)];
- end
- n=floor(nsample*n/sum(n)); % these are stratum sample sizes
- v=0;
- for i=1:m-1
  - F=fn(unifrnd(x(i),x(i+1),1,n(i)));
  - est=est+(x(i+1)-x(i))*mean(F);
  - v=v+var(F)*(x(i+1)-x(i))^2/n(i);
- end
Stratified Sample with more strata, optimal sample size

- Try 4 strata [0 .55] , [.55 .8], [.8 .94], [.94,1] with optimal sample sizes
  - 32717  78703  46420  42158
- \([\text{est}, v, n] = \text{stratified}([0 .55 .80 .94 1],500000)\)
- estimator 0.4617 variance = 3.7e-008
- Efficiency now around 24 times that of Crude. Try also:
  - \([\text{est}, v, n] = \text{stratified}([.47 .62 .75 .87 .96 1],500000)\) (var=1.4e-8)
Conditioning as a Variance Reduction Tool

- Example: suppose we wish to find the area of a region, say the area under the graph of a function $h(x)$.
- Crude Monte Carlo: Find a probability density function $g(x)$ such that

\[ cg(x) \geq h(x) \text{ for all } x. \]

Then generate random points $(X_i, Y_i)$ uniformly distributed under the graph of $cg(x)$: $X_i$ has p.d.f. $g(x)$ and $Y_i = cg(X_i)U_i$.

The proportion of these points which are also under the graph of $h(x)$

\[ \frac{\text{area under } h(x)}{\text{area under } cg(x)} = \frac{\text{area under } h(x)}{c} \]
Conditioning (II)

• Leads to the estimator:

\[ \hat{\theta}_{CR} = \frac{c \times \text{Number points under } h}{\text{Total number of points}} = \frac{c}{n} \sum_{i=1}^{n} I(Y_i \leq h(X_i)) = \frac{c}{n} \sum_{i=1}^{n} Z_i, \text{ say} \]

Consider an alternative estimator

\[ \hat{\theta}_{CO} = \frac{c}{n} \sum_{i=1}^{n} E[Z_i \mid X_i] = \frac{c}{n} \sum_{i=1}^{n} \frac{h(X_i)}{cg(X_i)} = \frac{1}{n} \sum_{i=1}^{n} \frac{h(X_i)}{g(X_i)} \]

We obtained this new estimator by taking the conditional expectation of the old. Does this always give an unbiased estimator with smaller variance?
Some Properties of Conditional Expectation

1. \( E\{E[Y \mid X]\} = E(Y) \)
2. \( \text{Var}(E[Y \mid X]) \leq \text{var}(Y) \) with equality if and only if \( Y \) is a function of \( X \).

\( \hat{\theta} \) is an unbiased estimator and for some random variable \( X \) we can compute \( E(\hat{\theta} \mid X) \) then this is another unbiased estimator, typically with smaller variance. Therefore

\[ \text{var}(Z_i) \geq \text{var}(E(Z_i \mid X_i)) \]

and the second estimator

\[ \frac{c}{n} \sum_{i=1}^{n} E[Z_i \mid X_i] \]

is better than the first \( \frac{c}{n} \sum_{i=1}^{n} Z_i \).
Example of Conditioning: Estimating pi

If we generate \((U, V)\) uniform on \([0,1]\), and if

\[
Z = I(U^2 + V^2 \leq 1) \quad \text{then} \quad E(Z) = \frac{\pi}{4}
\]

\[
E(Z \mid V) = (1 - V^2)^{1/2}
\]

Compare the crude estimator

\[
\frac{4}{n} \sum_{i=1}^{n} Z_i \quad \text{with the conditional expectation} \quad \frac{4}{n} \sum_{i=1}^{n} E(Z_i \mid V_i)
\]

\[
u = \text{rand}(2,200000); \quad z = 4 \times (\text{sum}(u.^2) < 1); \quad \text{theta1} = \text{mean}(z); \quad \text{var}(z) / \text{length}(z)
\]

\[
z2 = 4 \times \text{sqrt}(1 - u(2,:).^2); \quad \text{theta2} = \text{mean}(z2); \quad \text{var}(z2) / \text{length}(z2)
\]

