ANTITHETIC VARIABLES, CONTROL VARIATES

Variance Reduction Background: the simulation error estimates for some parameter $\theta \approx \bar{X}$, depend on $Var(\bar{X}) = Var(X)/n$, so the simulation can be more efficient if $Var(X)$ is reduced.

Antithetic Variables

- Key idea: if X_1 and X_2 are id RVs with mean θ,

$$Var(\frac{X_1 + X_2}{2}) = \frac{1}{4}(\ Var(X_1) + Var(X_2) + 2\ Cov(X_1, X_2)\ ),$$

  so variance is reduced if X_1 and X_2 have $Cov(X_1, X_2) \leq 0$.

- For many simulations, a θ estimator is $X_1 = h(U_1, \ldots, U_n)$ (with uniform U_i's) for some h; so consider the **antithetic** estimator $X_2 = h(1 - U_1, \ldots, 1 - U_n)$.

  The **combined estimator** is $(X_1 + X_2)/2$.

  a) the simplest example has $h(x) = x$, so if $X_1 = U$, then $X_2 = 1 - U$, and $(X_1 + X_2)/2 = 1/2$, with $Var((X_1 + X_2)/2) = 0$ (perfect negative correlation).

  b) Theorem: if $h(X_1, \ldots, X_n)$ is monotone for each variable, then

$$Cov(h(U_1, \ldots, U_n), h(1 - U_1, \ldots, 1 - U_n)) \leq 0.$$

  c) if $h(\mathbf{Y})$ is monotone, and Y_is are not iid uniform, but $Y_i = F_i^{-1}(U_i),\ i = 1, \ldots, n$, then $X(\mathbf{U}) = h(F_1^{-1}(U_1), \ldots, F_n^{-1}(U_n))$ is monotone in U_i's, so use

$$W = (X(\mathbf{U}) + X(1 - \mathbf{U}))/2.$$

ANTITHETIC VARIABLES CONTINUED

Antithetic Variable Examples

1. Estimate the integral $V = \int_0^\infty e^{-x} \ln(1+x^2)dx$ with MC. Use $t = 1 - e^{-x}$:

Matlab for MC method with antithetic variates:

```
N = 1000; g = @(t)log(1+log(1-t).^2); T = rand(1,N); X = g(T);
disp([mean(X) std(X) 2*std(X)/sqrt(N)])   % simple MC
    0.66848      0.73278     0.046345
T = rand(1,N/2); X = ( g(T) + g(1-T) )/2;
disp([mean(X) std(X) 2*std(X)/sqrt(N/2)]) % antithetic vars
    0.67881      0.30933     0.027667
```

Could also use $x = t/(1-t)$, with $dx = dt/(1-t)^2$.

Matlab for MC method with antithetic variates:

```
f = @(x)log(1+x.^2).*exp(-x); h = @(t)f(t./(1-t))./(1-t).^2;
T = rand(1,N); X = h(T);
disp([mean(X) std(X) 2*std(X)/sqrt(N)])   % simple MC
    0.6695      0.71206     0.045034
T = rand(1,N/2); X = ( h(T) + h(1-T) )/2;
disp([mean(X) std(X) 2*std(X)/sqrt(N/2)]) % antithetic vars
    0.67492      0.46416     0.041516
```

ANTITHETIC VARIABLE EXAMPLES CONT.

2. Estimate $V = \int_0^{\pi/4}\int_0^{\pi/4} x^2y^2\sin(x+y)\ln(x+y)dxdy$ with MC; notice integrand is monotone increasing in x and y. Use $x = \pi u_1/4$, $y = \pi u_2/4$:

Matlab for MC method with antithetic variates:

```
f = @(x)prod(x).^2.*sin(sum(x)).*log(sum(x));
N = 1000; U = rand(2,N); X = pi^2*f(pi*U/4)/16;
disp([mean(X) std(X) 2*std(X)/sqrt(N)])   % simple MC
    0.0037452     0.012534   0.00079271
U = rand(2,N/2); X = pi^2*( ( f(pi*U/4) + f(pi*(1-U)/4) )/2 )/16;
disp([mean(X) std(X) 2*std(X)/sqrt(N/2)]) % antithetic vars
    0.0037885    0.0085332   0.00076323
```

ANTITHETIC VARIABLE EXAMPLES CONT.

