Lecture 301

Produced by Dr. Worldwide

Welcome to the 305

Simulation for Continuous

- Many times we want to sample from a continuous distribution e.g. normal
- Suppose we want to simulate a random variable X having a cumulative distribution function (CDF)

$$F(x) = P(X \leq x)$$

- Then, we compute its inverse function $\mathrm{F}^{-1}(\mathrm{u})$ i.e. the function satisfying

$$F(F^{-1}(x)) = F^{-1}(F(x)) = x$$

- If U is a uniform $Uniform[0,1]$ random variable , then the random variable $F^{-1}(U)$ has the same distribution as X
- This method is called the inverse transform

Exponential Simulation

- An exponential random variable has cdf

$$F(x) = 1 - e^{-\lambda x}, \qquad x \geq 0$$

- With $\lambda > 0$ a parameter known as its "rate"
- Exponentials are often used to model the time between random arrivals
- To compute $F^{-1}(U)$, we set $u = F(x)$ and solve for x

$$u = 1 - e^{-\lambda x} \iff e^{-\lambda x} = 1 - u \iff$$

$$-\lambda x = \ln(1-u) \iff x = -\frac{1}{\lambda}\ln(1-u)$$

- If $U \sim Uniform[0,1]$, the random variable $X = -\frac{1}{\lambda}\ln(1-U)$ is an exponentially distributed random variable with rate λ

Exponential Simulation

- Note that if U is uniformly distributed in $[0,1]$, then $1 - U$ is too
- For the inverse transform method, we can replace U by $1 - U$ when convenient
- In the exponential example, we can set

$$X = -\frac{1}{\lambda} \ln U$$

Uniform Simulation

- Function RAND() samples $U \sim Uniform[0,1]$
- Q: How can we use RAND() to sample from $Uniform[a, b]$?
- If $\mathrm{U} \sim Uniform[0,1]$ and $X = (b - a)U + a$, then $\mathrm{X} \sim Uniform[a, b]$
- Q: What happens when $\mathrm{U} = 0$ or $\mathrm{U} = 1$?
- In Excel the formula is, $(b - a)RAND() + a$

Normal Simulation

- Most popular continuous distribution is the Normal distribution
- If $X \sim Normal(\mu, \sigma^2)$, then we can use the following pdf

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad x \in (-\infty, \infty)$$

- If $X \sim Normal(\mu, \sigma^2)$, then it can be written as $X = \sigma Z + \mu$ where $Z \sim N(0,1)$
- If we can simulate Standard Normal Z, then we can simulate any Normal X
- We will first focus on standard normal random variables

Normal Simulation

- Although there are more efficient methods for simulating Normal random variables, we could use the inverse transform method

- Set $Z = \Phi^{-1}(U)$ where $U \sim Uniform[0,1]$

- In Excel, the function $NORM.INV(u, \mu, \sigma)$ computes $F^{-1}(u)$ for the CDF function $F(x)$ where

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z-\mu)^2}{2\sigma^2}} dz$$

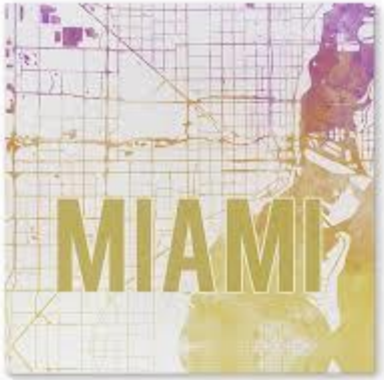
- Typically, we use NORM.INV to find percentiles (see Link 1 on course website)

- Therefore, the random number $NORM.INV(RAND(), \mu, \sigma)$ is $Normal(\mu, \sigma^2)$

Ex: Arrival Process

- We want to model the # of customers that come to a coffee shop during a day
- Since people walk into the coffee shop at random times, we want to use a model that reflects this fact
- We assume the times between consecutive arrivals are independent and identically distributed (i.i.d.) random variables
- Specifically, if we let τ_i be the time of arrival between the $(\mathrm{i}-1)$th and ith, then we can assume that the $\{\tau_i : i \geq 1\}$ are i.i.d.
- The set $\{\tau_i : i \geq 1\}$ are called interarrival times
- This set contains a random sample from a continuous distribution
- An assumption must be made about the distribution having CDF $F(x)$

Ex: Arrival Process



- An assumption must be made about the distribution with cdf
- Suppose that $F(x)$ is invertible (we can algebraically find $F^{-1}(u)$)
- Let $N(t)$ denote the number of arrivals in the interval $[0, t]$
- In our example,
  - $N(10) = Number\ of\ customers\ who\ visit\ coffee\ shop\ in\ 10\ minutes$
  - $N(60) = Number\ of\ customers\ who\ visit\ coffee\ shop\ in\ 1\ hour$
  - $N(1440) = Number\ of\ customers\ who\ visit\ coffee\ shop\ in\ 1\ day$
- Q: What is the mean of $N(t)$?
- Q: What is the standard deviation of $N(t)$?

