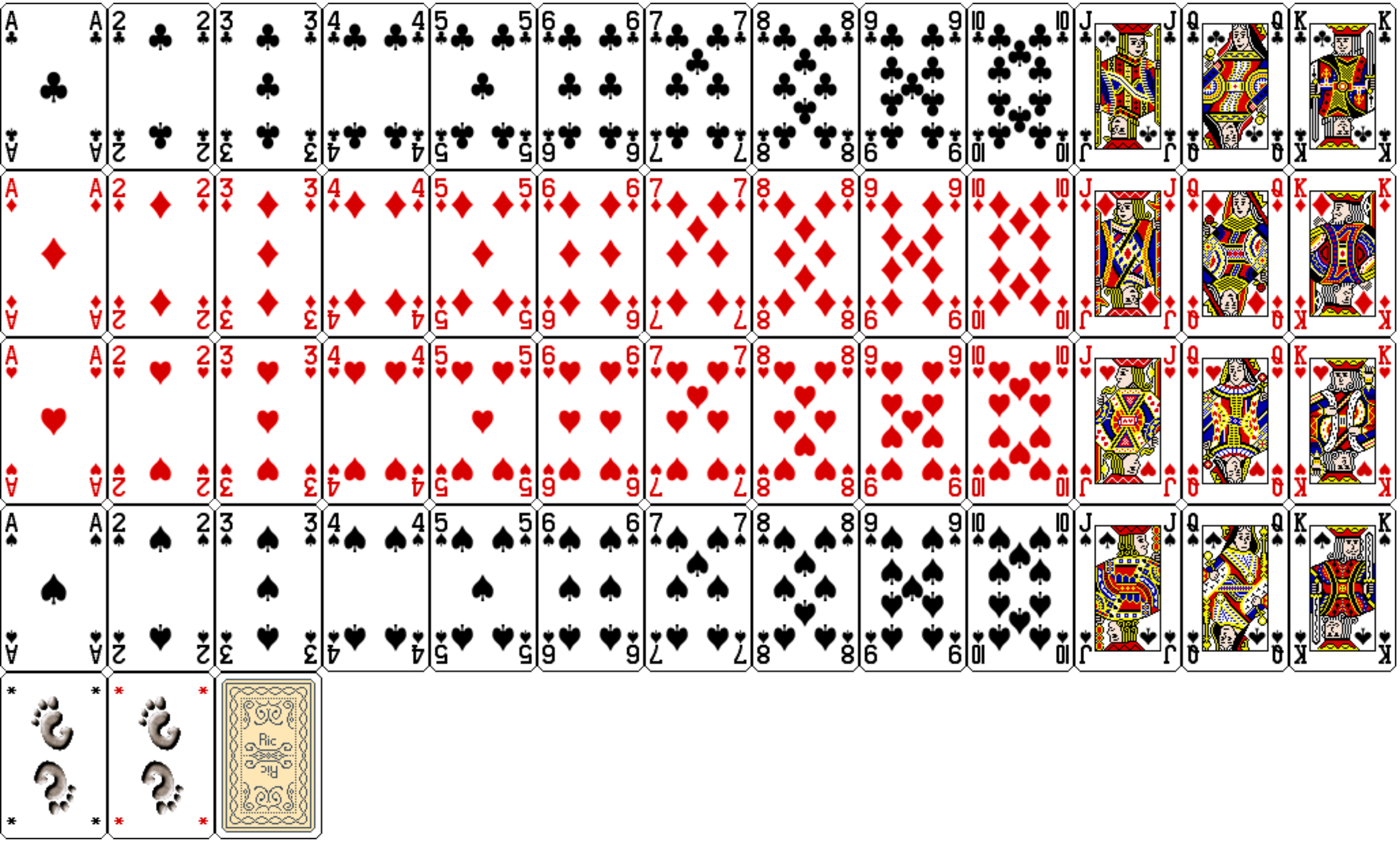
Random Variables & Probability Distributions

Geog 210B
Winter 2018

Why we need all this

- Theoretical ideas about the world in more formal mathematical operations
- Procedures to state hypotheses about the real world
- Testing of these hypotheses and reach conclusions
- Many abstract concepts with direct application in real world!
- Also in modeling direct mapping to the objective and subjective worlds



Clubs

Diamonds

Hearts

Spades

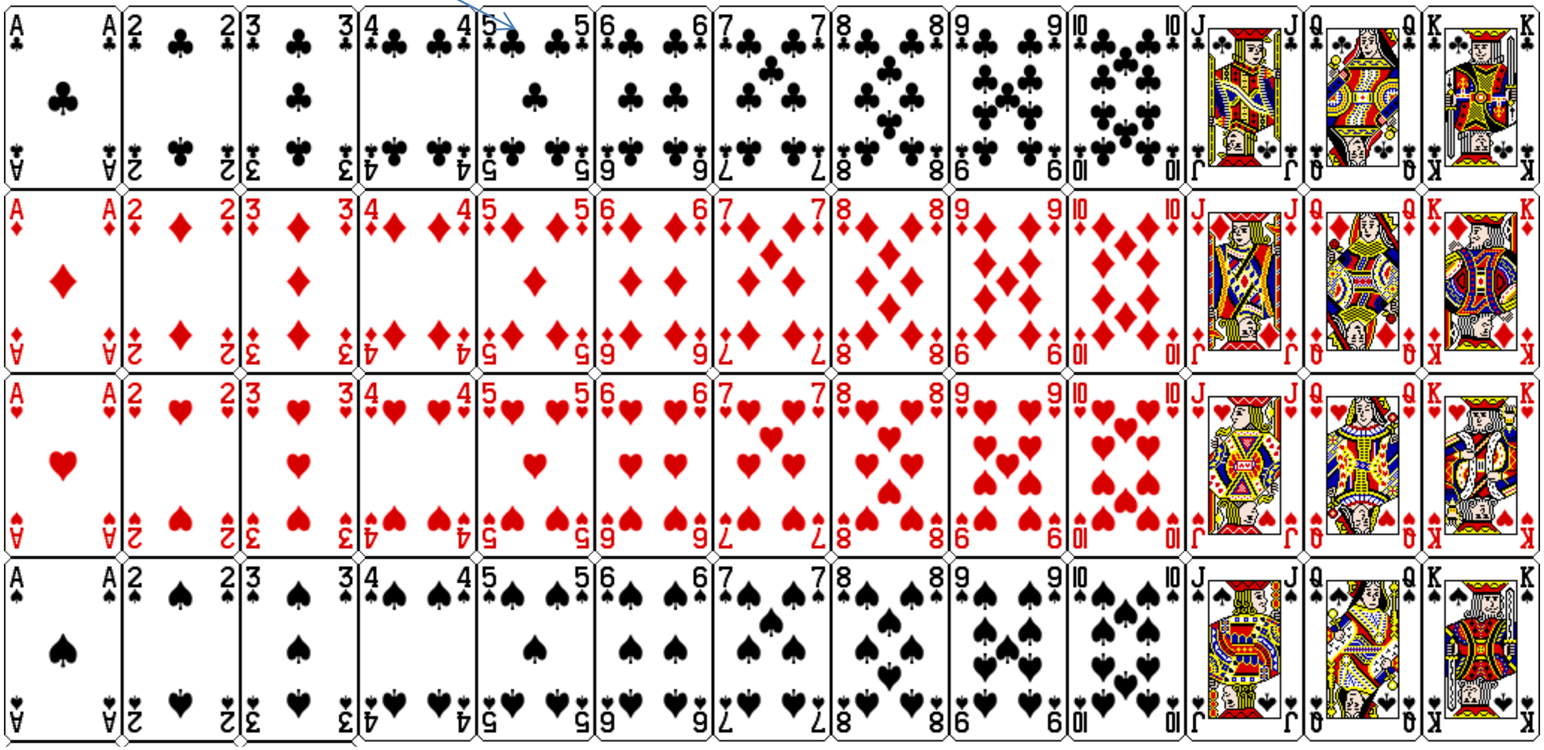
PLAY CARDS
(SIMPLE - NO JOKERS)

Event vs. Outcome

- 52 Playing cards
- ***Sample space*** is made by 52 elements
- Each individual card is an **outcome** (possible)
- **Event** is any combination of the 52 elements
- Event is also all 52 cards
- Event is also none of the 52 cards (the null event)
- The totality of all elementary outcomes is the sample space
- We can also define outcomes in different ways using an experiment – **statistical experiment** is an activity in which one outcome from many possible outcomes occurs