3. Simulating a queueing system with N arrivals.
   Two-servers in parallel example.
   Inter-arrival times for N arrivals are $-\ln(1-U_i)/\lambda$, and server times are $G_1^{-1}(V_i)$ or $G_2^{-1}(V_i)$ for $i = 1, \ldots, N$, with $U_i, V_i \sim Uniform(0,1)$.
   Output (e.g. average time in system) is a function $D(U_1, \ldots, U_N, V_1, \ldots, V_N)$, which should be monotone increasing in Us and Vs.
   Matlab results for 2-server parallel queue, ave. time in system, with "prllsv" function modified to allow U's and V's as inputs:

```
N = 1000; M = 100;
for i = 1 : M, U = rand(1,N); V = rand(1,N);
  D(i) = prllsvp(N,U,V,6,4,3);
end
disp([mean(D) std(D) 2*std(D)/sqrt(M)])% simple MC
      0.99208      0.36936     0.073871
for i = 1 : M/2, U = rand(1,N); V = rand(1,N);
  A(i)= ( prllsvp(N,U,V,6,4,3) +  prllsvp(N,1-U,1-V,6,4,3) )/2;
end
disp( [mean(A) std(A) 2*std(A)/sqrt(M/2)])
       1.0254      0.20094     0.056835
```

CONTROL VARIATES

Control Variates

- Background: assume the desired simulation quantity is $\theta = E[X]$;
  but there is another simulation RV Y with *known* $\mu_Y = E[Y]$.
  For any c, the RV $Z = X + c(Y - \mu_Y)$, is an unbiased estimator of θ,
  because $E[Z] = E[X] + c(E[Y] - \mu_Y) = \theta$.
  Now $Var(Z) = Var(X + cY) = Var(X) + c^2 Var(Y) + (2c)Cov(X, Y)$
  is minimized when $c = c^* = -Cov(X, Y)/Var(Y)$, and

$$Var(Z) = Var(X + c^*Y) = Var(X) - Cov(X, Y)^2/Var(Y).$$

  Y is called a **control variate** for X, with $Var(Z) \leq Var(X)$.
  In order to enhance variance reduction, choose a Y correlated with X.

- Implementation: $Cov(X, Y)$, $Var(Y)$ and $c*$, are estimated from the data.

$$Cov(X, Y) \approx \hat{Cov}(X, Y) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}),$$

$$Var(Y) \approx \hat{Var}(Y) = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2, \quad c^* \approx \hat{c}^* = -\frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2},$$

- Note: the goal is to choose Y so that Y is correlated with X,
  with Y easy to simulate and μ_Y easy to find.

CONTROL VARIATES CONTINUED

Control Variate Examples

1. Simple example: to estimate $\theta = \int_0^1 e^x dx$,
   try $Y = U$, so $\mu_Y = 1/2$, $Var(U) = 1/12$; also
   $Var(e^U) = \int_0^1 e^{2x} dx - (\int_0^1 e^x dx)^2 = (e^2 - 1)/2 - (e-1)^2 \approx .242$;
   $Cov(e^U, U) = \int_0^1 x e^x dx - \int_0^1 x dx \int_0^1 e^x dx = 1 - (e-1)/2 \approx .141086$
   $Var(X + c^* Y) = Var(X) - Cov(X,Y)^2/Var(Y) \quad \approx .0039.$

   Matlab for MC method with control variates:

```
N = 1000; U = rand(1,N); X = exp(U);
disp([mean(X) std(X) 2*std(X)/sqrt(N)])   % simple MC
      1.7488      0.49766     0.031475
Y = U; muY = 1/2; Xb = mean(X); Yb = mean(Y);
cs = -sum( (X-Xb).*(Y-Yb) )/sum( (Y-Yb).^2 );
Z = X + cs*( Y - muY );
disp([mean(Z) std(Z) 2*std(Z)/sqrt(N)])   % control variate MC
      1.7189     0.063119     0.003992
```

CONTROL VARIATES CONTINUED

2. Estimate $V = \int_0^2 e^{-x^2} dx = 2\int_0^1 e^{-(2u)^2} du$. Try $Y = 2e^{-2U}$:

Matlab for MC method with control variates:

```
N = 1000; U = rand(1,N); X = 2*exp(-(2*U).^2);
disp([mean(X) std(X) 2*std(X)/sqrt(N)])  % simple MC
     0.91666      0.68842      0.04354
Y = 2*exp(-2*U); muY = 1 - exp(-2);
Xb = mean(X); Yb = mean(Y);
cs = -sum( (X-Xb).*(Y-Yb) )/sum( (Y-Yb).^2 );
Z = X + cs*( Y - muY );
disp([mean(Z) std(Z) 2*std(Z)/sqrt(N)])  % control variate MC
     0.88474      0.13402    0.0084762
```

Note: actual $V \approx 0.88208$.

