Statistics for Health Care

Unit 3:

Overview/Teasers

Overview

- Introduction to probability and conditional probability; Bayes' Rule; diagnostic testing

Teaser 1, Unit 3

- A classic statistics problem: The Monty Hall Problem

Teaser 1, Unit 3

- You are on the Monty Hall show. You are presented with 3 doors (A, B, C), one of which has a valuable prize behind it (the other two have nothing). You choose door A; Monty Hall opens door B and shows you that there is nothing behind it. Then he gives you the choice of sticking with A or switching to C. Do you stay or switch; or does it make no difference?

www.cellulitedx.com/en-us/CelluliteDXTest.aspx?view=2

Free Hotmail   Windows   Windows Marketplace   Windows Media   Bookmarks   Imported From IE   Natura

HairDX

CelluliteDX   Patent Pending   Genetic Test For   Clinical   How To
                                Moderate To Severe Cellulite   Studies   Get The Te

Now you can find
are genetically pr
moderate to seve

What is the CelluliteDX Genetic Test?

Teaser 2, Unit 3

- http://www.cellulitedx.com/en-us/

The Cellulite DX genetic test is an actual genetic test that purports to predict a woman's risk of developing moderate to severe cellulite.

Here's the company's pitch:

"A patient who tests positive has approximately a 70% chance of developing moderate to severe cellulite."
"A patient who tests negative has approximately a 50% chance of not developing moderate to severe cellulite."

Is this a useful test?

Statistics for Health Care

Module 1:

Basic probability

Probability

- Probability – the chance that an uncertain event will occur (always between 0 and 1)

Symbols:

P(A) = "the probability that event A will occur"

P(red card) = "the probability of a red card"

P(~A) = "the probability of NOT getting event A" [complement]

P(~red card) = "the probability of NOT getting a red card"

P(A & B) = "the probability that both A and B happen" [joint probability]

P(red card & ace) = "the probability of getting a red ace"

Probability example:

You draw one card from a deck of cards. What's the probability that you draw an ace?

$$P(\text{draw an ace}) = \frac{\# \text{ of aces in the deck}}{\# \text{ of cards in the deck}} = \frac{4}{52} = .0769$$

Probability example:

- A = draw a red card
- B = draw an ace

P(A) =
P(B) =
P(~A) =
P(~B) =
P(A&B) =

Assessing Probability

1. Theoretical/Classical probability—based on theory (*a priori* understanding of a phenomena)

 e.g.: theoretical probability of rolling a 2 on a standard die is 1/6

 theoretical probability of drawing an ace in a standard deck is 1/13

2. Empirical probability—based on empirical data

 e.g.: empirical probability of an Earthquake in Bay Area by 2032 is .62 (based on historical data)

 empirical probability of a lifetime smoker developing lung cancer is 15 percent (based on empirical data)

From empirical data on blood type:

Out of 100 donors . . . . .

84 donors are RH+	16 donors are RH-
38 are O+	7 are O-
34 are A+	6 are A-
9 are B+	2 are B-
3 are AB+	1 is AB-

Source: AABB.ORG

1. **What's the probability that a random donor will be AB+?**
2. **What's the probability that a random donor will be O?**

From empirical data on blood type:

Out of 100 donors . . . . .

84 donors are RH+	16 donors are RH-
38 are O+	7 are O-
34 are A+	6 are A-
9 are B+	2 are B-
3 are AB+	1 is AB-

Source: AABB.ORG

1. **What's the probability that a random donor will be AB+? 3%**
2. **What's the probability that a random donor will be O? 45%**

Cancer probabilities: U.S. men

- P (developing cancer, lifetime) = 44.8%
- P (dying from cancer) = 23.1%

- P (developing prostate cancer, lifetime) = 16.2%
- P (dying from prostate cancer) = 2.8%

- P (developing lung cancer, lifetime) = 7.8%
- P (dying from lung cancer) = 6.7%

http://www.cancer.org/cancer/cancerbasics/lifetime-probability-of-developing-or-dying-from-cancer (ACS)

Cancer probabilities: U.S. women

- P (developing cancer, lifetime) = 38.2%
- P (dying from cancer) = 19.4%

- P (developing breast cancer, lifetime) = 12.4%
- P (dying from breast cancer) = 2.8%

- P (developing lung cancer, lifetime) = 6.4%
- P (dying from lung cancer) = 5.0%

Counting methods for calculating theoretical probability

- These are problems where all outcomes are equally likely. For example, drawing cards out of a deck or rolling dice.

