Ryan Oliveira

Statistical Genomics (CROPS 545), Spring 2017 Homework #1

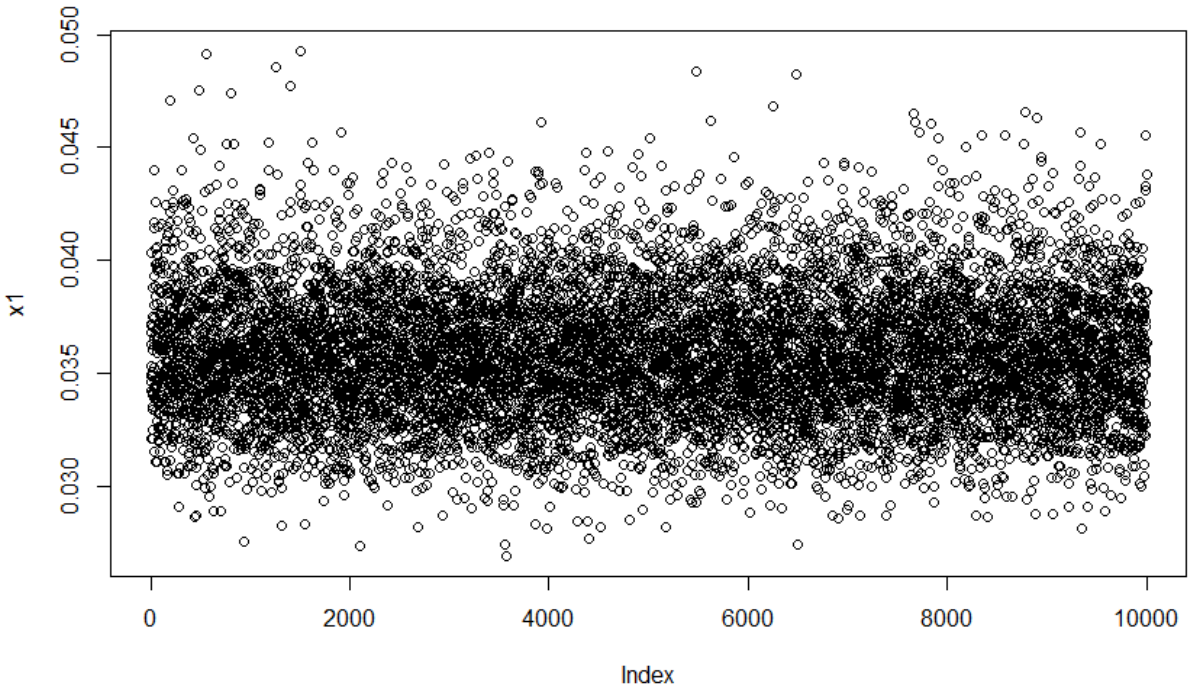
(1) and (2):

I used an underlying standard normal distribution with variance and standard deviation = 1 centered around mean = 0. Samples from this distribution were used as variable "r" in the Inverse Square Law, which states sound intensity (I) = P / π*4*r^2, where P is the power of the sound source and r is distance from the source.

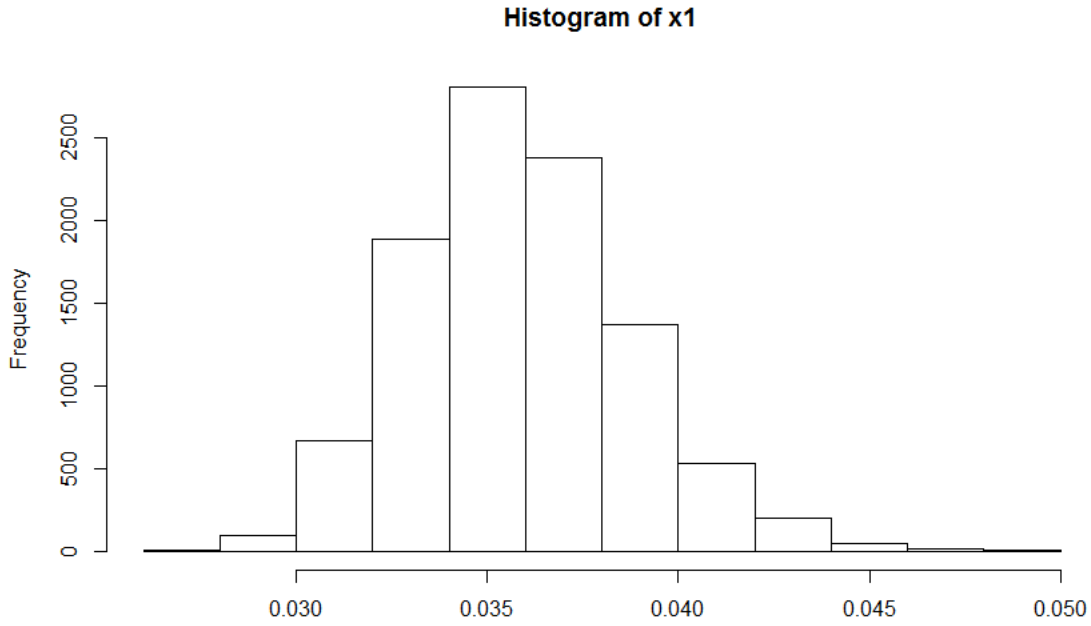
In each iteration, *df* number of observations are taken from the normal distribution, averaged, and the process is repeated *n* number of times to generate the distribution. This function is termed "rOliveira".

