FALL 2018 EE 511 SIMULATION

PROJECT #6

TAMOGHNA CHATTOPADHYAY

## 1. INTRODUCTION

We have an attached dataset which represents a set of $n = 100$ independent samples from some population. We calculate the sample mean and variance of the dataset and store them. We also generate a discrete approximation to the Cumulative Distribution function.

Then we split the data into equal size intervals and generate a discrete approximation to the distribution and determine the values of the PMF for the discrete approximation.

Using the bootstrap technique, we generate M bootstrap sets of samples based on the empirical distribution found, with each set containing n independent samples from the Empirical Distribution (repitition is allowed). We compute the sample mean and sample variance for each sample set. Let them be $m_i^*$ and $s_i^{2*}$ for $i = 1, \ldots, M$.

## 2. THEORY

$m_i^*$ and $s_i^{2*}$ should be very close to value of $m$ and $s^2$, where,

$m \rightarrow$ mean of empirical distribution

$s^2 \rightarrow$ variance of empirical distribution.

To find an estimate of MSE of the sample mean for overall population distribution, we can calculate:

$$MSE(m^*) = \frac{1}{M} \sum_{i=1}^{M} (m_i^* - m)^2$$

Similarly, to find an estimate of MSE of the sample Variance for overall population distribution, we can calculate :

$$MSE(s^{*2}) = \frac{1}{M} \sum_{i=1}^{M} (s_i^{*2} - s^2)^2$$

3. SIMULATION METHODOLOGY

I made use of the 'histogram' function in MATLAB to plot the PMF. I set the edges as

h.BinEdges = [0 : 5 : 50];

and using 'normalization' and 'Probability' inside the function.

The mean and variance functions are used to calculate the sample mean and variance.

For generating the new bootstrap samples, I randomly selected the numbers from the array to be in the bootstrap sample (with replacement). Then the mean and variance of the samples were calculated.

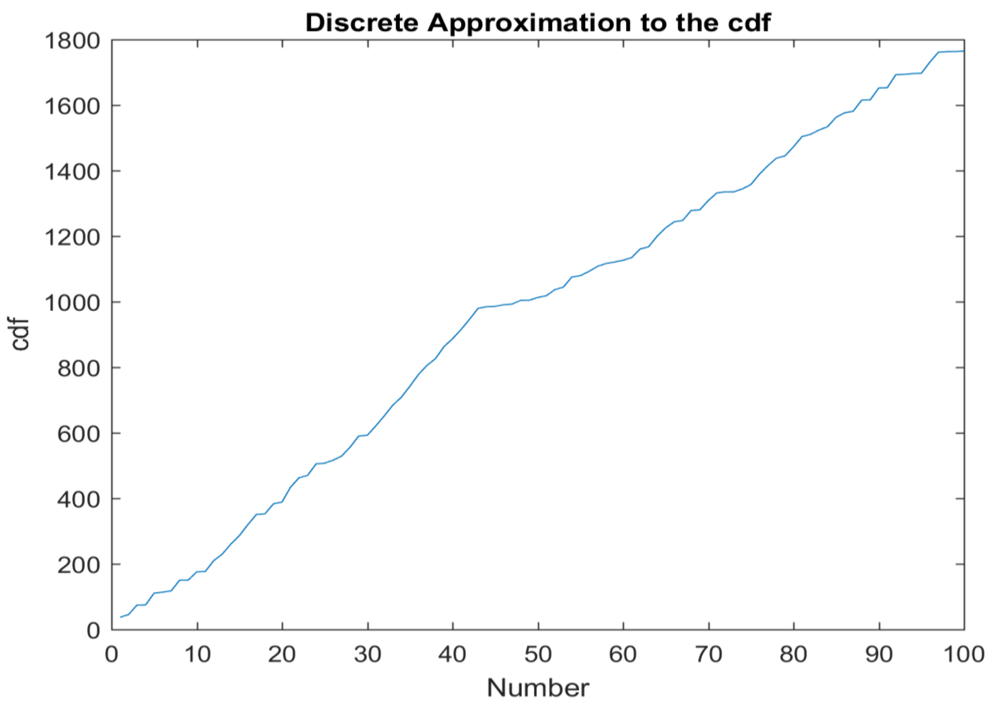
Then, using the formulae mentioned in theory, we calculated the MSE for mean and variance for the bootstrap samples.