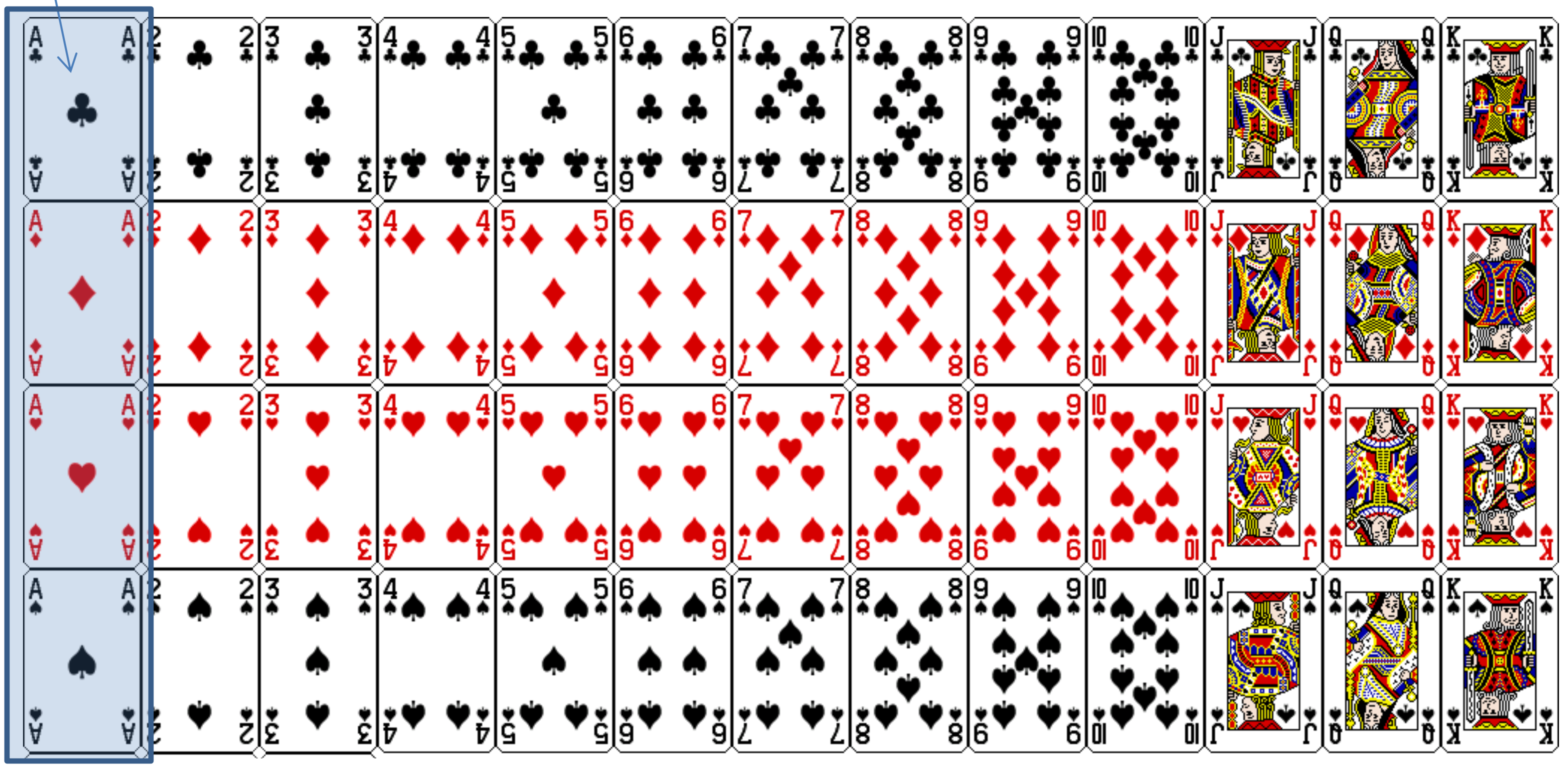
Elementary outcome

Sample Space



Elementary outcome = ace of clubs

Event: 4 Aces (shaded)



What is Probability?

- Chance of something happening?
- The odds?
- The likelihood?
- The relative frequency?
- A belief?
- Degree of certainty?
- Do we need more concepts to define it?

Review

BASIC PROBABILITY & STATISTICS CONCEPTS

HTTP://WWW.MATHPAGES.COM/HOME/IPROBABI.HTM

Concepts Outline

- Random variables and operators
- Probability distribution function (PDF) of a random variable
- Basic probability rules, sets, and outcomes & events
- Construct PDFs
- Compute PDF moments
- Illustrate some important PDFs
- Compute the sample mean, variance, and whatever else we need

Conceptually

- **Random variable**: a variable whose values are uncertain until we collect data about it. Each value this variable takes has an associated probability. We use the probability to depict this **uncertainty** in the values.
- Remember we divide variables into **discrete** (set of outcomes is either finite or countably infinite) and **continuous** (set of outcomes infinitely divisible and therefore not countable)
- Probability Distribution (f(y)) = A **listing** of all the values taken by variable Y and their associated probabilities

Example 1 (Toss a coin twice)

- Possible outcomes: {head, head}, {tail, tail}, {tail, head}, {head, tail}
- Probability of each outcome = ¼
- Let's call random variable the number of heads = Y
- Possible outcomes (values) of Y are 0, 1, 2 (denoted by lower case y)
- This is a Discrete Random Variable

Anyang Coin

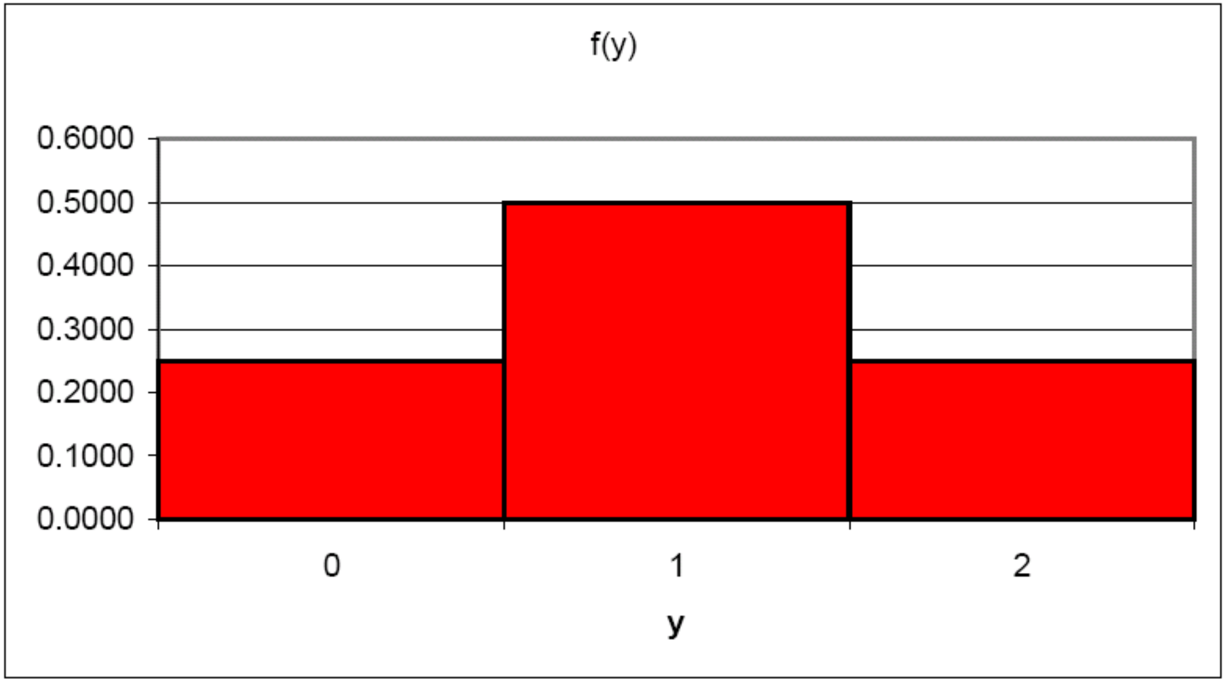
Lydian Coin

Terminology (again)

- Statistical experiment or random trial=an activity in which one outcome out of many possible outcomes occurs (tossing coins)
- Elementary outcome = each different outcome of our experiment – head or tail
- Sample space=set of all elementary outcomes
- Event = collection of elementary outcomes

Probability Mass Function (f(y))

Y = number of heads	f (y)	Elementary Outcomes
0	¼	{tail, tail}
1	½	{tail, head}, {head, tail}
2	¼	{head, head}



Other Possible Tossing Coins Events

- Event A – at least one H: A={(HH), (HT), (TH)}
- Event B- two of the same: B={(HH), (TT)}
- Event C – at least one T: C={(HT), (TH), (TT)}

http://americanhistory.si.edu/collections/numismatics/doubleea/doubleea.htm

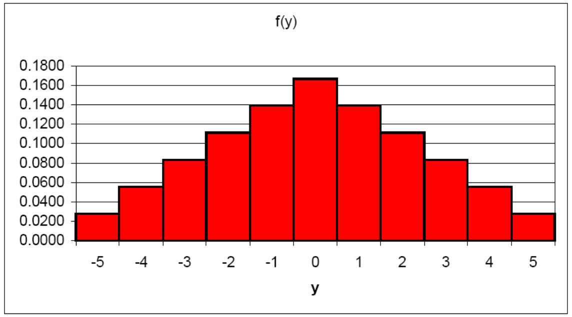
Example 2 (cast 2 dice)

- The sum of numbers (y)
- What are the possible numbers?
- How many 0?
- How many 1?
- How many 3?
- …..

- How many 12?

Two dice

y	f(y)	Elementary outcomes	Possible times
2	1/36	(1,1)	1
3	2/36	(1,2) (2,1)	2
4	3/36	(1,3) (2,2) (3,1)	3
5	4/36	(1,4) (2,3) (3,2) (4,1)	4
6	5/36	(1,5) (2,4) (3,3) (4, 2) (5,1)	5
7	6/36	(1,6) (2,5) (3,4) (4,3) (5,2) (6,1)	6
8	5/36	(2,6) (3,5) (4,4) (5,3) (6,2)	5
9	4/36	(3,6) (4,5) (5,4) (6,3)	4
10	3/36	(4,6) (5,5) (6,4)	3
11	2/36	(5,6) (6,5)	2
12	1/36	(6,6)	1



