HW 1. Math 504. Spring 2008. CSUF

by Nasser Abbasi

1 Problem description

A number is chosen at random from the interval $[0,1]$. This value is placed in a box, and twice this value is placed in a second box. One of these boxes is selected at random and opened to reveal the number inside. Given this observed value, what is the probability that this number is the smaller of the two.

1. Suppose that when we select a box, and observe the value, we have an opportunity to switch to the other box. The result above suggests that if we observe a value between 0 and 1, then we should switch, and otherwise, hold the value we have. Let the random variable Y denote the reward using such a strategy. Write a simulation program (in MATLAB, say) to estimate the expected value of Y. Use a 95% confidence interval, and determine the sample size so that the relative accuracy of your estimate is about one percent. In your report, explain how you determined your sample size. Also, compare theory and practise; that is, did your confidence interval include the true value.

2 Purpose and design of project

The purpose of this project is to estimate the expected value of a random variable (called Y) which is generated by an experiment that is described in the above problem statement. Each experiment generates one random variable y. The experiment is described well in the above problem statement and no need to repeat it here again.

In addition, we are asked to determine the interval over which we are 95% confident the estimated expected value will lie within. We are asked that the interval should not be wider than 1% of the true mean from either side of the estimated expected value.

The simulation involve a two stage process. In the first stage, an initial simulation was made for $20,000$ experiments in which we obtained an estimate of the population standard deviation s and estimate of the population mean given by the sample mean $\bar{X}$. These 2 values are used to determined the sample size (number of experiments) needed for the second simulation performed to meet the above stated requirement for relative accuracy in expected value of Y. Therefore, once the first simulation is completed, the sample size for the second simulation was found by solving for n (sample size) by setting the expression for the standard error to be 1% of the population mean (in which we are using an estimate of which is $\bar{X}$ as generated by the first simulation). Therefore, we solve for n from

$$1.96\frac{s}{\sqrt{n}} = 0.01\bar{X}$$

Finally, the second simulation was now run using the above computed n, and the confidence interval was found from

$$C.I. = \left\{ \bar{X} - 1.96\frac{s}{\sqrt{n}} \quad \cdots \quad \bar{X} + 1.96\frac{s}{\sqrt{n}} \right\}$$

Where in the above equation the s and $\bar{X}$ are the sample standard deviation and the sample mean resulting from this second simulation (and not the first simulation run used to estimate n).

Next, the histogram Y was plotted to obtain an estimated of the probability density function of Y.

3 Summary of numerical results

For the initial simulation run, we used $20,000$ experiments and obtained the following estimate of the standard deviation and the population mean

$$\begin{aligned} s &= 0.625760 \\ \bar{X} &= 0.931717 \end{aligned}$$

Now solve for n from

$$1.96\frac{s}{\sqrt{n}} = 0.01\bar{X}$$

we found

$$\boxed{n = 17328}$$

Running the second stage simulation now to estimate the expected value of Y we obtain the following result that the estimate of the expected value of Y is

$$\boxed{\bar{X} = 0.9311632}$$

and the 95% confidence interval was found to be

$$\boxed{\{0.92190 \quad \cdots \quad 0.94043\}}$$

4 Discussion of numerical results

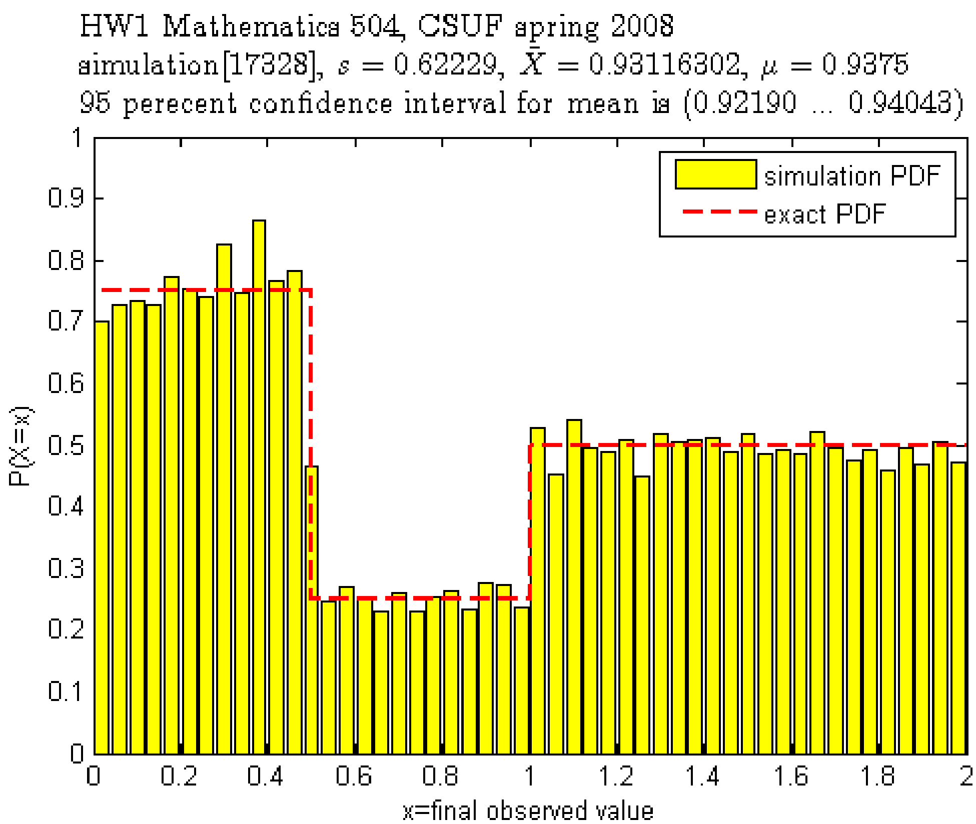
Since we know that the true value of $E(Y) = \frac{15}{16} = 0.937\,5$, we see that the

$$\boxed{\text{95\% confidence interval did contain the true value}}$$

We also notice that the relative error in $\bar{X}$ (the estimate of the expected value) when compared to the true mean $\mu = \frac{15}{16}$ is calculated as $\frac{\mu - \bar{X}}{\mu} = \frac{0.9375 - 0.9311632}{0.9375} = 0.0067593 \simeq 0.7\%$ which is little below the 1% requirement.

We note that the value of the relative error did not come out exactly 1% because we used an estimate of the true mean in order to find the sample size needed for the calculation.

The result of the simulation is the estimate of the PDF of Y which is shown in the plot below. The number of bins used is 50. This was determined by trial and error to obtain the most pleasing looking histogram.



We note that the true PDF is given below (derived in the class) and we see from the above plot that the estimated PDF is very close to the analytical PDF.

$$P(Y = y) = \begin{cases} \frac{3}{4} & 0 \leq y \leq 0.5 \\ \frac{1}{4} & 0.5 \leq y \leq 1 \\ \frac{1}{2} & 1 < x \leq 2 \end{cases}$$