## 4. Results and Observations

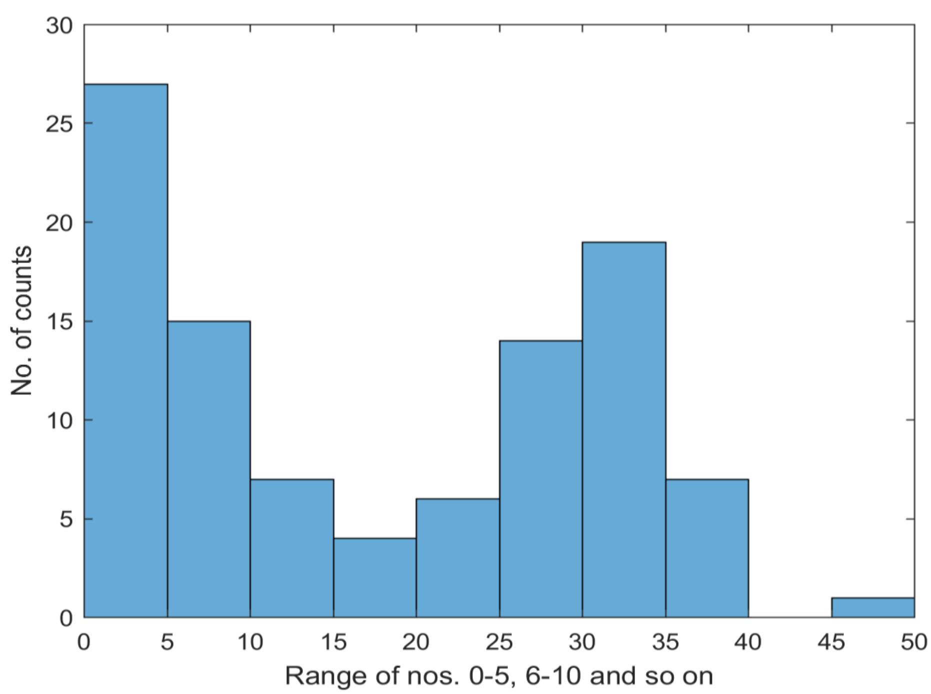
For part (a), Sample Mean = 17.6471

Sample Variance = 177.2323

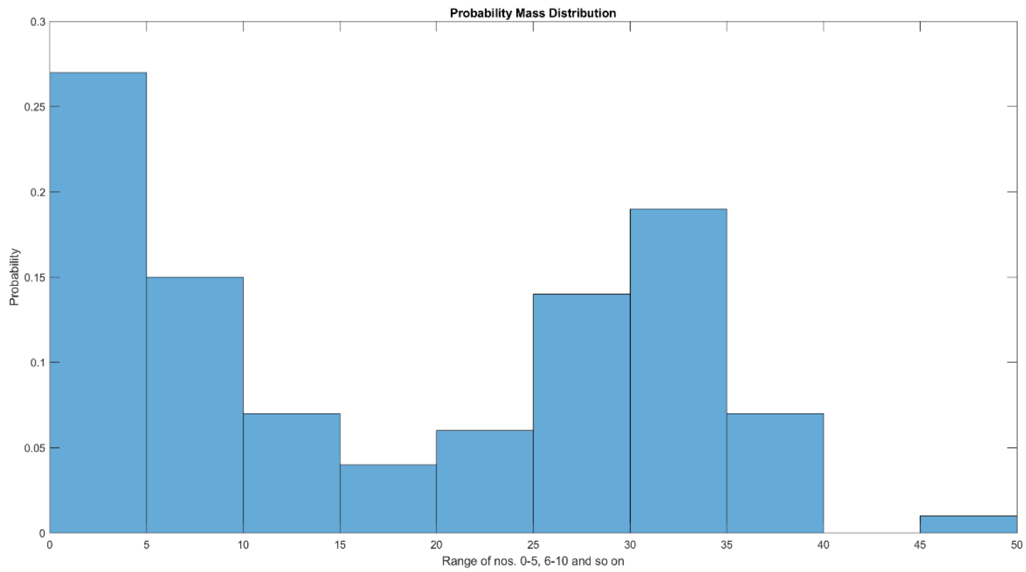
For part (b),




For part (c),



The histogram looks like above of the dataset provided. It shows how much of the data falls in every range of equal size. The PMF, normalized between 0 and 1, looks like below:




For part (d),
I got 100 values of sample means and sample variances for M = 100 and similarly, 50 values of sample means and sample variances for M = 50.

For part (e),
For sample mean,
MSE for M = 50, MSE = 1.4587
MSE for M = 100, MSE = 2.2385

For part (f),
For sample variance,
MSE for M = 50, MSE = 102.4106
MSE for M = 100, MSE = 118.7877

For my experiment, as value of M increases, i.e., as the number of bootstrap samples increase, there is an increase in MSE for both the mean and variance.

For every bootstrapping sample of 100 values, the mean and variance of that sample is very close to the original mean and variance.

## 5. Code

```
clc;
clear all;
```

```
% Part a
Nos = [37.12 8.45 28.96 0.27 36.22 2.78 3.98 32.79 0.14 24.87 1.33 33.25
19.91 30.43 25.84 33.55 31.10 1.86 30.57 5.34 45.39 28.67 7.12 35.38 1.92
9.25 12.55 27.49 33.72 2.30 28.32 30.92 32.62 24.10 33.56 35.62 27.88 20.71