Efficiency gain is about 3.3
Combining conditioning and stratified sample

We estimated

\[ \pi = 4 \int_0^1 (1 - u^2)^{1/2} \, du \]  using the conditional expectation estimator

\[ \frac{4}{n} \sum_{i=1}^{n} (1 - V_i^2)^{1/2} \]. What if we use a stratified sample with one observation in each interval \(( \frac{i - 1}{n}, \frac{i}{n} ), i = 1, 2, ..., n \). Result is \( \frac{4}{n} \sum_{i=1}^{n} (1 - (\frac{i - 1 + V_i}{n})^2)^{1/2} \)

\[ n=500000; \text{u=rand}(1,n); \text{w}=((0:((n-1)))+u)/n; \]
\[ z3= 4*\text{sqrt}(1-w.^2); \text{theta3=mean(z3)}; \pi=\text{theta3} \]

Accurate to 8 decimals!
IMPORTANCE SAMPLING

• Idea: Why not weight observations in different parts of the sample space differently?

• e.g. interested in behaviour of a communication system or network under heavier than usual loads. Generate simulation under heavy loads and adjust the results for the difference.
Importance Sampling

• Can I increase the probability of the samples that most effect my estimator without creating bias?

• Use importance sampling when:
  – we want to find \( \int f(u)du \)
  – We can find a probability density function \( g(u) \) such that it is easy to generate random variables with this probability density function and \( g \) is \textit{roughly proportional to} \( f \) so that \( f(u)/g(u) \) is reasonably close to constant.
Importance II

• to find an appropriate importance sampling distribution, start with a function $g$ similar to $f$ that is non-negative and easily integrated. Then multiply this function by a constant to obtain a p.d.f. Then:

$$\theta = \int f(u)du = \int \frac{f(z)}{g(z)}g(z)dz = E\left[\frac{f(Z)}{g(Z)}\right]$$

Where $Z$ is a random variable having probability density function $g(z)$. We should therefore estimate $\theta$ using

$$\theta_{IM} = \frac{1}{n} \sum_{i=1}^{n} \frac{f(Z_i)}{g(Z_i)}$$

where the random variable $Z_i$ are generated with probability density function $g(z)$. For example $Z_i = G^{-1}(U_i)$ where $G$ is the c.d.f. of $Z$. 
Variance of Importance sample Estimator

\[ \theta = \int f(u)du = \int \frac{f(z)}{g(z)} g(z)dz \]

\[ \theta_{\text{IM}} = \frac{1}{n} \sum_{i=1}^{n} \frac{f(Z_i)}{g(Z_i)} \] where \( Z_i \) independent with p.d.f. \( g(z) \).

\[ \text{Var}(\theta_{\text{IM}}) = \frac{1}{n} \text{var}\left( \frac{f(Z_i)}{g(Z_i)} \right) \]

More generally to estimate \( E[h(X)] \) where \( X \) has p.d.f. \( f(x) \), use

\[ E[h(X)] = E[h(Z) \frac{f(Z)}{g(Z)}] \] where \( Z \) has p.d.f. \( g(z) \). The estimator is

\[ \theta_{\text{IM}} = \frac{1}{n} \sum_{i=1}^{n} h(Z_i) \frac{f(Z_i)}{g(Z_i)} \] with \( Z_i \) independent, p.d.f. \( g(z) \).

\[ \text{Var}(\theta_{\text{IM}}) = \frac{1}{n} \text{var}[h(Z_i) \frac{f(Z_i)}{g(Z_i)}] \]
Example: importance sampling