CONTROL VARIATE EXAMPLES CONT.

3. Estimate $V = \int_0^{\pi/4}\int_0^{\pi/4} x^2y^2\sin(x+y)\ln(x+y)dxdy$;
   Use x^2y^2 to "approximate" the integrand.

Matlab for MC method with control variates:

```
N = 1000; U = rand(2,N);
f = @(x)prod(x).^2.*sin(sum(x)).*log(sum(x));
X = pi^2*f(pi*U/4)/16; Xb = mean(X);    % simple MC
disp([mean(X) std(X) 2*std(X)/sqrt(N)])
   0.0035267     0.011023   0.00069718
g = @(x)prod(x).^2; muY = (pi/4)^6/9;
Y = pi^2*g(pi*U/4)/16; Yb = mean(Y);
cs = -sum( (X-Xb).*(Y-Yb) )/sum( (Y-Yb).^2 );
Z = X + cs*( Y - muY );
disp([mean(Z) std(Z) 2*std(Z)/sqrt(N)])  % control variate MC
   0.0037289    0.0043002   0.00027197
```

CONTROL VARIATE EXAMPLES CONT.

4. Asian option: this has $S_m = S_{m-1}e^{(r-\frac{\sigma^2}{2})\delta+\sigma\sqrt{\delta}Z}$, with $\delta = T/M$, $Z \sim Normal(0,1)$ and expected profit $E[e^{-rT}\max(\frac{1}{M}\sum_{i=1}^{M} S_i(\mathbf{Z}) - K, 0)]$.
   Control variate is $Y = e^{-rT}\max((\prod_{i=0}^{M} S_i(\mathbf{Z}))^{\frac{1}{M+1}} - K, 0)]$,
   with $\mu_Y = e^{-rT}(\ e^{(r-\frac{\sigma^2}{6})\frac{T}{2}}S_0\Phi(d) - K\Phi(d - \sigma\sqrt{\frac{T}{3}})\ )$,
   $d = (\ln(S_0/K) + (r + \sigma^2/6)T/2)/(\sigma\sqrt{T/3})$.

Matlab for MC method with control variates:

```
S0 = 50; K = 50; M = 16; T = 1; dlt = T/M; r = 0.05; s = 0.1;
rd = ( r - s^2/2 )*dlt; N = 10000; Z = randn(M,N);
S = S0*exp(cumsum(rd + s*sqrt(dlt)*Z)); % simple MC
X = exp(-r*T)*max( mean(S)-K, 0 );
disp([mean(X) std(X) 2*std(X)/sqrt(N)])
     1.9522       2.2445     0.044889
d = ( log(S0/K) + (r+s^2/6)*T/2 )/(s*sqrt(T/3));
Y = exp(-r*T)*max((S0*prod(S)).^(1/(M+1))-K,0);
Xb = mean(X); Yb = mean(Y); cs = -sum((X-Xb).*(Y-Yb))/sum((Y-Yb).^2);
muY = exp(-r*T)*( S0*exp((r-s^2/6)*T/2)*normcdf(d) ...
                        - K*normcdf(d-s*sqrt(T/3)) );
C = X + cs*( Y - muY );
disp([mean(C) std(C) 2*std(C)/sqrt(N)]) % control variate MC
    1.9375      0.03389    0.0006778
```

ASIAN OPTION RESULTS CONT.

Matlab for MC method with control variates for M = 200:

```
S0 = 50; K = 50; M = 200; T = 1; dlt = T/M; r = 0.05; s = 0.1;
rd = ( r - s^2/2 )*dlt; N = 10000; Z = randn(M,N);
S = S0*exp(cumsum(rd + s*sqrt(dlt)*Z)); % simple MC
X = exp(-r*T)*max( mean(S)-K, 0 );
disp([mean(X) std(X) 2*std(X)/sqrt(N)])
     1.8269       2.1632     0.043264
d = ( log(S0/K) + (r+s^2/6)*T/2 )/(s*sqrt(T/3));
SP = [S0*ones(1,N); S].^(1/(M+1)); Y = exp(-r*T)*max(prod(SP)-K,0);
Xb = mean(X); Yb = mean(Y); cs = -sum((X-Xb).*(Y-Yb))/sum((Y-Yb).^2);
muY = exp(-r*T)*( S0*exp((r-s^2/6)*T/2)*normcdf(d) ...
                        - K*normcdf(d-s*sqrt(T/3)) );
C = X + cs*( Y - muY );
disp([mean(C) std(C) 2*std(C)/sqrt(N)]) % control variate MC
       1.8303     0.030711   0.00061421
```