MATH 416/516 MIDTERM EXAM 2 Solution Key

1. (10 pts.) To estimate $\theta = E[X]$, $X_1, X_2, \ldots X_{100}$ have been simulated with sample variance $S^2 = 1$. Approximately how many more simulation runs would be needed if we want the standard deviation of the estimator for $E[X]$ to be less than .05. In this case, what would an approximate 95% confidence interval for θ be?
Answer: using $S = 1$, you need the sample standard deviation $\frac{S}{\sqrt{n}} = \frac{1}{\sqrt{n}} \approx .05$, or $n \approx 20^2 = 400$, so approximately 300 more runs are needed. If the sample mean with $n = 400$ is $\bar{X}$, an approximate 95% confidence interval for θ is $[\bar{X} - 2(.05), \bar{X} + 2(.05)] = [\bar{X} - .1, \bar{X} + .1]$.

2. (12 pts.) Explain in detail how antithetic variables would be used for a simulation to estimate
$$\theta = \int_0^1 \int_0^1 e^x \sin((x^2 + y^3)/2) dx dy.$$
Explain why antithetic variables should reduce variance for this problem, compared to raw simulation.
Answer: letting $f(x, y) = e^x \sin((x^2 + y^3)/2)$, an antithetic variables estimator for θ is
$$\theta \approx \frac{1}{N} \sum_{i=1}^{N} \frac{f(U_i, V_i) + f(1 - U_i, 1 - V_i)}{2},$$
with $U_i, V_i \sim Uniform(0, 1)$. This estimator should have smaller variance than the raw estimator because $f(x, y)$ is monotonic increasing in both variables, for $0 \le x \le 1, 0 \le y \le 1$.

3. (20 pts.) Suppose Y is a uniform(0,1) random variable, and conditional on $Y = y$, X is a normal random variable with mean y and variance 9.

   - Explain in detail how raw simulation would be used to estimate $\theta = P\{X > 2\}$.
   Answer: a normal RV X with mean Y and variance 9 is $X = Y + 3Z$ where $Z \sim Normal(0, 1)$, so raw simulation would use
   $$\theta \approx \frac{1}{N} \sum_{i=1}^{N} I(Y_i + 3Z_i > 2),$$
   with $Y_i \sim Uniform(0, 1), Z_i \sim Normal(0, 1)$, and $I(\cdot)$ is the indicator function.
   - Explain in detail how conditional expectation would be used to obtain an improved estimator for θ.
   Answer: $P\{X > 2\} = P\{Y + 3Z > 2|Y\} = P\{Z > \frac{2-Y}{3}|Y\} = 1 - \Phi(\frac{2-Y}{2}$, so conditional simulation would use
   $$\theta \approx \frac{1}{N} \sum_{i=1}^{N} \Big(1 - \Phi\Big(\frac{2 - Y_i}{3}\Big)\Big),$$
   with $Y_i \sim Uniform(0, 1)$. An alternate (integral) derivation uses
   $$\theta = \int_0^1 \int_2^\infty \frac{e^{-\frac{(x-y)^2}{2(9)}}}{3\sqrt{2\pi}} dx dy,$$
   with the standardized $z = \frac{x-y}{3}$, to show
   $$\theta = \int_0^1 \int_{\frac{2-y}{3}}^\infty \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}} dz dy = \int_0^1 \Big(1 - \Phi\Big(\frac{2 - y}{3}\Big)\Big) dy.$$

4. (25 pts.) Assume simulation is used to estimate $\theta = E[X]$ for some output variable X, and another output variable Y is available with $E[Y] = \mu_Y$ known.

