Choosing the best fit distribution

Once we have the estimated parameters of our candidate distributions, -- which in our case are the exponential, normal, gamma, lognormal and Weibull distributions -- we find the distribution with the best fit.

AIC
Akaike's Information Criterion is a method of choosing the best fitting distribution among several candidate distributions.

To compute AIC for each of our distributions, we used the Matlab function:

```
value = aic(model1,...,modeln)
```

Among the distributions we compared, the best AIC values were obtained for the exponential and gamma distributions.

Distribution	AIC (MLE)	ΔAIC	AIC (MoM)	ΔAIC
Normal	1122.165	726.592	1122.165	63.452
Exponential	**395.568**	**0**	**1058.713**	**0**
Gamma	**396.9113**	**1.3433**	**1060.545**	**1.832**
LogNormal	1074.17	678.602	1137.914	79.201
Weibull	1060.141	664.573	1074.17	15.457

Here, we see that the Exponential and Gamma distributions are very close in their AIC values, in the case of MLE as well as MoM, though the Exponential distribution is slightly better.

We now use the KS test to finalize our distribution.

KS test

The Kolmogorov Smirnov test is a non-parametric test for comparing the fit of two distributions using the CDFs of the distributions.

In our case, we are comparing our interevent data with each of our candidate distributions to find the best fit. As we have only one distribution to compare to known distributions, only a single-distribution ks test is required.

For this, we used Matlab's ks_test function:

```
[h,p,ksstat] = kstest(__)
```

The table below summarizes our results for the goodness-of-fit test. The parameters of each distribution are those estimated using Maximum Likelihood Estimation and Method of Moments.

The p-value is compared to an alpha value of 0.025 since we used a 95% confidence interval. All distributions except for the normal distribution fail to reject the null hypothesis that proposes that the distributions are equal (represented by h = 0).

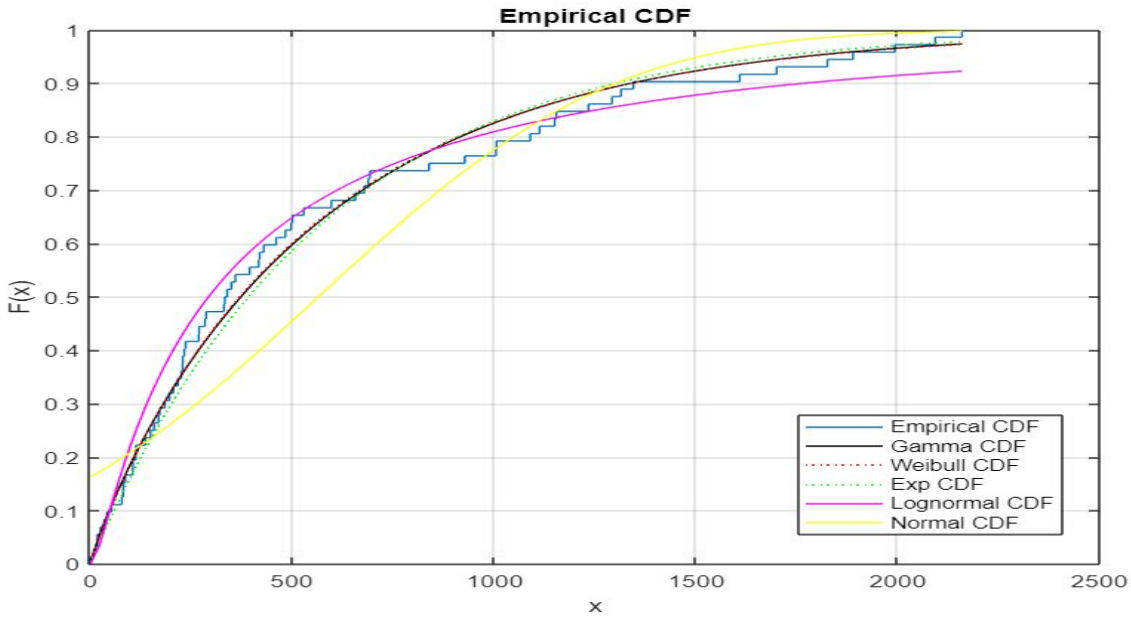
	MLE parameters			MoM parameters		
Distribution	p-value	ks_stat	h	p-value	ks_stat	h
Normal(mu=565.823,sigma=574.262)	0.0061	0.1976	1	0.0060	0.1979	1
Exponential(mu=565.823)	0.7866	0.0749	0	0.7866	0.0749	0
Gamma(a=0.890632,b=635.305)	**0.9222**	**0.0628**	**0**	**0.8275**	**0.0717**	**0**