Counting methods

Great for gambling! Fun to compute!

If <u>outcomes are equally likely to occur</u>…

$$P(A) = \frac{\#\,\text{of ways A can occur}}{\text{total}\,\#\,\text{of outcomes}}$$

<u>Note:</u> these are called "counting methods" because we have to **<u>count</u>** the number of ways A can occur and the number of total possible outcomes.

Example 1: Calculate the following probabilities ("with replacement"):

- What is the probability of getting 1 six when rolling a die?

- What is the probability of getting 2 sixes when rolling two dice?

- What is the probability of getting a sum of 6 when rolling two dice?

Example 2: Calculate the following probability ("without replacement"):

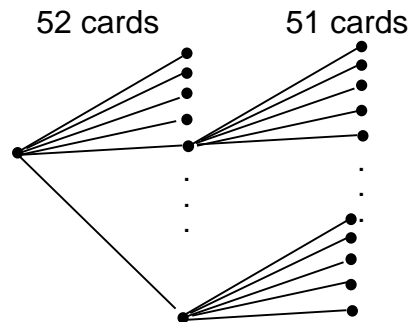What is the probability of drawing two aces from a standard deck of cards?

Calculate the following probability ("without replacement"):

$$P(\text{draw 2 aces}) = \frac{\#\ of\ \text{ways you can draw ace, ace}}{\#\ of\ \text{different 2 - card sequences you could draw}}$$

Numerator: A♣A♦, A♣A♥, A♣A♠, A♦A♥, A♦A♠, A♦A♣, A♥A♦, A♥A♣, A♥A♠, A♠A♣, A♠A♦, or A♠A♥ = 12

Denominator = 52x51 = 2652 -- why?

52 cards   51 cards

$$\therefore P(\text{draw 2 aces}) = \frac{12}{52x51}$$

OR, ignore order:

$$P(\text{draw 2 aces}) = \frac{\#\text{ of pairs of aces}}{\#\text{ of different two - card hands you could draw}}$$

Numerator: A♣A♦, A♣A♥, A♣A♠, A♦A♥, A♦A♠, A♥A♠ = 6

Denominator = $\dfrac{52x51}{2} = 1326$ ←————————— Divide out order!

$\therefore P(\text{draw 2 aces}) = \dfrac{6}{\frac{52x51}{2}} = \dfrac{12}{52x51}$

Alternatively, use multiplication rule (see Module 2):

What's the probability that you draw 2 aces when you draw two cards from the deck?

$$P(\text{draw ace on first draw}) = \frac{\# \text{ of aces in the deck}}{\# \text{ of cards in the deck}} = \frac{4}{52}$$

$$P(\text{draw an ace on second draw too}) = \frac{\# \text{ of aces in the deck}}{\# \text{ of cards in the deck}} = \frac{3}{51}$$

$$\therefore P(\text{draw ace AND ace}) = \frac{4}{52} \text{ x } \frac{3}{51}$$

In probability, AND means multiply. *More like this in module 2.*

Example 3: What's your probability of winning the lottery?

A lottery works by picking 6 numbers from 1 to 49. What's your probability of winning?

Numerator: How many ways can you win? 1

Denominator: How many combinations of 6 numbers could you choose? …

Choosing function (combinations!)

If r objects are taken from a set of n objects without replacement and ignoring order, how many different samples are possible?

$$\binom{n}{r} = \frac{n!}{(n-r)!\,r!}$$

Lottery combinations

How many different combinations of 6 numbers can be formed from the integers 1 through 49?

$$\binom{49}{6} = \frac{49!}{(49-6)!6!} = \frac{49 x 48 x 47 x 46 x 45 x 44 x 43!}{43! x 6!}$$

$$= \frac{49 x 48 x 47 x 46 x 45 x 44}{6!} =$$

$$\frac{49 x 48 x 47 x 46 x 45 x 44}{6 x 5 x 4 x 3 x 2 x 1} = 13,983,816$$

Lottery probability

Thus, your probability of winning the lottery is: 1/13,983,816!

Statistics for Health Care

Calculating probabilities: Permutations and Combinations

Summary of Counting Methods

1. **Permutations (order matters) with replacement**

2. **Permutations (order matters) without replacement**

3. **Combinations (order doesn't matter) without replacement**

Permutations—Order matters!

A permutation is an ordered arrangement of objects.

