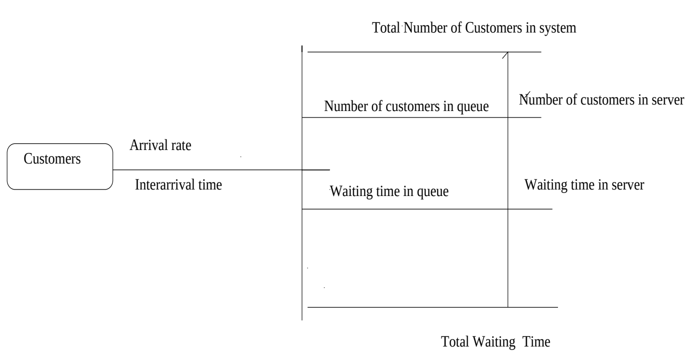
Activity 1. Simulating a queueing system where service time is Uniform, arrivals are Poisson, and there is one server

Queueing theory is a very important modeling technique that represents a computer system as a network of service centers, each of which is treated as a queueing system. That is, each service center has an associated queue or waiting line where customers who cannot be served immediately queue (wait) for service. The customers are, of course, part of the queueing network. Customer is a generic word used to describe workload requests such as CPU service, I/O service requests, requests for main memory, etc. All these arrive at random to the service facility. Queueing theory models are often used to determine the effects of changes in the configuration of a computer system.

The simplest queueing theory model is the `M/M/1` model, which assumes: (a) that customers arrive in accordance with a Poisson process with average rate λ and thus the interarrival times are exponentially distributed with mean $\frac{1}{\lambda}$. (b) Service time by the server is assumed exponential with parameter μ. Expected service time is then $\frac{1}{\mu}$. (c) The customers are served one at a time by a single server. If the server is busy upon the customer's arrival, then the customer waits in the queue. Another simple model is one where the service time is uniform, instead of exponential. The set of random variables involved in either case is summarized in the diagram displayed below, which is an adaptation of a diagram in Allen(1990), p. 251.

Queueing Theory random variables. A queueing system.



For these models we are usually interested in determining, among other things, the average number of customers in the system (or in the queue) and the average amount of time a customer spends in the system, among other things.

We simulate first the queueing system where the service time is uniform. We will leave it as an exercise to simulate the M/M/1 system.

Activity: Simulation of the M/Uniform/1 system

Suppose that customers arrive at a single server according to a Poisson process with arrival rate $\lambda = 4$ per hour (=1/15 per minute). Assume that the service time of a customer (in minutes) is uniformly distributed on the interval (5, 15).Goodman(1988). To simulate the operation of the system, we proceed as follows:

Generate a sequence of exponential random variables with parameter $\lambda = 1/15$

```
it =c(rep(0,1000))
lambda = 1/15
it =rexp(1000,lambda)
```

Generate now a sequence of uniform random variables in the interval (5,15)

```
s= c(rep(0,1000))
s=runif(1000,5,15)
```

Now create a new variable "arrival times" -at- that contains the cumulated interarrival times. This will give the time of arrival of each customer. That is, if, for example, the first four values of *it* are 15, 16, 17, 8, the first four values of the variable at should be: 15, 31, 56 and 63. That is, customer 1 in this example arrived at minute 15, customer 2 arrived at minute 31, and so on.

```
at= c(rep(0,1000))
at=cumsum(it)
```

Combine the information about arrival times, at, and service time, s, to describe the simulated operation of the queueing system as follows:

Create a variable for the queueing time, qt, and another for the exit time, exit

```
qt = c(rep(0,1000))
exit = c(rep(0,1000))
```

Now determine the queueing time and the exit time for each customer.

```
for(i in 1:999) {
if(at[i] +s[i] + qt[i] <= at[i+1]){
    qt[i+1]=0
    exit[i]= at[i]+s[i]+qt[i]
  }
 else if(at[i] + s[i]+qt[i] > at[i+1]){
 exit[i]=at[i]+s[i]+qt[i]
 qt[i+1] =  at[i]+s[i]+qt[i] -at[i+1]
}
}
exit[1000]=at[1000]+qt[1000]+s[1000]
```

We know the service time, the queueing time and the exit time for each customer. So we can now figure out the total time spent by each customer in the system. Let's call it tts

```
tts = c(rep(0,1000))
tts = exit - at
```

We would also like to know the average number of customers in the system. For that, we first need to put the operation of the system in real time.

```
at=matrix(at)
exit = matrix(exit)
atexit = cbind(at,exit)
tatexit = t(atexit)
realtime=matrix(tatexit,2000,1,byrow=T)
```

Odd row numbers of this last matrix give us the arrival time, even numbers give us the exit time. So we can add one column to the matrix, which tells us that at every arrival there is a new customer, at every departure one customer less.

```
customer = matrix(rep(1,2000),2000,1)
realtime=cbind(realtime,customer)
for(i in 1:1000) {
   realtime[2*i,2]=-1
     }
```

```
oo = order(realtime[,1])
trace = realtime[oo,]
```

Now we would like to focus on the number of customers. So we create a variable that tells us how many customers are in the system at each real time of entry or exit.

```
tracecum= cumsum(trace[,2])
```

Question 1.- Find the distribution of the total time spent in the system, the distribution of queueing time, and the summary statistics for both. Explain what you see in these plots and summary statistics that is relevant to explain the performance of this queueing system. To what family of distributions do the distributions that you got belong?

Question 2.- Find the distribution of the number of customers in the system. Find also a box plot. Describe the distribution and associate it to some family. Find the summary statistics. Explain your graphical and numerical results, in terms of the performance of the system.

Question 3.- To see whether the behavior of the variables analyzed in question 1 and 2 changes when we change the assumption about service time, repeat the analysis done in this activity, but under the assumption that the service time is exponential with parameter $\lambda = 1/15$.

References

Allen, A.O.(1990). Probability, Statistics, and Queueing Theory with Computer Science Applications, 2nd edition. Academic Press.

Goodman, R. (1988). Introduction to Stochastic Models. Benjamin/Cummings Publixhing Co., Inc.