Probability Distribution Function (PDF) of the discrete random variable Y

PDF mathematically

$$f(Y) = P(Y = y_i) \quad for\ i = 1,2,\ldots,n$$

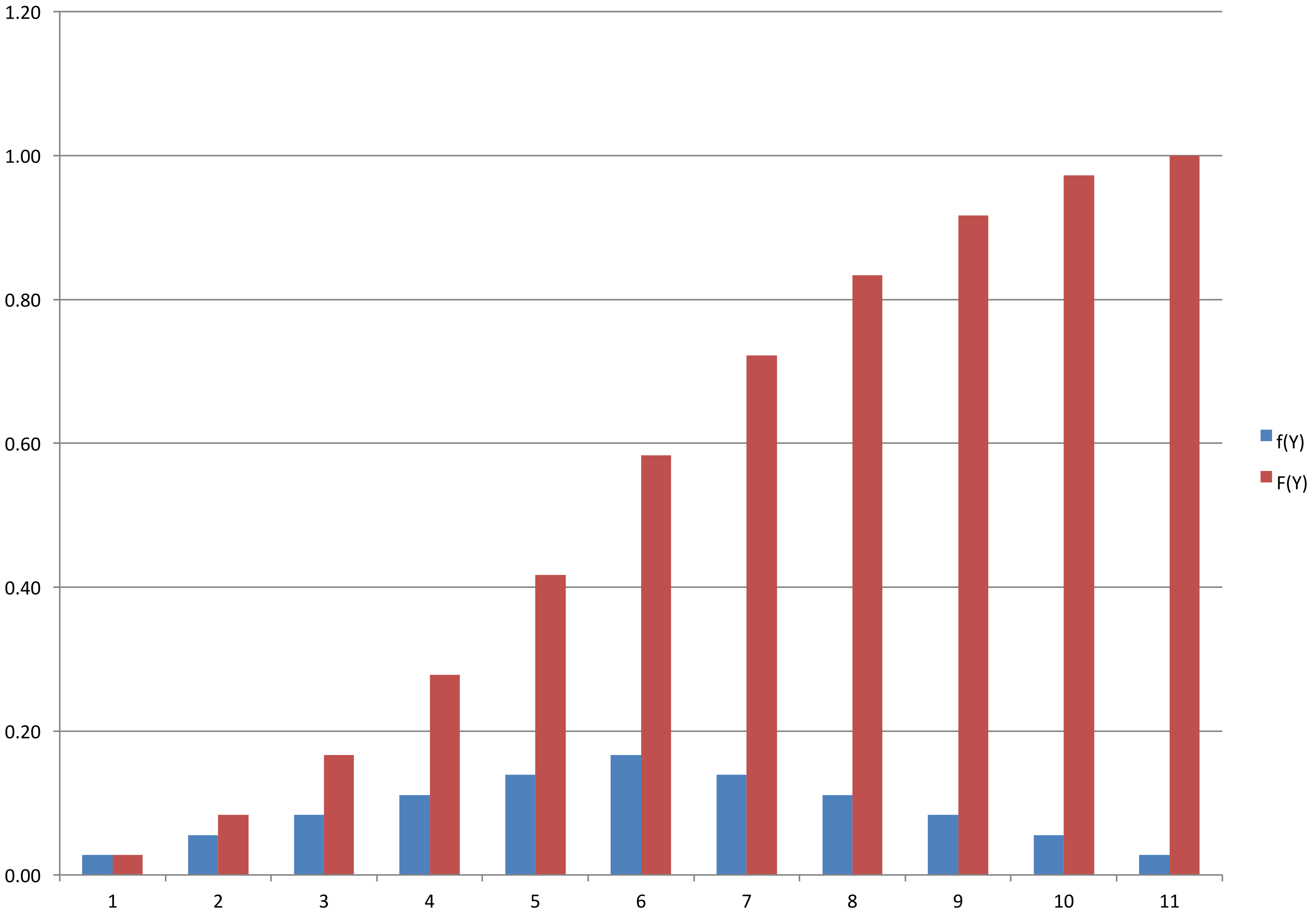
$$= 0 \quad for Y \neq y_i$$

Define also the CUMULATIVE Distribution Function

$$F(Y) = P(Y \le y) = \sum_{i=1}^{y} f(Y)$$

Value considered	PDF	Value considered	CDF	CDF
y	f(Y)		F(Y)	F(Y)
2	1/36	y=2	1/36	0.03
3	2/36	y≤3	1/36+2/36	0.08
4	3/36	y≤4	1/36+2/36+3/36	0.17
5	4/36	y≤5	1/36+2/36+3/36+4/36	0.28
6	5/36	y≤6	1/36+2/36+3/36+4/36 +5/36	0.42
7	6/36	y≤7	Etc	0.58
8	5/36	y≤8	Etc	0.72
9	4/36	y≤9	Etc	0.83
10	3/36	y≤10	Etc	0.92
11	2/36	y≤11	Etc	0.97
12	1/36	y≤12	1	1.00

Relationship between CDF and PDF



Es are the elementary outcomes and their assigned probabilities are P(E)

$$0 \leq P(E_i) \leq 1 \;\; for \;\; i = 1,2,\ldots,n$$

Impossible outcome => P(E) = 0
Certain outcome (outcome for sure)=> P(E) =1

If Event A is made of elementary outcomes k

$$P(A) = \sum_{k \in A} P(E_k) \;\; for \;\; i = 1,2,\ldots,n$$

$$0 \leq P(A) \leq 1$$

Quick Summary

- Measurements on any variable will always vary
- The pattern of variation of a variable is called its ***distribution***
- It can be described both mathematically and graphically
- The distribution records (lists) all possible numerical values of a variable and how often each value occurs (its frequency).

Distributions

- They are functions of random variables
- People studied them in the past and now we know a lot about them (tabulated values)
- Integrate them, take derivatives, partition them and so forth
- We also use them to derive other functions
- This knowledge allows us to combine variables of known distributions (mixing)

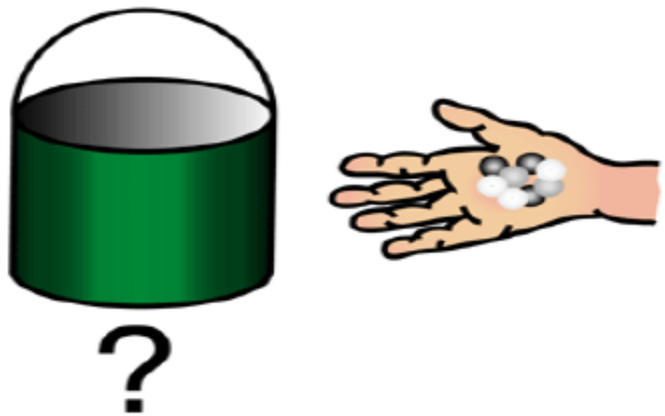
Definitions part 1

- **Population** is a collection of all units of interest.
- A **sample** is a subset of a population that is actually observed.
- A measurable property or attribute associated with each unit of a population is called a **variable**.
- **Parameter** is a numerical characteristic of a population.
- **Statistic** is a numerical characteristic of a sample.

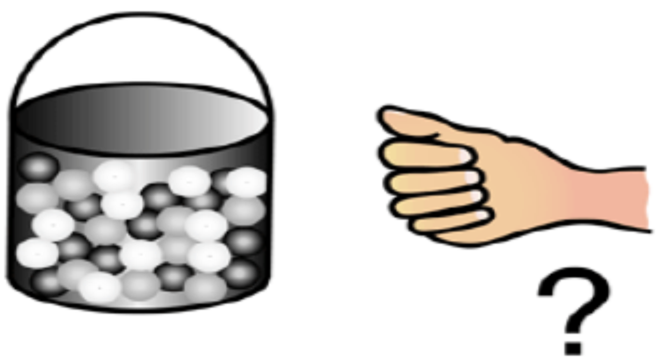
- We use statistics to infer the values of parameters.
- A random sample gives a non-zero chance to every unit of the population to enter the sample.

Definitions part 2

STAT



PROB



- Given what you have in your hand. What can you infer about the bucket?
  - In statistics, we assume that the population and its parameters are **unknown and the sample is used to infer the values of the parameters.**
- Given what is in the bucket, what is in your hand?
  - In probability, we assume that the population and its parameters are **known and compute the probability of drawing a particular sample.**

Norma Gilbert, Statistics

Probability is “deductive” (general -> particular)
Statistics is “inductive” (particular -> general)

Definitions part 3

- Remember two basic questions asked by inferential statistics:
  - How close is the value of a statistic to the corresponding parameter of the entire population.
    - If we have a sample of 874 elements (sample) that we assumed comes from a population of mean age 43.5 years. How far is the sample average from the mean of the entire population?
  - Someone hypothesized a particular value for the parameter of a population or some relationship between the parameters of two or more populations.
    - For example, past studies show that the mean age of all people is 43.5 and someone may want to test whether a particular diet leads to higher mean age or if mean age of men is lower than that of women.

Definitions part 4

- Different samples give different estimates of population parameters (called sampling variability).
- Sampling variability leads to “sampling error” = difference between observed and imaginary population value
- Often we work with a transformed statistic that follows a *theoretical distribution* and we use it to draw conclusions – this theoretical distribution is named *sampling distribution* (more later and in the lab)

What is a probability?

- The usual coin (chance of 1/2 for each side)
- The usual die (1/6 for each face)
- Occurrence over total possibilities

- Impossibility -> probability = 0
- Absolute certainty -> probability = 1
- Everything else ->probability = anywhere in between 0 and 1

More definitions

- **Random experiment =** procedure whose outcome cannot be predicted in advance (tossing a coin a few times)

- **Sample Space (S) =** The finest grain, mutually exclusive, collectively exhaustive listing of all possible outcomes
  - **sample space for tossing two coins *S={H,H},{H,T},{T,H},{T,T}***

- **Event (A) a set of outcomes (subset of S):** No heads A={T,T}

- **Union (or):** A=heads on first, B=heads on second A U B= {H,T},{H,H},{T,H}

- **Intersection (and):** A= heads on first, B=heads on second $A \cap B$ = {H,H}

- **Complement of Event A – set of all outcomes not in A:** A={T,T}, A^c={H,H},{H,T},{T,H}

Play cards (52 total)

- Associated with each event A in S is the probability of A
- P(A) Axioms:
  1. $P(A) \geq 0$
  2. $P(S) = 1$ where S is the sample space
  3. $P(A \cup B) = P(A) + P(B)$ if A and B are mutually exclusive
- Play cards:

  P(ace) = 4/52

  P(red) =26/52.

  P(ace and black) = 2/52

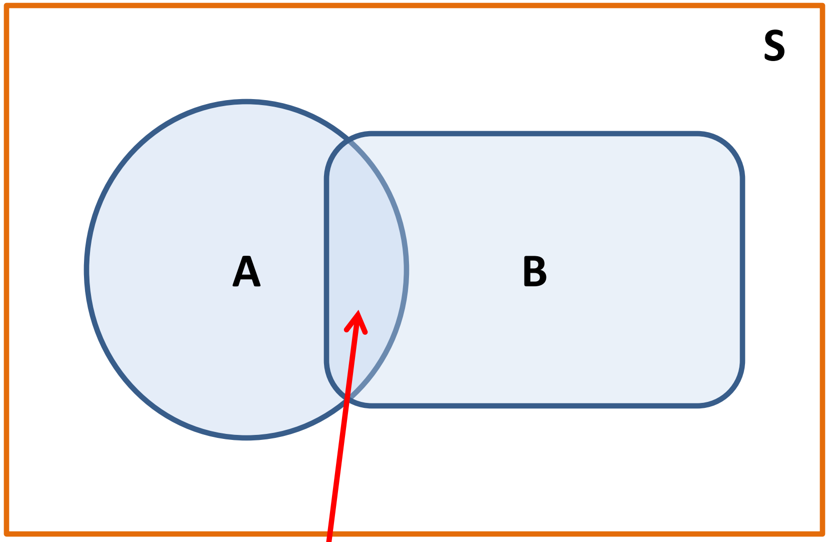
  P(king) = ?

