Math 350 - Homework 4 - Solutions

1. *If $x_0 = 5$ and*
$$x_n \equiv 3x_{n-1}(mod\ 150)$$
find $x_1, \ldots, x_{10}$.

   Note that this is equivalent to $x_n \equiv 3^n 5(\text{mod } 150)$, $n = 1, \ldots, 10$, which is the remainder in $\{0, 1, \ldots, 149\}$ after dividing $3^n 5$ by 150. These are
$$x_1 = 15, x_2 = 45, x_3 = 135, x_4 = 105, x_5 = 15, x_6 = 45, x_7 = 135, x_8 = 105, x_9 = 15, x_{10} = 45.$$
   Notice that this sequence repeats itself every 4 steps. (In fact, $105 \equiv 3^4 105(\text{mod } 150)$.) It would make for a very poor pseudo-random number generator!

2. *This problem refers to the integral*
$$I = \int_0^\infty x(1+x^2)^{-2} dx.$$
   (a) *Find the exact value of I.*

   The change of variables $u = 1 + x^2$ gives
$$I_R = \int_0^R x(1+x^2)^{-2}\, dx = \frac{1}{2}\int_1^{1+R^2} u^{-2}\, du = \left[-\frac{1}{2u}\right]_1^{1+R^2} = \frac{1}{2} - \frac{1}{2(1+R^2)}.$$
   Therefore,
$$I = \lim_{R\to\infty} I_R = \frac{1}{2}.$$
   (b) *Estimate the error made by approximating I by*
$$I_R = \int_0^R x(1+x^2)^{-2} dx.$$
   How big should R be if we wish $|I - I_R|$ to be less than, say, 10^{-2}?

   It follows from part (a) that the error is given by
$$|I - I_R| = \frac{1}{2(1+R^2)}.$$
   If we want this to be less than 10^{-2} we need (after solving the inequality for R)
$$R > \sqrt{\frac{10^2}{2} - 1} = \sqrt{49} = 7.$$
   Also note (with a view towards part (c)) that if $R = 20$ the error is $|I - I_R| = 1.2 \times 10^{-3}$.

(c) *Use Monte Carlo simulation to approximate the integral and compare your estimate with the exact value. I suggest taking $R = 20$.*

We do this as follows. Write $g(x) = x(1+x^2)^{-2}$ and let $X, X_1, \ldots, X_n$ be independent, uniformly distributed random variable over the interval $[0, R]$, where n is large. (We will discuss later the error incurred in taking n finite.) Then by the law of large numbers

$$\frac{1}{R}\int_0^R x(1+x^2)^{-2}\,dx = E[g(X)] \approx \frac{g(X_1) + \cdots + g(X_n)}{n}.$$

Write $X_j = RV_j$, where $V_1, \ldots, V_n$ are independent random variables uniform over $[0, 1]$. Then

$$I \approx R(g(X_1) + \cdots + g(X_n))/n = R(g(RV_1) + \cdots + g(RV_n))/n.$$

I execute this in Matlab with the commands (I've chosen $n = 10^6$)

```
n=10^6;
R=20;
V=rand(1,n);
X=R*V;
I=R*sum(X.*(1+X.^2).^(-2),2)/n
```

which gives (a typical) value: 0.4977 for the Monte Carlo approximation of I_{20}.

3. *Use Monte Carlo to approximate the value of the double integral*

$$\int_0^1\int_0^1 e^{(x+y)^2}\,dx\,dy.$$

Let X and Y be independent uniform random variables over $[0, 1]$. Then (X, Y) is a uniform random variable over the square $[0, 1] \times [0, 1]$. The Monte Carlo method is based on the identity

$$\int_0^1\int_0^1 e^{(x+y)^2}\,dx\,dy = E\left[e^{(X+Y)^2}\right] \approx \frac{\sum_{j=1}^n e^{(X_j+Y_j)^2}}{n},$$

where (X_j, Y_j) are independent random variables with the same distribution as (X, Y).

This can be implemented in Matlab as

```
m=10^6;
X=rand(1,m);
Y=rand(1,m);
I=sum(exp((X+Y).^2))/m
```

A typical value is $I = 4.8939$.

We can compare this with a numerical approximation of the integral obtained by simple Riemann sum approximation. The following does this. (Look for the very useful command `repmat(A, n, m)` in the Matlab Help facility.)

```
m=10^3;
x=0:1/m:1;
```

```
k=length(x);
X=repmat(x,k,1);
Y=repmat(x',1,k);
I=sum(sum(exp((X+Y).^2)))*(1/m)^2
```

The above gives $I = 4.9156$.