- Find $\int_0^1 x^{1.6} e^{-x} \, dx$. The true value is around 0.19113.
- Crude: $u=\text{rand}(1,200000);$
  - $\text{mean}(u.^{1.6}.*\exp(-u)) \% = 0.1916$
  - $\text{var}(u.^{1.6}.*\exp(-u))/200000 \% = 6.7676e-008$
- Importance:

\[
g(x) = 2.6x^{1.6} \text{ for } 0 < x < 1. \quad G(x) = x^{2.6} \quad \text{and } G^{-1}(U) = U^{1/2.6}
\]

\[
\theta_{IM} = \frac{1}{n} \sum_{i=1}^{n} \frac{f(Z_i)}{g(Z_i)} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{2.6} e^{-Z_i} \quad \text{where } Z_i = U_i^{1/2.6}.
\]
Importance sampling example

• $u = \text{rand}(1,200000)$;
• $\theta_{\text{IM}} = \exp(-u.^(1/2.6))./2.6$;
• $\text{mean}(\theta_{\text{IM}})$
• $\text{var}(\theta_{\text{IM}})/200000 \approx 9.3300e-009$
• (about 7 times as efficient as crude)
Importance Sampling in Option Pricing example

• For the call option, try and approximate $f$ with a linear function starting at $(.47, 0)$.

$$g(x) = k(x - .47)^+ \text{ where we choose } k \text{ so this is a p.d.f.}$$

$$c.d.f. \ G(x) = \frac{k}{2} [(x - .47)^+]^2 \text{ so } k = 7.12$$

$$G^{-1}(U) = .47 + .53\sqrt{U} \text{ generates } Z \text{ from the importance sampling density.}$$

Why not use a density function (e.g. a quadratic density) that more closely approximates the function $f$?
Matlab Code for Importance Sample. Stock Option Example.

function [est,v]=importance(f,g,Ginv,u)
% runs a simulation on the function "f" using importance density "g"(both character strings) and inverse c.d.f. "Ginverse"
e.g. [est,v]=importance('fn','importancedens','Ginverses',rand(1,200000));
% outputs all estimators (use mean(est)) and variance of estimator.
% IM is the inverse c.f of the importance distribution c.d.f.
IM= eval(strcat(Ginv,'(u)'));  %=.47+.53*sqrt(U);
%IMdens is the density of the importance sampling distribution at IM
IMdens=eval(strcat(g,'(IM)'));       %2*(IM-.47)/(.53)^2;
FN=eval(strcat(f,'(IM)'));
est=FN./IMdens;       %  mean(est) provides the estimator
v=var(FN./IMdens)/length(IM);   % this is the variance of the estimator

%GIVES  mean(est) = 0.4614  v = 6.4329e-008
EFFICIENCY GAIN OVER CRUDE IS AROUND 35.
Exponential tilting and importance sampling: Estimating the probability of rare events

Suppose we wish to estimate an extreme quantile such as VAR 99.9%. Then we need to estimate the probability that a random variable exceeds some large value. i.e. when

\[ S_{10} > a \quad \text{where} \quad S_{10} = \sum_{i=1}^{10} X_i \quad \text{for independent random variables} \]

\( X_i \) all have probability density function \( f(x) \). This is \( E\{I(S_{10} > a)\} \) and is estimate by the corresponding sample mean.

If we conduct the simulation under this distribution very few of the sums will exceed \( a \). What if we generate \( X_i \) as independent random variables under a different (tilted) probability density function \( g(x) = \frac{e^{tx}}{k} f(x) \) where \( k \) is chosen so that this is a density function and then use importance sampling.
For example, suppose that \( f(x) = e^{-x} \). Then the tilted distribution is
\[
g_t(x) = \frac{1}{k} e^{tx-x} = e^{-x(1-t)} \times (1-t)
\]
since this is the multiple that makes this function a p.d.f. Another exponential distribution.

