ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Elements of Statistics and Probability

LECTURE 2 –Introduction to Probability

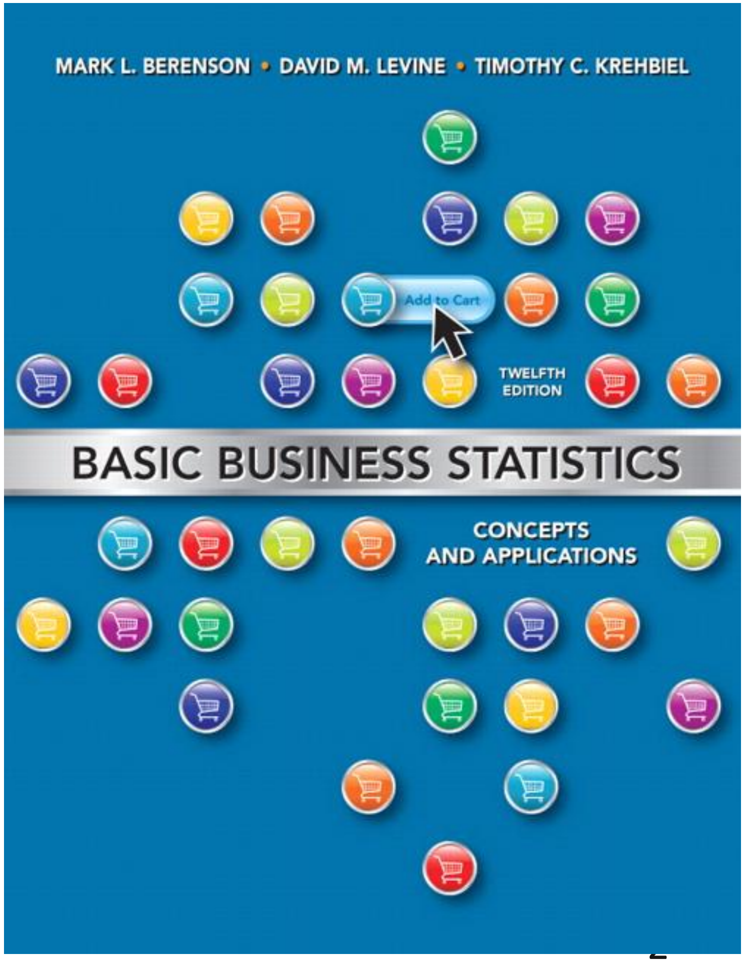
Xanthi Pedeli
Assistant Professor, xpedeli@aueb.gr
Department of Statistics, AUEB

Notes by Ioannis Ntzoufras, *Professor*
Department of Statistics, AUEB

1.2. Measuring uncertainty & risk using probabilities

The main material of this lecture is taken by

Berenson, M.L., Levine, D.M. και Krehbiel, T.K. (2011). Basic Business Statistics. 12th edition. Pearson.



What a Manager Needs to Know About Statistics

- To Know How to Properly Present Information
- To Know How to Draw Conclusions about Populations Based on Sample Information
- To Know How to Improve Processes
- To Know How to Obtain Reliable Forecasts

The Growth and Development of Modern Statistics

Stat-istics: collection of population and economic information vital to the state

↓

The development of the mathematics of probability theory

↓

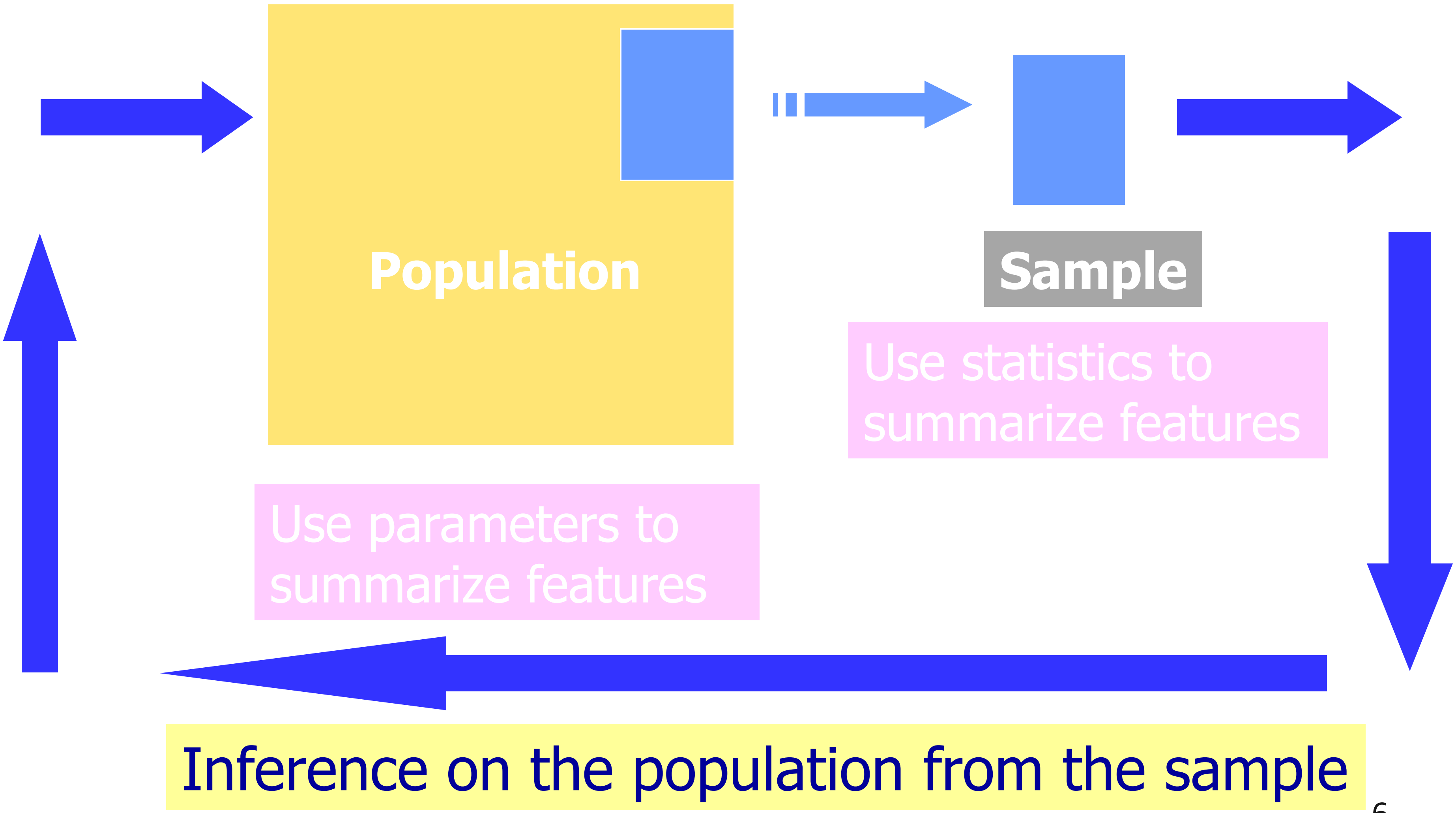
The advent of the computer

Some Important Definitions

ΟΠΑ AUEB

- A Population (Universe) is the Whole Collection of Things Under Consideration
- A Sample is a Portion of the Population Selected for Analysis
- A Parameter is a Summary Measure Computed to Describe a Characteristic of the Population
- A Statistic is a Summary Measure Computed to Describe a Characteristic of the Sample

Population and Sample

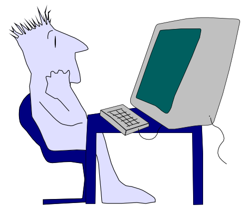


Statistical Methods

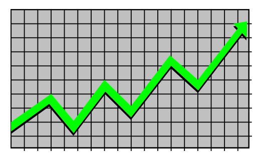
- Descriptive Statistics
  - Collecting and describing data
- Inferential Statistics
  - Drawing conclusions and/or making decisions concerning a population based only on sample data

Descriptive Statistics

- Collect Data
  - E.g., Survey



- Present Data
  - E.g., Tables and graphs



- Characterize Data
  - E.g., Sample Mean = $\dfrac{\sum X_i}{n}$

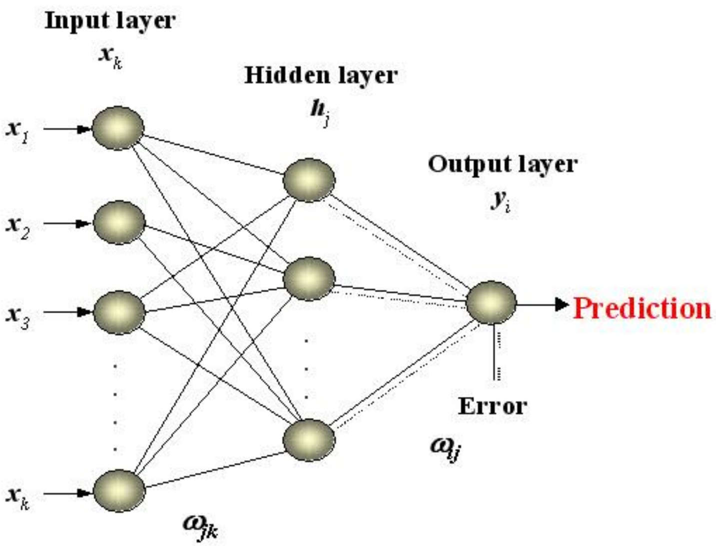
Inferential Statistics

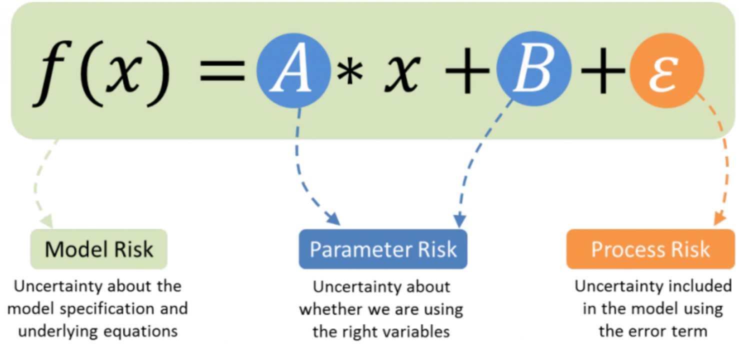
- Estimation
  - E.g., Estimate the population mean weight using the sample mean weight
- Hypothesis Testing
  - E.g., Test the claim that the population mean weight is 120 pounds

Drawing conclusions and/or making decisions concerning a population based on sample results.

