

# FUNDAMENTALS OF QUANTITATIVE MODELING

Richard Waterman

*Module 3: Probabilistic models*



Wharton  
UNIVERSITY of PENNSYLVANIA

ONLINE

---

## Module 3 content

- What are probabilistic models?
- Random variables and probability distributions -- the building blocks
- Examples of probabilistic models
- Summaries of probability distributions: means, variances and standard deviation
- Special random variables: Bernoulli, Binomial and Normal
- The Empirical Rule

---

# Probabilistic models

- These are models that incorporate ***random variables*** and ***probability distributions***
- Random variables represent the potential outcomes of an uncertain event
- Probability distributions assign probabilities to the various potential outcomes
- We use probabilistic models in practice because realistic decision making often necessitates recognizing uncertainty (in the inputs and outputs of a process)

---

# Key features of a probabilistic model

- By incorporating ***uncertainty*** explicitly in the model we can measure the uncertainty associated with the outputs, for example by giving a range to a forecast, which is a more realistic goal
- In a business setting incorporating ***uncertainty*** is synonymous with understanding and quantifying the ***risk*** in a business process, and ideally leads to better management decisions

# Oil prices



If you run an energy intensive business, an airline for example, then the price of oil is a key determinant of your profitability



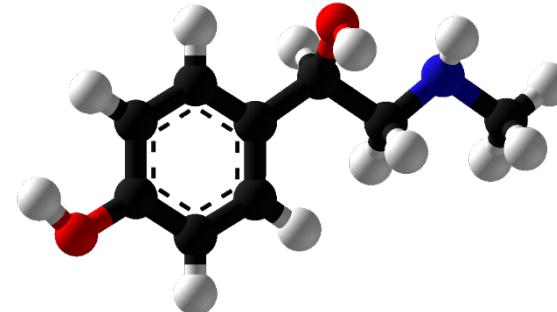
For medium or long-term investment planning (buying new planes) the future price of oil is an important consideration



But who knows the price of oil in ten years? No-one. But we may be able to put a probability distribution around the future price and incorporate the uncertainty into the decision making process

# Valuing a drug development company

- A company has 10 drugs in a development portfolio
- Given a drug has been approved, you have predicted its revenue
- But whether a drug is approved or not is an uncertain future event (a random variable). You have estimated the probability of approval
- You only wish to invest in the company if the company's expected total revenue for the portfolio is over \$10B in 5 years time
- You need to calculate the ***probability distribution*** of the total revenue to understand the investment risk



---

# Some examples of probabilistic models

- *Regression models* (module 4)
- *Probability trees*
- *Monte Carlo* simulation
- *Markov models*

# Regression models

- $E(Price | Carats) = -259.6 + 3721 \times Carats$
- The gray band gives a prediction interval for the price of a diamond taken from this population



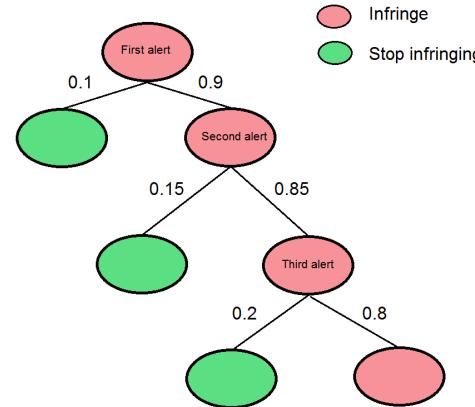
---

# Regression models

- Regression models use data to estimate the relationship between the mean value of the outcome(Y) and a predictor variable(X)
- The intrinsic variation in the raw data is incorporated into forecasts from the regression model
- The less noise in the underlying data the more precise the forecasts from the regression model will be

# Probability trees

- Probability trees allow you to propagate probabilities through a sequence of events