  P (king of hearts) = ?

  P (red or black) = ?

  P(spades) = ?

VENN DIAGRAMS FOR SETS & PROBABILITY THEOREMS/RULES



Everything within the BLUE borders is the union between A and B (A U B)

Outcome in A or B

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Subtract the intersection to eliminate double counting

Example of spades = event A -> P(A) = 13/52
Get an Ace = event B -> P(B)=4/52

Get an Ace of Spades = event A ∩ B = 1/52

Get Ace or Spades = 13/52 + 4/52 – 1/52

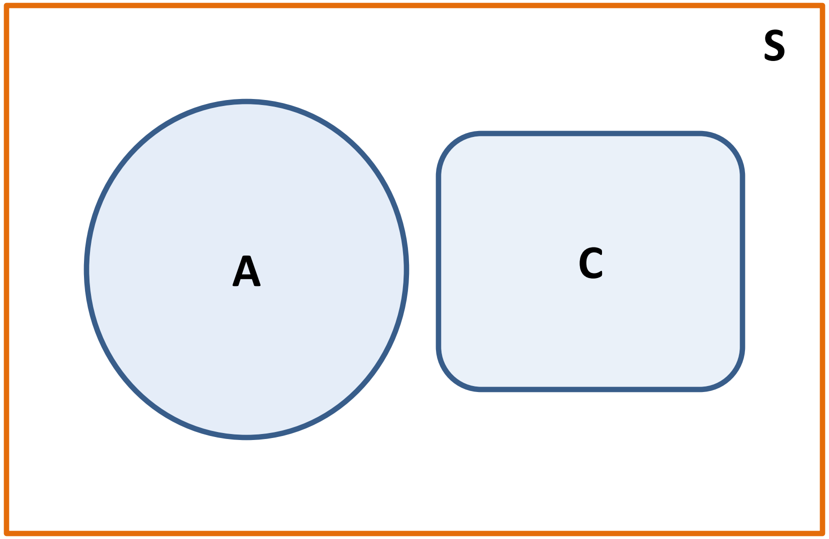
16 cards are Aces or Spades =
13 spades + 3 non-spade Aces

Shaded area is the intersection between A and B (A ∩ B)

Outcome in A and B

General Addition Rule for Non-Mutually Exclusive Events

- When events are not mutually exclusive, there is some overlap.
- When P(A) and P(B) are added, the probability of the intersection (and) is added twice. To compensate for that double addition, the intersection needs to be subtracted once.
- P(A or B) = P(A) + P(B) - P(A and B)
- Always valid
- P(ace or black) = P(ace) + P(black) – P(ace and black)= 4/52 + 26/52 – 2/52 = 28/52 = 7/13



When A & C are mutually exclusive (disjoint)

$$P(A or C) = P(A \cup C) = P(A) + P(C)$$

$$P(A and C) = P(A \cap C) = 0$$

P(red card and black card) = ?

Two events are mutually exclusive (disjoint) if they cannot occur at the same time.

If two events are disjoint, then the probability of them both occurring at the same time is 0.

P(red card or black card) = P(red card) + P(black card) = 26/52 + 26/52 = P(S) = 1

Conditional Probabilities

- Conditional Probability: $P(A|B) = P(A \cap B)/P(B)$
- $P(A \cap B) = P(A|B)P(B)$
- Example: Drawing a card from a deck of 52 cards, P(Hearts)=1/4.
- However, if it is known that the card is red, P(Hearts | Red) = ½.
- Sample space has been reduced to the 26 red cards.
- Selecting hearts from the red partition:
  - P(Hearts|Red) = P(Hearts and Red)/P(red) =(1/4)/(1/2)=2/4=1/2

Independence

- Independence $P(A|B)=P(A)$
- There are situations in which knowing that event B occurred gives no information about event A,
- Example: knowing that a card is black gives no information about whether it is an ace.
- $P(\text{ace} \mid \text{black}) = 2/26 = 4/52 = P(\text{ace})$.
- If two events are independent then $P(A\cap B)=P(A)P(B)$
- $P(A\cap B)=P(A|B)P(B)=P(A)P(B)$

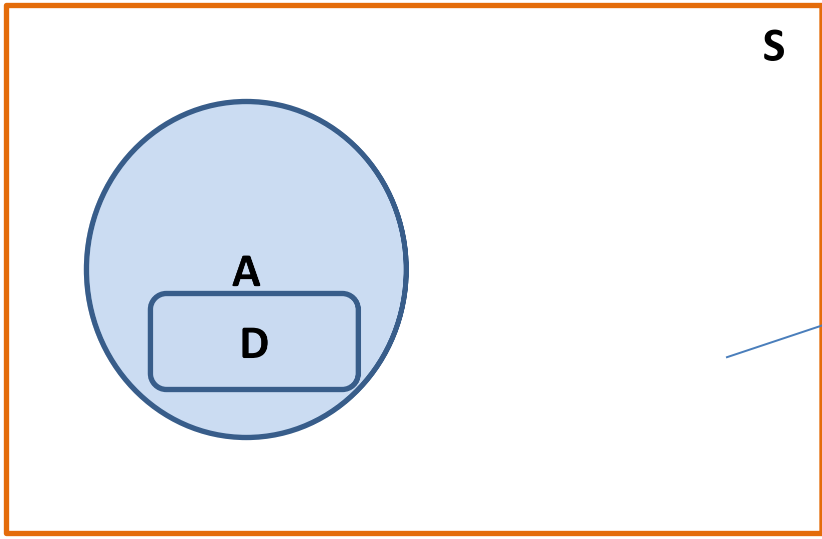
- $P(\text{ace of hearts}) = P(\text{ace}) * P(\text{hearts}) = 4/52 * 13/52 = 1/52$
- Card is red means can't be black-> $P(A|B)=0$.
- Independent events are not the same as disjoint events (an ace can be an ace of hearts).

Summary: Independence vs Disjoint

- If A and B are disjoint:
  - $P(A \cup B) = P(A) + P(B)$
  - $P(A \cap B) = 0$
- If A and B are independent:
  - $P(A \cap B) = P(A) * P(B)$
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

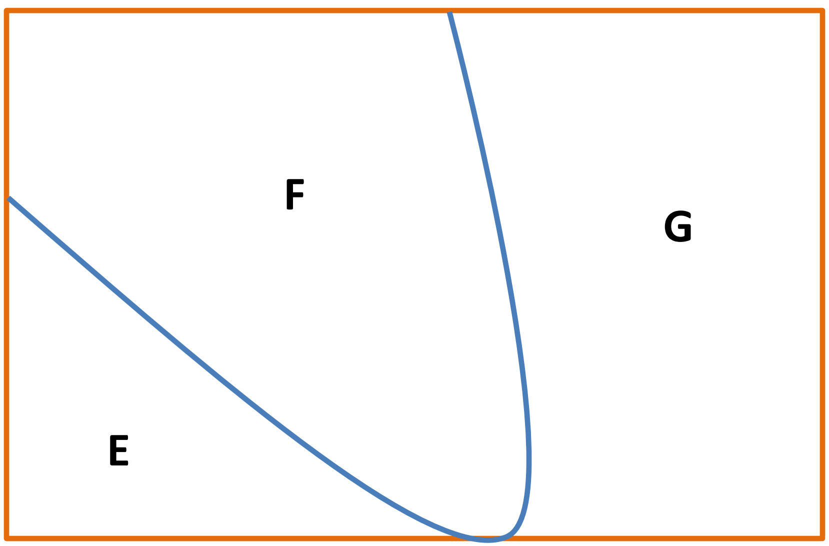
Conditionals

- $P(A \cap B) = P(A|B)\,P(B) = P(B|A)\,P(A)$
- $P(B|A) = P(A|B)\,P(B) / P(A)$
- $P(B)$ = prior probability
- $P(B|A)$ = posterior probability
- P(hearts | red)=P(red | hearts) * P(hearts) / P(red) = 1* 0.25 / 0.5 = 0.5