LogNormal(mu=5.68103,sigma=1.40321)	0.6588	0.0840	0	0.0546	0.1556	0
Weibull(a=548.344,b=0.93302)	0.9102	0.0641	0	0.7128	0.0802	0

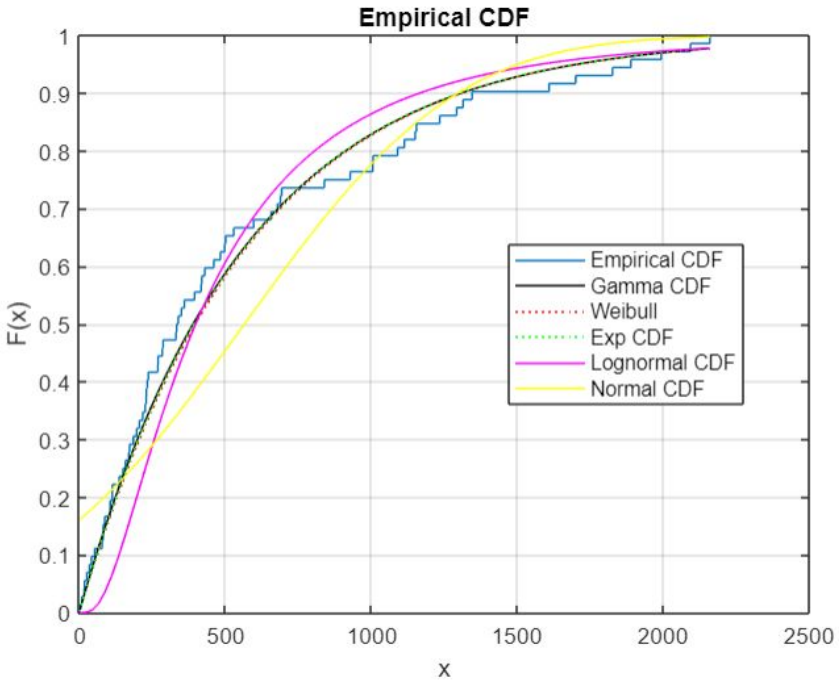
We then consider the value of ks_stat. The distribution with the lowest ks_stat, or correspondingly, the highest p-value is the distribution with the best fit.

Clearly, the Gamma Distribution with parameters a=0.890632 and b=635.305 has the lowest ks_stat value.

On plotting the empirical CDF of our data with each of the candidate distributions, we get the following graphs, each one corresponding to our method of parameter estimation.



KS-Test for parameters using MLE



KS-Test for parameters using MOM

We prefer the KS Test results over AIC since… **// unsure of why**

We conclude that the Gamma distribution with parameters a=0.890632 and b=635.305 gives the best fit for our interevent time data.

References:

https://towardsdatascience.com/kolmogorov-smirnov-test-84c92fb4158d
https://in.mathworks.com/help/ident/ref/aic.html
https://in.mathworks.com/help/stats/kstest.html
http://www.mit.edu/~6.s085/notes/lecture5.pdf
https://www.sciencedirect.com/topics/medicine-and-dentistry/akaike-information-criterion
https://www.researchgate.net/post/Model_selection_by_The_Akaikes_Information_Criterion_AIC_what_is_common_practice