4. *Use simulation to approximate Cov(U, e^U), where U is uniform on $(0,1)$. Compare your approximation with the exact answer.*

   Recall that the covariance of random variables X and Y, having means $\mu_X = E[X]$ and $\mu_Y = E[Y]$, is defined by
$$\text{Cov}(X,Y) := E[(X-\mu_X)(Y-\mu_Y)] = E[XY] - \mu_X\mu_Y.$$

   Before obtaining this by simulation, let us obtain the covariance by direct calculation of the integrals involved. We have $\mu_U = \int_0^1 u\,du = 1/2$, $\mu_{e^U} = \int_0^1 e^u\,du = e-1$, and
$$\text{Cov}(U,e^U) = \int_0^1 ue^u\,du - \frac{e-1}{2} = \frac{3-e}{2} \approx 0.1409.$$

   To do this by simulation, we let $U_1,\dots,U_n$ be independent uniform r.v. over $(0,1)$ and approximate μ_U by $(U_1+\cdots+U_n)/n$, μ_{e^U} by $(e^{U_1}+\cdots+e^{U_n})/n$, and the covariance by $\frac{1}{n}\sum_i^n(U_j-\mu_U)(e^{U_j}-\mu_{e^U})$. This can be implemented in Matlab as follows:

```
n    = 10^6;
U    = rand(1,n);
V    = exp(U);
mu_u = sum(U)/n;
mu_v = sum(V)/n;
J    = ones(1,n);
cov  = sum((U-mu_u*J).*(V-mu_v*J))/n
```

   A typical value of `cov` is 0.1405.

5. *For independent uniform $(0,1)$ random variables $U_1, U_2, \dots$ define*
$$N = \textit{Minimum}\left\{n : \sum_{i=1}^n U_i > 1\right\}.$$

   That is, N is equal to the number of random numbers that must be summed to exceed 1.

   (a) *Estimate $E[N]$ by generating 100 values of N.*

   (b) *Estimate $E[N]$ by generating 1000 values of N.*

   (c) *Estimate $E[N]$ by generating 10000 values of N.*

   (d) *What do you think is the value of $E[N]$?*

   The following Matlab program obtains an approximation of $E[N]$ for m values of N:

```
m=10^3;           %Number of trial values of N
N=zeros(1,m);     %This vector will collect the values of N
```

```
for i=1:m
    L=0;
    n=0;
    while L<=1;
        U=rand(1);
        L=L+U;
        n=n+1;
    end
    N(i)=n;
end
EN=sum(N)/m      %Approximation of E[N] by the trial average.
```

We can run this program for the values $m = 10^2, 10^3, 10^4$. These are (a) $E[N] \approx 2.7700$, (b) $E[N] \approx 2.7260$, (c) $E[N] \approx 2.7292$. (I checked also for $m = 10^7$, which gave the value $E[N] \approx 2.7187$, while $e \approx 2.7183$.)

A reasonable (but surprising) guess is that the exact value is e. This guess turns out to be correct, and it is not too hard to prove. The key point is the following observation. Let $U_1, U_2, \dots$ be independent uniform r. v. over $(0, 1)$, $S_n = U_1 + \cdots + U_n$, and define $F_n(s) := P(S_n \leq s)$. Then the probability density function of S_n is $f_n(s) = F_n'(s)$. I claim that if $0 \leq s \leq 1$ then

$$F_n(s) = \frac{s^n}{n!}.$$

This can be shown by induction, based on the identity

$$P(S_{n+1} \leq s) = \int_0^s P(S_{n+1} \leq s | S_n = t) f_n(t)\, dt,$$

where the conditional probability under the integral sign is given by

$$P(S_{n+1} \leq s | S_n = t) = P(S_{n+1} - S_n \leq s - t) = s - t.$$

We now calculate $E[N] = \sum_{n=2}^{\infty} nP(N = 2)$. Observe that $\{N = n\} = \{S_n > 1\} \cap \{S_{n-1} \leq 1\}$, so

$$P(N = n) = P(S_n > 1, S_{n-1} \leq 1) = \int_0^1 P(S_n > 1, S_{n-1} \leq 1 | S_{n-1} = z) f_{n-1}(z)\, dz.$$

The conditional probability under the integral sign is

$$P(S_n > 1, S_{n-1} \leq 1 | S_{n-1} = z) = P(S_n - S_{n-1} > 1 - z) = z$$

and $f_{n-1}(z) = F_{n-1}'(z) = \frac{z^{n-2}}{(n-2)!}$. Therefore,

$$P(N = n) = \int_0^1 \frac{z^{n-1}}{(n-2)!}\, dz = \frac{n-1}{(n!)}.$$

Finally,

$$E[N] = \sum_{n=2}^{\infty} nP(N = n) = \sum_{n=2}^{\infty} \frac{n(n-1)}{(n-2)!} = \sum_{n=0}^{\infty} \frac{1}{n!} = e.$$