- If $W = X + c(Y - \mu_Y)$, show that $E[W] = E[X]$.
  Answer: $E[W] = E[X+c(Y-\mu_Y)] = E[X]+c(E[Y]-\mu_Y) = E[X]+c(\mu_Y-\mu_Y) = E[X]$.
- By considering $Var(W) = Var(X + c(Y - \mu_Y))$, show that
  $Var(W) = Var(X) + c^2Var(Y) + 2cCov(X, Y)$.
  Hint: for RVs X, Y, and constants a, b, you may assume
  $Var(aX + bY) = a^2Var(X) + b^2Var(Y) + 2abCov(X, Y)$.
  Answer: using the hint and the fact that $Var(X + a) = Var(X)$ for constant a
  $Var(W) = Var(X+cY-c\mu_Y) = Var(X+cY) = Var(X)+c^2Var(Y)+2cCov(X, Y)$.
- Show that $Var(W)$ is minimized when $c = c^* = -Cov(X, Y)/Var(Y)$.
  Answer: let $f(c) = Var(X) + c^2Var(Y) + 2cCov(X, Y)$. Then
  $f'(c) = 2cVar(Y) + 2Cov(X, Y) = 0$, when $c = c^* = -Cov(X, Y)/Var(Y)$;
  $f(c^*)$ is minimum because $f''(c^*) = Var(Y) > 0$.
- If $\theta = \int_0^1 e^{-x^2}\cos(x)dx$ describe in detail how raw simulation could be used to estimate θ, and how the control variable $Y = cos(X)$ would be used with simulation to provide a reduced variance estimator for θ.
  Answer: raw simulation would use $\theta \approx \frac{1}{N}\sum_{i=1}^N e^{-U_i^2}\cos(U_i)$, with $U_i \sim Uniform(0, 1)$.
  For control variates with $Y = cos(X)$, $\mu_Y = \int_0^1 cos(x)dx = sin(1)$; using $X_i = e^{-U_i^2}\cos(U_i)$, $Y_i = \cos(U_i)$, c^* is estimated from the data using $c^* = -\frac{\sum_{i=1}^N (X_i-\bar{X})(Y_i-\bar{Y})}{\sum_{i=1}^N (Y_i-\bar{Y})^2}$, and then the control variate estimate for θ is $\theta \approx \bar{W} = \bar{X} + c^*(\bar{Y} - \sin(1))$.

5. (16 pts.) Explain in detail how stratified sampling with 10 equal-width strata and $n = 10000$ total samples would be used with a simulation for $\theta = \int_0^1 e^x \sin(x^2)dx$.
Answer: first divide $[0, 1]$ into 10 subintervals $[\frac{i-1}{10}, \frac{i}{10}], i = 1, \ldots, 10$. In each subinterval i determine 1000 samples $V_{ij} = \frac{i-1}{10} + \frac{U_{ij}}{10}, j = 1, \ldots, 1000$. with all $U_{ij} \sim Uniform(0, 1)$.
Then the stratified sampling estimator for θ is $\Theta = \frac{1}{10}\sum_{i=1}^{10} \bar{X}_i$, with $\bar{X}_i = \frac{1}{1000}\sum_{j=1}^{1000} e^{V_{ij}}\sin(V_{ij}^2)$.
The sample variance $Var(\Theta)$ for this estimator is estimated using

$$Var(\Theta) = \sum_{i=1}^{10}\frac{1}{10^2}Var(\bar{X}_i) \approx \frac{1}{10}\sum_{i=1}^{10}\frac{1}{10}\frac{S_i^2}{1000} \text{ with } S_i^2 = \frac{1}{999}\sum_{j=1}^{1000}(e^{V_{ij}}\sin(V_{ij}^2) - \bar{X}_i)^2.$$

6. (17 pts.) Explain in detail how raw simulation with exponential (mean 1) random variables would be used to estimate $\theta = \int_0^\infty \frac{1}{1+(x-1)^2}e^{-x}dx$. Explain in detail how importance sampling could be used to provide an estimator for θ with reduced variance (hint: use a tilted density).
Answer: for raw simulation $\theta \approx \frac{1}{N}\sum_{i=1}^N \frac{1}{1+(Y_i-1)^2}$, with $Y_i = -\ln(1-U_i), U_i \sim Uniform(0, 1)$.
Importance sampling with a tilted density would sample from $Ce^{tx}e^{-x}$ instead of from e^{-x} so

$$\theta = \int_0^\infty \frac{e^{-xt}}{C(1+(x-1)^2)}(Ce^{-x(1-t)})dx, \text{ with } 1 = C\int_0^\infty e^{-x(1-t)}dx = \frac{C}{1-t}, \text{ and } C = (1-t), t < 1.$$

Then $\theta \approx \frac{1}{N}\sum_{i=1}^N \frac{e^{-Y_it}}{(1-t)(1+(Y_i-1)^2)}$, with $U_i = (1-t)\int_0^{Y_i} e^{-x(1-t)}dx$, so $Y_i = -\ln(1-U_i)/(1-t)$.
The tilting parameter $t < 1$ should be chosen to make $Var(\frac{e^{-xt}}{(1-t)(1+(x-1)^2)})$ small.