Lecture 31

Produced by Dr. Worldwide

Welcome to the 305

Poisson Process

- When the set of interarrival times $\{\tau_i : i \geq 1\}$ follow an exponential distribution with rate λ, we have a Poisson process with rate λ

- Poisson process is a classic way to model random arrivals

- Parameter λ is called the rate, since it corresponds to the average number of arrivals per unit of time

- The cdf for an $EXP(\lambda)$ is $F(x) = 1 - e^{-\lambda x}$

- In simulation, $\tau_i = -\frac{1}{\lambda}\ln(U_i)$ where $\{U_1, U_2, U_3, \cdots\}$ are $i.i.d.\ Uniform[0,1]$

Ex: Coffee Shop Revenue

- A coffee shop is open for 8 hours Monday through Friday
- Customers arrive according to a Poisson process with a rate 5 per hour
- The time between arrivals $\tau_i \sim EXP(5)$
- The amount a customer spends can be approximated using a $Normal(2.5,1)$
- This means the average customer spends $2.50
- Q: What is the problem with using the Normal distribution for the amount spent?
- Any customer who arrives before closing will be served
- We want to simulate the revenue of the coffee shop for one day (8 hours)

Ex: Coffee Shop Revenue

- We must simulate both the arrival times of customers and the amount of money they will spend

- To generate interarrival times we use the inverse transform method

$$\tau_i = -\frac{1}{5}\ln(U_i)$$

where $U_i \sim Uniform[0,1]$

- To generate the amounts spend by different customers, we use the inverse transform method based on the NORM.INV function in Excel with a mean of 2.5 and a standard deviation of 1

- We simulate the times when customers arrive as well as the amounts they spend, then we add the expenses of all the customers who arrived during the time interval $[0,8]$

Ex: Coffee Shop Revenue

- Download CoffeeShop.xlsx from the link Sheet 1 on the course website
- Focus on tab named Revenue
- Descriptions of columns
  - Column A contains interarrival times
  - Column B contains arrival times (notice the calculation)
  - Column C contains the amount spent (notice the use of MAX() function
  - Column D checks to see if the customer made it by closing
- Simulated revenue for a single day

Revenue for the day:		
Sum of column C up to the last arrival in [0,8]		
120.062456		

Ex: Coffee Shop Queue

- Same coffee shop from previous example
- We want to simulate the customer queue
- Assume there is only one person at the cash register, and that this person both takes the order and prepares the coffee
- Moment of silence for this poor worker
- Each customer takes a random amount of time to be served, which we model as an exponential random variable with mean of 5 minutes (1/12 hour)

$$\frac{1}{\lambda} = 5\,min = \frac{1}{12}hr \quad \Rightarrow \quad \lambda = 12$$

- Customers are served in a first-come-first-serve basis

Ex: Coffee Shop Queue

- We want to use simulation to determine the maximum waiting time experienced by a customer during the day, and the number of customers still present at the coffee shop (either in queue or being served) at closing time

- Generate interarrival times $\{\tau_i : i \geq 1\}$ as before, making sure that we have enough to cover the 8-hour interval

- Let χ_i denote the service time of customer i

- To generate service times we use the inverse transform method

$$\chi_i = -\frac{1}{12}\ln(V_i)$$

  where $V_i \sim Uniform[0,1]$

- Status of coffee shop changes whenever a customer arrives or leaves

Ex: Coffee Shop Queue

- The number of customers in the coffee shop is the state of the system
- The arrival (A_i) and departure times (D_i) are known as events
- In simulation, we keep track of interarrival times, service times, arrival times, and departure times
- The only random numbers are interarrival times and the service times
- Customer *i* will start his/her service at two potential times
  - At the arrival time if no one is there
  - At the departure time of customer $i - 1$ if there is somebody

$$\max\left\{A_i, D_{i-1}\right\}, \qquad i = 2, 3, \ldots,$$

where the first customer starts always at time A_1

Ex: Coffee Shop Queue

- The number of customers in the coffee shop at the end of the day is

$$Number\ of\ Arrivals\ in\ [0,8] - Number\ of\ Departures[0,8]$$

- Download CoffeeShop.xlsx from the link Sheet 1 on the course website
- Focus on tab named Queue
- Notice that departure time is the sum of arrival time and service time
- Column H shows the waiting times which is the difference between the start of service and arrival time

