Homework 3
CS626 Data Analysis and Simulation

Name: Sidi Chang
Due: 11:59 p.m., Tuesday, 5/31

1 Question 1

I have finished the work of making all corrections to HW2 and return them for a final grade.

2 Question 2

I use two traces in Homework 2, the first one is $wdev_0$, the second one is $proj_3$. Comparing the two traces, we can find that $wdev_0$ data set is harder than the data set of $proj_3$, so I will use $proj_3$ to finish the job.

We need to consider Read and Write separately, so I will consider write first. Both read and write are all right-censored and both are Type-II censoring with $n = r$, we can get get the MLE as $\sum n_i/n$ for exponential from CSCI 688 class. But as we all know, for this kind of simulation, the best way to guess is Gamma distribution. So I will use Gamma, Exponential, Log linear distribution to solve this kind of problem.

2.1 Write part

2.1.1 Gamma Distribution

a. MLE shows that $a = 0.470863, b = 820726$. So we can draw the original graph as well as fitted:
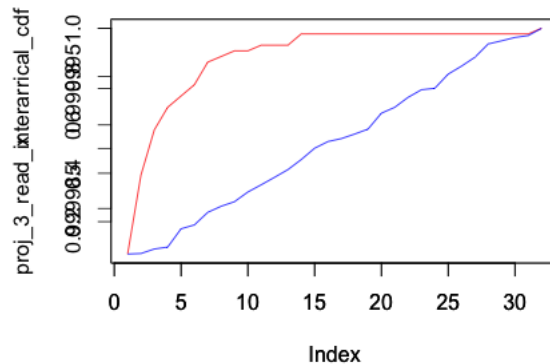
We can see from the above that the graph fits does for the empirical distribution. It is common cause gamma distribution is not a good fit for this.

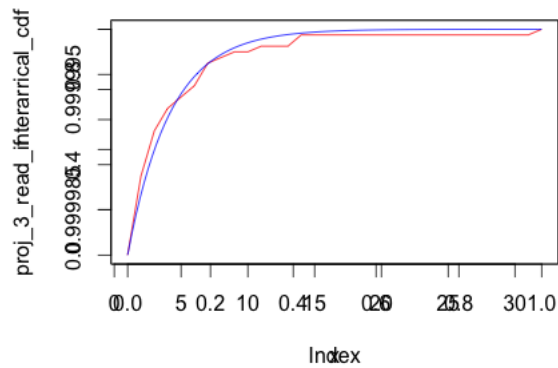b. Compare statistics of queuing when driving a FIFO queue.

After running the program, I got the result as below: There is a big difference from Gamma and Original trace, so we can conclude Gamma is not a good fit for this one.

Table 1: metric: Gamma verus Original

	$\bar{d}$	$\bar{q}$	$\bar{x}$
Gamma	342.2	0.0002	0.00015
Original	1413.7	0.00008	0.00015

2.1.2 Exponential Distribution

a. Using MLE we find that $u = 17085318$ and we can get the exponential distribution based on that. Then we can plot the graph as below:

From the graph above, we can easily see that the result is much better than the Gamma one.

b. Compare statistics of queuing when drive a FIFO queue. The difference on $\bar{q}$ is much closer
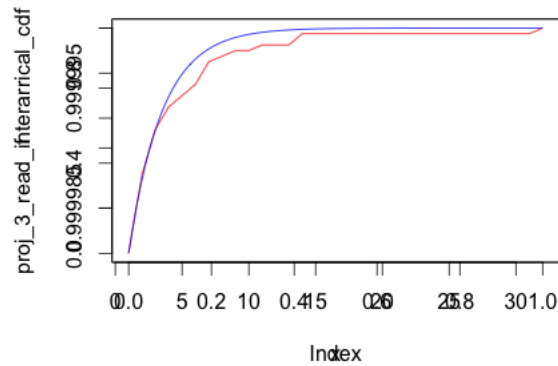
Table 2: metric: Exponential verus Original

	$\bar{d}$	$\bar{q}$	$\bar{x}$
Exponential	0.2942	1.7e-8	0.00015
Original	1413.7	0.00008	0.00015

though $\bar{d}$ is bigger, but the simulated fitted line is better than previous one.