Inferential Statistics

- Model based inference
  - Explain causal relationships ⇒ leads to preventive strategies (i.e. change the future)
  - Predict future behavior (when no intervention is possible)

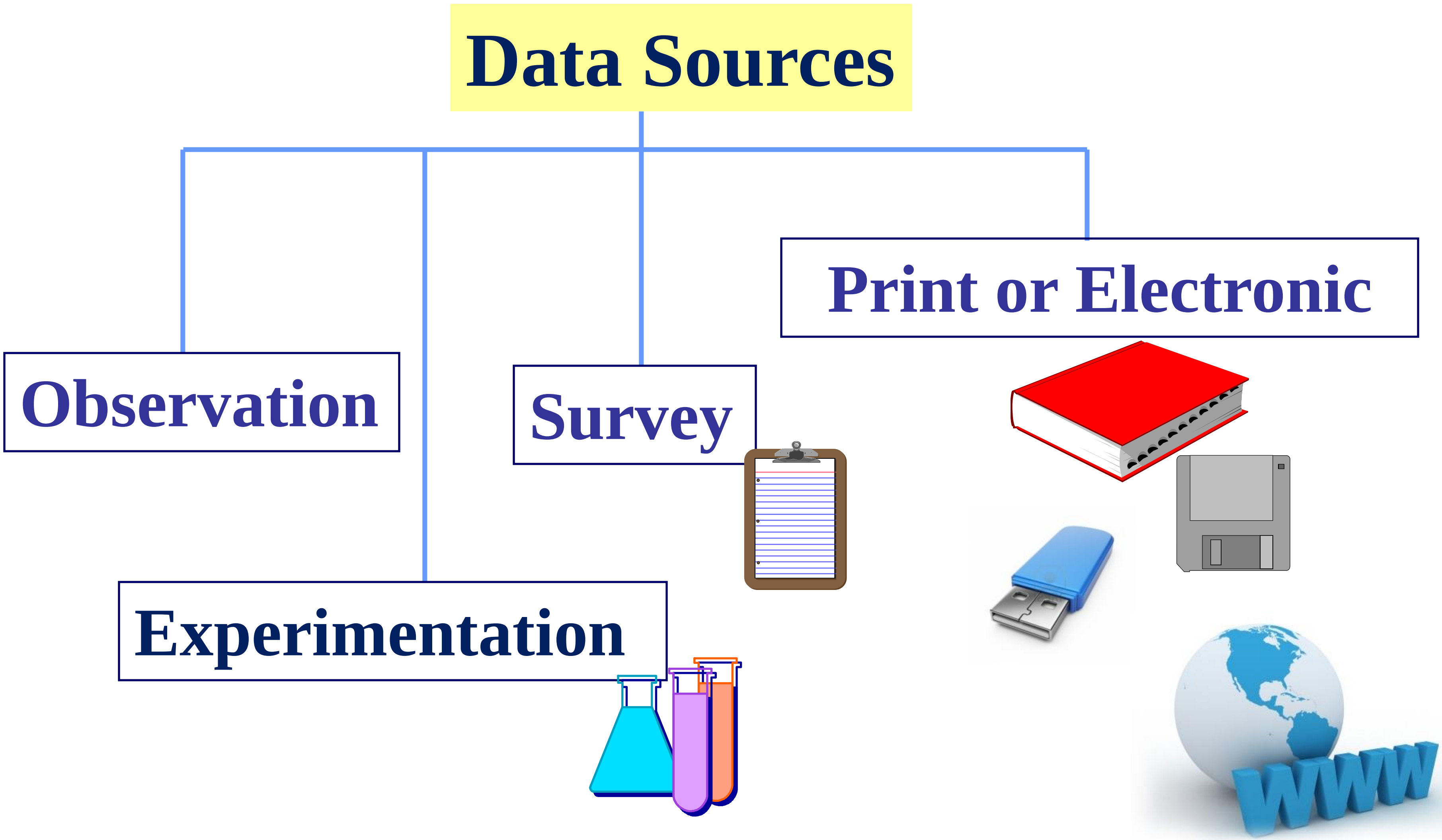




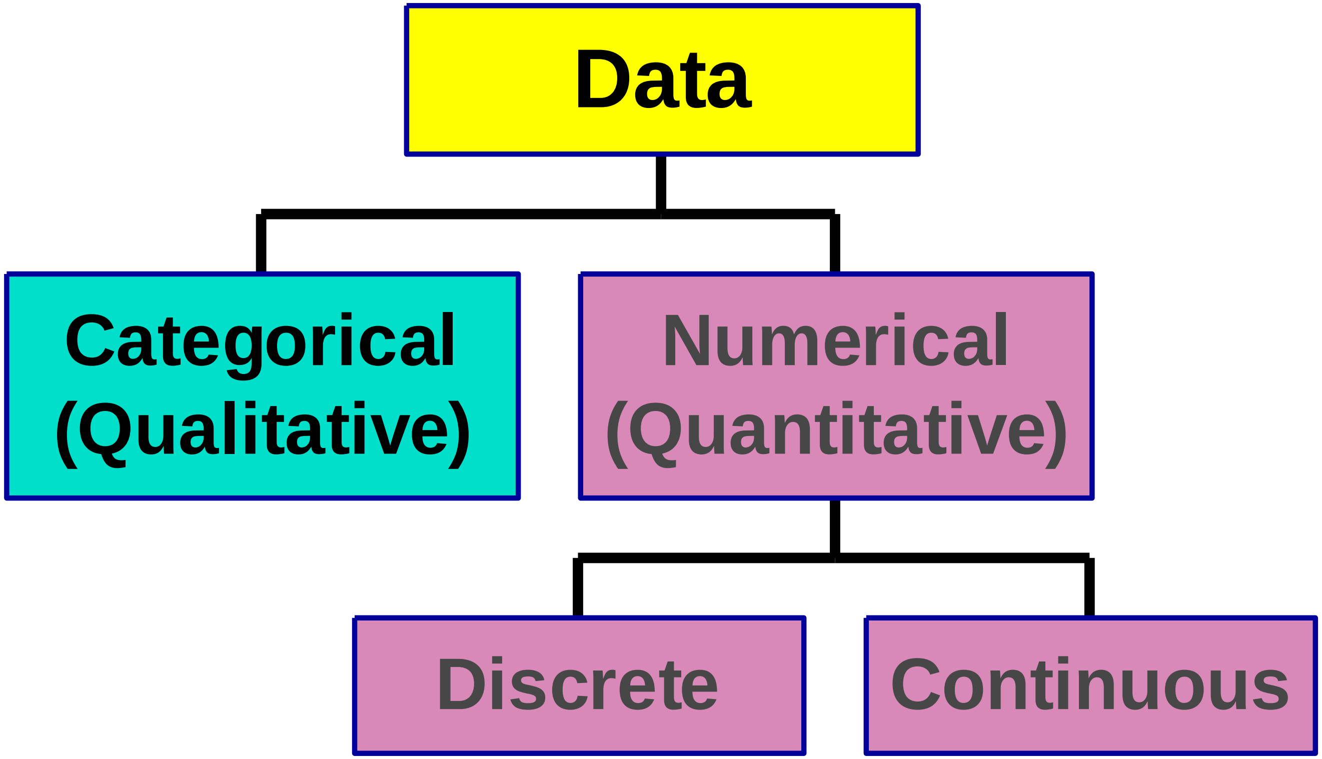
Why We Need Data

- To Measure Performance of Ongoing Service or Production Process
- To Evaluate Conformance to Standards
- To Assist in Formulating Alternative Courses of Action
- To Satisfy Curiosity

Data Sources



Types of Data

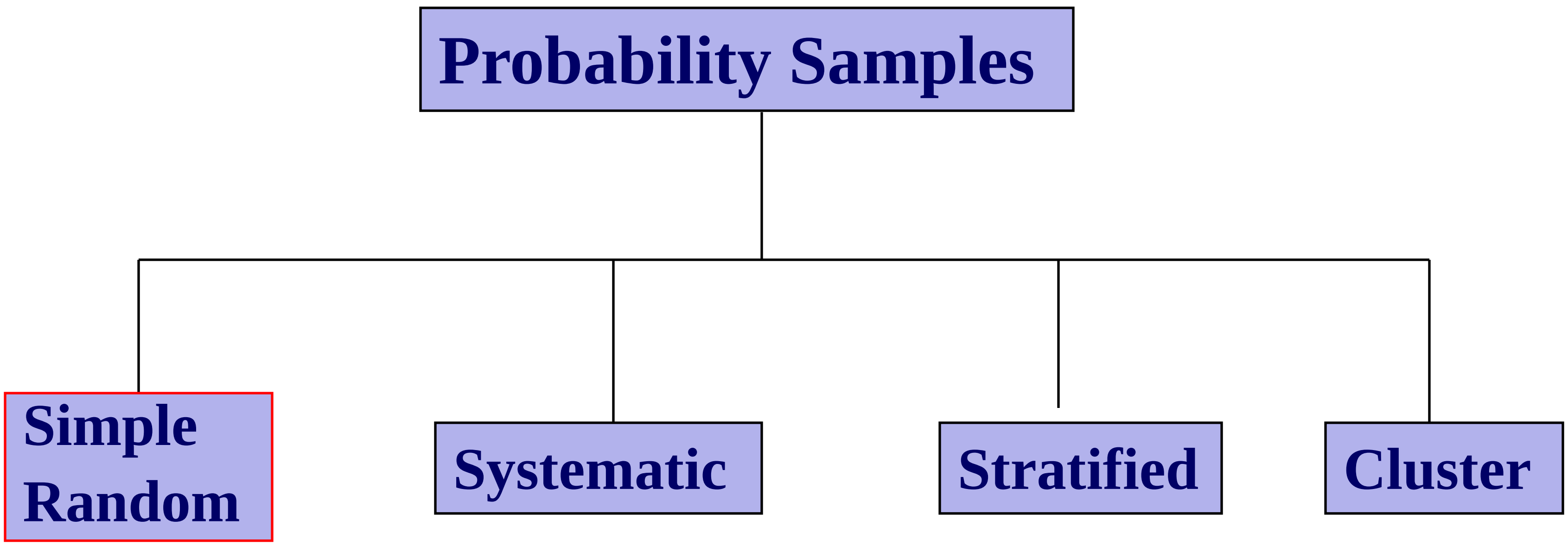


Reasons for Drawing a Sample

- Less Time Consuming Than a Census
- Less Costly to Administer Than a Census
- Less Cumbersome and More Practical to Administer Than a Census of the Targeted Population