36.62 24.03 28.00 31.44 33.32 5.01 1.30 4.56 2.28 11.33 0.24 8.53 5.27 18.52
7.63 31.03 4.06 12.83 15.43 8.75 4.65 5.21 7.90 26.48 6.81 32.20 25.69 18.18
4.48 30.33 1.68 28.44 23.26 3.35 0.17 8.90 13.29 31.54 26.16 22.79 6.89 27.92
30.99 6.93 13.27 10.08 28.95 13.40 4.57 34.10 0.76 36.40 0.60 39.74 1.11 2.40
1.05 34.10 29.95 1.94 0.16 1.43];
Exp_val = mean(Nos);
Variance = var(Nos);

% Part b
E = zeros(1,100);
E(1) = Nos(1);
for i = 2:100
    E(i) = E(i-1) + Nos(i);
end

figure
plot(E);
xlabel('Number');
ylabel('cdf');
title('Discrete Approximation to the cdf');

% Part c
figure
edges = [0:5:50];
h = histogram(Nos, edges);
Val = h.Values;
xlabel('Range of nos. 0-5, 6-10 and so on');
ylabel('No. of counts');


Val = Val/100;
figure
h = histogram(Nos, 'Normalization', 'probability');
h.BinEdges = [0:5:50];
xlabel('Range of nos. 0-5, 6-10 and so on');
ylabel('Probability');
title('Probability Mass Distribution');

% Part d
M = 50;
sample1 = zeros(M,100);

for i = 1:M
    for j = 1:100

        pos = randi(length(Nos));
        sample1(i,j) = Nos(pos);

    end
end

sample_mean1 = mean(sample1,2);
```

```
sample_var1 = var(sample1.').';

% Part e
sm1 = (sample_mean1 - Exp_val).^2;
S1 = sum(sm1);
MSE1 = S1/M;

% Part f
sv1 = (sample_var1 - Variance).^2;
S2 = sum(sv1);
MSE2 = S2/M;
```

% For M = 50, just change the value of M in the code.