When *n* = 10000 observations are plotted, the mean = 0.0359, the variance = 8.17 x 10^{-6}, the standard deviation = 6.94 x 10^{-4}, the skew = 0.384, and the kurtosis = 3.36.

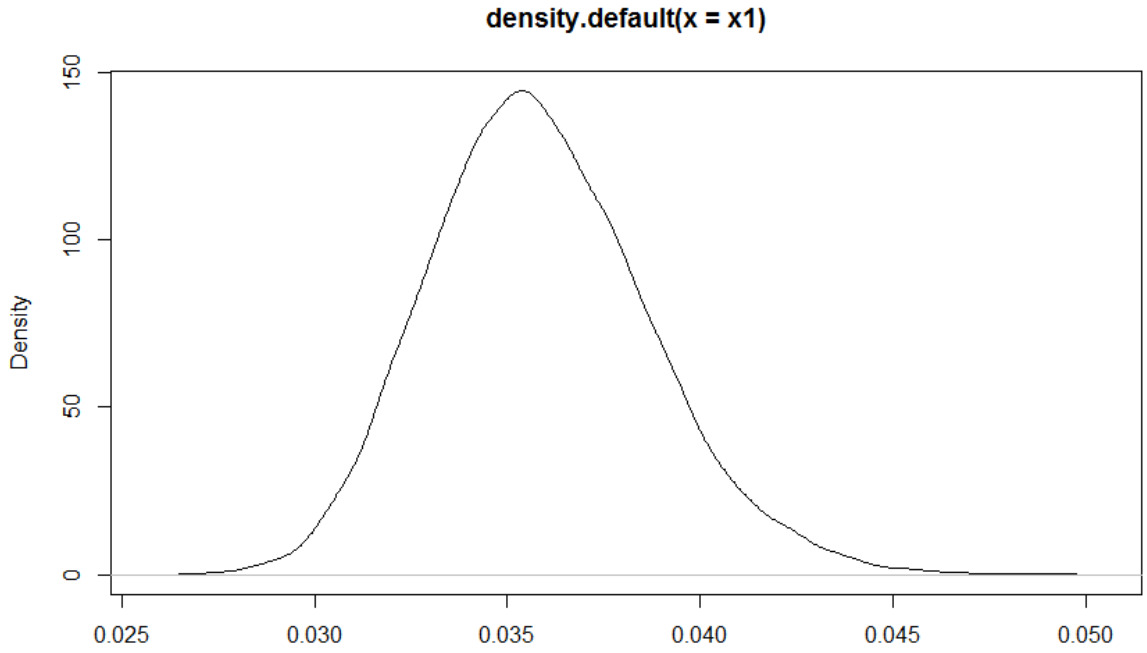
First, we plot the distribution:



Next, we can observe a histogram of the distributed values, illustrating the mean of 0.0358. The data appears roughly symmetric, which is consistent with the skew of 0.384 (between -0.5 and +0.5 being considered approximately symmetric). Similarly, the height and sharpness of the peak appears normal, in accordance with how the kurtosis of 3=3.36 is close to the normal distribution's kurtosis of 3.