Probability Sampling

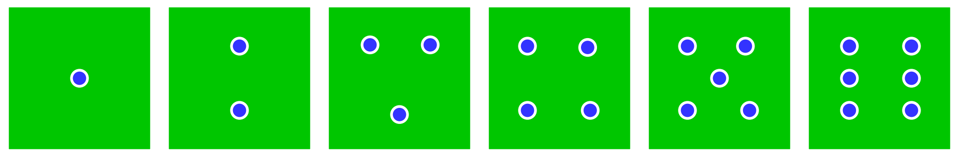
- Subjects of the Sample are Chosen Based on Known Probabilities



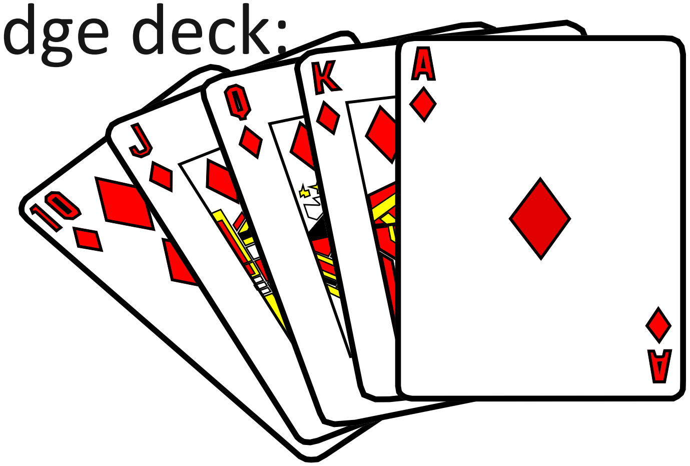
- Basic Probability Concepts
  - Sample spaces and events, simple probability, joint probability
- Conditional Probability
  - Statistical independence, marginal probability
- Bayes' Theorem

Sample Spaces

- Collection of All Possible Outcomes
  - E.g., All 6 faces of a die:



  - E.g., All 52 cards of a bridge deck:



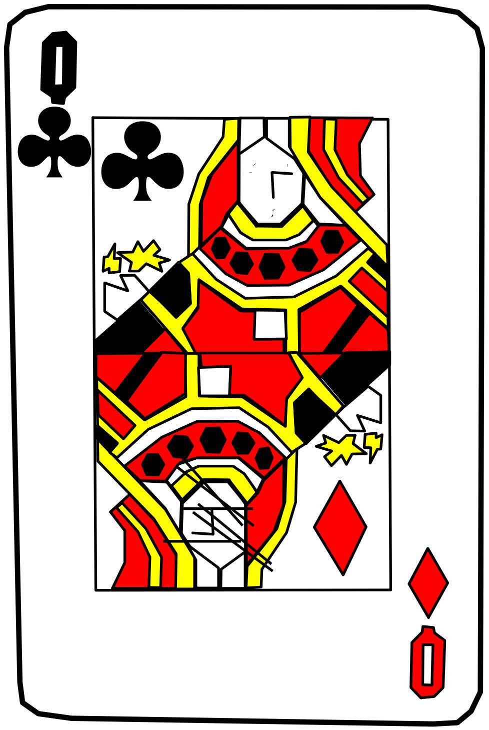
Events

- Simple Event
  - Outcome from a sample space with 1 characteristic
  - E.g., a Red Card from a deck of cards
- Joint Event
  - Involves 2 outcomes simultaneously
  - E.g., an Ace which is also a Red Card from a deck of cards

Special Events

- Impossible Event
  - Impossible event
  - E.g., Club & Diamond on 1 card draw
- Complement of Event
  - For event A, all events not in A
  - Denoted as A’
  - E.g., A: Queen of Diamond
    A’: All cards in a deck that are not Queen of Diamond

Impossible Event



Special Events

(continued)

- Mutually Exclusive Events
  - Two events cannot occur together
  - E.g., A: Queen of Diamond; B: Queen of Club
- Collectively Exhaustive Events
  - One of the events must occur
  - The set of events covers the whole sample space
  - E.g., A: All the Aces; B: All the Black Cards; C: All the Diamonds; D: All the Hearts
    - Events A, B, C and D are collectively exhaustive
    - Events B, C and D are also collectively exhaustive

Probability

Probability words

There's a good chance it will rain today.

It is quite likely...

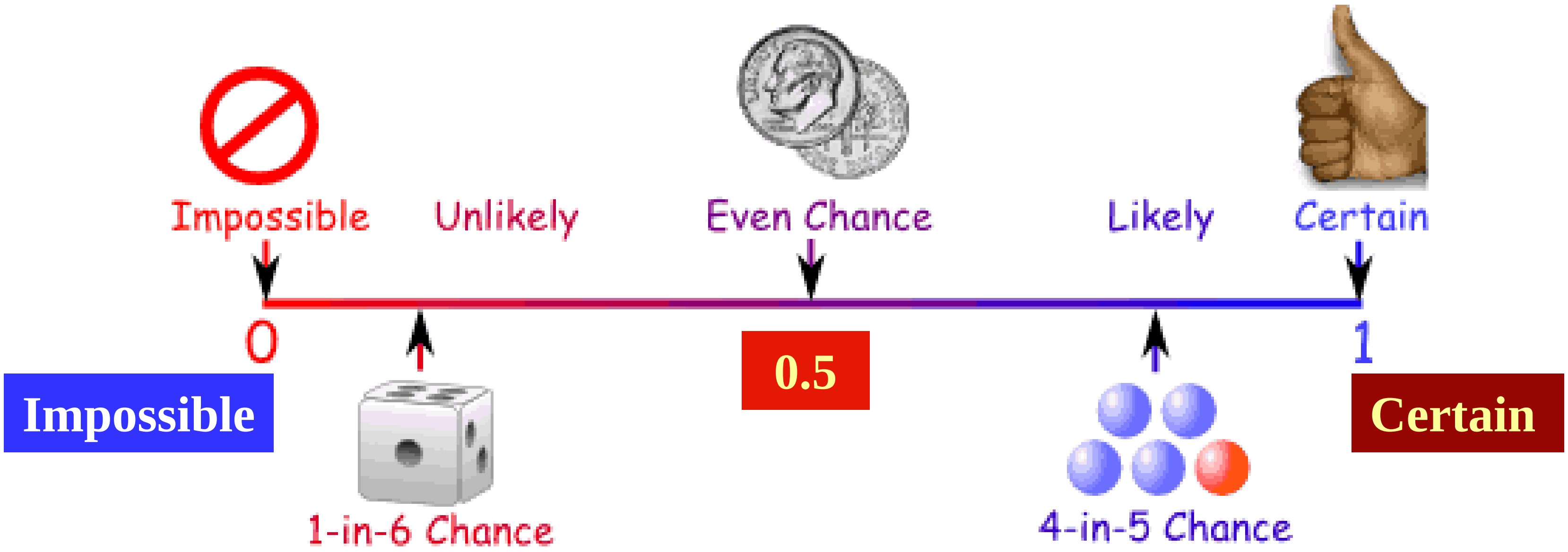
It will definitely...

There's a good chance...

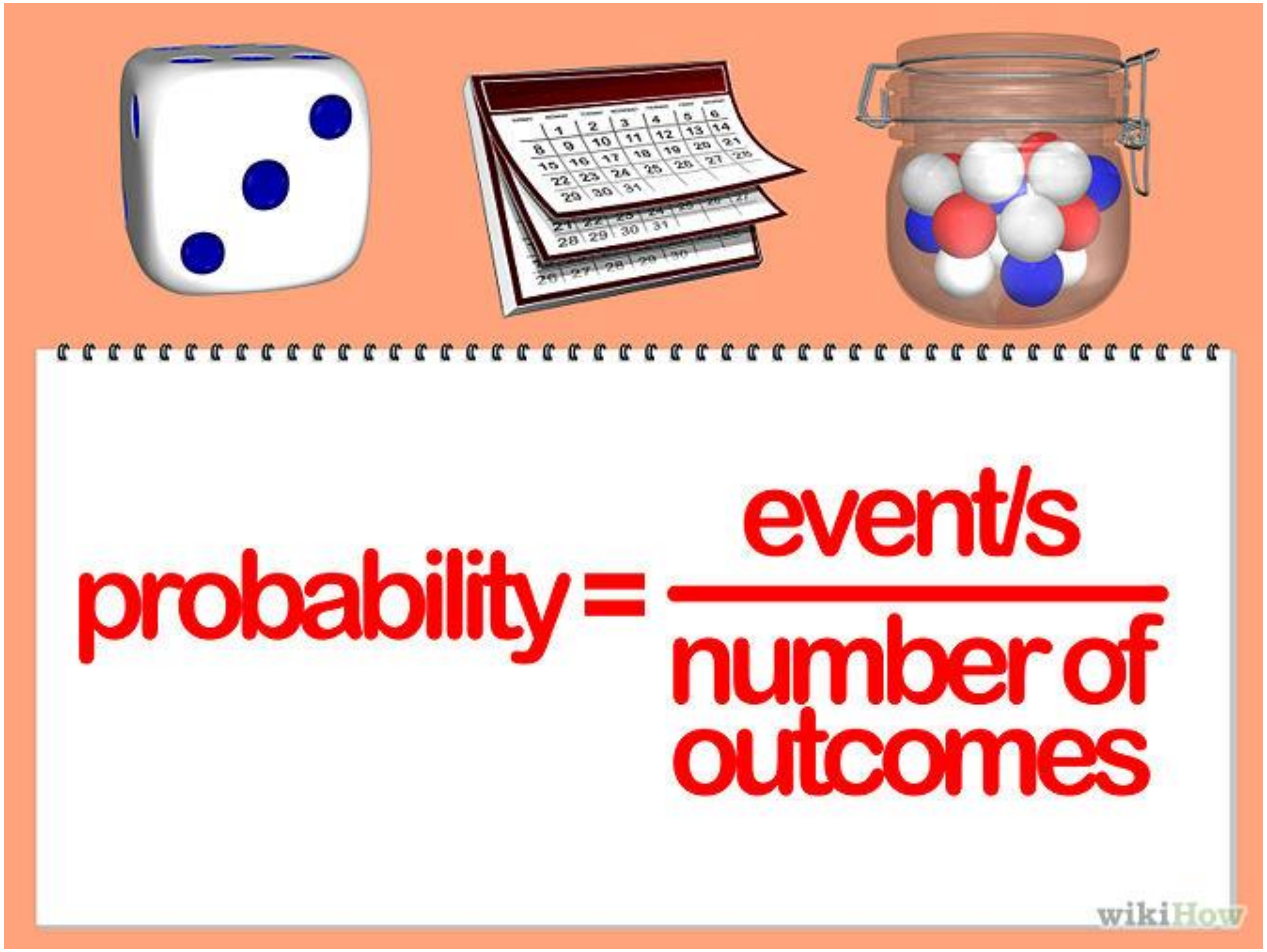
Probability

Probability

- Probability is the Numerical Measure of the Likelihood that an Event Will Occur. Value is between 0 and 1
- Sum of the Probabilities of all Mutually Exclusive and Collective Exhaustive Events is 1



