Bootstrap Resampling Project

Jordan Turley

Dependencies:

numpy: http://www.numpy.org/
Installation: pip3 install numpy

Run:

python3 final_project.py

Summary:

For this assignment I wrote a program to test the performance of bootstrap resampling for six different distributions: normal, uniform, exponential, Erlang, binomial, and Poisson. For these six distributions, I used bootstrap resampling to generate an 80% confidence interval for the skewness and excess kurtosis of the distribution. I used Python with numpy to sample from these distributions, generate bootstrap resamples, and calculate the confidence interval.

Distributions, Results, and Analysis

To get an initial feel for our results, I generated a sample of size 100 and generated 1000 bootstrap resamples to generate an 80% confidence interval for the skewness and excess kurtosis. I repeated this 100 times, checking each time if the true skewness or excess kurtosis is in the range of our confidence interval. From this, I can see how close or far our true confidence level is from 80%. After this, I modified the parameters of the distribution to see how this affects the confidence level.

In all our results, we see the same general phenomenon occurring. For all distributions other than Exponential and Erlang, we see the confidence level for skewness is very close to 80%, or even higher than 80% in one case. However, for the excess kurtosis, other than the uniform distribution, our results are significantly lower than 80%. This can be explained because the kurtosis of the distribution measures the heaviness of the tails. However, to do this accurately, we need information from the tails, which means that we need enough data to see what occurs in the tails, since it is much less likely to get a sample point in a tail. Even for n = 100, we will see that our confidence intervals could have been improved significantly. We explore increasing n later and see how this affects our confidence intervals.

Normal

I used the standard normal with $\mu = 0$ and $\sigma^2 = 1$. My results can be see in Table 1.

Table 1 – Normal(0, 1) Results

Statistic	Correct/Total
Skewness	78/100
Excess Kurtosis	63/100

From our results, we can see that the skewness performs almost exactly like we would expect it to. We wanted our confidence interval to be correct 80% of the time, and it was correct 78% of the time. The other two percent could simply be due to randomness. To see if or how changing the parameters affects, we explored several combinations of parameters. First, we explored changing the mean while holding the variance constant. Then, we changed the variance while holding the mean constant. These results can be seen in Table 2 and Table 3.

Table 2 –Results of Varying Mean of Normal

Parameters	**Skewness**	**Excess Kurtosis**
$\mu = 5, \sigma^2 = 1$	69/100	61/100
$\mu = 10, \sigma^2 = 1$	75/100	58/100
$\mu = 100, \sigma^2 = 1$	71/100	64/100
$\mu = -50, \sigma^2 = 1$	72/100	61/100

Table 3 – Results of Varying Variance of Normal

Parameters	**Skewness**	**Excess Kurtosis**
$\mu = 0, \sigma^2 = 4$	69/100	61/100
$\mu = 0, \sigma^2 = 16$	65/100	64/100
$\mu = 0, \sigma^2 = 25$	73/100	59/100
$\mu = 0, \sigma^2 = 100$	79/100	58/100

As we can see, changing the parameters does not have any significant affect. There is some variation, but this is most likely due to randomness. This makes sense; the skewness and excess kurtosis of the normal distribution are defined as constants, so it would make sense that the skewness and excess kurtosis estimates do not change based on the mean or variance.

Uniform

I used the standard uniform with $\theta_1 = 0$ and $\theta_2 = 1$. My results can be seen in Table 4.

Table 4 – Uniform(0, 1) Results

Statistic	**Correct/Total**
Skewness	85/100
Excess Kurtosis	79/100

We can see that bootstrap resampling performs exceptionally well for the Uniform distribution. 85% of our confidence intervals contained the true skewness, exceeding our 80% confidence level, which was a matter of luck. However, 79% of our excess kurtosis confidence intervals contained the true excess kurtosis. There is an explanation to why this occurs only with the uniform distribution. As I said before, to learn about the excess kurtosis, we need a lot of data to learn about the tails. However, the uniform distribution is unique; we are equally as likely to see a value in a tail as we are in the middle. Because of this, our sample points are uniformly spread from θ_1 to θ_2 and we can learn more about the tails in this distribution than we are any of the other distributions. We explore changes in θ_1 and θ_2 in Table 5.

Table 5 – Varying Uniform Results

Parameters	Skewness	Excess Kurtosis
$\theta_1 = 0$, $\theta_2 = 100$	88/100	82/100
$\theta_1 = -10$, $\theta_2 = 10$	83/100	78/100
$\theta_1 = -25$, $\theta_2 = 75$	86/100	82/100
$\theta_1 = -100$, $\theta_2 = 100$	87/100	74/100

We can see that changing the parameters of the distribution has no real impact on the accuracy of the confidence interval, which makes sense, since the skewness and excess kurtosis of the uniform distribution are defined as constants. Any uniform distribution is basically a rescaled version of the standard uniform distribution, so changing these parameters should not have any effect.