2.1.3 Log linear Distribution

a. Applying MLE we can get $a = 18.4484$, $b = 3.5803$, then we can plot the two plots as below:

This one is also better than Gamma but not as good as the Exponential.

b. Compare statistics when running a FIFO queue:

Table 3: metric: log linear verus Original

	$\overline{d}$	$\overline{q}$	$\overline{x}$
Log linear	22.41	0.0001	0.00015
Original	1413.7	0.00008	0.00015

2.2   Read Part

2.2.1   Gamma Distribution

Applying MLE and we can get $a = 0.222591$, $b = 767565000$, the graph is also weird for the empirical data. The statistics of queuing is as below:

Table 4: metric: Gamma verus Original

	$\overline{d}$	$\overline{q}$	$\overline{x}$
Gamma	6981.4	0.017	0.05
Original	3115.6	0.008	0.05

We can see the difference between $\overline{d}$ is still big.

2.2.2   Exponential Distribution

Using MLE we can get $u = 386178$, and the graph is similar to the Write part. The statistics of queuing is as below:
Also, exponential seems to be the best fit for this trace.

Table 5: metric: Exponential verus Original

	$\overline{d}$	$\overline{q}$	$\overline{x}$
Gamma	2235.3	0.006	0.05
Original	3115.6	0.008	0.05

2.2.3 Log linear Distribution

As above, we can using MLE to get $a = 12.0341$, $b = 1.93794$. The same method can be implemented and then we can get

Table 6: metric: Log Linear verus Original

	$\overline{d}$	$\overline{q}$	$\overline{x}$
Log Linear	5134.2	0.006	0.05
Original	3115.6	0.008	0.05

Above all, I still think the exponential is the best fit for the trace $proj_3$.

3 Question 3

Yes of course. From the data set above, we can see the fitted plot is almost the same result for read and write. Also, there is just a single server, which means the server can only handle one read or one write at the same time. So there is no need to consider them separately. Just like Homework 2, we can think them together which is a easier way to do data analysis.

4 Question 4

i. Using ssq4 and making some changes to it, we can deal with the SJF problems. Using the method of batch means with $k = 64$ batches with each size 1024. we can get the confidence interval of this problem:

Table 7: metric: Batch means

$\overline{d}$	$\overline{w}$	$\overline{l}$	$\overline{q}$	$\overline{x}$
(0.12282,0.13617)	(0.20311,0.21596)	(2.03001,2.17279)	(1.23333, 1.36667)	(0.97501,0.80788)

ii. Changing the code to FIFO just as Homework 2, we can get another confidence interval with 95%:

iii. For this question, we need to have $\lambda < \mu$, and we can see all the result satisfies this constraint. So I believe my answer is the right answer for this question.

Table 8: metric: Batch means

$\overline{d}$	$\overline{w}$	$\overline{l}$	$\overline{q}$	$\overline{x}$
(0.19287,0.22117)	(0.27331,0.30101)	(2.74444,3.07321)	(1.94409, 2.26011)	(0.79803,0.81410)

iv. So I use hist\$density to draw the histograms of the waiting queue lengths for the two queues as below:
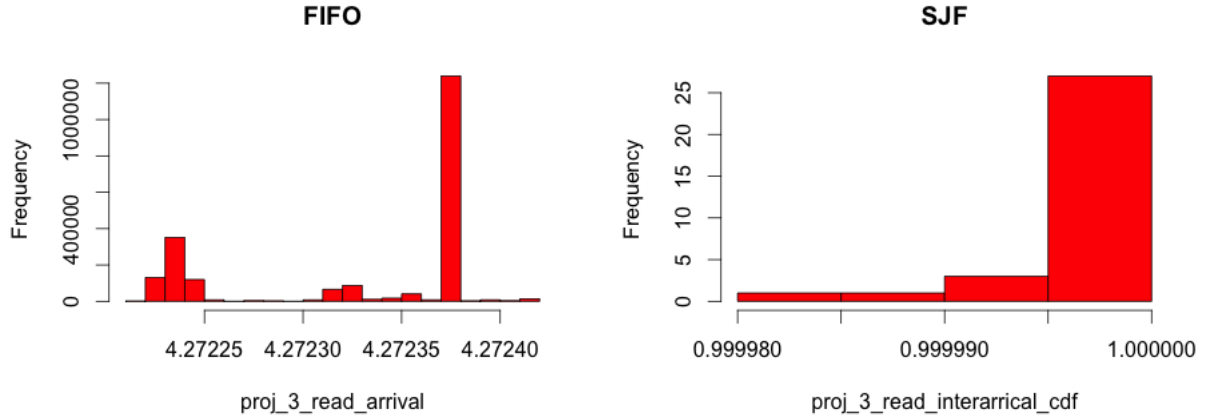
Obviously, we can get the waiting queue length time is less than the waiting queue length of FIFO since the length is from (0,2) in SJF comparing with (0,4) in FIFO.