Probability



Joint Probability Using Contingency Table

Event	Event: B_1	Event: B_2	Total
A_1	$P(A_1$ and $B_1)$	$P(A_1$ and $B_2)$	$P(A_1)$
A_2	$P(A_2$ and $B_1)$	$P(A_2$ and $B_2)$	$P(A_2)$
Total	$P(B_1)$	$P(B_2)$	1

Joint Probability

Marginal (Simple) Probability

Computing Compound Probability

- Probability of a Compound Event, A or B:

$$P(A \text{ or } B) = \frac{\text{number of outcomes from either A or B or both}}{\text{total number of outcomes in sample space}}$$

E.g. $P(\text{Red Card or Ace})$

$$= \frac{\text{4 Aces + 26 Red Cards - 2 Red Aces}}{\text{52 total number of cards}}$$

$$= \frac{28}{52} = \frac{7}{13}$$

Compound Probability (Addition Rule)

$$P(A_1 \text{ or } B_1) = P(A_1) + P(B_1) - P(A_1 \text{ and } B_1)$$

Event	Event		Total
	B_1	B_2	
A_1	$P(A_1 \text{ and } B_1)$	$P(A_1 \text{ and } B_2)$	$P(A_1)$
A_2	$P(A_2 \text{ and } B_1)$	$P(A_2 \text{ and } B_2)$	$P(A_2)$
Total	$P(B_1)$	$P(B_2)$	1

For Mutually Exclusive Events: $P(A \text{ or } B) = P(A) + P(B)$

Computing Conditional Probability

- The Probability of Event A Given that Event B Has Occurred:

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}$$

E.g.

$$P(\text{Red Card given that it is an Ace})$$

$$= \frac{2 \text{ Red Aces}}{4 \text{ Aces}} = \frac{1}{2}$$

Conditional Probability Using Contingency Table

Type	Color		Total
	Red	Black	
Ace	2	2	4
Non-Ace	24	24	48
Total	26	26	52

Revised Sample Space

$$P(\text{Ace} \mid \text{Red}) = \frac{P(\text{Ace and Red})}{P(\text{Red})} = \frac{2/52}{26/52} = \frac{2}{26}$$

Conditional Probability and Statistical Independence

- Conditional Probability:

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}$$

- Multiplication Rule:

$$P(A \text{ and } B) = P(A|B)\ P(B)$$

$$= P(B|A)\ P(A)$$

Conditional Probability and Statistical Independence

(continued)

- Events A and B are Independent if

$$P(A \mid B) = P(A)$$

$$\text{or } P(B \mid A) = P(B)$$

$$\text{or } P(A \text{ and } B) = P(A)P(B)$$

- Events A and B are Independent When the Probability of One Event, A, is Not Affected by Another Event, B

Bayes' Theorem

$$P(B_i \mid A) = \frac{P(A \mid B_i)P(B_i)}{P(A \mid B_1)P(B_1) + \bullet\bullet\bullet + P(A \mid B_k)P(B_k)}$$

$$= \frac{P(B_i \text{ and } A)}{P(A)}$$

Same Event

Adding up the parts of A in all the B's

Bayes’ Theorem Example

- A drilling company estimates that there is 40% probability to extract oil for their new oil well.
- For more information it is scheduled a detailed inspection. Historically, 60% of successful oil wells and 20% of unsuccessful oil wells have detailed audits.
- Given that this oil well has scheduled for detailed inspection, which is the probability of being successful?

Bayes' Theorem Example (cont.)

- We assume A+ = successful oil well

  A- = unsuccessful oil well
- P(A+) = 0.4, P(A-) = 1-P(A+)=0.6
- Let B the event of detailed audit
- Conditional probabilities:

  $P(B|A+) = 0.6$ $P(B|A-) = 0.2$
- Aim to compute P(A+|B)

Bayes' Theorem Example (cont.)

- Apply Bayes' Theorem:

$$P(A+|B) = \frac{P(B|A+)P(A+)}{P(B|A+)P(A+) + P(B|A-)P(A-)}$$

$$= \frac{0.6 \cdot 0.4}{0.6 \cdot 0.4 + 0.2 \cdot 0.6}$$

$$= \frac{0.4}{0.6}$$

$$= 0.667$$

Enumeration Rules

- In many cases, there is a large number of possible outcomes.
- Enumeration rules can be used in these cases to help calculate probabilities.

Enumeration Rules

- Enumeration Rule 1:
  - If any of *k* mutually exclusive and collectively exhaustive events can occur in each of *n* trials, then the number of possible outcomes is equal to

$$k^n$$

  - **Example:** If we roll a fair dice 3 times, then there are 6^3=216 possible outcomes

Enumeration Rules

- Enumeration Rule 2:
  - If there are k_1 events in the first trial, k_2 events in the second trial, … and k_n events in the nth trial, then the number of possible outcomes is equal to

$$k_1 k_2 \dots k_n$$

  - **Example:** You want to go to a park, eat at a restaurant, and watch a movie. There are 3 parks, 4 restaurants, and 6 movie options. How many possible combinations are there?
  - **Answer:** You have 3*4*6=72 different options

Enumeration Rules

- Enumeration Rule 3:
  - The number of ways that n elements can be placed in order is

$$n! = n(n-1)\cdots 1$$

  - **Example:** You have five books to put on a shelf. In how many ways these books can be placed on the shelf?
  - **Answer:** 5!=5*4*3*2*1=120 different ways

Enumeration Rules

- Enumeration Rule 4:
  - **Permutations:** The number of arrangements of X elements that have been selected out from n objects is

$$ {}_{n}P_{X} = \frac{n!}{(n-X)!} $$

  - **Example:** You have five books, and you are going to put three on a shelf. In how many ways these books can be placed on the shelf?
  - **Answer:** ${}_{n}P_{X} = \frac{5!}{(5-3)!} = \frac{120}{2} = 60$ different ways

Enumeration Rules

- Enumeration Rule 5:
  - **Combinations:** The number of arrangements of X elements that have been selected out from n objects, regardless of order, is

$$ {}_{n}C_{X} = \frac{n!}{X!\,(n-X)!} $$

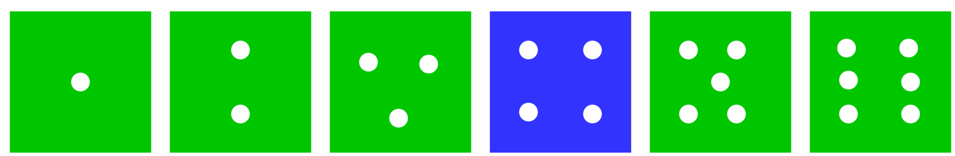
  - **Example:** You have five books, and you are going to choose three for reading. How many different combinations are there, if we ignore the order in which they are selected?
  - **Answer:** ${}_{n}C_{X} = \frac{5!}{3!(5-3)!} = \frac{120}{12} = 10$ different options

1.2.3. Some Discrete Probability Distributions

- The Probability of a Discrete Random Variable
- Covariance and Its Applications in Finance
- Binomial Distribution
- Poisson Distribution
- Hypergeometric Distribution

Random Variable

- Random Variable
  - Outcomes of an experiment expressed numerically
  - E.g., Toss a die twice; count the number of times the number 4 appears (0, 1 or 2 times)



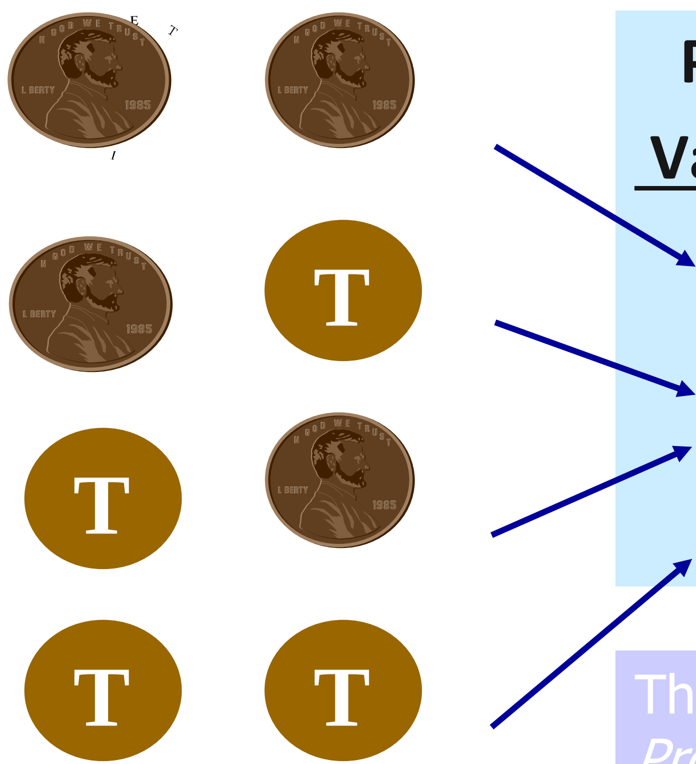
•E.g., The number of cars in a parking lot. A parking lot can only hold a certain number of cars.