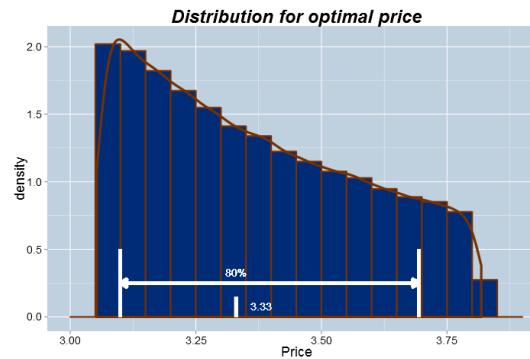
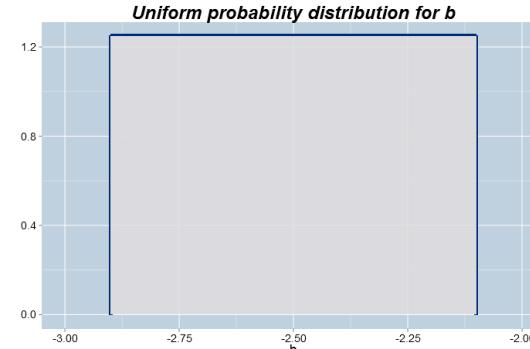
- $P(\text{Stop infringing}) = 0.1 + 0.9 \times 0.15 + 0.9 \times 0.85 \times 0.2 = 0.388.$

# Monte Carlo simulation

- From the demand model:  
Quantity = 60,000 Price<sup>-2.5</sup>
- The optimal price was  $p_{opt} = \frac{c b}{1+b}$ , where b = -2.5, c is the cost, c = 2, and  $p_{opt} \approx 3.33$
- But what if b is not known exactly?
- Monte Carlo simulation replaces the number -2.5 with a random variable, and recalculates  $p_{opt}$  using different realizations of this random variable from some stated probability distribution

# Input and output from a MC simulation

- Input:  $b$  from a uniform distribution between -2.9 and -2.1
- Output:  $p_{opt} = \frac{c b}{1+b}$
- 100,000 replications
- Interval = (3.1,3.7)



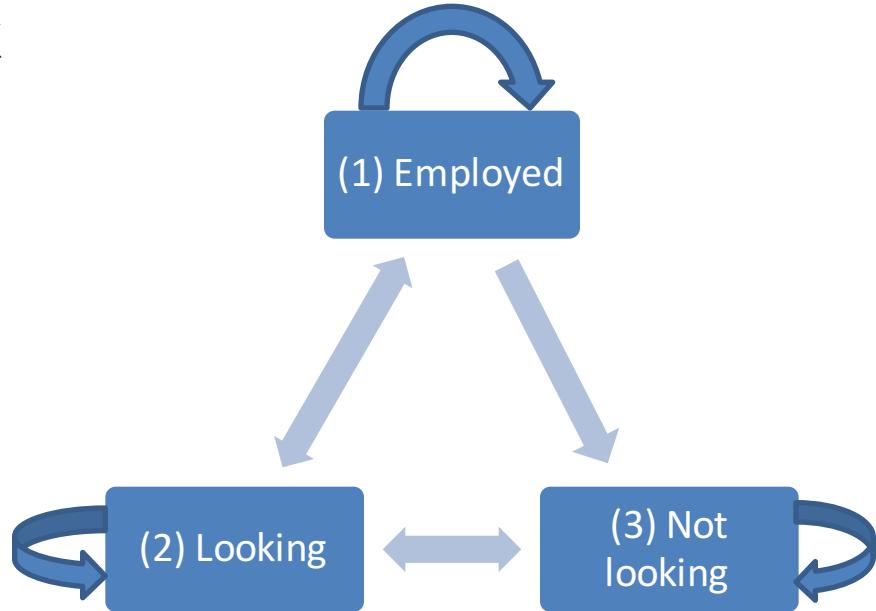
---

# Markov chain models

- Dynamic models for discrete time state space transitions
- Example: employment status (the state of the chain)
- Treat time in 6 month blocks
- Model states:
  1. Employed
  2. Unemployed and looking
  3. Unemployed and not looking

# Probability transition matrix

$$\begin{array}{c} \text{Next state} \\ \begin{pmatrix} 1 & 2 & 3 \end{pmatrix} \\ \text{Current state} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.2 & 0.7 & 0.1 \\ 0 & 0.3 & 0.7 \end{pmatrix} \end{array}$$



Markov property (lack of memory): transition probabilities only depend on the current state, not on prior states. Given the present, the future does not depend on the past

---

# Building blocks of probability models

- Random variables (discrete and continuous)
- Probability distributions
- Random variables represent the potential outcomes of an uncertain event
- Probability distributions assign probabilities to the various potential outcomes