D is a subset of A $D \subset A$

White region is the complement of A in S

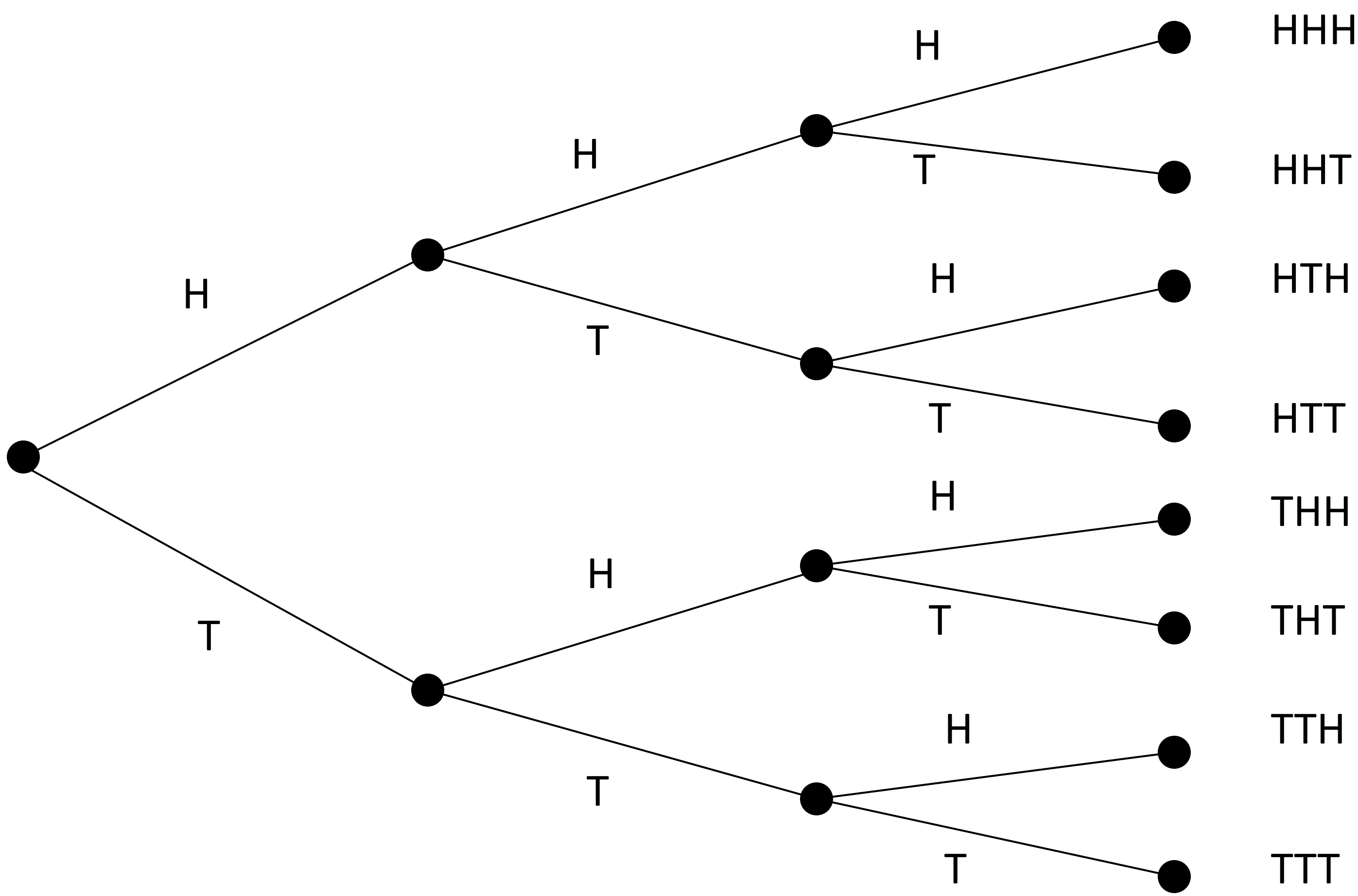


E, F, and G are partitions of S

Tree Diagrams of Events

- Start with first event
  - Draw tree with possible outcomes
- For each branch of tree:
  - Draw tree with possible outcomes of second event
- Continue for as many events as you wish to include in the sequence
- Flip a coin 3 times

Visualize event by a tree: Flipping a coin three times



PLAY CARDS (POKER)

Factorials

- If n is a positive integer, then
- $n! = n\,(n-1)\,(n-2) \ldots (3)(2)(1)$
- $n! = n\,(n-1)!$
- A special case is 0!
- $0! = 1$
- What is 3! = 3 X 2 X 1 = 6
- What is 715! = use a calculator

Poker

- The probability of each type of 5-card hand can be computed by calculating the proportion of hands of that type among all possible hands
- All possible hands is computed using a special operation (the binomial coefficient)

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \binom{52}{5} = \frac{52!}{5!(52-5)!} = 2598960$$

Hand	Frequency	Approx. Probability	Approx. Cumulative	Approx. Odds	Mathematical expression of absolute frequenc
Royal flush	4	0.000154%	0.000154%	649,739 : 1	$\binom{4}{1}$
Straight flush (excluding royal flush)	36	0.00139%	0.00154%	72,192.33 : 1	$\binom{10}{1}\binom{4}{1}-\binom{4}{1}$
Four of a kind	624	0.0240%	0.0256%	4,164 : 1	$\binom{13}{1}\binom{4}{4}\binom{48}{1}$
Full house	3,744	0.144%	0.170%	693.2 : 1	$\binom{13}{1}\binom{4}{3}\binom{12}{1}\binom{4}{2}$
Flush (excluding royal flush and straight flush)	5,108	0.197%	0.367%	507.8 : 1	$\binom{13}{5}\binom{4}{1}-\binom{10}{1}\binom{4}{1}$
Straight (excluding royal flush and straight flush)	10,200	0.392%	0.76%	253.8 : 1	$\binom{10}{1}\binom{4}{1}^5-\binom{10}{1}\binom{4}{1}$
Three of a kind	54,912	2.11%	2.87%	46.3 : 1	$\binom{13}{1}\binom{4}{3}\binom{12}{2}\binom{4}{1}^2$
Two pair	123,552	4.75%	7.62%	20.03 : 1	$\binom{13}{2}\binom{4}{2}^2\binom{11}{1}\binom{4}{1}$

Expectation of a Discrete Random Variable

- Expectation is the average of a random variable

For a discrete random variable with values $x_1, x_2, x_3, \ldots, x_k$

$$E(x) = \sum_{i=1}^{k} x_i P(x_i)$$

What is the expectation of variable that represents coin tossing?
x_1=1 (head), x_2=2 (tails)

Casting a die?

Value of x

Probability of value x

Expectation is a Linear Operator (x, y are random variables, a,b, and c are constants)

- $E(x+c)= E(x) + c$ when c is a constant
- $E(x+y)=E(x) + E(y)$
- $E(ax)=aE(x)$
- $E(a+bx)=E(a)+E(bx)=a+bE(x)$

Variance of a Discrete Random Variable

- Variance is the squared difference of the value of the variable minus its Expectation multiplied by the probability of occurrence of the value

$$\text{For a discrete random variable with values } x_1, x_2, x_3, \ldots, x_k$$

$$V(x) = \sum_{i=1}^{k} (x_i - E(x_i))^2 P(x_i)$$

What is the variance of variable that represents coin tossing?
x_1=1 (head), x_2=2 (tails)

Casting a die?

Uniform Distribution (figures form pages 221-222 BBR)

$$P(x)=\frac{1}{k} \quad x=1,2,3,\ldots,k$$

$$E(x)=\frac{1}{k}\sum_{x=1}^{k}x_i=\frac{1}{k}\left[\frac{k(k+1)}{2}\right]=\frac{k+1}{2}$$

$$Var(x)=\frac{k^2-1}{12}$$

$$E(X)=\sum_{x=1}^{k}xP(x)=\sum_{i=1}^{k}x\left(\frac{1}{k}\right) \tag{5-24}$$

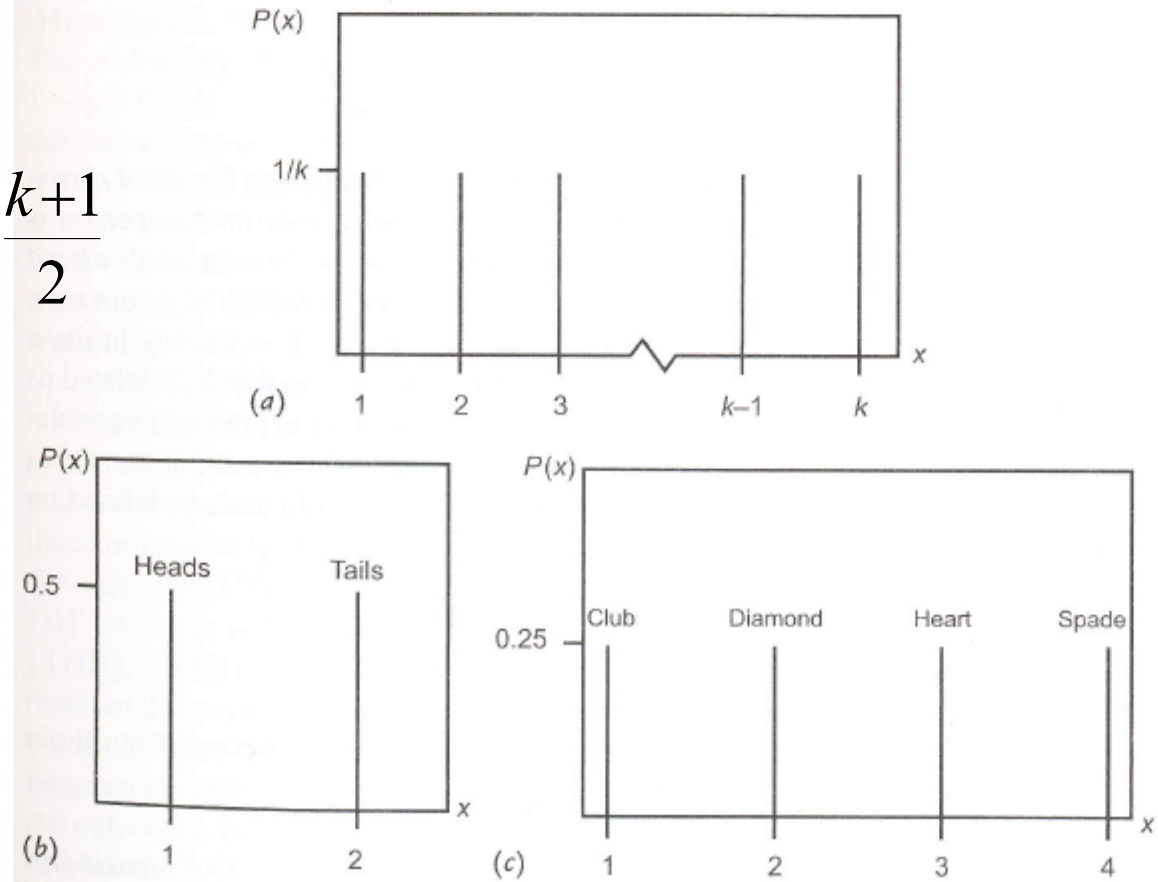


FIGURE 5-6. Discrete uniform random variables.