Discrete Random Variable

- Discrete Random Variable
  - Obtained by counting (0, 1, 2, 3, etc.)
  - Usually, a finite number of different values
  - E.g., Toss a coin 5 times; count the number of tails (0, 1, 2, 3, 4, or 5 times)

Discrete Probability Distribution Example

Event: Toss 2 Coins Count # Tails



Probability Distribution

Values	Probability
0	1/4 = .25
1	2/4 = .50
2	1/4 = .25

This is using the *A Priori Classical Probability* approach.

Discrete Probability Distribution

- List of All Possible [X_j , $P(X_j)$] Pairs
  - X_j = Value of random variable
  - $P(X_j)$ = Probability associated with value
- Mutually Exclusive (Nothing in Common)
- Collective Exhaustive (Nothing Left Out)

$$0 \leq P\left(X_j\right) \leq 1 \qquad \sum P\left(X_j\right) = 1$$

Summary Measures

- Expected Value (The Mean)
  - Weighted average of the probability distribution
  - $$\mu = E(X) = \sum_j X_j P(X_j)$$
  - E.g., Toss 2 coins, count the number of tails, compute expected value:

$$\mu = \sum_j X_j P(X_j)$$

$$= (0)(.25) + (1)(.5) + (2)(.25) = 1$$

Summary Measures

(continued)

- Variance
  - Weighted average squared deviation about the mean
  - $$\sigma^2 = E\left[(X-\mu)^2\right] = \sum \left(X_j - \mu\right)^2 P\left(X_j\right)$$
  - E.g., Toss 2 coins, count number of tails, compute variance:

$$\sigma^2 = \sum \left(X_j - \mu\right)^2 P\left(X_j\right)$$

$$= (0-1)^2 (.25) + (1-1)^2 (.5) + (2-1)^2 (.25)$$

$$= .5$$

Sum of Two Random Variables

- The expected value of the sum is equal to the sum of the expected values

$$E(X+Y)=E(X)+E(Y)$$

- The variance of the sum is equal to the sum of the variances plus twice the *covariance*

$$Var(X+Y)=\sigma_{X+Y}^2=\sigma_X^2+\sigma_Y^2+2\sigma_{XY}$$

Portfolio Expected Return and Risk

- The portfolio expected return for a two-asset investment is equal to the weighted average of the two assets

$$E(P) = wE(X) + (1-w)E(Y)$$

where

w = portion of the portfolio value assigned to asset X

- Portfolio risk

$$\sigma_P = \sqrt{w^2\sigma_X^2 + (1-w)^2\sigma_Y^2 + 2w(1-w)\sigma_{XY}}$$

Important Discrete Probability Distributions

Discrete Probability Distributions

- **Bernoulli**
- **Binomial**
- **Geometric**
- **Discrete uniform**
- **Poisson**
- **Negative binomial**

Discrete uniform distribution

- Outcomes: $x \in \{1,2,3, \ldots N\}$
- All outcomes have equal probability

  $P(X=x) = 1/N$
- $E(X)=(N+1)/2$
- $V(X)=(N^2-1)/12$
- Example: toss of a fair dice
  - $N=6$
  - $P(X=x)=1/6$

Discrete uniform distribution

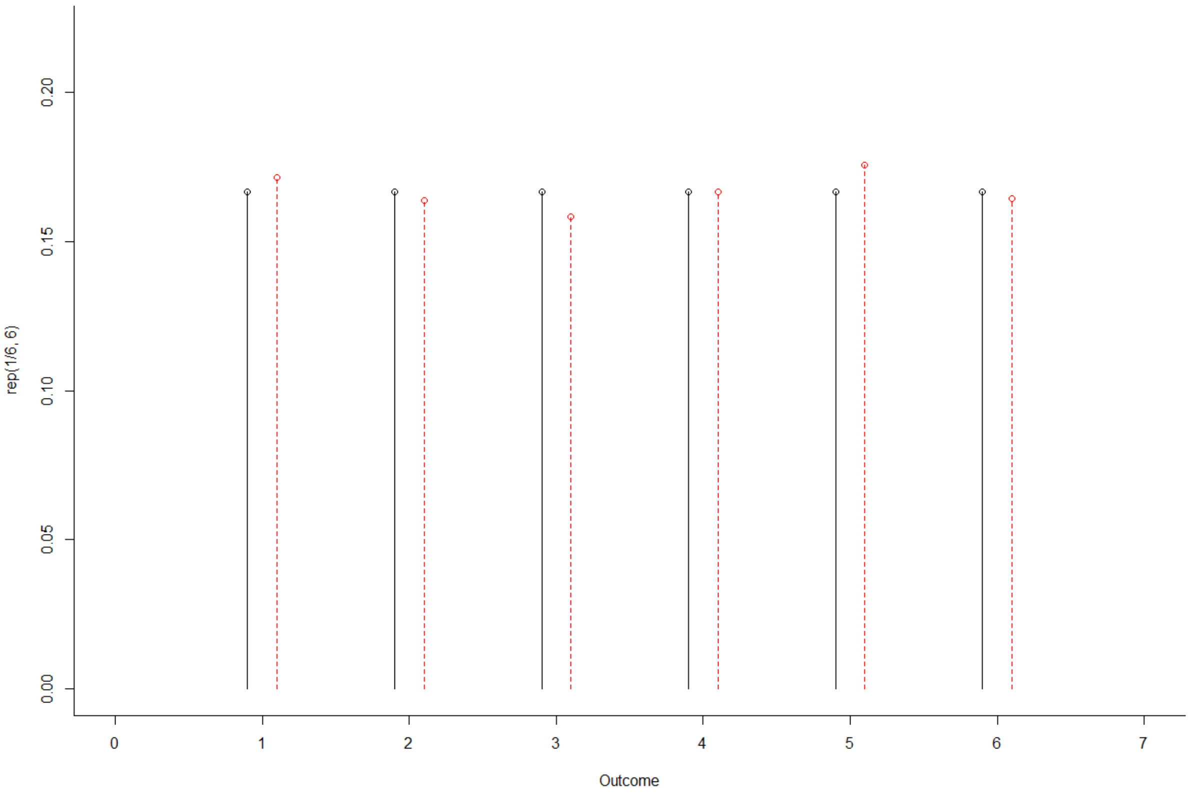
Generating random numbers from the discrete uniform in R

```
x<-sample(1:6, 10000, replace=T)
table(x)/length(x)
1/6
```

```
x
     1      2      3      4      5      6 
0.1714 0.1636 0.1583 0.1667 0.1755 0.1645 
> 
> 1/6
[1] 0.1666667
```

Discrete uniform distribution

Comparison of true and estimated probabilities



Bernoulli distribution

- Outcomes: $x \in \{0,1\}$ (failure/success)

  $P(X=1) = p$

  $P(X=0) = (1-p)$
- $E(X)= p$
- $V(X)= p(1-p)$
- Example: toss of a fair coin (also discrete uniform)
  - $p=1/2$
  - In this case the variance/uncertainty is maximized

Binomial Probability Distribution

- ‘n’ Identical Bernoulli Trials
  - E.g., 15 tosses of a coin; 10 light bulbs taken from a warehouse
- Trials are Independent
  - The outcome of one trial does not affect the outcome of the other
- Constant Probability for Each Trial
  - E.g., Probability of getting a tail is the same each time we toss the coin
- 2 Sampling Methods
  - Infinite population without replacement
  - Finite population with replacement

Binomial Probability Distribution Function

$$P(X)=\frac{n!}{X!(n-X)!}p^X(1-p)^{n-X}$$

$P(X):$ probability of X successes given n and p

$X:$ number of "successes" in sample $(X=0,1,\cdots,n)$

$p:$ the probability of each "success"

$n:$ sample size

Tails in 2 Tosses of Coin

X	P(X)
0	1/4 = .25
1	2/4 = .50
2	1/4 = .25

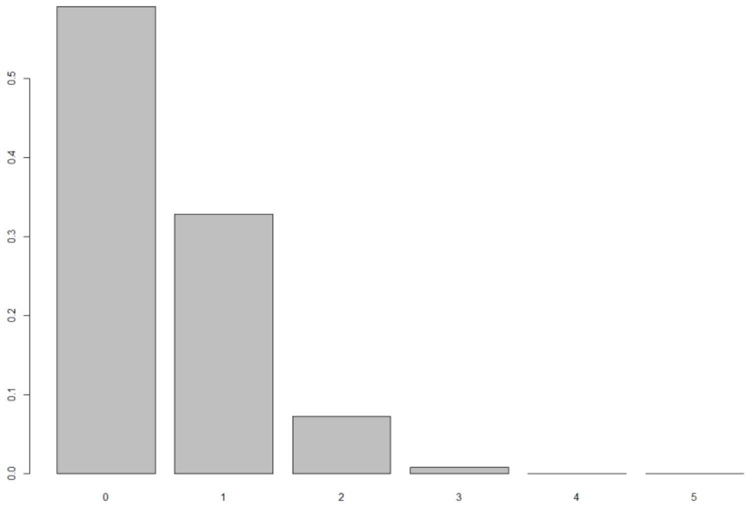
```
> dbinom(0:2,2,0.5)
[1] 0.25 0.50 0.25
```

Binomial Distribution Characteristics

- Mean
  - $\mu = E(X) = np$
- Variance and Standard Deviation
  - $\sigma^2 = np(1-p)$

    $\sigma = \sqrt{np(1-p)}$

$n = 5 \quad p = 0.1$



barplot(dbinom(0:5,5,0.1), names=0:5)

$$\mu = np = 5(.1) = .5$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{5(.1)(1-.1)} = .6708$$

Binomial Distribution Characteristics

Estimating the mean and the variance using random samples

$$\mu = np = 5(.1) = .5$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{5(.1)(1-.1)} = .6708$$

```
> x<-rbinom(10000, 5, 0.1)
> mean(x)
[1] 0.5031
> sd(x)
[1] 0.6775222
```

Example: Binomial Distribution

A mid-term exam has 30 multiple choice questions, each with 5 possible answers. What is the probability of randomly guessing the answer for each question and passing the exam (i.e., having guessed at least 18 questions correctly)?