Exponential

I used the exponential with $\mu = 1$. My results can be seen in Table 6.

Table 6 – Exponential(1) Results

Statistic	Correct/Total
Skewness	56/100
Excess Kurtosis	43/100

In contrast to the Normal and Exponential distributions, we see that the true confidence level of skewness is much lower than our expected 80%. In the case of the exponential, which is heavily one-sided, we only get data in one side of the distribution, and most of the data is near the mean. This does not let us learn as much as we would like about both the skewness and excess kurtosis.

Table 7 – Varying Exponential Results

Parameters	Skewness	Excess Kurtosis
$\mu = 1/100$	46/100	38/100
$\mu = 1/10$	38/100	32/100
$\mu = 2$	54/100	39/100
$\mu = 5$	41/100	26/100
$\mu = 10$	46/100	32/100
$\mu = 50$	58/100	36/100

We can see that varying μ does not significantly change the results. They are all close to each other within some random error. We can see that we still need to collect more data to get a better feel for the skewness and excess kurtosis. Skewness and excess kurtosis is defined as a constant for the exponential distribution, so it should be expected that changing the parameter does not change the confidence.

Erlang

I used the Erlang with $\alpha = 1$ and $\beta = 1$. My results can be seen in Table 8.

Table 8 – Erlang(1, 1) Results

Statistic	Correct/Total
Skewness	51/100
Excess Kurtosis	41/100

Like the Exponential, we see that our true confidence level for skewness and excess kurtosis is much lower than 80%. The same thing occurs; Erlang(1, 1) only has one tail, so we are unable to learn about the skewness. We explored several changes to alpha and beta to see how this affects our confidence level. We held beta constant and changed alpha in Table 9, and held alpha constant and changed beta in Table 10.

Table 9 – Results of Varying Alpha of Erlang

Parameters	Skewness	Excess Kurtosis
$\alpha = 5,\ \beta = 1$	61/100	45/100
$\alpha = 10,\ \beta = 1$	58/100	38/100
$\alpha = 25,\ \beta = 1$	70/100	52/100
$\alpha = 50,\ \beta = 1$	70/100	50/100

Table 10 – Results of Varying Beta of Erlang

Parameters	Skewness	Excess Kurtosis
$\alpha = 1,\ \beta = 5$	48/100	38/100
$\alpha = 1,\ \beta = 10$	45/100	31/100
$\alpha = 1,\ \beta = 25$	38/100	34/100
$\alpha = 1,\ \beta = 50$	45/100	33/100

Like the other distributions, as we change beta, we do not see any significant change in our confidence level. However, this is not the case with alpha. We can see that, as alpha increases, we are able to be more and more confident in our confidence intervals. As beta is held constant and alpha is changed, this moves the center of the distribution. When alpha and beta are both equal to one, the distribution is heavily skewed right. However, once we get to alpha = 50, the distribution is much more 'normal' looking, so we get sample points in both tails, so we can estimate the skewness with more confidence. The same thing happens with the kurtosis. However, this confidence increase is not as significant, because we still do not have enough data to estimate it as accurately as we would like.

Binomial

I used the binomial with $n = 100$ and $p = 0.5$. The results can be seen in Table 11.

Table 11 – Binomial(100, 0.5) Results

Statistic	Correct/Total
Skewness	80/100
Excess Kurtosis	59/100

Like most of the other distributions, skewness performs just as we would expect, but the excess kurtosis does not get enough data in the tails to give is a good enough estimate. In Table 12, I explore how changing n affects the results, and in Table 13, I explore how changing p affects the results.

Table 12 – Results of Varying n of Binomial

Parameters	Skewness	Excess Kurtosis
n = 2, p = 0.5	83/100	78/100
n = 5, p = 0.5	75/100	76/100
n = 10, p = 0.5	79/100	69/100
n = 50, p = 0.5	75/100	67/100
n = 200, p = 0.5	71/100	67/100

Table 13 – Results of Varying p of Binomial

Parameters	Skewness	Excess Kurtosis
n = 100, p = 0.01	64/100	38/100
n = 100, p = 0.1	76/100	56/100
n = 100, p = 0.25	84/100	66/100

As n gets smaller, we see that we can be more confident in our confidence intervals. When n becomes small and p = 0.5, it is very possible to get random samples in the tails, which are very short and fat. We learn a lot about the tails. However, as n increases and gets larger, we get fewer samples in the tails, so we do not learn as much.