# A discrete random variable

- Roll a fair die



x	1	2	3	4	5	6
$P(X = x)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$

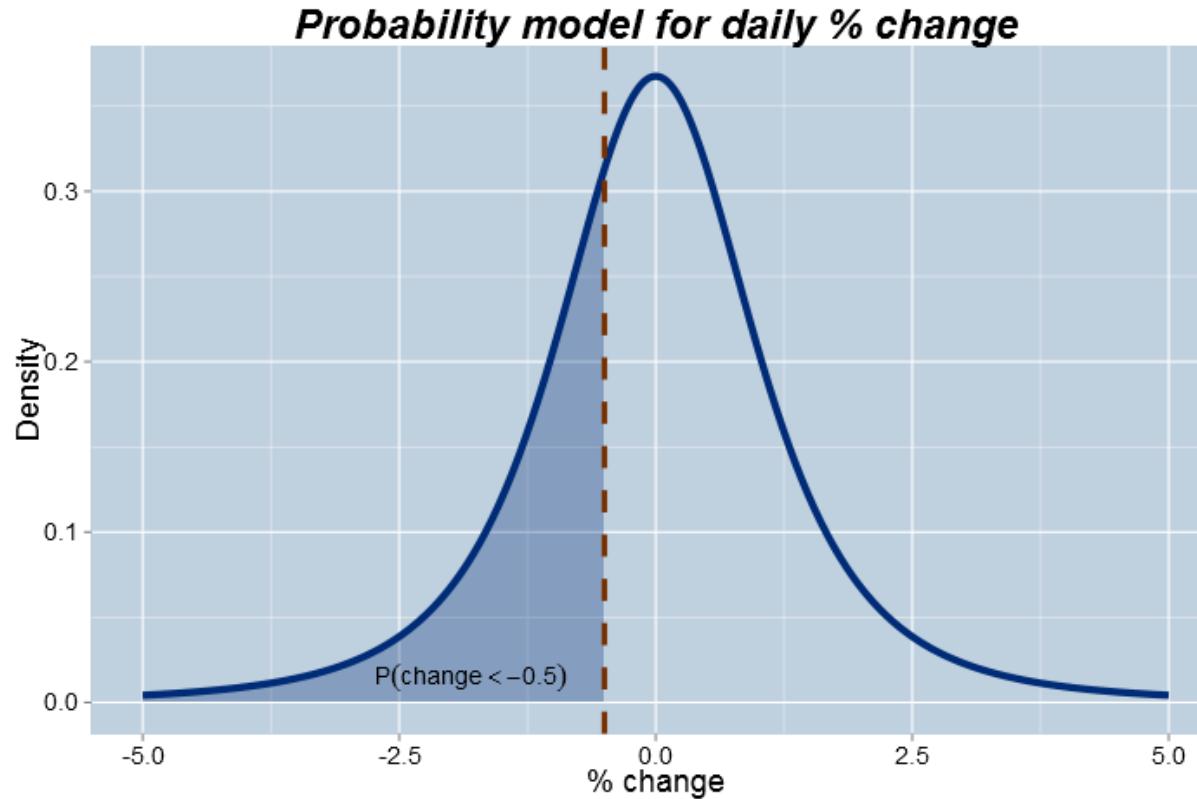
- Probabilities lie between 0 and 1 inclusive
- Probabilities add to 1

---

# A continuous random variable

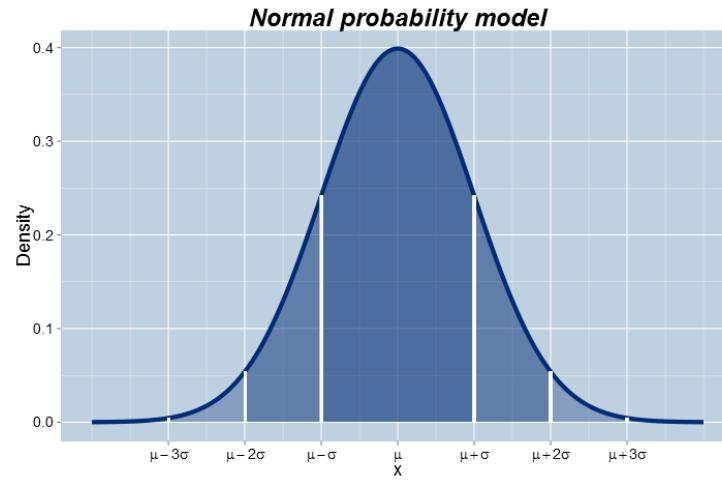
- The *percent change* in the S&P500 stock index tomorrow:  $100 \times \frac{p_{t+1} - p_t}{p_t}$  where  $p_t$  is the closing price on day t
- It can potentially take on **any** number between -100% and infinity
- For a continuous random variable probabilities are computed from areas under the *probability density function*