If we generate \( X_i \) from this p.d.f. it has

expected value \( \theta = \frac{1}{1-t} \). The importancesample estimator is
\[
E\{I(\sum_{i=1}^{n} Z_i > a) \prod_{i=1}^{n} \frac{f(Z_i)}{g_t(Z_i)} \} = E\{I(\sum_{i=1}^{n} Z_i > a) \prod_{i=1}^{n} \theta e^{-tZ_i} \} = \theta^n E\{I(\sum_{i=1}^{n} Z_i > a) e^{-t \sum_{i=1}^{n} Z_i} \}
\]

where \( Z_i \) are independent with \( \text{exponential}(\theta) \) p.d.f.

What is the best value of \( \theta \)? The variance of the importancesample estimator is small when we choose \( t \) so that \( \sum_{i=1}^{n} Z_i \) is approximately \( a \). Choose \( t = 1 - 1/\theta \) so that

\[
E(Z_i) = \theta = a/n
\]
so choose \( \theta \) approximately \( a/n \).
Matlab Code for Example, a=22.6, n=10

- N=500000
- X=exprnd(1,10,N);
- est1=(sum(X)>22.6);
- mean(est1) % 0.001
- var(est1) % 8.8e-004

Using Importance sampling
- Z=exprnd(2.5,10,N);
- S=sum(Z); theta=2.5
- est2=theta^10*exp(-(1-1/theta)*S).*(S>22.6);
- mean(est2) % =0.001
- var(est2) % 5.5e-6, and SE about 0.000003

EFFICIENCY GAIN FROM IMPORTANCE SAMPLING IS ABOUT 180.
Common Random Numbers: estimating a difference

• Suppose we wish to estimate the difference between two systems or values: e.g.
  – Estimate the slope of a function at a point
  – Estimate the difference in performance between two pieces of equipment
  – Estimate the difference between a call option price for $r=0.05$ and $r=0.06$. 
Importance Sampling, Tilting and the Saddlepoint Approximation

The Edgeworth Expansion of the probability density function of the sample mean of i.i.d. random variables takes the form

\[ n(x; \mu, \sigma^2) \{1 + \frac{\kappa}{6\sqrt{n}} [z^3 - 3z] + O(1/n) \} \]

where \( z = \sqrt{n} (x - \mu) / \sigma \) and \( \kappa = E(X - \mu)^3 \), and

\( n(x; \mu, \sigma^2) \) is the usual normal density function approximation.

Notice that at \( x = \mu \), the correction is zero and the normal approximation is accurate to \( O(1/n) \).

Consider tilting the original distribution so as to achieve this...

If we wish to approximate the probability density function \( f(x) \) of \( X \) at the point \( x \) and \( X \) has cumulant generating function \( K(t) = \ln E\{\exp(Xt)\} \), then
Saddlepoint (cont)

the density \( f_\theta(x) = e^{\theta x - K(\theta)} f(x) \) has mean given by \( K'(\theta) \) so if we choose \( \theta \) so that \( K'(\theta) = x \) the mean is \( x \) and the variance \( K''(\theta) \).

The Edgeworth approximation to the density \( f_\theta \)

is \( n(K'(\theta), K'(\theta), K''(\theta)) = \frac{1}{\sqrt{2\pi K''(\theta)}} \)

Therefore the (saddlepoint) approximation to \( f(x) \) is

\[
\frac{1}{\sqrt{2\pi K''(\theta)}} e^{K(\theta) - \theta x}, \text{ where } K'(\theta) = x.
\]
Two steps of saddlepoint

• Saddlepoint approximation has two steps
  – Tilt the original density so that mean of tilted distribution is around the point of interest $x$
  – Use the normal approximation to the tilted distribution.

• Importance sampling from the tilted density avoids the second step.
Tilting and large deviations (Robert and Casella p 132)

Large deviations results concern the tail of a distribution of partial sums. e.g.

\[ P[S_n > x] \] when this probability is very small.

