Lecture 29

Produced by Dr. Worldwide

Welcome to the 305

Stochastic Simulation

- Simulation is used to “simulate” the operations of various kinds of real-world systems or processes with the aid of a computer
- We make assumptions about how a system works
- These assumptions constitute a model
- Models are used to gain some understanding of how a system behaves
- In simple models, we can use analytical (mathematical) methods to obtain exact information on questions of interest
- In complex models, use simulation to evaluate the model numerically

Monte Carlo Process

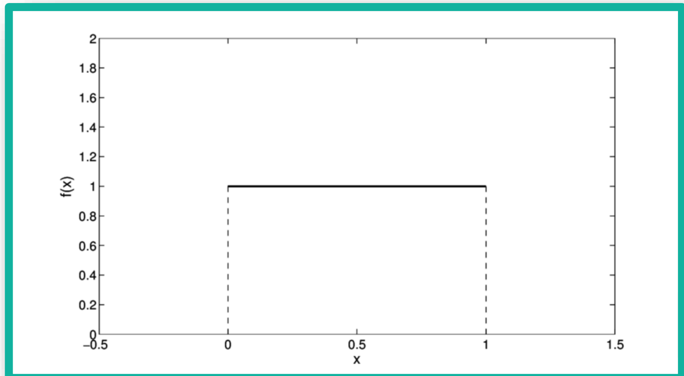
- Large proportion of applications of simulations are for probabilistic models
- Monte Carlo is a technique for selecting numbers randomly from a probability distribution
- Monte Carlo is not a simulation model, but a mathematical process used within a simulation
- Monte Carlo simulation can also be thought of as multiple probability simulation
- Read Link 1 on course website for more description about Monte Carlo

Random Number Generator

- To run a simulation, we may need "observations" from a random process
- If we "know" the distribution of these observations, we might be able to generate them artificially
- Historically, random numbers have been generated physically
  - Throwing dice
  - Dealing cards
  - Drawing balls from urns
- Not until mid 1950's where electronic random number generators were used
- Two kinds of random number generators: physical and numerical
- Computer are equipped with numerical random number generators

Random Number Generator

- Q: Other than my teaching, is anything truly random?
- Numerical random number generators are not really random
- The most basic type of random numbers that we can artificially generate are from the Uniform[0,1] distribution where numbers between 0 and 1 are equally likely
- RAND() function in Excel samples from this distribution
- We can generate random numbers from any other distribution from the RAND() function



Discrete Random Variable

- Variance of a discrete random variable X, denoted Var[X], is given by

  $Var[X] = E[(X - E[X])^2] = E[X^2] - (E[X])^2$

- Properties of the variance
  - $\text{Var}[X] \geq 0$ for all random variable X
  - $Var[cX] = c^2 \times Var[X]$ for any scalar $c \in \mathbb{R}$
  - $Var[\sum_{i=1}^{n} c_i X_i] = \sum_{i=1}^{n} c_i^2 Var[X_i]$ for any scalars $c_1, c_2, \cdots, c_n \in \mathbb{R}$ only if the X_i's are independent (uncorrelated)

- Standard deviation of a random variable X, denoted SD[X], is given by $\sqrt{Var[X]}$

- The standard deviation has the same units as the mean and the original variable

Discrete Random Variable

- Greek notation for the three parameters
  - Mean = μ
  - Variance = σ^2
  - Standard deviation = σ
- A parameter is any numerical quantity that characterizes a population

Ex: Tossing Coin

- Suppose $X = Number\ of\ Heads\ in\ 5\ Coin\ Tosses$
- The pmf $\mathrm{p}(X)$ can be calculated according to the binomial distribution

x	0	1	2	3	4	5
$p(x)$	$(0.5)^5$	$5(0.5)^5$	$10(0.5)^5$	$10(0.5)^5$	$5(0.5)^5$	$(0.5)^5$

- Q: What is the mean μ_{x}, variance σ_{x}^2, and standard deviation σ_x of X?
- Calculation of μ_{x}

$$
\begin{aligned}
E[X] &= 0p(0) + 1p(1) + 2p(2) + 3p(3) + 4p(4) + 5p(5) \\
&= (1)5(0.5)^5 + (2)10(0.5)^5 + (3)10(0.5)^5 + (4)5(0.5)^5 + 5(0.5)^5 \\
&= (0.5)^5\,(5 + 20 + 30 + 20 + 5) = (0.5)^5(80) = \frac{5}{2}
\end{aligned}
$$

Ex: Tossing Coin

- For variance, we first calculate $\mathrm{E}[X^2]$

$$E[X^2] = 0^2p(0) + 1^2p(1) + 2^2p(2) + 3^2p(3) + 4^2p(4) + 5^2p(5)$$
$$= (1)5(0.5)^5 + (4)10(0.5)^5 + (9)10(0.5)^5 + (16)5(0.5)^5 + 25(0.5)^5$$
$$= (0.5)^5\,(5 + 40 + 90 + 80 + 25) = (0.5)^5(240) = \frac{15}{2}$$

