Statistical inference: exponentials simulation

Phil Allen

25 December 2015

Overview

We investigate the exponential distribution in R. By simulation, we compare the mean of independent identically variables with the asymptotic result given by the Central Limit Theorem.

We find that that Central Limit Theorem predicts the variance and distribution of the means fairly well.

Simulations

```
# Set the pseudo-random seed and name the constants given in the question.
set.seed(42)
l=0.2
SS=40
nosim=1000
```

We use a rate $\lambda = 0.2$ so that the population mean and standard deviation are both 5. We investigate the distribution of averages of 40 exponentials, with 1000 trials.

We first sample the exponential distribution into a matrix with 1000 rows and 40 columns. Then apply mean across the columns of each row to obtain a vector of means of 40 exponentials of length 1000.

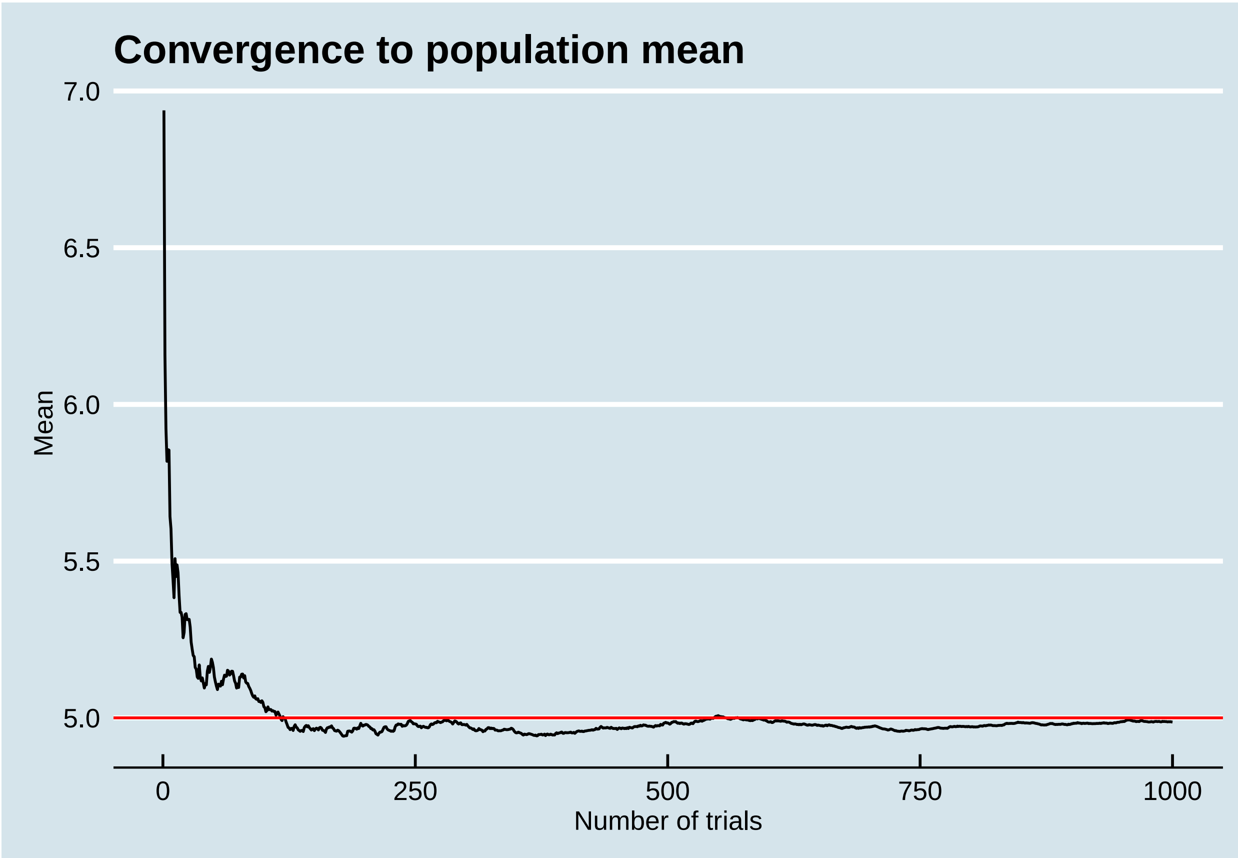
```
obs <- matrix(rexp(SS*nosim,l), nosim) # nosim rows of SS observations
sims <- apply(obs,1,mean)
```

Results

Sample Mean versus Theoretical Mean

The sample mean across the simulation trials is 4.9865, whereas the population mean is 5. It is interesting to see how the sample mean converges to the population mean as the number of trials increases from 1 to 1000.