Binomial Distribution

- Statistical experiment = Bernoulli process with n independent trials
- The same pair of outcomes (e.g., head – tail).
- Probability of each outcome is the same across trials (π success 1- π failure)
- Random variable x is the number of successes among the n trials

TABLE 5-7
Possible Outcomes of the Coin Toss Experiment

$n = 1$				T		H		
$n = 2$			TT		TH			
					HT		HH	
$n = 3$		TTT		TTH				
				THT				
				HTT		THH		
					HTH			
					HHT		HHH	
$n = 4$	TTTT		TTTH		TTHH		THHH	HHHH
			TTHT		THTH		HTHH	
			THTT		HTTH		HHTH	
			HTTT		THHT		HHHT	
				HTHT				
				HHTT				

Trials and Binomial Distribution

TABLE 5-9
Binomial Distributions

Number of trials n	Number of heads	Probability of number of heads
1	0	$(1-\pi)$
	1	π
2	0	$(1-\pi)^2$
	1	$2\pi(1-\pi)$
	2	π^2
3	0	$(1-\pi)^3$
	1	$3\pi(1-\pi)^2$
	2	$3\pi^2(1-\pi)$
	3	π^3
4	0	$(1-\pi)^4$
	1	$4\pi(1-\pi)^3$
	2	$6\pi^2(1-\pi)^2$
	3	$4\pi^3(1-\pi)$
	4	π^4

$$P(T)=1-\pi$$
$$P(H)=\pi$$
$$P(TT)=P(T)P(T)=(1-\pi)(1-\pi)$$

Binomial in R

- Run sample experiments and study outcomes
- Draw numbers out of the binomial distribution

CONTINUOUS RANDOM VARIABLES

Sample of tossing 1 coin 1000 times

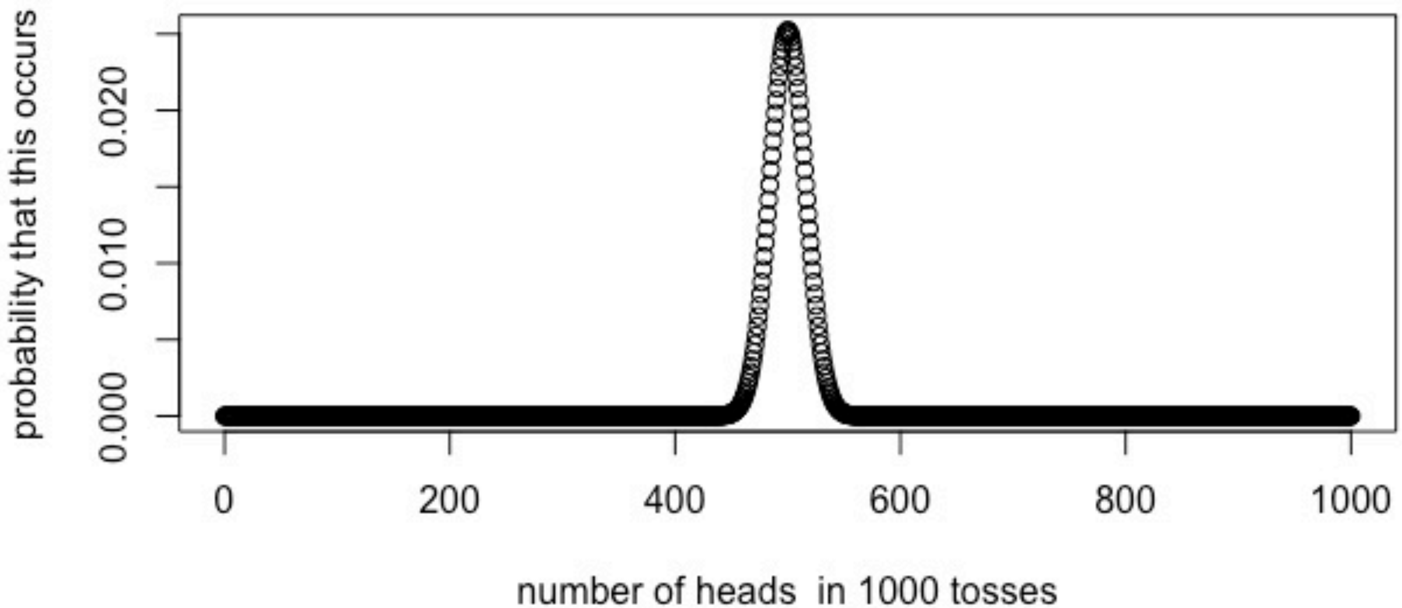
```
coin = sample(c("Heads","Tails"),1000, replace=T, prob=c(0.500,0.500))
coinsum(coin == "Heads")
sum(coin == "Tails")
table(coin)
```

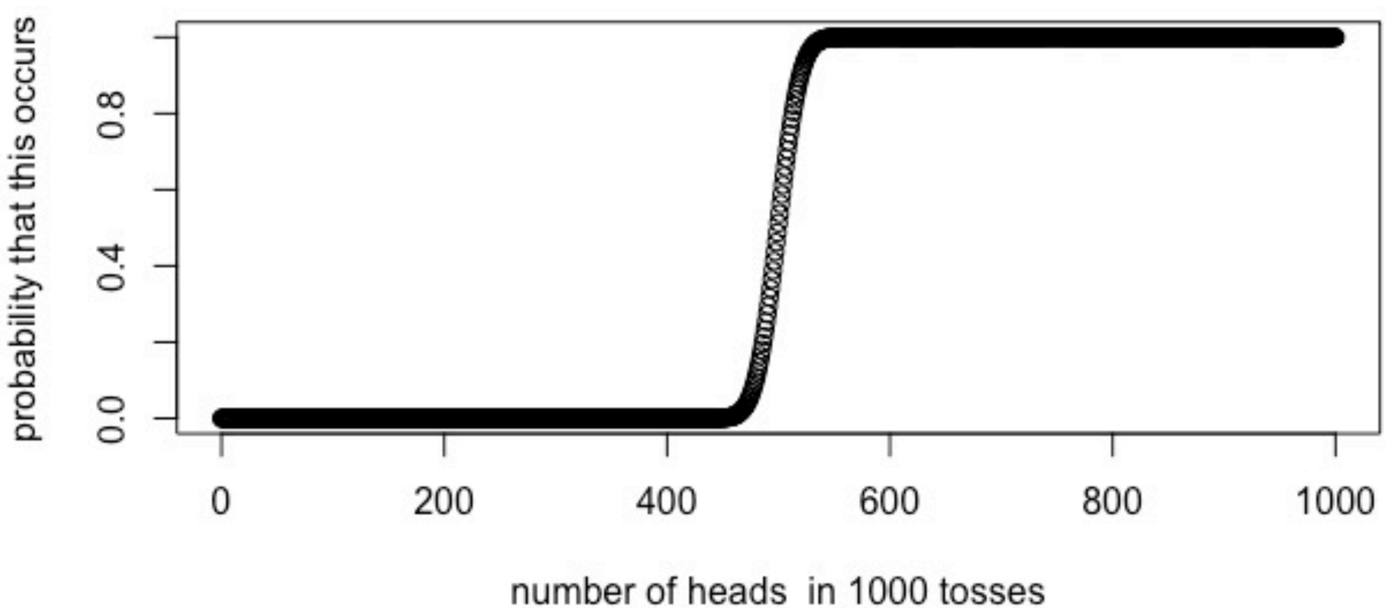
The outcome of this is:

Heads Tails   485   515

Same but from its theoretical distribution

```
# set the number of tosses to be 1000

numcoins = 1000

# let the number of heads vary from 0 to total number of coin tossing

numheads = 0:bcoins

# we can get the probability of each number of heads occurrence

dbinom(numheads,size=numcoins,prob=0.5)

# and finally we can plot the probability of each of the outcomes using plot
# type p just points, type l just lines, type b lines and points, type h histogram

plot(numheads,dbinom(numheads,size=numcoins,prob=0.5),type="p",  xlab = "number of heads  in 1000 tossins",       ylab = "probability that this occurs")

# cumulative binomial probability

plot(numheads,pbinom(numheads,size=numcoins,prob=0.5),type="p",     xlab = "number of heads  in 1000 tossings",     ylab = "probability that this occurs")
```


probability that this occurs
0.000
0.010
0.020
0
200
400
600
800
1000
number of heads in 1000 tosses


probability that this occurs
0.0
0.4
0.8
0
200
400
600
800
1000
number of heads in 1000 tosses

Continuous Variables

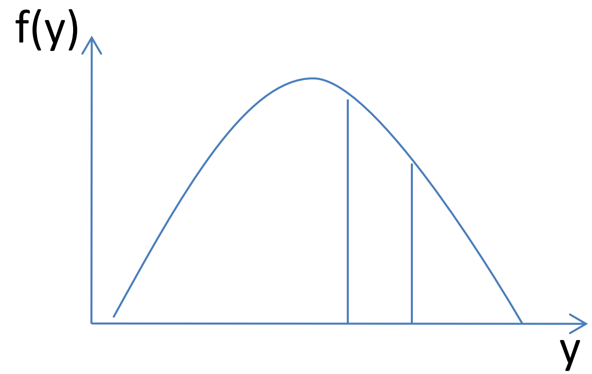
- Instead of finite values we have infinitely many values within an interval
- Instead of a frequency histogram – we see a continuous line
- Instead of Sums (Σ) we work with integrals ($\int$)

- The most famous is the Normal distribution

For a continuous random variable probability of a point is zero and we work only with intervals of values using the **probability density function (f(y))**.

$$\Pr ob(a \le y \le b) = \int_a^b f(y)dy \ge 0$$

$$\int_{-\infty}^{+\infty} f(y)dy = 1$$



If x takes values only between a and b then:

$$\int_a^b f(y)dy = 1$$

Expectation and Variance of a Continuous Random Variable

f(x) is called probability density function. You can think of f(x)dx as an elementary probability