v. When comparing, we can see that SJF is better. I think the main reason is because that there are some huge jobs in the queue which causes FIFO waiting queue becomes longer. It is the same as why exponential is better than others since others may have more chance of getting huge jobs.

5 Question 5

From Homework 2, it is known that $proj_3$ is a good trace cause the autocorrelation is smooth, so I choose the $wdev_0$ trace to solve this problem.

i. I will also use Erlang(4,0.02) random variable as service time. With $k = 64$ and $b = 1024$.
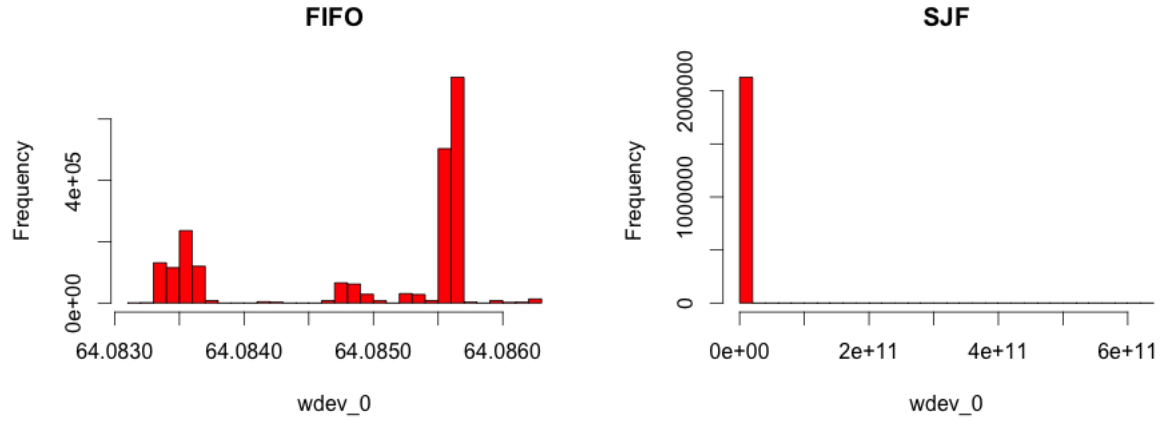
Table 9: metric: Batch means

$\overline{d}$	$\overline{w}$	$\overline{l}$	$\overline{q}$	$\overline{x}$
$(5.92, 11.34)10^8$	$(5.81, 11.23)10^8$	$(300, 780)$	$(300, 780)$	$(0.2, 0.3)$

ii. Using FIFO we can get

iii. Also we need $\lambda < \mu$ and all the metrics are satisfied.

Table 10: metric: Batch means

$\bar{d}$	$\bar{w}$	$\bar{l}$	$\bar{q}$	$\bar{x}$
$(6.11, 12.34)10^8$	$(5.81, 11.23)10^8$	$(300, 780)$	$(303, 780)$	$(0.2, 0.3)$

iv. Applying the same step as Question 4, we can get

v. I think for this trace, the answer is not very obvious. But I do not find a good k for this problem, I think if the k is suitable, the answer will be easy to figure out. Also when the batch increase, it should be accurate but when drawing histogram, we may need to drop some data.