# Probability distribution of S&P500 daily % change



# Key summaries of probability distributions

- Mean ( $\mu$ ) measures centrality
- Two measures of spread:
  - Variance ( $\sigma^2$ )
  - Standard deviation ( $\sigma$ )



# The Bernoulli distribution

- The random variable  $X$  takes on one of two values:
  - $P(X = 1) = p$
  - $P(X = 0) = 1-p$
- Often viewed as an experiment that takes on two outcomes, success/failure. Success = 1 and failure = 0
- $\mu = E(X) = 1 \times p + 0 \times (1 - p) = p$
- $\sigma^2 = E(X - \mu)^2 = (1 - p)^2 p + (0 - p)^2 (1 - p) = p(1 - p)$
- $\sigma = \sqrt{p(1 - p)}$
- For  $p = 0.5$ ,  $\mu = 0.5$ ,  $\sigma^2 = 0.25$  and  $\sigma = 0.5$

---

## Example: drug development

- Will a drug under development be approved?
- $X = \begin{cases} Yes = 1 \\ No = 0 \end{cases}$
- $P(X = Yes) = 0.65$
- $P(X = No) = 0.35$
- If drug is approved then the projected revenue is \$500m, 0 otherwise
- $\text{Expected(Revenue)} = \$500m \times 0.65 + \$0 \times 0.35 = \$325m$

# The Binomial distribution

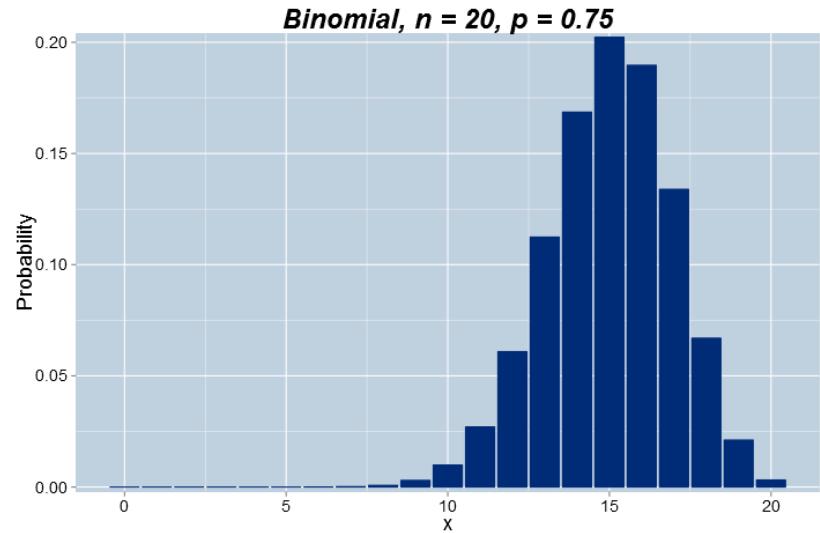
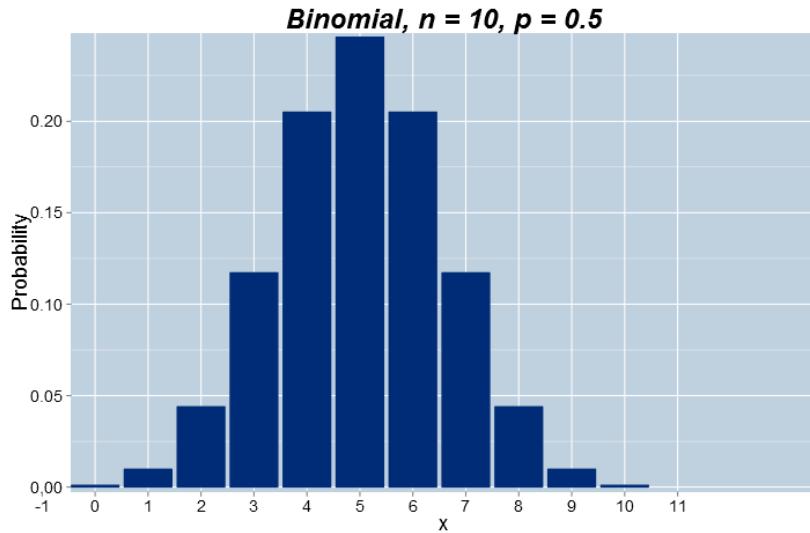
---