Are the assumptions for the binomial distribution met?

$n = 30 \qquad p = 0.2$

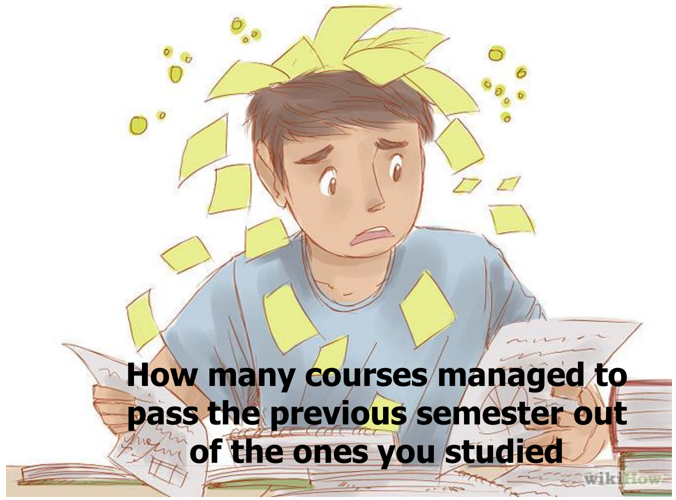
$P(X \geq 18) = 1.84245(10)^{-6}$

```
> 1-pbinom( 17, 30, 0.2 )
[1] 1.842448e-06
```

```
> n<-50000000
> x<-rbinom(n, 30, 0.2)
> sum(x>17)/n
[1] 1.84e-06
```

Examples of binomial random variables

Number of successful shout outs in 5 trials



Poisson Distribution

Siméon Poisson

Poisson Distribution

- Discrete events ("successes") occurring in a given area of opportunity ("interval")
  - "Interval" can be time, length, surface area, etc.
- The probability of a "success" in a given "interval" is the same for all the "intervals"
- The number of "successes" in one "interval" is independent of the number of "successes" in other "intervals"
- The probability of two or more "successes" occurring in an "interval" approaches zero as the "interval" becomes smaller
  - E.g., # customers arriving in 15 minutes
  - E.g., # calls received per minute at a call centre

Poisson Probability Distribution Function

$$P(X)=\frac{e^{-\lambda}\lambda^{X}}{X!}$$

$P(X)$: probability of X "successes" given λ

X: number of "successes" per unit

λ: expected (average) number of "successes"

e: 2.71828 (base of natural logs)

E.g., Find the probability of 4 customers arriving in 3 minutes when the average number of customers is 3.6 per 3 minutes.

$$P(X=4)=\frac{e^{-3.6}3.6^{4}}{4!}=.1912$$

```
> dpois(4,3.6)
[1] 0.1912223
```

Poisson Distribution in R

Poisson {stats}

The Poisson Distribution

Description

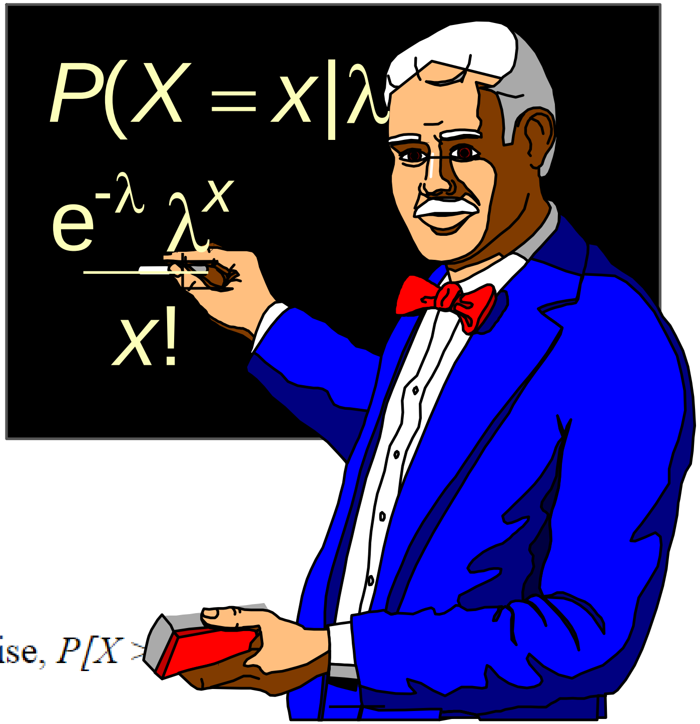
Density, distribution function, quantile function and random generation for the Poisson distrib

Usage

```
dpois(x, lambda, log = FALSE)
ppois(q, lambda, lower.tail = TRUE, log.p = FALSE)
qpois(p, lambda, lower.tail = TRUE, log.p = FALSE)
rpois(n, lambda)
```

Arguments

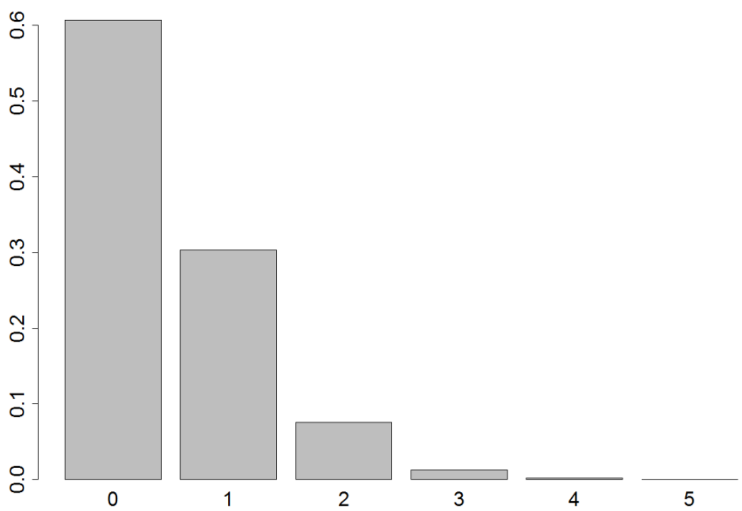
`x`	vector of (non-negative integer) quantiles.
`q`	vector of quantiles.
`p`	vector of probabilities.
`n`	number of random values to return.
`lambda`	vector of (non-negative) means.
`log, log.p`	logical; if TRUE, probabilities p are given as log(p).
`lower.tail`	logical; if TRUE (default), probabilities are $P[X \le x]$, otherwise, $P[X$



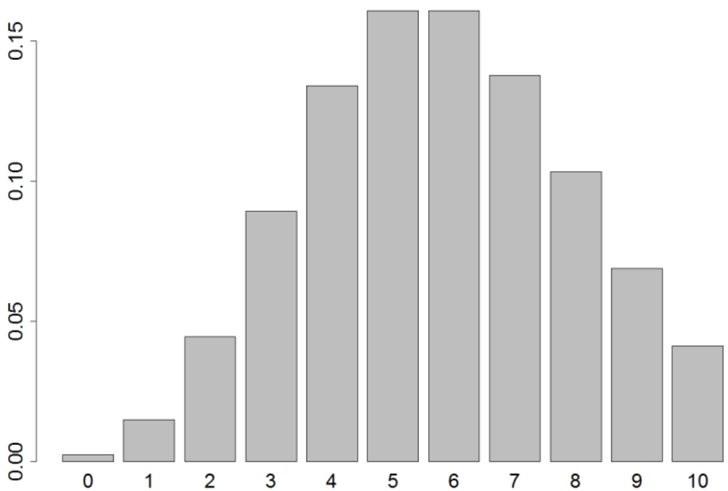
Poisson Distribution Characteristics

- Mean $\mu = E(X) = \lambda$
- Variance $\sigma^2 = \lambda$

λ = 0.5



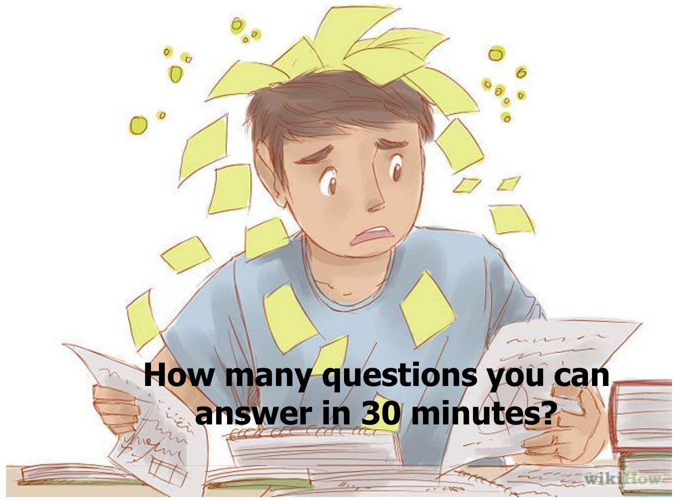
λ = 6




```
barplot(dpois(0:5,0.5),names=0:5,cex.axis=2, cex.names=2)
barplot(dpois(0:10,6),names=0:10,cex.axis=2, cex.names=2)
```

Examples of Poisson random variables

Number of successful points in 10 minutes of play



Number of clients arriving in a queue next 60 minutes waiting for iPad/iPod etc

1.2. Measuring uncertainty & risk using probabilities
1.2.4. The Normal Distribution

- The Normal Distribution
- The Standardized Normal Distribution
- Evaluating the Normality Assumption

Continuous Probability Distributions

- Continuous Random Variable
  - Used to model continuous quantities
  - Values from interval of numbers
  - Absence of gaps
  - E.g. X : weight, in kg, of a person selected at random
- Probability distributions
  - We can calculate $P(X \leq k)$ or $P(a \leq X \leq b)$
  - We don't calculate $P(X = k)$. Actually, $P(X = k)$=0. Why?

