Math 350 - Homework 8 - Solutions

1. *For any set of numbers* $x_1, \ldots, x_n$*, prove algebraically that*

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$$

where $\bar{x} = \sum_{i=1}^{n} x_i/n$.

$$\begin{aligned}\sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}^2) &= \sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + n\bar{x}^2 \\ &= \sum_{i=1}^{n} x_i^2 - 2n\bar{x}^2 + \bar{x}^2 n \\ &= \sum_{i=1}^{n} x_i^2 - n\bar{x}^2.\end{aligned}$$

2. *Write a program that uses the recursions given by Equations (7.6) and (7.7) to calculate the sample mean and sample variance of a data set.*

The following program does it.

```
function [s_mean, s_var]=sample_mv(X)
%Imput:  X is a vector of data values. Either a row or column vector.
%
%Output: s_mean is the sample mean of X
%        s_var  is the sample variance of X
%
%Note:   I assume that X has length at least 2.

s_mean = (X(1)+X(2))/2;
s_var  = (X(2)-X(1))^2/2;
m         = length(X);

for j=3:m
    s_mean_old = s_mean;
    s_mean     = s_mean + (X(j) - s_mean)/j;
    s_var      = (1-1/(j-1))*s_var + j*(s_mean - s_mean_old)^2;
end
```

We can test this program by comparing with the mean and variance functions used by Matlab. For example, let X consist of 1000 independent random variables uniformly distributed over $(0, 1)$. Then

```
 X=rand(1,1000);
[s_mean, s_var]=sample_mv(X)

s_mean =

    0.4995

s_var =

    0.0801
```

These are exactly the mean and variance obtained using the Matlab commands `mean(X)` and `var(X)`.

3. *Estimate* $\int_0^1 \exp(x^2)\,dx$ *by generating random numbers. Generate at least* 100 *values and stop when the standard deviation of your estimator is less than* 0.01*.*

   If X is a random variable uniformly distributed over $(0, 1)$, then $E[\exp(X^2)] = \int_0^1 \exp(x^2)\,dx$. By the strong law of large numbers, if $X_1, X_2, \ldots$ are independent random variables with the same distribution as X,

   $$\lim_{n\to\infty} \frac{e^{X_1^2} + \cdots + e^{X_n^2}}{n} = E[\exp(X^2)]$$

   holds with probability 1. Let $Y_i = \exp(X_i^2)$, $\overline{Y}_n = (Y_1 + \cdots + Y_n)/n$, and $\theta = \int_0^1 \exp(x^2)\,dx$. Then $\overline{Y}_n$ is an unbiased estimator of θ.

   We set $d = 0.01$. In the next program, we calculate $\overline{Y}_n$ and the sample standard deviation S_n successively until n becomes big enough that $S_n/\sqrt{n} < d$ and $n \geq 100$. For such an n, the value of $\overline{Y}_n$ satisfies:

   $$|\overline{Y}_n - \theta| < 1.96 \times 0.01 \approx 0.02$$

   with 95% certainty.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
d  = 0.01;
Y  = exp(rand(1)^2);
S2 = 0;
n=1;
while n<100 | S2/n>d^2
    Y_old = Y;
    Y     = Y + (exp(rand(1)^2) - Y)/n;
    S2    = (1-1/n)*S2+(n+1)*(Y-Y_old)^2;
    n=n+1;
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

One run of the program gave the estimate $Y = 1.4673$ of the integral. Therefore, we know that the integral is less than 0.02 away from the value 1.46 with 95% certainty. For this precision, the program generated $n = 2277$ values of Y.

For comparison, the same integral can be approximated by simple Riemann sum as

```
x=0:0.001:1;
sum(exp(x.^2))*0.001

ans =

    1.4645
```

4. *It can be shown that if we add random numbers until their sum exceeds* 1*, then the expected number added is equal to* e*. That is*

$$N = \min\left\{ n : \sum_{i=1}^{n} U_i > 1 \right\}$$

then $E[N] = e$*.*

   (a) *Use this idea to estimate* e*, using* 1000 *simulation runs.*

   (b) *Estimate the variance of the estimator in (a) and give a* 95 *percent confidence interval estimate of* e*.*

```
N = 0;
A = 0;
while A<=1
    A=A+rand(1);
    N=N+1;
end
X  = N; %This is the value of Xbar_1
S2 = 0; %This is the value of S2_1
for k=1:1000
    N = 0;
    A = 0;
    while A<=1
        A=A+rand(1);
        N=N+1;
    end
    X_old = X;
    X     = X + (N-X)/(k+1);
    S2    = (1-1/k)*S2+(k+1)*(X-X_old)^2;
end
% Estimator value:
X
% Estimator variance:
S2/1000
% 95% confidence interval:
[X-1.96*sqrt(S2/1000),X+1.96*sqrt(S2/1000)]
```

A typical set of values is:

```
Estimator: X = 2.7063

Estimator variance: 7.5965e-04

95% confidence interval: [2.6523, 2.7603]
```

5. *Use the approach that is presented in Example 3a of Chapter 3 to obtain an interval of size less than* 0.1*, which we can assert, with* 95 *percent confidence, contains* π*. How many runs were necessary?*

```
X  = 0;
n  = 1;
d  = 1;
while 4*1.96*d>0.05 | n<100
    V1 = 2*rand(1)-1;
    V2 = 2*rand(1)-1;
    R2 = V1^2+V2^2;
    J  = (R2<1);
    X  = X + (J-X)/n;
    d  = sqrt(X*(1-X)/n);
    n  = n + 1;
end
%Estimate of pi:
4*X
%Number of runs:
n
```

One run of this program gave the following values: 95% confidence interval: 3.1581 ± 0.05; and the number of runs: $n = 4087$.