- Then, we calculate $\sigma_{\mathrm{x}}^2 = \mathrm{Var[X]} = \mathrm{E}[X^2] - (E[X])^2$

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = \frac{15}{2} - \left(\frac{5}{2}\right)^2 = \frac{5}{4}$$

- Download TossingCoin.xlsx from link Sheet 1 on course website
- Calculation of mean, variance, and standard deviation in Excel

Ex: Tossing Coin

- Use BINOM.DIST() to find individual probabilities for each value of X
- Check sum of probabilities
- Notice equivalent ways to calculate variance σ_x^2

X=# of Heads in 5 Coin Flips				
x	**p(x)=P(X=x)**	**x*p(x)**	**(x-2.5)^2*p(x)**	**x^2*p(x)**
0	0.03125	0	0.1953125	0
1	0.15625	0.15625	0.3515625	0.15625
2	0.3125	0.625	0.078125	1.25
3	0.3125	0.9375	0.078125	2.8125
4	0.15625	0.625	0.3515625	2.5
5	0.03125	0.15625	0.1953125	0.78125
Sum	*1*	*2.5*	*1.25*	*7.5*
	Total Prob.	Mean	Variance	
Alternative Calculation of Variance	E[X^2]	E[X]^2	E[X^2]-E[X]^2	
	7.5	6.25	1.25	
Standard Deviation	1.11803399			

Continuous Random Variable

- The probability density function f(x) (pdf) is a function that assigns a weight to every possible value of a continuous random variable such that

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

- It is important to note that $f(x)$ is not a probability and can take values greater than 1

- To find probability $\mathrm{P}(\mathrm{X} \in B)$, we calculate the following integral

$$P(X \in B) = \int_B f(x)\,dx$$

- Since there are an uncountable number of possible values, $\mathrm{P}(\mathrm{X} = x) = 0 \;\; \forall x \in \mathbb{R}$

Ex: Light Bulb Lifespan

- Lifetime of a light bulb (measured in years) is a continuous random variable having density $f(x) = e^{-x}$ for $x \geq 0$

- Q: What is the probability a random light bulb will burn out during the first month of use?

- Let $X = life\ of\ light\ bulb$ and $X \in [0, \infty)$

- Calculation of probability

$$
\begin{aligned}
P(X \leq 1/12) &= \int_0^{1/12} f(x)\, dx \\
&= \int_0^{1/12} e^{-x} dx \\
&= -e^{-x} \Big|_0^{1/12} \\
&= 1 - e^{-1/12}
\end{aligned}
$$

Ex: Waiting Time

- Train passes exactly every 15 minutes through the central station
- Waiting time for train (measured in minutes) is a continuous random variable
- Q: What is the probability if you walk into the station at random (without knowing when the last train passed), what is the probability you will have to wait less than 5 minutes to catch a train?
- Let $X = waiting\ time$ and $X \in [0,15]$
- Q: What is the probability $\mathrm{P}(X = 5)$?
- Makes sense to assume uniform distribution $f(x) = \frac{1}{15}$ which is like assuming that all values in the interval $[0,15]$ are equally likely.

Ex: Waiting Time

- Calculation of probability

$$\begin{aligned} P(X > 10) &= \int_{10}^{15} f(x)\, dx \\ &= \int_{10}^{15} \frac{1}{15}\, dx \quad \text{(since every point in } [0,15] \text{ is equally likely)} \\ &= \frac{x}{15}\Big|_{10}^{15} \\ &= \frac{1}{3} \end{aligned}$$

Normal Distribution

- The most famous continuous random variable is the normal distribution

- The normal distribution is a good approximation for many different natural phenomena, and in general, is what we obtain when we average many random variables

- The normal distribution $N(\mu, \sigma^2)$ has two parameters
  - Mean μ controls the center
  - Variance σ^2 controls the spread

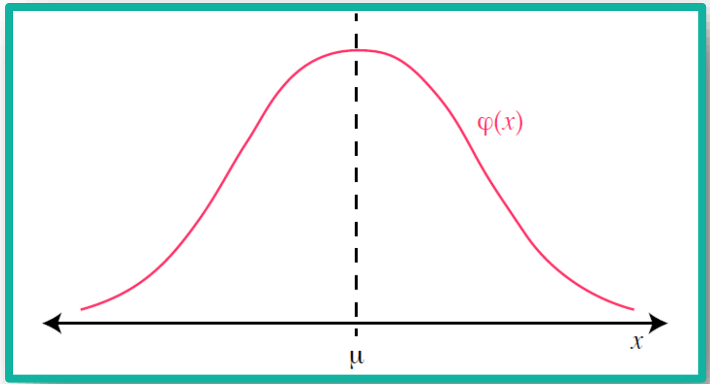
- The density function of a normal with parameters (μ, σ^2) is