To demonstrate this, take cumulative means of the subsequences of the sequence of simulations: first value; mean of first and second value: mean of first, second and third value, and so on. The length of subsequence, which is the number of trials, is plotted on the x-axis in the following plot.



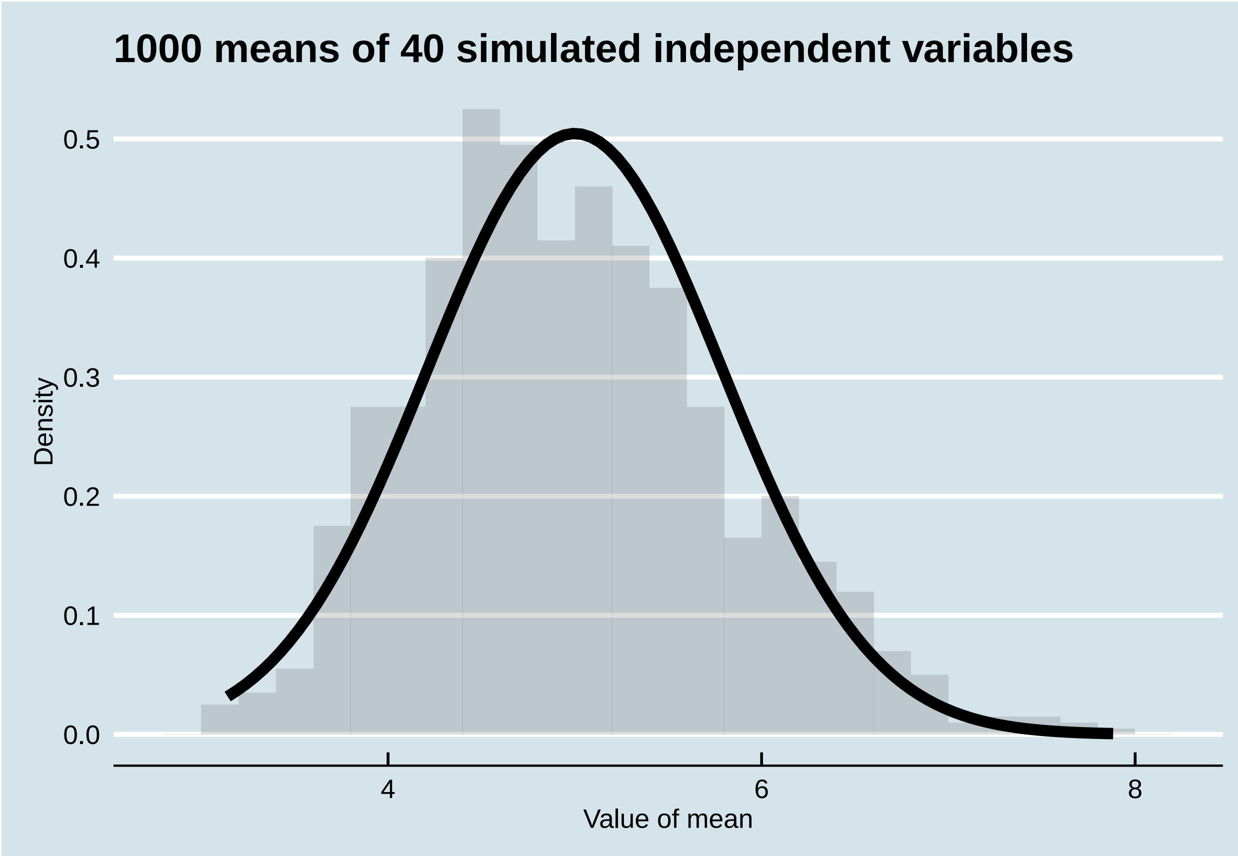
In this case, there is marked divergence from the theortical mean (shown by red horiztonal line) during the first 100 trials.

Sample Variance versus Theoretical Variance

Applying the formula given in the lectures, the *population* variance of the average of 40 independent $exp(0.2)$ variables is $\frac{(1/0.2)^2}{40}$, which is 0.625. In this simulation, the sample variance across the trials is 0.679.

Distribution

We can demonstrate that the distribution of the means is approximately normal by plotting a density histogram of the simulation data, and superimposing the density function of the corresponding normal distribution, $N(5, 5/\sqrt{40})$, as shown by the thick black line.



However, there is some remaining left skew, reflecting the density function of a single exponential variable, which is $\lambda e^{-\lambda x}$ for $x \geq 0$.