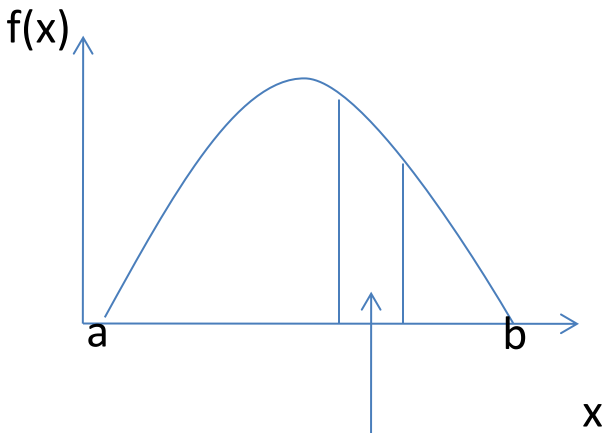
$$E(x) = \int_a^b x f(x)\,dx$$

$$Var(x) = \int_a^b [x - E(x)]^2 f(x)\,dx$$

$$\text{Prob}(a < x < b) = \int_a^b f(x)\,dx$$

$$f(x) \geq 0,\ a \leq x \leq b$$
$$\textit{if all x values between a and b}$$

$$\text{Prob}(a \leq x \leq b) = \int_a^b f(x)\,dx = 1$$

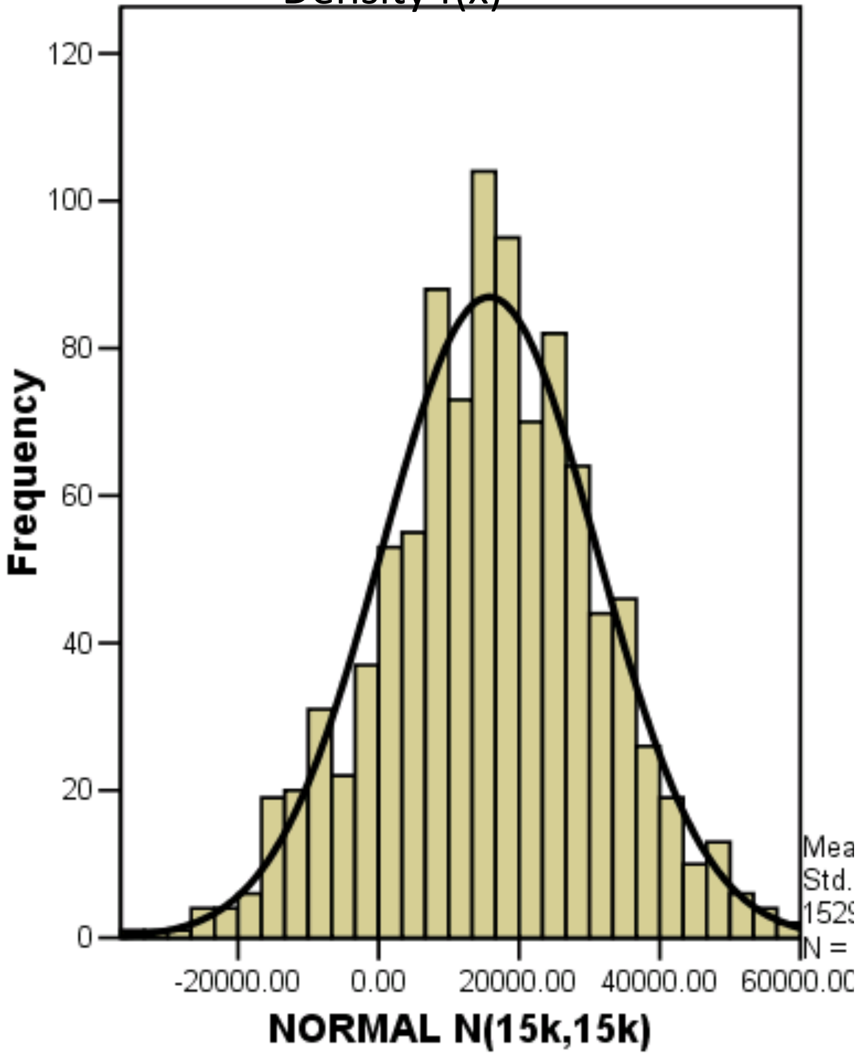


The probability of an occurrence defined over an interval of values

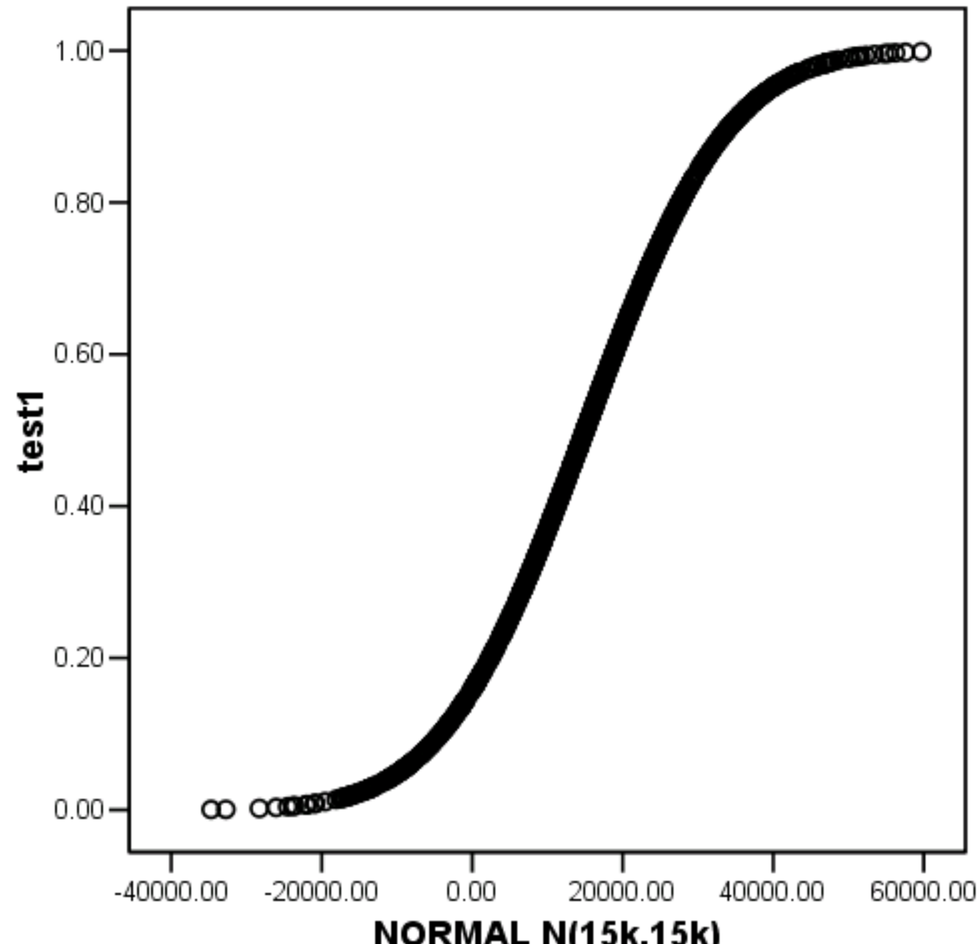
Page 251 BBR – numerical example

$\frac{1}{\sigma\sqrt{(2\pi)}} e^{\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)}$ Normal Distribution

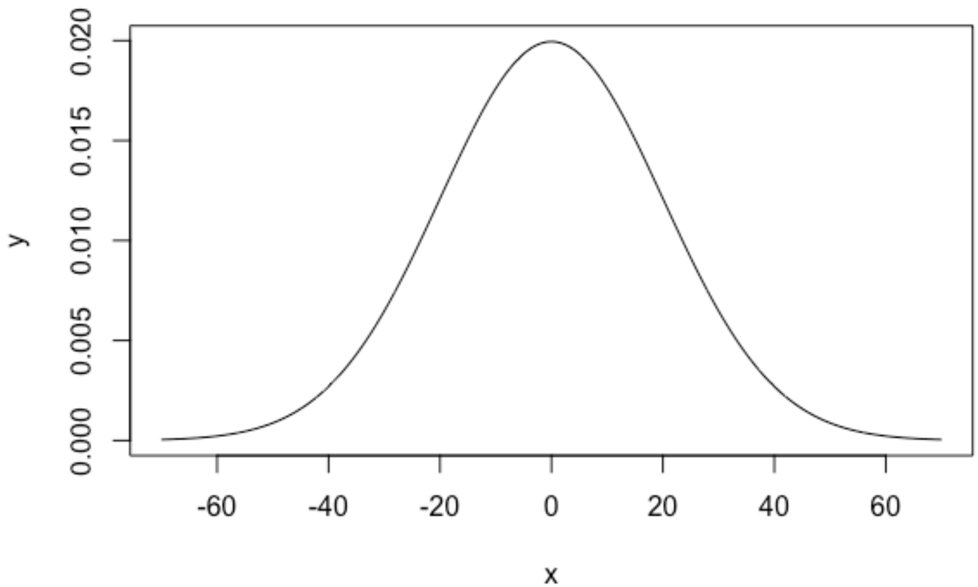
Density f(x)



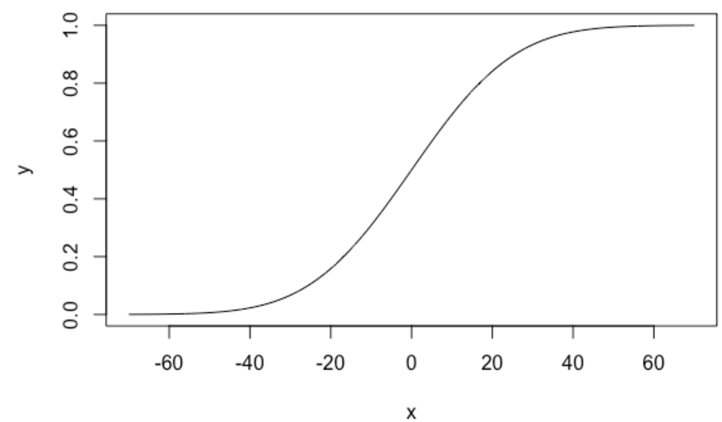
Cumulative F(x)