- A Binomial random variable is the number of success in  $n$  ***independent*** Bernoulli trials
- Independent means that  $P(A \text{ and } B) = P(A) \times P(B)$
- Independence means that knowing that A has occurred provides no information about the occurrence of B
- Independence is a common simplifying assumption in many probability models and makes their construction and subsequent calculations much easier

# The Binomial distribution

- Example: toss a fair coin 10 times and count the number of heads (call this  $X$ )
- Then  $X$  has a Binomial distribution with parameters  $n = 10$  and  $p = 0.5$ .
- In general:  $P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$ , where  $\binom{n}{x}$  is the **binomial coefficient**:  $\frac{n!}{x!(n-x)!}$
- $\mu = E(X) = np$ ,  $\sigma^2 = E(X - \mu)^2 = np(1 - p)$

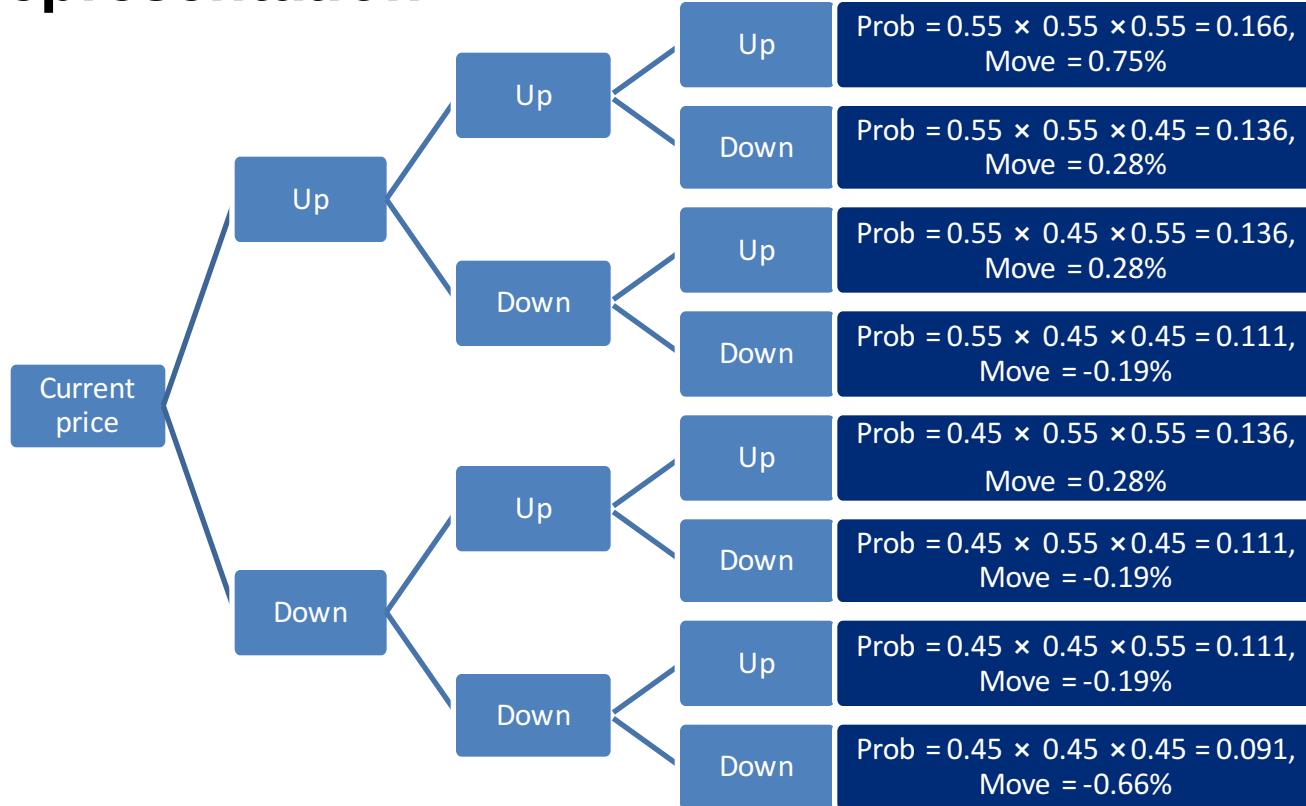
# Binomial probability distributions



## Example: Binomial models for markets (oil for example)

- Assume that the market either goes up or down each day
- It goes up  $u\%$  with probability  $p$  and down  $d\%$  with probability  $1-p$