Maximum waiting time	
0.42893819	hours
25.7362916	minutes

Ex: Coffee Shop Queue

- In the coffee shop it may be unrealistic to assume that customers arrive at the same rate throughout the day

- To make our models more realistic, we can change the rate during different periods of the day

- Assume that we can divide the interval $[0,8]$ into k periods $[0, t_1], (t_1 t_2], \cdots, (t_{k-1}, 8)]$ such that the arrival rate during each period is constant λ_i

- Simulate the arrivals in each period using the corresponding rate

- In periods with large λ_i , arrivals will be closer to each other, while in periods with small λ_i, they will be more spread out

Sample Mean and Variance

- Consider a simulation study where we have generated observations i.i.d. observations $\{X_1, X_2, \cdots, X_n\}$ of some random phenomenon

- $\{X_1, X_2, \cdots, X_n\}$ is a sample from some distribution F

- We can compute the sample mean and sample variance

$$\overline{X}(n) = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad S^2(n) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}(n))^2$$

- The random variables $\bar{X}(n)$ and $S^2(n)$ satisfy

$$E\left[\overline{X}(n)\right] = \mu \qquad E\left[S^2(n)\right] = \sigma^2$$

Where μ is the mean of F and σ^2 is the variance of F

Ex: Tres Simulations

- Download LLN.xlsx from link Sheet 2 on course website
- Simulate values for three different distributions
  - $Uniform[1,3]$: $f(x) = \frac{1}{2}$ for $1 \leq x \leq 3$
  - $6 - Sided\ Die$: $P(X = x) = \frac{1}{6}$ for $x \in \{1,2,3,4,5,6\}$
  - $Normal(5,4)$: $f(x) = \frac{1}{\sqrt{8\pi}} e^{-(x-5)^2/8}$ for $-\infty < x < \infty$
- For each of the examples, we want to compute true mean and true variance to sample mean and sample variance
- As we increase the number of simulated values, the sample mean and variance get closer to the true mean and variance

Ex: Tres Simulations

- For the $Uniform[1,3]$

$$\mu = E[X_1] = \int_{-\infty}^{\infty} x f(x) dx = \int_1^3 \frac{x}{2} dx = \left.\frac{x^2}{4}\right|_1^3 = \frac{9-1}{4} = 2$$

$$E[X_1^2] = \int_{-\infty}^{\infty} x^2 f(x) dx = \int_1^3 \frac{x^2}{2} dx = \left.\frac{x^3}{6}\right|_1^3 = \frac{27-1}{6} = \frac{13}{3}$$

$$\sigma^2 = E[X_1^2] - (E[X_1])^2$$

$$\sigma^2 = \frac{13}{3} - (2)^2 = \frac{13-12}{3} = \frac{1}{3}$$

Ex: Tres Simulations

- For the $6 - \textit{Sided Die}$

$$\mu = E[X_1] = \sum_{x=0}^{\infty} xP(X = x) = \sum_{x=1}^{6} \frac{x}{6} = \frac{1+2+3+4+5+6}{6} = \frac{7}{2}$$

$$E[X_1^2] = \sum_{x=0}^{\infty} x^2P(X = x) = \sum_{x=1}^{6} \frac{x^2}{6} = \frac{1+4+9+16+25+36}{6} = \frac{91}{6}$$

$$\sigma^2 = E[X_1^2] - (E[X_1])^2$$

$$\sigma^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$

- For the $Normal(5,4)$, $\mu = 5$ and $\sigma^2 = 4$

Ex: Tres Simulations

- For the Uniform sample, $n = 44$, for the 6-sided die, $n = 52$, for the Normal sample, $n = 53$

- Try increasing the sample sizes of all of these to 100, 1000, 10000 in the tabs named Uniform(1,3), 6-sided die, and Normal(5,4)

- By increasing the size of the sample *n*, we can clearly see that

$$\overline{X}(n) = \frac{1}{n}\sum_{i=1}^{n} X_i \approx \mu$$

- Q: How quickly does this convergence happen?