<u>With replacement</u>=once an event occurs, it can occur again (after you roll a 6, you can roll a 6 again on the same die).
<u>Without replacement</u>=an event cannot repeat (after you draw an ace of spades out of a deck, there is 0 probability of getting it again).
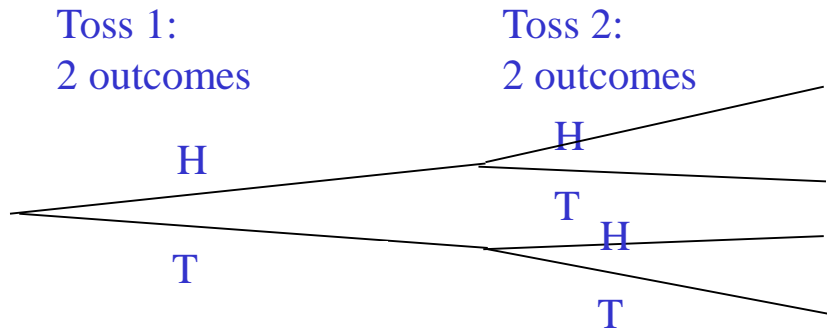
1. Permutations—with replacement

With Replacement – Think coin tosses, dice, and DNA.
 "memoryless" – After you get heads, you have an equally likely chance of getting a heads on the next toss (unlike in cards example, where you can't draw the same card twice from a single deck).
 E.g.:What's the probability of getting two heads in a row ("HH") when tossing a coin?

Toss 1:
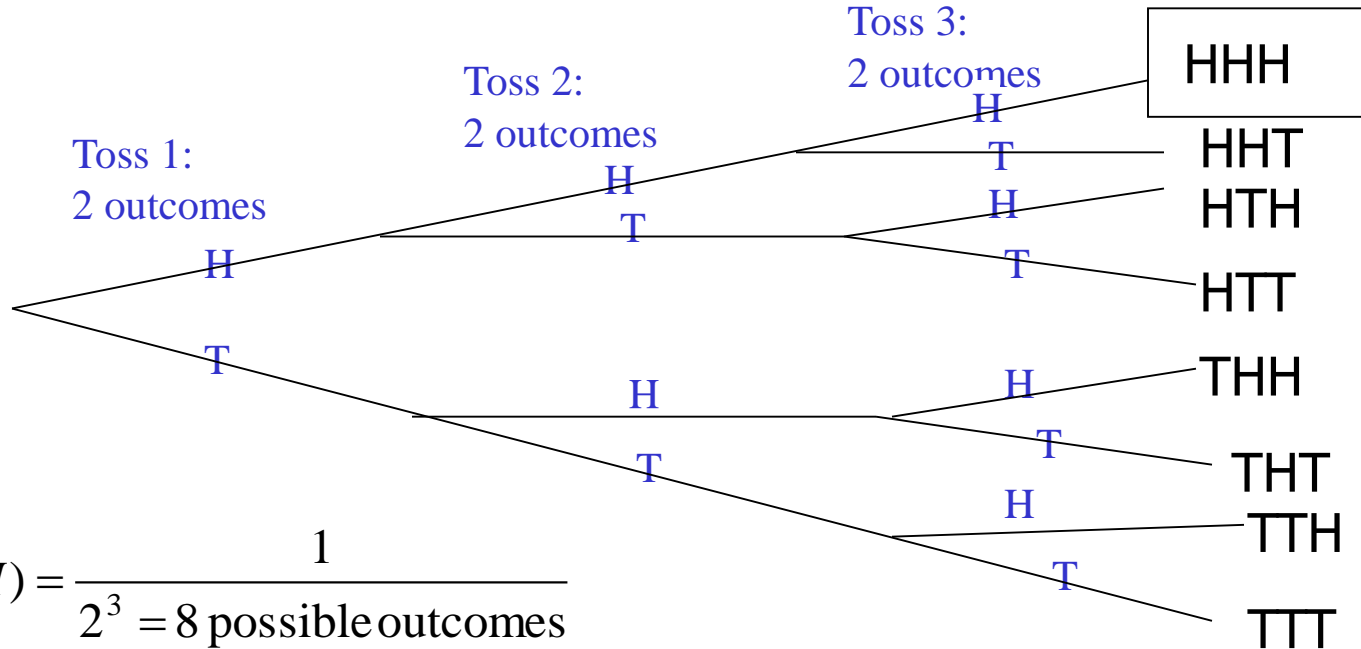2 outcomes

Toss 2:
2 outcomes

H

T

H

T

H

T

2^2 total possible outcomes: HH, HT, TH, TT

$$P(HH) = \frac{1 \text{ way to get HH}}{2^2 \text{ possible outcomes}}$$

1. Permutations—with replacement

What's the probability of 3 heads in a row?