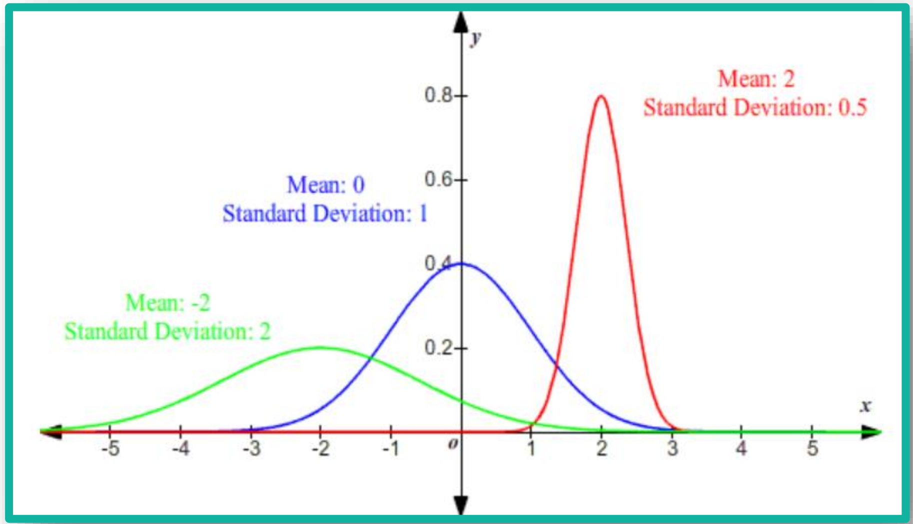
$$\varphi(x) = \frac{e^{-(x-\mu)^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}}, \qquad x \in \mathbb{R}$$

- For contribution to formula, this is also called the Gaussian distribution

Normal Distribution

- Special case when $\mu = 0 \; and \; \sigma^2 = 1$ is called the standard normal distribution
- Normal density function is symmetric with respect to its mean, and if the variance increases, the curve becomes “flatter”





Normal Distribution

- Calculating probabilities for normally distributed variables using integration is not possible since the function $\varphi(x)$ does not have an "anti-derivative"

- We can use numerical tables to calculate probabilities, but since we are using Excel we are not going to do this nonsense

- Let $X \sim N(\mu, \sigma^2)$ and let $x \in (-\infty, \infty)$
  - $P(X \leq x) = NORM.DIST(x, \mu, \sigma, TRUE)$ (Area to left of *x*)
  - $P(X \geq x) = 1 - NORM.DIST(x, \mu, \sigma, TRUE)$ (Area to right of x)

- Suppose $x_1, x_2 \in (-\infty, \infty)$ and $x_1 \leq x_2$,
  $P(x_1 \leq X \leq x_2) = NORM.DIST(x_2, \mu, \sigma, TRUE) - \mathrm{NORM.DIST}(\mathrm{x}_1, \mu, \sigma, TRUE)$

Ex: Carpet Store

- Super Shag is a carpet store and based off historical data the average weekly demand for Super Shag is 4,200 yards of carpet and the standard deviation of the weekly demand is 1,400 yards of carpet
- Store manager believes the weekly demand is approximately normal
- Let $X = weekly\ demand\ in\ yards$ and assume $X \sim N(4200, 1400^2)$
- Q: What is the probability the demand for Super Shag next week will exceed 3,000 yards?
- Q: What is the probability the demand next week will be 5,000 yards or less?
- Q: What is the probability next week's demand will be between 3,000 and 5,000 yards?

Ex: Carpet Store

- Download Shaggy.xlsx from link Sheet 2 on course website
- Observe the formulas for the calculation of our 3 probabilities

Probability	Calculation
Greater than 3,000	0.804317
Less than 5,000	0.7161454
Between 3,000 and 5,000	0.5204624

Sample Statistics

- We often assume the Normal distribution to find probabilities about our population; however, we never know the population mean and variance

- Suppose we conduct an experiment many times and take note of all our observations $X_1, X_2, \cdots, X_n$

- We assume these observations are distributed according to some distribution

- Our goal is to estimate the mean and variance of the population from which are sample comes from

- To estimate the true mean μ, we use the sample mean $\bar{X}$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Sample Statistics

- To estimate the true variance σ^2, we use the sample variance S^2

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

- Both the sample mean and the sample variance are random variables

- The statistics $\bar{X}$ and S^2 are unbiased estimators for μ and σ^2

- Calculation of these in Excel
  - Sample Mean = AVERAGE(data)
  - Sample Variance = VAR.S(data)
  - Sample Standard Deviation = STDEV.S(data)

- Create Excel file with data 64,63,69,70,68,75,69,66,67,62 and calculate the sample mean and sample variance

The End

Dale