Continuous Probability Distributions

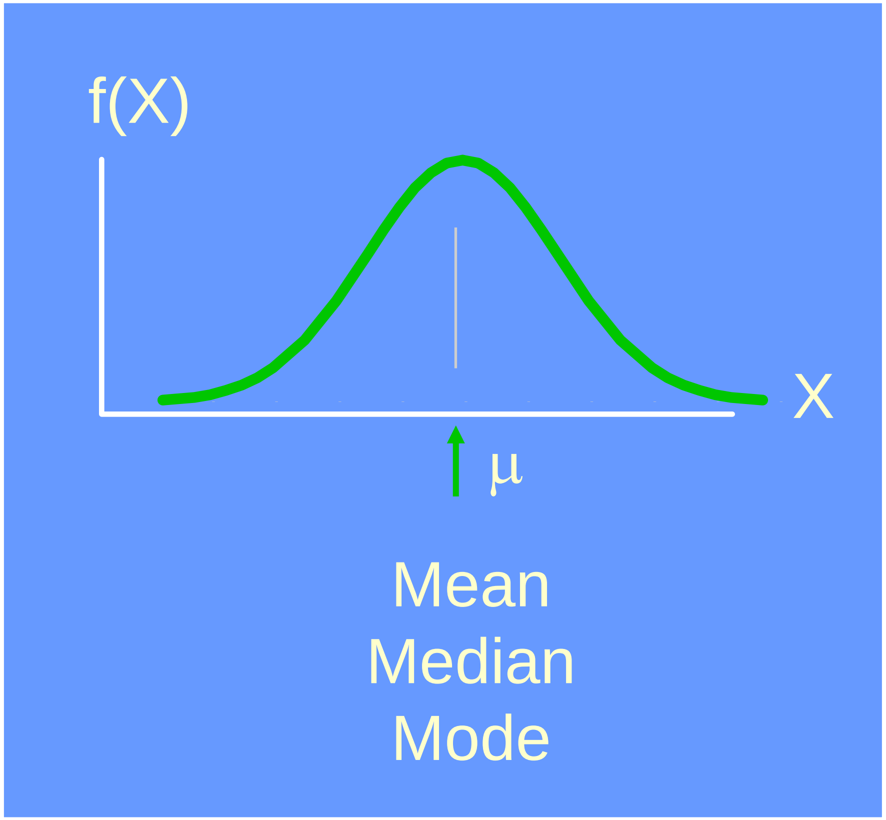
- Calculating Probabilities
  - Probability density function (PDF) $f(x)$
  - Cumulative density (distribution) function (CDF) $F(x)$
  - $P(a \leq X \leq b) = \int_a^b f(x)dx = F(b) - F(a)$

Continuous Probability Distributions

- Properties of the PDF
  - $f(x) \geq 0, x \in \mathbb{R}$
  - $\int_{-\infty}^{+\infty} f(x)dx = 1$
- Properties of the CDF
  - $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$
  - $P(X \leq x) = P(X < x)$ and $P(X \geq x) = P(X > x)$

The Normal Distribution

- "Bell Shaped"
- Symmetrical
- Mean, Median and Mode are Equal
- Interquartile Range Equals 1.33 σ
- Infinite Range



The Mathematical Model

$$f(X) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(1/2)\left[(X-\mu)/\sigma\right]^2}$$

$f(X)$: density of random variable X

$\pi \approx 3.14159; \qquad e \approx 2.71828$

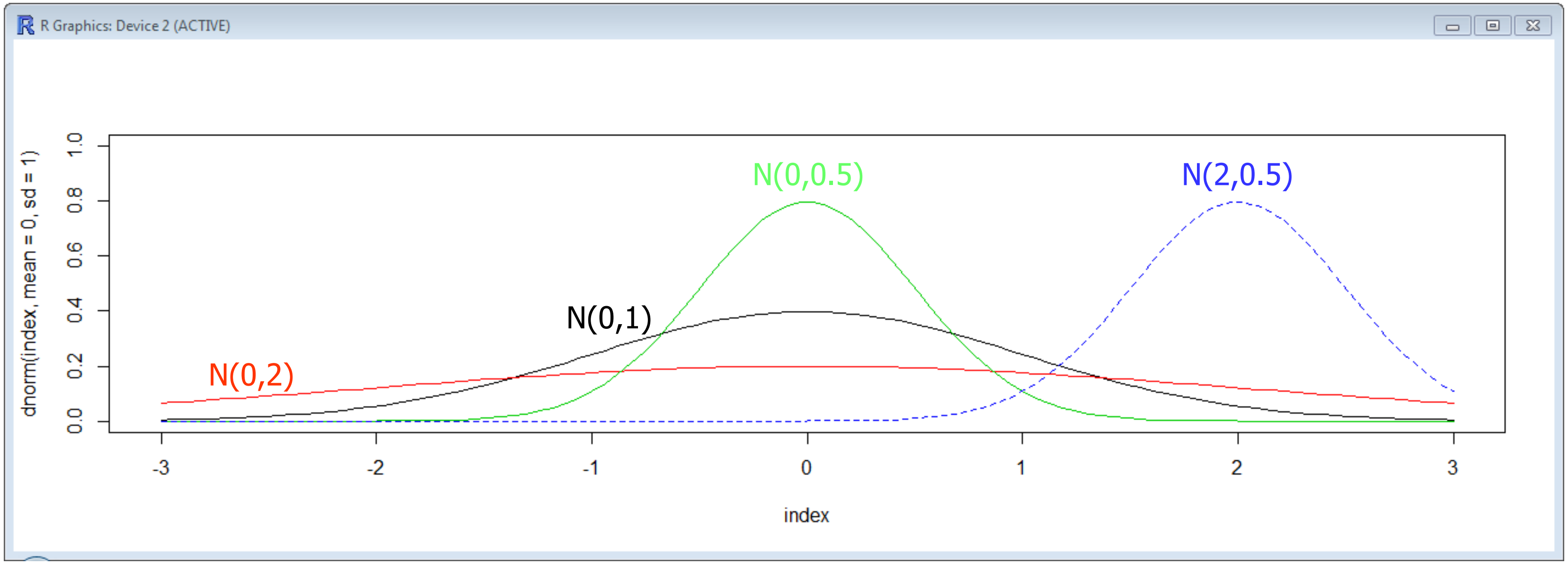
μ: population mean

σ: population standard deviation

X: value of random variable $(-\infty < X < \infty)$

Many Normal Distributions

There is an Infinite Number of Normal Distributions



Varying the Parameters σ and μ, we obtain different Normal Distributions

The Standardized Normal Distribution

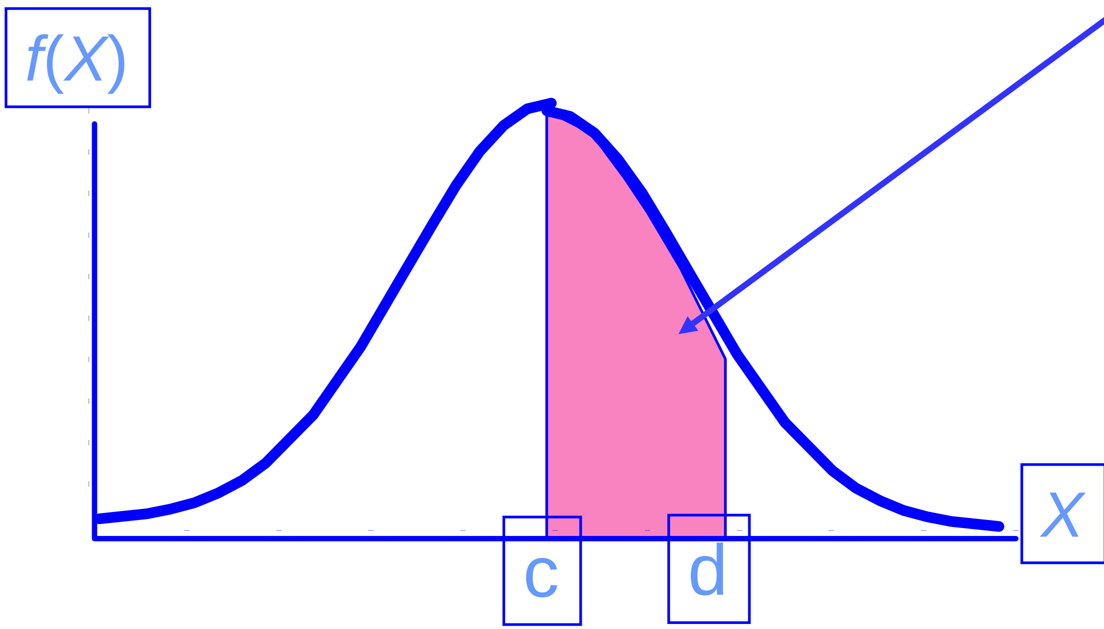
$Z = \dfrac{X - \mu}{\sigma}$ follows a standardized (normalized) normal distribution with a mean 0 and a standard deviation 1.

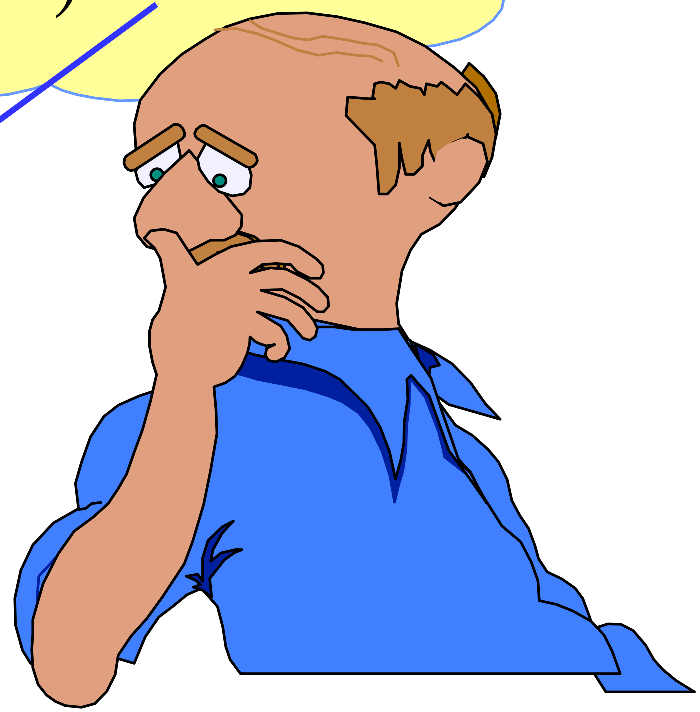
We calculate all probabilities with regard to the standardized normal distribution

Finding Probabilities

Probability is the area under the curve!

$$P(c \le X \le d) = ?$$





Assessing Normality

- Not All Continuous Random Variables are Normally Distributed
- It is Important to Evaluate How Well the Data Set Seems to Be Adequately Approximated by a Normal Distribution

Assessing Normality

(continued)

- Observe the Distribution of the Data Set
  - Do approximately 68% of the observations lie between mean $\pm$ 1 standard deviation?
  - Do approximately 95% of the observations lie between mean $\pm$ 2 standard deviations?
  - Do approximately 99.7% of the observations lie between mean $\pm$3 standard deviations?