Central Limit Theorem

- One of the most famous results in probability states that if n is large, then

$$\sqrt{n}\left(\frac{\overline{X}(n)-\mu}{\sigma}\right)=\frac{1}{\sqrt{n\sigma^2}}\sum_{i=1}^{n}(X_i-\mu)$$

  has approximately the same distribution as a $Normal(0,1)$ random variable

- This is known as the Central Limit Theorem (CLT)

- This can be also written as

$$\bar{X}(n) \sim Normal\left(\mu, \sigma^2/\sqrt{n}\right) \qquad (approximately)$$

- This is true no matter what the distribution of X

Ex: CLT

- To see what the CLT says using simulation, we can generate *m* samples, each of size *n*, from distribution *F*

- For each of the *m* samples we compute the statistic

$$T = \sqrt{n}\left(\frac{\overline{X}(n) - \mu}{\sigma}\right)$$

- We plot the histogram for the sample $\{T_1, T_2, \cdots, T_m\}$, which should look like the histogram of a $Normal(0,1)$

- See tab named CLT in LLN.xlsx which contains such an example for the case when the distribution *F* corresponds to a $Uniform[1,3]$

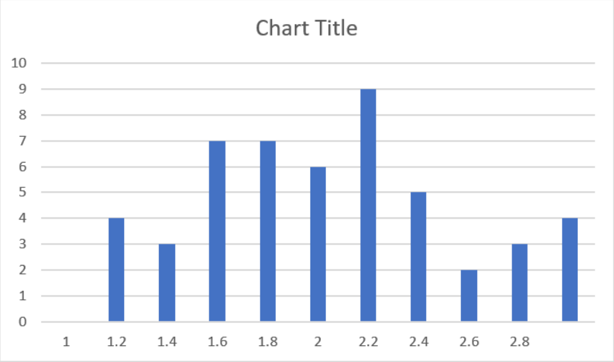
- In this example, $m = 20$ and $n = 50$

Ex: CLT

- Histogram for the first sample

Uniform[1,3] from Sample 1

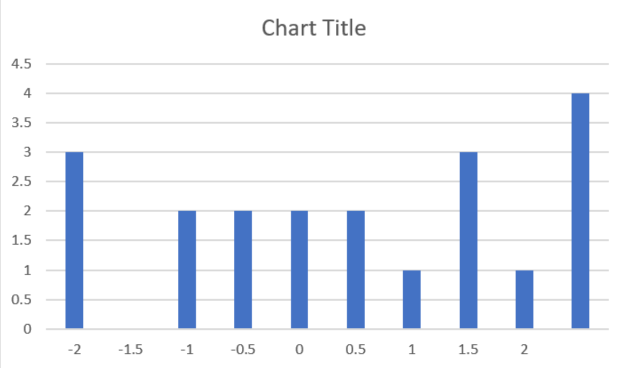
Bins	Frequency
1	0
1.2	4
1.4	3
1.6	7
1.8	7
2	6
2.2	9
2.4	5
2.6	2
2.8	3
	4



- Histogram for $\{T_1, T_2, \cdots, T_m\}$

T statistic

Bins	Frequency
-2	3
-1.5	0
-1	2
-0.5	2
0	2
0.5	2
1	1
1.5	3
2	1
	4



- Ideally, we need a lot larger *m* and need to compare different values of *n*

The End

Dale