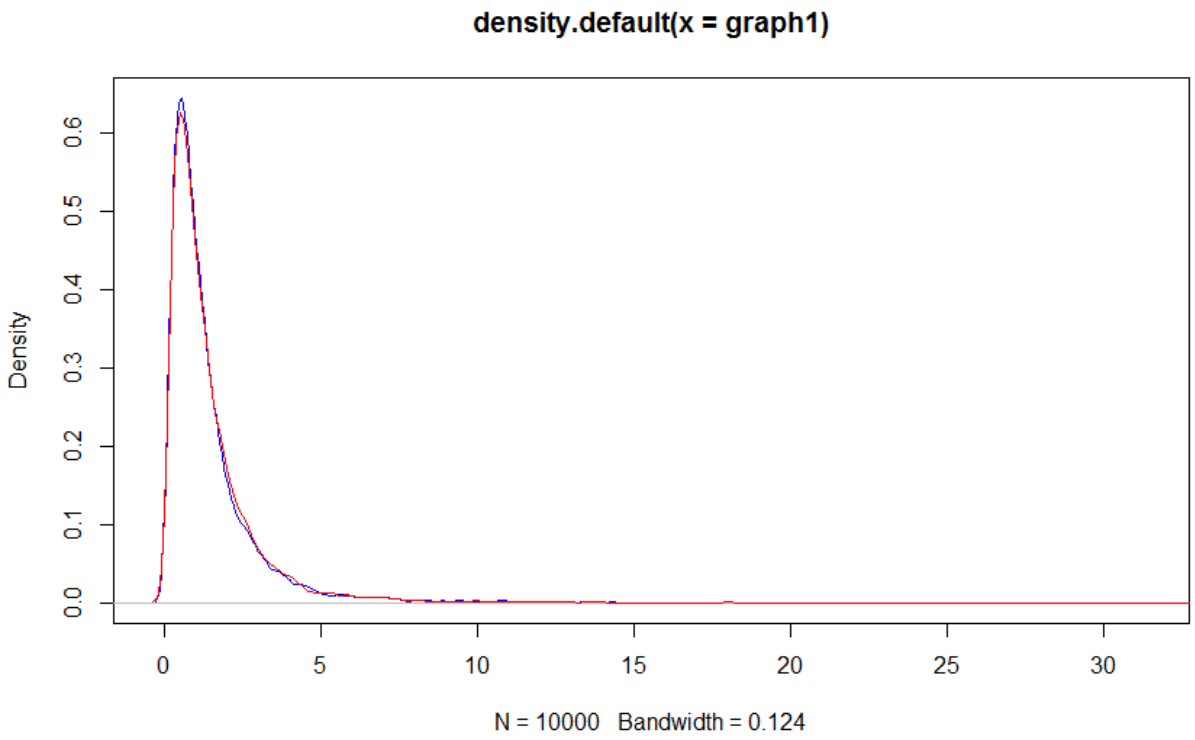


The density presents a similar visual story:



As one example of how function “rOliveira” could be implemented, let us take a series of private concerts given by a talented musician. Let us assume the concerts play at a relatively uniform source power of P=100 Watts. Let *x* designate the distance at which people sit in meters away from the concertist. Naturally, people sit as close to hear the talented musician on either side (designating positive and negative distances); let us assume a normal seating distance distribution around the musician. Using rOliveira allows one to predict the average sound intensity (in W/m^2) that reaches an audience of size *df* over *n* concerts. Populations that fall at a range below the 5% quantile on the left are hearing a nonrandom level of quietude, e.g. the instrument needs tuning or they don’t like the music!

(3): Please see the accompanying R source code for the full equation with comments. As the graphs show, there is strong identity between the derived equation (blue) and rf function (red):



(4): The following table shows the *p* values obtained using Studen'ts *t*-test to determine if the mean of "*n*" observations from the derived function for F distribution differs significantly from the expected mean of $df_2/(df_2-2)$.

n	*p* value	Conclusion
10	0.476	The observed mean does not significantly differ from the expected mean.
100	0.966	The observed mean does not significantly differ from the expected mean.
1000	0.626	The observed mean does not significantly differ from the expected mean.
100,000	0.278	The observed mean does not significantly differ from the expected mean.

This table shows the *p* values obtained using a generated empirical null distribution of 10,000 means generated from *n* = x observations.

n	*p* value	Conclusion
10	0.4491	The observed mean does not significantly differ from the expected mean.
100	0.4775	The observed mean does not significantly differ from the expected mean.
1000	0.4966	The observed mean does not significantly differ from the expected mean.

The number of operations required to generate an empirical null distribution for n=100,000 requires prohibitive running time and was not performed.

(5): The following table shows the *p* values obtained using a chi-squared test to determine if the variance of "*n*" observations from the derived function for F distribution differs significantly from the expected variance of $(2n_2^2(n_1+n_2-2))/(n_1(n_2-2)^2(n_2-4))$.

n	*p* value	Conclusion
10	0.497	The observed variance does not significantly differ from the expected variance.
100	0.999	The observed variance does not significantly differ from the expected variance.
1000	0.315	The observed variance does not significantly differ from the expected variance.
100,000	0.891	The observed variance does not significantly differ from the expected variance.

This table shows the *p* values obtained using a generated empirical null distribution of 10,000 variances generated from *n* = x observations.

n	*p* value	Conclusion
10	0.276	The observed variance does not significantly differ from the expected variance.
100	0.3593	The observed variance does not significantly differ from the expected variance.
1000	0.4192	The observed variance does not significantly differ from the expected variance.

Similar to the above, the number of operations required to generate an empirical null distribution for n=100,000 requires prohibitive running time and was not performed.