Simulation 2

-Quantifying the World-

Lecturer: Darren Homrighausen, PhD

Overview

This session will focus on queuing theory and testing for distributional assumptions

- Queuing theory: Using stochastic processes to answer business-related questions such as:

  How many servers should I hire in my restaurant?

- Distributional assumptions: Allows us to formally answer the question:

  Is there evidence that a distribution doesn't fit my data?

Queuing Theory

General queue

In order to model a process, we need to specify

- Arrival time distribution: The inter-arrival times of customers (A)
- Service time distribution: How long it takes to service a customer (B)
- The number of servers: How are customers getting processed (C)
- Others: Whether the queue is allowed to be infinite or finite, system capacity, queuing discipline (e.g. FIFO, LIFO, ...)

The classic notation is to write it as a "$A/B/C$ queuing model"

Poisson process

By far the most popular approach is to specify:

- Arrival time distribution: Independent and memory-less
- Service time distribution: Independent and memory-less
- The number of servers: How are customers getting processed (C)

This is written as a "$M/M/C$ queuing model"

Famous probability theorem: The only "Independent and memory-less" process is the Poisson process

Poisson process in a restaurant

Let's suppose we are analyzing a restaurant

We assume that $C = 1$ as there is only one chef
(This is the easiest component to generalize)

We seat customers as they arrive, and they leave when they are done

Once all the seats are taken, then customers have to wait

Of course, customers won't wait if there is too long of a line

We can use statistics & probability to analyze this situation

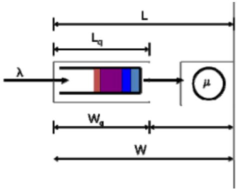
Poisson process in a restaurant

Let

- λ be the customer arrival rate
- μ be the service rate
- $\rho = \lambda/\mu$ be the utilization factor
- $(1-\rho)\rho^n$ be the probability of having n customers

We can compute various other quantities under this model:

- $L = \rho/(1-\rho)$ is the average number of customers
- $W = 1/(\mu - \lambda)$ is the average time spent in the restaurant
- $W_q = \rho/(\mu - \lambda)$ is the average time spent in the queue



Model implications

Example: Suppose based on previous experience and looking at other restaurants in the area we expect 400 people to come to the restaurant over three hours

$\rightarrow$ the arrival time is $\lambda = 400/180 = 2.22$ customers/minute

$\rightarrow$ we expect each customer to be there for $W = 55$ minutes

What is μ and ρ?

What is the probability of no one being in the restaurant?

Model implications

Suppose the capacity of the restaurant is 120 people

We expect a customer to leave if there more than 10 people already in line

What is the probability a customer leaves without eating at the restaurant?

Data

Now, suppose we have instead been running the store for a while and we observe a typical Sunday dinner rush

(Data: `arrivals.Rdata` and `numberInRestaurant.Rdata`)

We have the time between arrivals of customers and the number of customers in the restaurant at various points during the rush

How do we estimate the parameters?

(Do not proceed until you have a good idea about how to answer this question)

How can we get a measure of uncertainty about the parameter μ?

(After considering this question, go to `poissonProcessEstimation.R`)

Distributional Assumptions

Kolmogorov-Smirnov Test

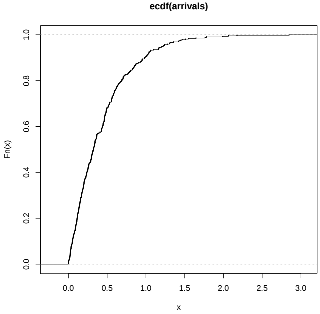
A very general hypothesis test for whether a distribution is appropriate is the Kolmogorov-Smirnov test

It compares the cumulative distribution function (CDF) for the data to see if they are "similar"

Let's look at an example

Kolmogorov-Smirnov Test

```
plot(ecdf(arrivals),cex=.2)
```



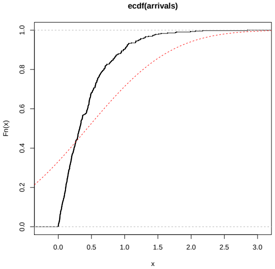
The hypothesis test:

$$H_0 : \text{The distributions are the same}$$
$$H_A : \text{The distributions are not the same}$$

Kolmogorov-Smirnov Test

Compare arrival times to a normal dist. with the same mean:

```
lines(seq(-1,4,length=1000),
   pnorm(seq(-1,4,length=1000),mean(arrivals)),lty=2,col='red')
```




```
out_ks = ks.test(arrivals,pnorm,mean(arrivals))
> print(out_ks)
One-sample Kolmogorov-Smirnov test
data:  arrivals
D = 0.33207, p-value < 2.2e-16
alternative hypothesis: two-sided
```

Kolmogorov-Smirnov Test

Run a Kolmogorov-Smirnov (KS) test comparing the arrivals times to the modeled distribution