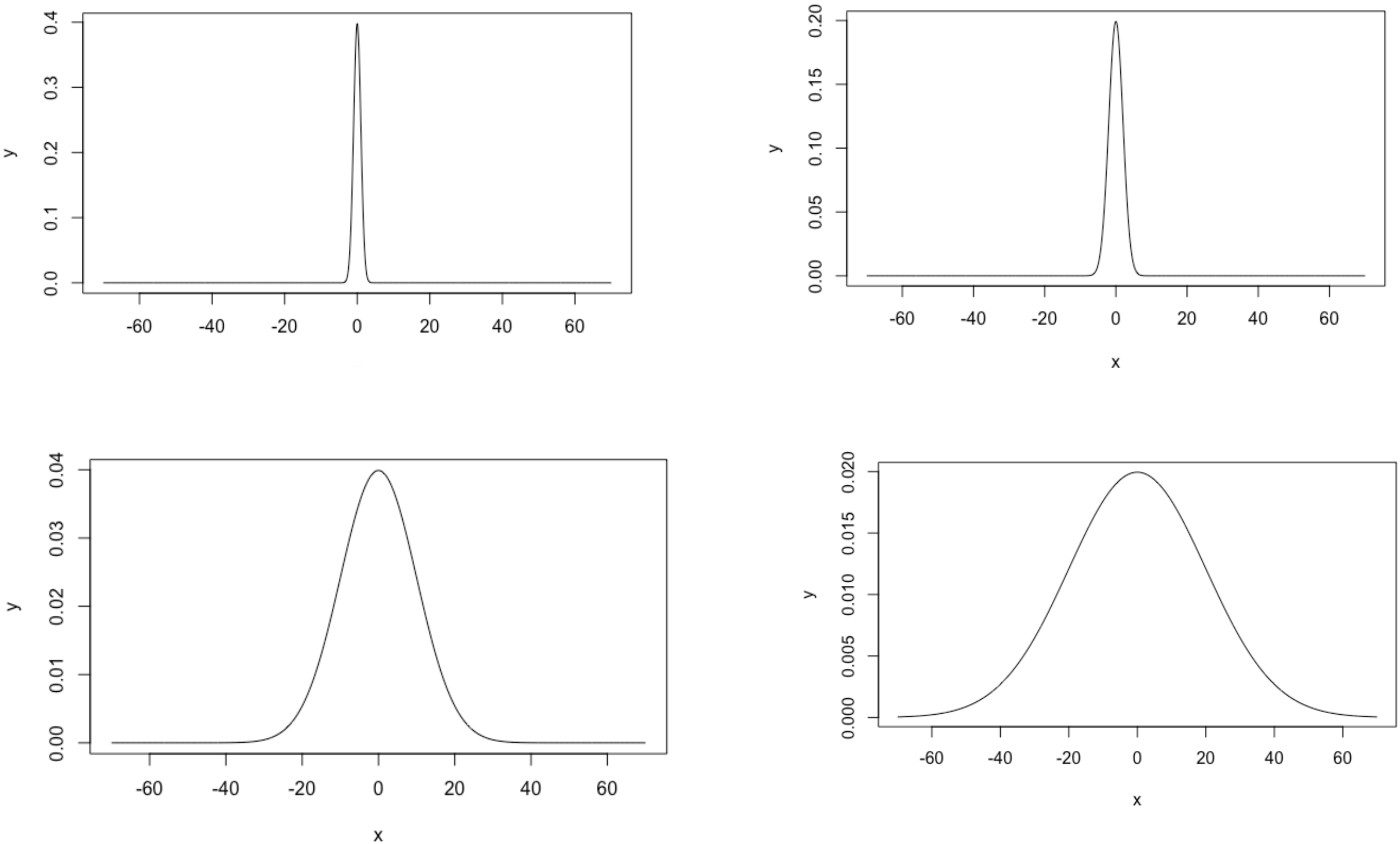
Same in R



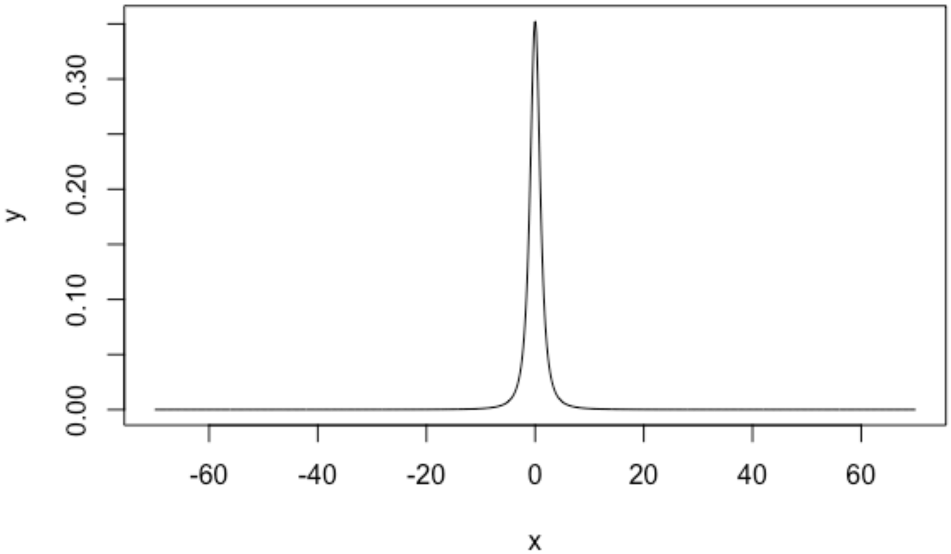

```
x   <- seq(-70,70,length=1000)
y   <- dnorm(x,mean=0.0, sd=20)
plot(x,y, type="l", lwd=1)
x   <- seq(-70,70,length=1000)
y   <- pnorm(x,mean=0.0, sd=20)
plot(x,y, type="l", lwd=1)
```



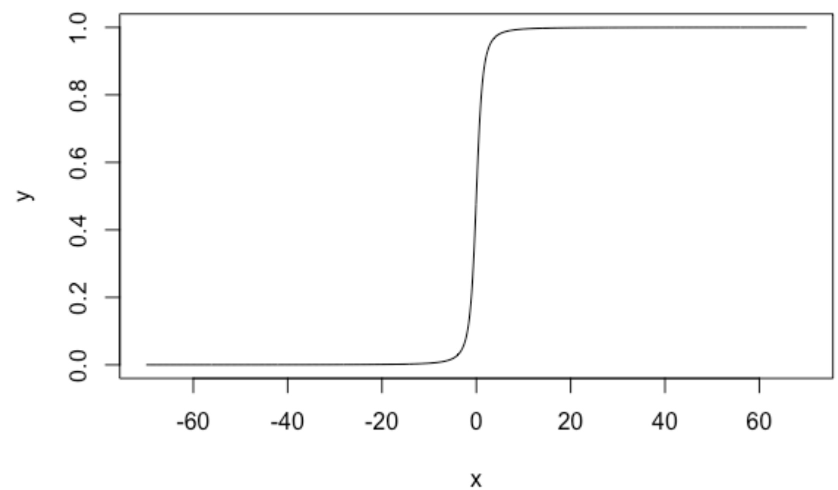
Normal Distributions with mean =0, and SD = 1, 2, 10, and 20




```
x   <- seq(-70,70,length=1000)
y   <- dnorm(x,mean=0.0, sd=1)
plot(x,y, type="l", lwd=1)
```

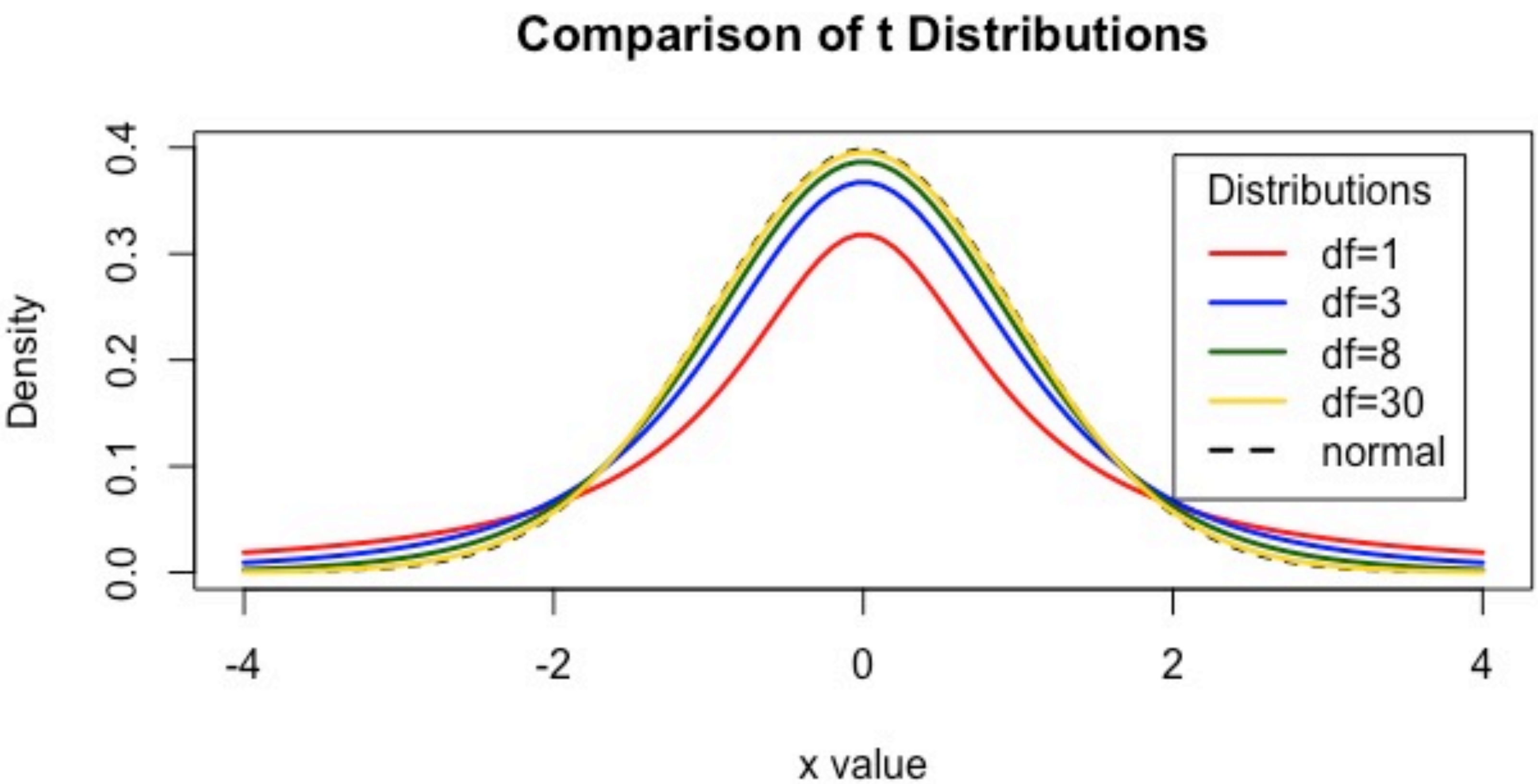


This is the t distribution that we will use often to check regression models



```
x <- seq(-4, 4, length=100)
hx <- dnorm(x)

degf <- c(1, 3, 8, 30)
colors <- c("red", "blue", "darkgreen", "gold", "black")
labels <- c("df=1", "df=3", "df=8", "df=30", "normal")

plot(x, hx, type="l", lty=2, xlab="x value",
   ylab="Density", main="Comparison of t Distributions")

for (i in 1:4){
 lines(x, dt(x,degf[i]), lwd=2, col=colors[i])
}

legend("topright", inset=.05, title="Distributions",
    labels, lwd=2, lty=c(1, 1, 1, 1, 2), col=colors)
```

http://www.statmethods.net/advgraphs/probability.html



Comparison of t Distributions
Density
0.0
0.1
0.2
0.3
0.4
-4
-2
0
2
4
x value
Distributions
df=1
df=3
df=8
df=30
normal