- Assume days are *independent*
- Example:  $p = 0.55$ ,  $1 - p = 0.45$ ,  $u = 0.25\%$ ,  $d = 0.22\%$
- Take a time horizon of 3 days
- There are 8 possible outcomes:
  - $\{\text{UUU}\}, \{\text{UUD}\}, \{\text{UDU}\}, \{\text{UDD}\}, \{\text{DUU}\}, \{\text{DUD}\}, \{\text{DDU}\}, \{\text{DDD}\}$
- For each outcome there will be an associated market move. For example, if we see  $(U,U,U)$  then the market goes up by a factor of  $1.0025 \times 1.0025 \times 1.0025 = 1.007519$ , that is a little over  $\frac{3}{4}$  of a percent.

# Tree representation



---

# The Normal distribution

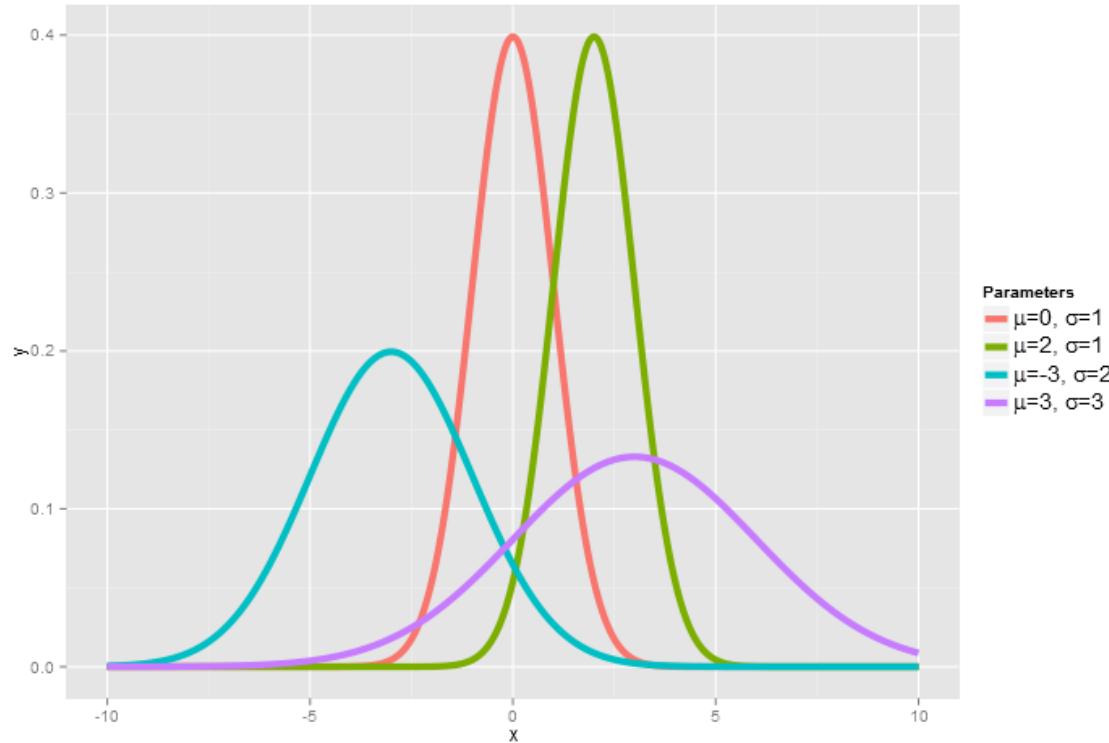
- The Normal distribution, colloquially known as the *Bell Curve*, is the most important modeling distribution
- Many disparate processes can be well **approximated** by Normal distributions
- There are mathematical theorems (the Central Limit Theorem) that tell us Normal distributions should be expected in many situations
- A Normal distribution is characterized by its mean  $\mu$  and standard deviation  $\sigma$ . It is symmetric about its mean