As p becomes very small, the distribution becomes very one-sided, as our samples will most likely contain very few successes. The same thing happens; we are not able to learn about the skewness as we only see data from one side at most, and even then, we do not see much data. Also, we do not see a lot of data in the tail, so we do not learn much about it. As p increases, we see more successes and see more samples in the tails. We do not see a lot but see enough to get a better estimate than when p is very small. The same thing would occur if p was very large.

Poisson

I used the Poisson with $\lambda = 2$. The results can be seen in Table 14.

Table 14 – Poisson(2) Results

Statistic	Correct/Total
Skewness	75/100
Excess Kurtosis	49/100

The Poisson distribution performs much like the other distributions we have seen. We can be pretty close to 80% confident in our interval for skewness but are much less confident in our interval for excess kurtosis. I explore changing the value of lambda in Table 15.

Table 15 – Results of Varying Lambda of Poisson

Parameters	Skewness	Excess Kurtosis

λ = 1	65 / 100	47/100
λ = 5	69/100	57/100
λ = 10	65/100	58/100
λ = 50	75/100	61/100

As lambda is increased, the confidence intervals for excess kurtosis seem to perform slightly better. This is because the Poisson distribution has a variance equal to lambda. Our confidence in the interval for skewness increases, as we see more data on each side of the distribution, so we can estimate the symmetry of the distribution more accurately. We also get more samples in the tails, which will increase our confidence in the excess kurtosis estimate, as we learn more about the tails.

Sample Size

After analyzing the results for each distribution at n = 100, I wanted to explore how the accuracy of our confidence intervals changes as we collect more data or less data. The results of varying the sample size can be seen in Table 16 and Table 17.

Table 16 – Results of Varying Sample Size on Skewness

Sample Size	Normal	Uniform	Exponential	Erlang	Binomial	Poisson
n = 5	61/100	84/100	12/100	17/100	70/100	52/100
n = 10	67/100	79/100	29/100	27/100	76/100	54/100
n = 50	71/100	79/100	44/100	42/100	66/100	62/100
n = 500	76/100	72/100	66/100	57/100	72/100	83/100
n = 1000	80/100	73/100	66/100	70/100	72/100	76/100

Table 17 – Results of Varying Sample Size on Excess Kurtosis

Sample Size	Normal	Uniform	Exponential	Erlang	Binomial	Poisson
n = 5	9/100	27/100	0/100	0/100	3/100	0/100
n = 10	35/100	51/100	3/100	4/100	28/100	17/100
n = 50	49/100	78/100	29/100	26/100	59/100	43/100
n = 500	68/100	85/100	54/100	47/100	73/100	68/100
n = 1000	77/100	85/100	58/100	62/100	72/100	67/100

As we can see, when n is very small, bootstrap resampling is not very effective. It still does well for something like the skewness of the uniform, but it does not do well for any of the distributions in estimating excess kurtosis. With only five data points, we do not have enough data to accurately estimate something like kurtosis, and in most cases do not have enough data to accurately estimate skewness. However, as n increases and gets very large, we are more and more confident in our intervals for all the distributions. This increase is much more drastic for excess kurtosis, but still applies to skewness as well. For example, at n = 5 we are only 17% confident in the Erlang skewness confidence interval, but once we reach n = 1000, we are 70% confident. This increase is even more drastic for excess kurtosis. At n = 5 none of our confidence intervals accurately estimate the excess kurtosis of the Erlang, but at n = 1000, we are 62%

confident in our interval. All the intervals seem to converge to some confidence level, which is at or slightly below our 80% target. This confirms what is always true in statistics: more data will always be better. At n = 5, n = 10 or even n = 50, we simply do not have enough data to learn what we could if we had 500 or 1000 data points.

Bootstrap Resamples

The last thing I wanted to analyze is how changing the number of bootstrap resamples affects our confidence. I used the original distribution parameters and a sample size of n = 100. My results can be seen in Table 18 and Table 19.

Table 18 – Results of Varying Bootstrap Resamples on Skewness

Bootstrap Resamples	Normal	Uniform	Exponential	Erlang	Binomial	Poisson
b.r. = 2	28	35	21	14	36	25
b.r. = 10	64	75	43	50	69	62
b.r. = 100	73	81	47	51	65	61
b.r. = 500	75	85	42	47	69	70
b.r. = 10000	70	78	48	51	66	68