- Evaluate Normal Probability Plot
  - Do the points lie on or close to a straight line with positive slope?

1.2.4. Other Continuous Distributions

- Continuous uniform distribution
- The exponential
- Chi-squared distribution
- Student t
- F-distribution
- Gamma
- Beta

The Uniform Distribution

- Properties:
  - The probability of occurrence of a value is equally likely to occur anywhere in the range between the smallest value *a* and the largest value *b*
  - It is used to obtain random variables from various distributions
  - $$\mu = \frac{(a+b)}{2}$$
  - $$\sigma^2 = \frac{(b-a)^2}{12}$$

The Uniform Distribution

(continued)

- The Probability Density Function

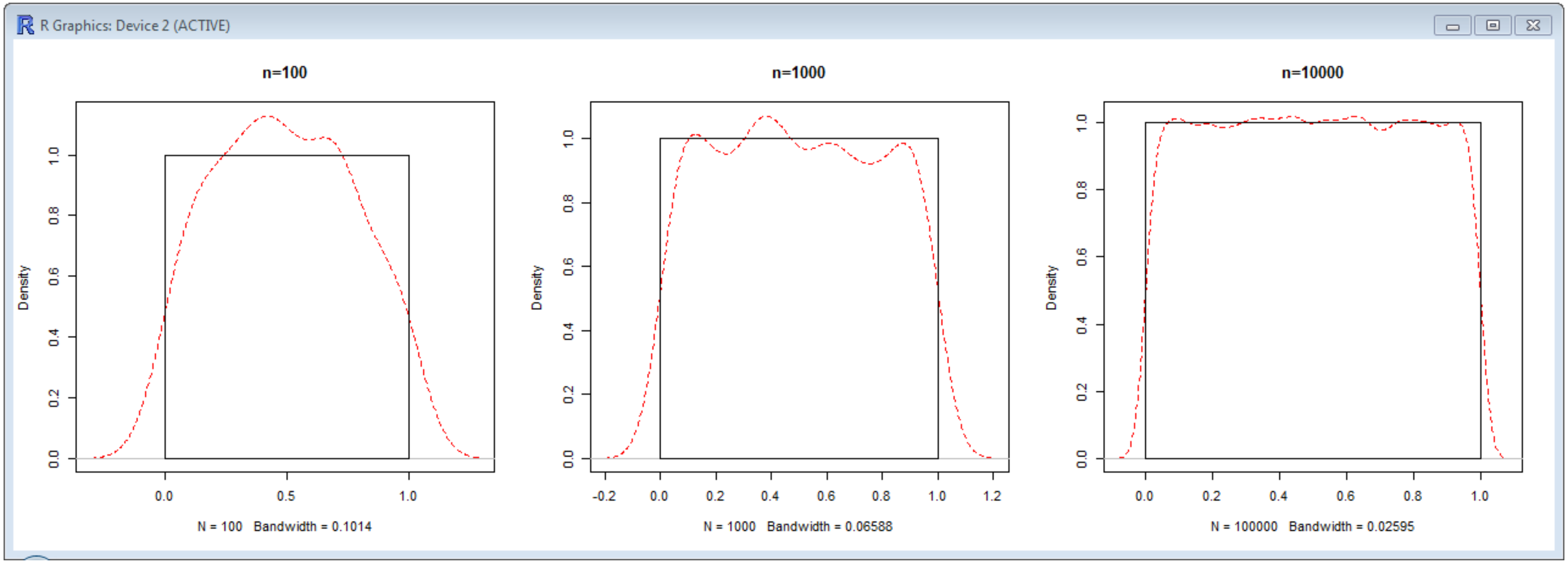
$$f(X) = \frac{1}{(b-a)} \quad \text{if } a \le X \le b$$

- Property: Intervals of equal length have equal probabilities
- Application: Selection of random numbers

```
dunif(x, min = 0, max = 1, log = FALSE)
punif(q, min = 0, max = 1, lower.tail = TRUE, log.p = FALSE)
qunif(p, min = 0, max = 1, lower.tail = TRUE, log.p = FALSE)
runif(n, min = 0, max = 1)
```

The Uniform Distribution

(continued)



Exponential Distributions

$P(\text{arrival time} < X) = 1 - e^{-\lambda X}$

$X \in [0, \infty)$

λ: the population average number of arrivals per unit of time

$1/\lambda$: average time between arrivals

$e = 2.71828$

E.g., Drivers arriving at a toll bridge; customers arriving at an ATM machine

Exponential Distributions

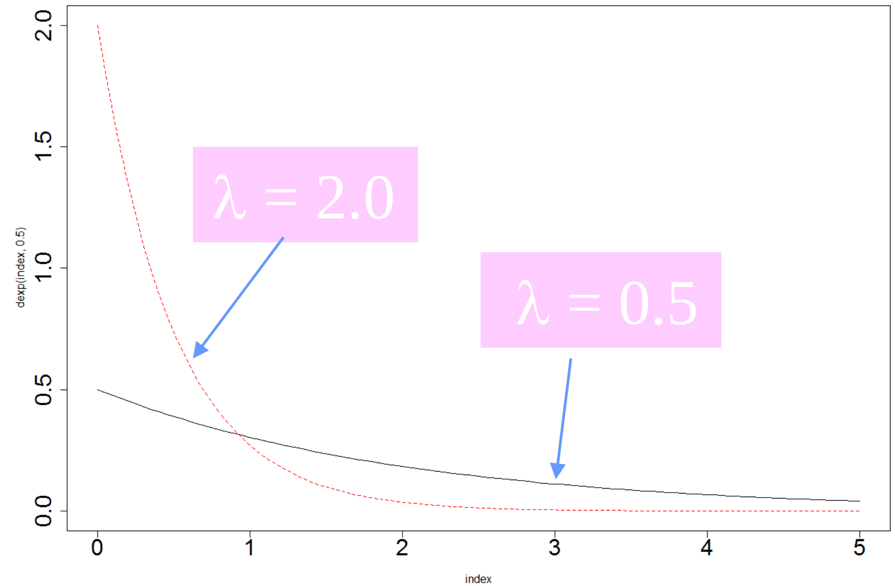
(continued)

- Describes Time or Distance between Events
  - Used for queues
- Density Function

$$f(x) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}}$$

- Parameters

$$\mu = \lambda \qquad \sigma = \lambda$$




```
index<-seq(0,5, length.out=100)
plot(index, dexp(index, 0.5), type='l',
        ylim=c(0,2), cex.axis=2)
lines(index, dexp(index, 2), col=2, lty=2)
```

Example

E.g., Customers arrive at the checkout line of a supermarket at the rate of 30 per hour. What is the probability that the arrival time between consecutive customers will be greater than 5 minutes?

$$\lambda = 30 \qquad X = 5/60 \text{ hours}$$

$$P(\text{arrival time} > X) = 1 - P(\text{arrival time} \le X)$$

$$= 1 - \left(1 - e^{-30(5/60)}\right)$$

$$= .0821$$

```
> 1-pexp(5/60,30)
[1] 0.082085
```

Chi-squared (x^2) distribution

- Positive valued
- Can be obtained as the sum of k independent squared standardized normal random variables
- Is the sampling distribution of the sample variance
- Mean= k
- Variance= 2k

```
dchisq(x, df, ncp = 0, log = FALSE)
pchisq(q, df, ncp = 0, lower.tail = TRUE, log.p = FALSE)
qchisq(p, df, ncp = 0, lower.tail = TRUE, log.p = FALSE)
rchisq(n, df, ncp = 0)
```

Student's t distribution

- It is symmetric but with fatter tails than the normal
- Arises at the t-test
- Can be obtained by Z/sqrt(X/k) with Z~N(0,1) and X~x^2 with k degrees of freedom
- Obtained as the sampling distribution in the test statistic in the t-test when the variance is estimated from the sample
- Mean=0
- Variance=k/(k-2)

```
dt(x, df, ncp, log = FALSE)
pt(q, df, ncp, lower.tail = TRUE, log.p = FALSE)
qt(p, df, ncp, lower.tail = TRUE, log.p = FALSE)
rt(n, df, ncp)
```

The F-distribution

- Positive valued
- Is defined as $(X_1/k_1)/(X_2/k_2)$ where X_1 and X_2 are chi-squared random variables with k_1 & k_2 dfs.
- Arises when considering a ratio of sample variances.
- Parameters k_1 & k_2

```
df(x, df1, df2, ncp, log = FALSE)
pf(q, df1, df2, ncp, lower.tail = TRUE, log.p = FALSE)
qf(p, df1, df2, ncp, lower.tail = TRUE, log.p = FALSE)
rf(n, df1, df2, ncp)
```

The Gamma distribution

- Positive valued
- Parameters a & b
- Big range of shapes
- Mean = a/b
- Variance = a/b^2

```
dgamma(x, shape, rate = 1, scale = 1/rate, log = FALSE)
pgamma(q, shape, rate = 1, scale = 1/rate, lower.tail = TRUE,
       log.p = FALSE)
qgamma(p, shape, rate = 1, scale = 1/rate, lower.tail = TRUE,
       log.p = FALSE)
rgamma(n, shape, rate = 1, scale = 1/rate)
```

The Beta distribution

- Defined in the (0,1) interval
- Parameters a,b
- Wide range of shapes
- Can be obtained as X/(X+Y) with X, Y being Gamma random variables with parameters (a,1) & (b,1).
- Mean=a/(a+b)
- Variance= p (1-p)/(a+b+1) with p=a/(a+b)

```
dbeta(x, shape1, shape2, ncp = 0, log = FALSE)
pbeta(q, shape1, shape2, ncp = 0, lower.tail = TRUE, log.p = FALSE)
qbeta(p, shape1, shape2, ncp = 0, lower.tail = TRUE, log.p = FALSE)
rbeta(n, shape1, shape2, ncp = 0)
```