**Cramer's Theorem.**

Put \( I(x) = \theta x - K(\theta) \) where again \( \theta \) is such that \( K'(\theta) = x \). Then

\[
\frac{1}{n} \ln(P[S_n > x]) \sim -I(x)
\]
Importance Sampling (Robert and Casella 85-87)

• Suppose $X$ as the student t distribution with 12 degrees of freedom. Estimate

$$E\left( \frac{X^5}{1+(X-3)^2} I(X \geq 0) \right)$$

using crude Monte Carlo and importance sampling and various choices of the importance distribution including Cauchy, $U[0,1/2.1]$, and normal distribution. Which do you prefer and why?
Reducing Variance when estimating the difference of Expected values

In general we wish to estimate

\[ E(h_1(X)) - E(h_2(Y)) \]

where \( X, Y \) have cumulative distribution functions \( F_X \) and \( F_Y \) respectively. Then

\[
\text{var}[h_1(X) - h_2(Y)] = \text{var}[h_1(X)] + \text{var}[h_2(Y)] - 2\text{cov}(h_1(X), h_2(Y))
\]

and this is small if \( \text{cov}(h_1(X), h_2(Y)) \) is large.

When functions \( h_1(X), h_2(Y) \) are both increasing functions we can arrange a large covariance if we use the SAME uniform to generate \( X \) and \( Y \) i.e.

\[ X = F_X^{-1}(U), \quad Y = F_Y^{-1}(U). \]

Opposite of antithetic.
Theorem: Common and Antithetic Random Numbers

Theorem: Suppose $h_1(x), h_2(y)$ are both increasing or both decreasing functions. Subject to the constraints that $X, Y$ have given cumulative distribution functions $F_X$ and $F_Y$ respectively, then the covariance $\text{cov}(h_1(X), h_2(Y))$ is maximized when we use common random numbers, i.e. $X = F_X^{-1}(U), \ Y = F_Y^{-1}(U)$ and minimized when we use antithetic random numbers: $X = F_X^{-1}(U), \ Y = F_Y^{-1}(1-U)$
Comparing estimators for Call Option Pricing Example

• script9
Combining Monte Carlo Estimators

• If I have many MC estimators, with/without various variance reduction techniques, which should I choose?
Combining Estimators

• Suppose I have \( m \) unbiased estimators all of the same parameter \( \theta \)
• Put these estimators in a vector \( Y \)

\[
Y = (Y_1, Y_2, \ldots, Y_m)'
\]

so that \( E(Y) = 1 \theta \) where 1 represents the vector \((1, 1, \ldots, 1)\) of length \( m \).

Any linear combination of these estimators with coefficients that add to one is also an unbiased estimator of the parameter \( \theta \).

Which such linear combination is best?
Best linear combination of estimators.

Q: What linear combination of these estimators does the best job of estimating $\theta$.

A: The best linear combination is $\sum_i b_i Y_i$ where $\sum b_i = 1$ and $b_i$ is proportional to $\sum_{j=1}^m C_{ij}$ and where $C = V^{-1}$ is the inverse of the covariance matrix $V$ of $Y$.

$V_{ij} = \text{cov}(Y_i, Y_j)$. 
Estimating Covariance

When we have many independent values of these estimators, (from \( n \) simulations, e.g. \( Y_{ij}, j = 1,2,...,n \) are replicated values of \( Y_i \)), we may estimate variance-covariance matrix using sample covariance:

Estimate \( V_{ij} \) by

\[
\frac{1}{n-1} \sum_{k=1}^{n} (Y_{ik} - \bar{Y}_i)(Y_{jk} - \bar{Y}_j)
\]
Theorem on Optimal Linear Combination of estimators

*Theorem:* (Best Linear Combination of Estimators)  
The linear combination of estimators $Y_i, i = 1, 2, \ldots, m$ of the form 

$$
\sum_{i=1}^{m} b_i Y_i \text{ where the vector } b \text{ is given by }
$$

$$
b' = (1'V^{-1}1)^{-1} 1'V^{-1}. \text{ Here } 1 \text{ is the column vector of } m \text{ ones.}
$$