Ex: Arrival Process

- Process for simulation
  - Step 1: Simulate a large enough sample of $\{\tau_i : i \geq 1\}$ based on cdf $\tau_i = F^{-1}(U_i)$ where $U_i \sim Uniform[0,1]$ such that $\sum \tau_i \geq t$
  - Step 2: Count the number of number of τ's that were able to "fit" into the interval $[0, t]$, i.e. find k such that

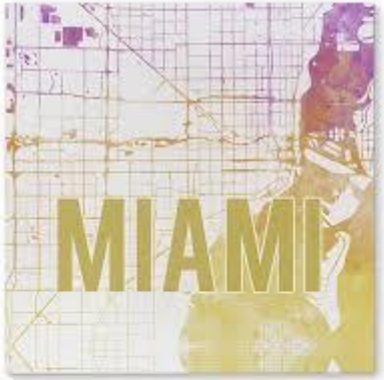
$$\sum_{i=1}^{k} \tau_i \leq t < \sum_{i=1}^{k+1} \tau_i$$

  - Step 3: Return $N(t) = k$
  - Step 4: Repeat steps 1-3

- For right now, we assume $\tau_i \sim Uniform[1,5]$

- Download ArrivalProcess.xlsx from the link Sheet 1 on the course website

Ex: Arrival Process



- We sample interarrival times according to the formula $(5-1)RAND()+1$
- We get actual arrival time of a customer adding the amount of time elapsed between this customer and the last customer (A8:A39) to the arrival time of the last customer (B8:B39)
- Simulated data

7	**Interarrivals**	**Arrival times**			
8	2.21529181	2.21529181			
9	2.37839429	4.5936861		*Number of arrivals*	
10	3.78539481	8.37908091		18	
11	4.18518667	12.5642676			
12	4.86816408	17.4324316			

Ex: Arrival Process

- The cell D10 contains a realization of $N(60)$ which counts the number of customers who arrive within the first 60 minutes

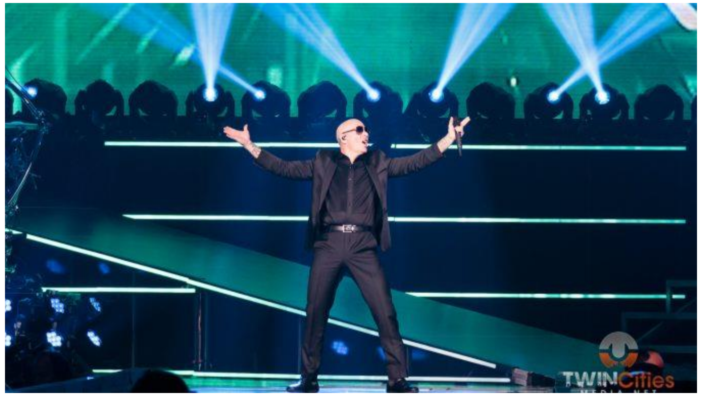
- Notice the Excel formula COUNTIF(B8:B39,”<60”)

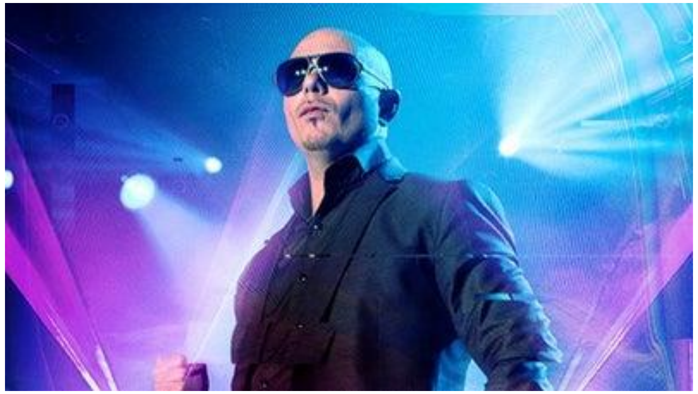
- Q: What is the problem with the $Uniform[1,5]$ distribution for interarrival times?

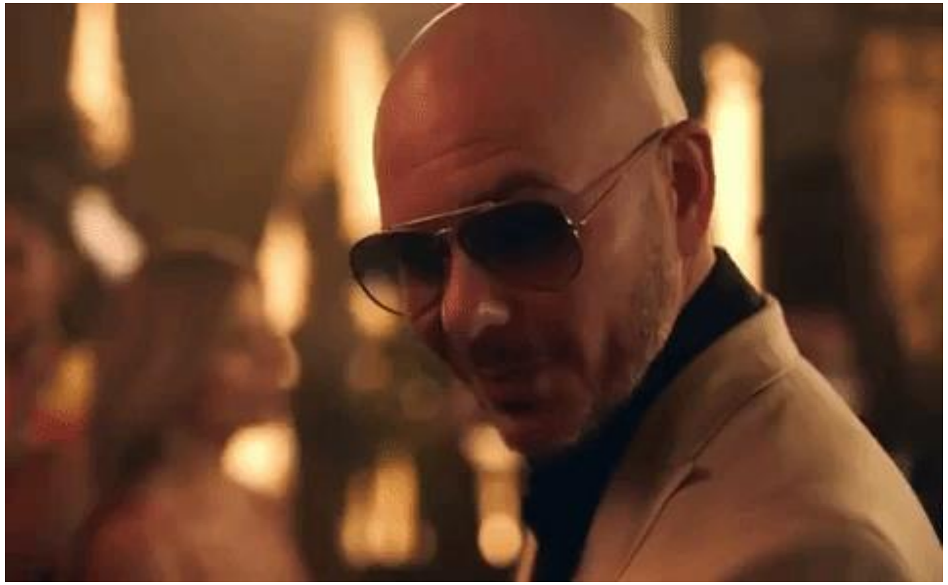
- Q: What is needed to estimate the mean and standard deviation of $N(60)$?

- Q: Does it matter if we change our Excel formula from COUNTIF(B8:B39,”<60”) to COUNTIF(B8:B39,”<=60”)?

The End







Dale