---

# Examples

- There is a *universality* to the Normal distribution
  - Biological: heights and weights
  - Financial: stock returns
  - Educational: exam scores
  - Manufacturing: the length of an automotive component
- It is therefore often used as a distributional assumption in Monte Carlo simulations (knowing the mean and standard deviation is enough to define a Normal distribution)

# Plots of various Normal distributions

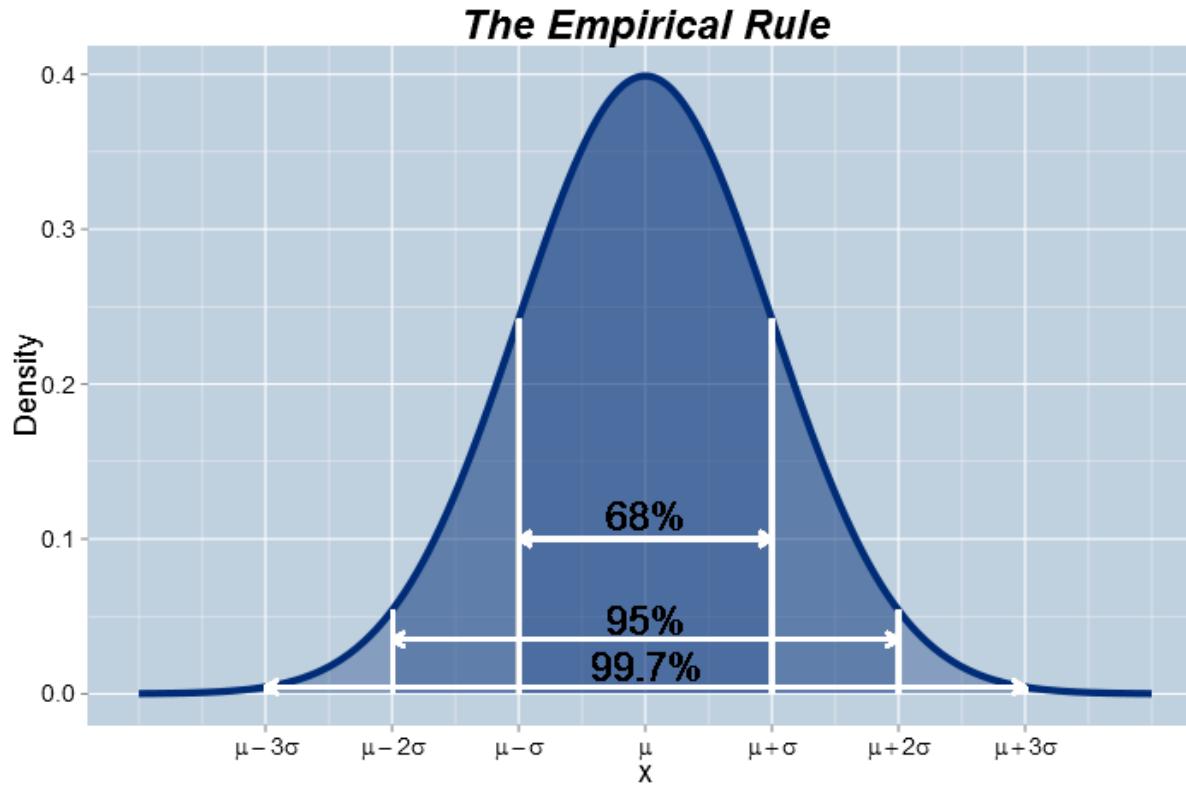


---

# The Empirical Rule

- The Empirical Rule is a rule for calculating probabilities of events when the underlying distribution or observed data is approximately Normally distributed
- It states
  - There is an approximate **68%** chance that an observation falls within **one** standard deviation from the mean
  - There is an approximate **95%** chance that an observation falls within **two** standard deviations from the mean
  - There is an approximate **99.7%** chance that an observation falls within **three** standard deviations from the mean