The variance of the resulting estimator is 

$$
b'V^{-1}b = \frac{1}{1'V^{-1}1}.
$$
Example: combining the estimators of the call option price

Consider the following estimators:

\[ Y_1 = \frac{0.53}{2} \left( f(0.47 + 0.53U) + f(1 - 0.53U) \right) \text{(antithetic)} \]

\[ Y_2 = \frac{0.37}{2} \left[ f(0.47 + 0.37U) + f(0.84 - 0.37U) \right] + \frac{0.16}{2} \left[ f(0.84 + 0.16U) + f(1 - 0.16U) \right] \]

This is stratified into [.47,.84] and [.84,1] and uses antithetic within strata, common U between strata.

\[ Y_3 = 0.37 f(0.47 + 0.37U) + 0.16 f(1 - 0.16U) \]

stratified, antithetic between strata
Example: (cont)

\[ Y_4 = \int g(u)du + [f(U) - g(U)] \text{(control variate)} \]

where \( g(u) = 6[(u - .47)^+]^2 + (u - .47)^+ \)

\[ Y_5 = \frac{f(Z)}{g(Z)} \text{ where } Z = .47 + .53\sqrt{U} \text{ (importance sampling)} \]

Generate simulated values of all five estimators \( Y_1, ..., Y_5 \) using the same uniform.

Do this repeatedly for \( n \) values of \( U \). Obtain the individual estimators and the covariance matrix \( V \). The best linear combination is

\[ \sum_{i=1}^{5} b_i \overline{Y_i} \]

where the vector \( b \) is given by \( b = \frac{1}{1'V^{-1}1} V^{-1}1 \). \( (b \) is proportional to the sum of the rows of \( V^{-1} \) - rescaled so \( \sum b_i = 1 \))
MATLAB function OPTIMAL

function [o,v,b,t1]=optimal(U)
% generates optimal linear combination of five estimators and outputs
% average estimator and variance.
% t1=cputime;
Y1=(.53/2)*(fn(.47+.53*U)+fn(1-.53*U)); t1=[t1 cputime];
Y2=.37*.5*(fn(.47+.37*U)+fn(.84-.37*U))+.16*.5*(fn(.84+.16*U)+fn(1-.16*U));
t1=[t1 cputime];
Y3=.37*fn(.47+.37*U)+.16*fn(1-.16*U); t1=[t1 cputime];
intg=2*(.53)^3+.53^2/2; Y4=intg+fn(U)-GG(U); t1=[t1 cputime];
Y5=importance('fn','importancedens','Ginverse',U); t1=[t1 cputime];
X=[Y1' Y2' Y3' Y4' Y5'];
mean(X)
V=cov(X); Z=ones(5,1); C=inv(V); b=C*Z/(Z'*C*Z);
o=mean(X*b); % this is mean of the optimal linear combinations
% t1=[t1 cputime];
v=1/(Z'*V*b);
t1=diff(t1); % these are the cputimes of the various estimators.
Results for option pricing

• \([o,v,b]=\text{optimal}(\text{rand}(1,100000))\)
• Estimators = 0.4619  0.4617  0.4618  0.4613  0.4619
• \(o = 0.46151 \% \text{ best linear combination (true value}=0.46150)\)
• \(v = 1.1183\text{e}-005 \% \text{variance per uniform input}\)
• \(b’ = -0.5503  1.4487  0.1000  0.0491  -0.0475\)
Efficiency of Optimal Linear Combination

- Efficiency gain based on number of uniform random numbers $0.4467/0.00001118$ or about 40,000.
- However, one uniform generates 5 estimators requiring 10 function evaluations.
- Efficiency based on function evaluations approx 4,000
- A simulation using 500,000 uniform random numbers; 13 seconds on Pentium IV (2.4 Ghz) equivalent to twenty billion simulations by crude Monte Carlo.