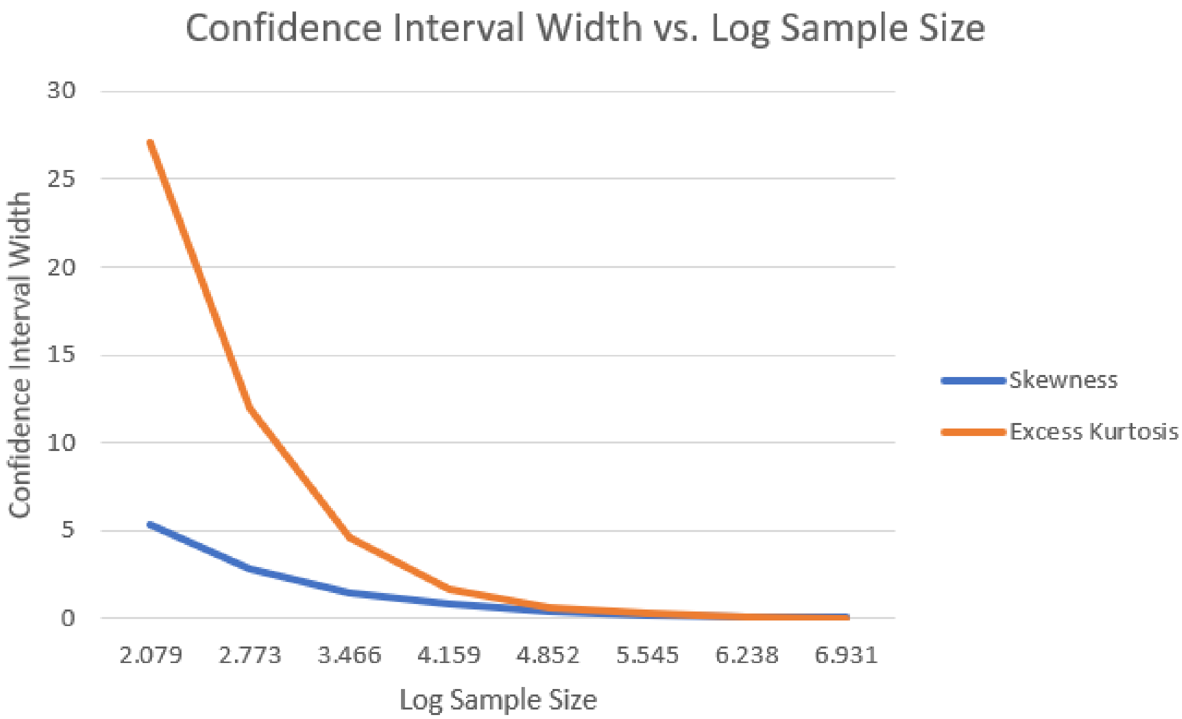
Table 19 – Results of Varying Bootstrap Resamples on Excess Kurtosis

Bootstrap Resamples	Normal	Uniform	Exponential	Erlang	Binomial	Poisson
b.r. = 2	30	28	11	17	26	22
b.r. = 10	52	70	30	37	65	56
b.r. = 100	58	72	39	32	60	49
b.r. = 500	52	78	34	31	63	51
b.r. = 10000	72	82	32	31	64	46

We can see that the convergence is quick as the number of bootstrap resamples increases. We see for a very small number of bootstrap resamples, like two, that our results could be a lot better. However, even at only ten our results are close to what they will be for 100. Even after we resample just 100 times, it seems like we have learned most of what we will learn from the data; the difference from 100 to 500 is not extremely significant for any of the distributions. This shows us that there is no practical reason for resampling thousands or tens of thousands of times. For my purposes, resampling 1000 times gave me results that were just as good as resampling 10000 times, and took one-tenth of the time to run. In practice, there is no point to resampling so many times that it takes several hours or days to run. Computational time can be saved by letting it run for an hour or overnight.

Width of Confidence Interval

The last thing I wanted to analyze was the width of confidence intervals as the sample size increases. I used the standard normal distribution and computed 100 confidence intervals at varying sample sizes: 8, 16, 32, 64, 128, 256, 512, and 1024. The results can be seen in the following graph.



For my x-axis, I used the log of the sample size so that it would be linear. As we would expect, the width of the confidence intervals decreases as the sample size increases. Confidence intervals for excess kurtosis are much wider than skewness, but once the sample size is about 256 or greater, the confidence intervals for both skewness and excess kurtosis are very narrow.

Conclusion

Overall, we can see that even if you have data from a distribution that you know nothing about, you can still create a very reasonable confidence interval for any parameter by using bootstrap resampling. If we have good reason to believe that data comes from a specific distribution, then we can obtain a more accurate confidence interval using that distribution, but in the real world, this is very often not the case, so bootstrap resampling is the best option. If your sample contains data that is in the tails, then you will be able to estimate skewness or kurtosis more effectively as this gives you data about the tails. More data is also better in any case; no matter what the distribution is, more data will give you a confidence level that you are able to put more confidence in, as well as a narrower confidence interval. The number of bootstrap resamples is important but suffers heavily from diminishing returns; after resampling a certain number of times, you have simply learned everything that you can learn from the data, and the only thing that could allow you to learn more is if you had more data.