# The Empirical Rule illustrated



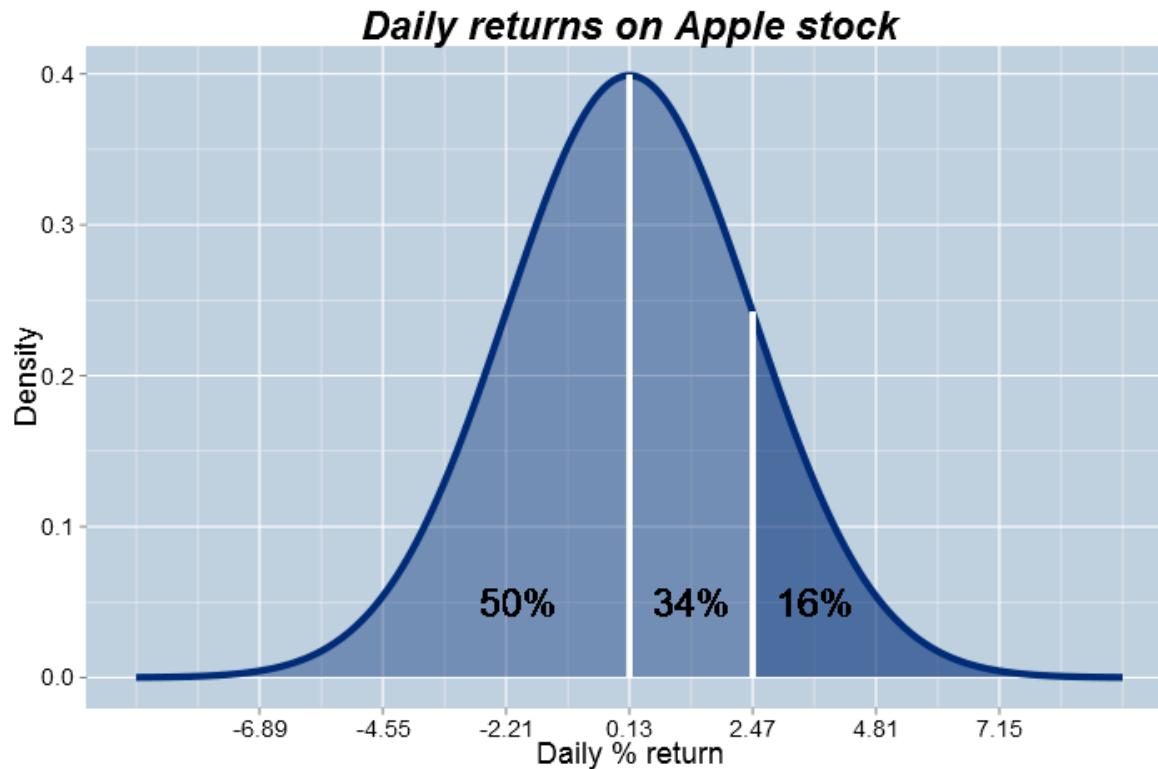
# Empirical Rule example

- Assume that the daily **return** on Apple's stock is approximately Normally distributed with mean  $\mu = 0.13\%$  and  $\sigma = 2.34\%$
- What is the probability that tomorrow Apple's stock price increases by more than 2.47%?
- Technique: count how many standard deviations 2.47% is away from the mean, 0.13%. Call this **counter** the **z-score**

$$Z = \frac{2.47 - 0.13}{2.34} = 1$$

- So, from the Empirical Rule the probability equals approximately 16%

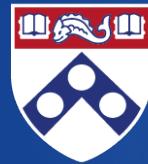
# Illustrating the answer



---

# Module 3 Summary

- What are probabilistic models?
- Random variables and probability distributions -- the building blocks
- Examples of probabilistic models
- Summaries of probability distributions: means, variances and standard deviation
- Special random variables: Bernoulli, Binomial and Normal
- The Empirical Rule



Wharton  
UNIVERSITY *of* PENNSYLVANIA



Wharton  
UNIVERSITY *of* PENNSYLVANIA

---

ONLINE