Toss 1:
2 outcomes

Toss 2:
2 outcomes

Toss 3:
2 outcomes

H
T

H
T

H
T
H
T
H
T
H
T

HHH
HHT
HTH
HTT
THH
THT
TTH
TTT

$$P(HHH) = \frac{1}{2^3 = 8 \text{ possible outcomes}}$$

1. Permutations—with replacement

When you roll a pair of dice (or 1 die twice), what's the probability of rolling 2 sixes?

$$P(6,6) \frac{1 \text{ way to roll } 6, 6}{6^2} = \frac{1}{36}$$

What's the probability of rolling a 5 and a 6?

$$P(5 \& 6) = \frac{2 \text{ ways}: 5,6 \text{ or } 6,5}{6^2} = \frac{2}{36}$$

Summary: permutation with replacement

"order matters" and "with replacement"→ use powers→

$$(\# \text{ possible outcomes per event})^{\text{the \# of events}} = n^r$$

Practice problem:

1. How many different codons (=3 nucleotide "word") of DNA are possible given that there are 4 nucleotides (A, G, C, T)?

Practice problem:

Answer:

2. Permutations—without replacement

Without replacement—Think cards (w/o reshuffling) and seating arrangements.

Example: You are moderating a debate of gubernatorial candidates. How many different ways can you seat the panelists in a row? Call them Arianna, Buster, Camejo, Donald, and Eve.

2. Permutation—without replacement

→ "Trial and error" method:

Systematically write out all permutations:

A B C D E

A B C E D

A B D C E

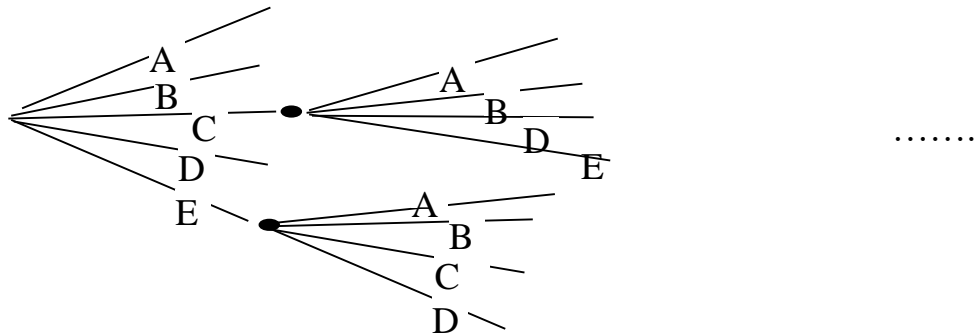A B D E C

A B E C D

A B E D C

Quickly becomes a pain!
Easier to figure out patterns using a the probability tree!

2. Permutation—without replacement

Seat One:
5 possible

Seat Two:
only 4 possible

Etc….
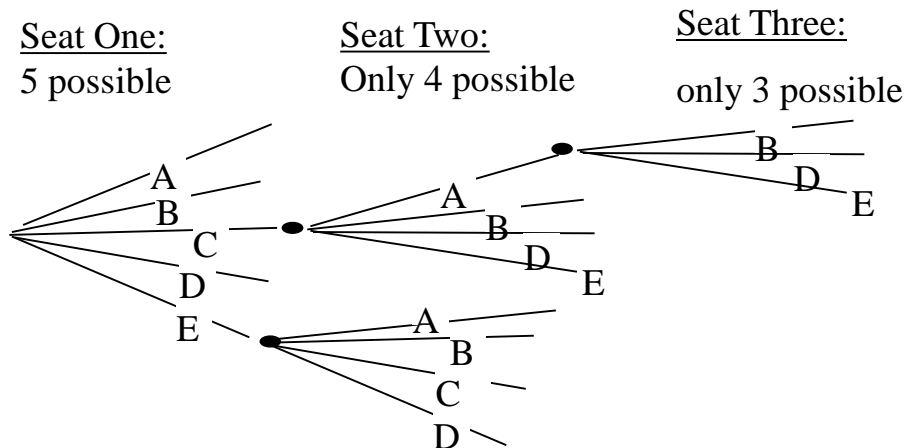
A
B
C
D
E

A
B
D
E

A
B
C
D

…….

# of permutations = 5 x 4 x 3 x 2 x 1 = 5!

There are 5! ways to order 5 people in 5 chairs (since a person cannot repeat)

2. Permutation—without replacement

What if you had to arrange 5 people in only 3 chairs (meaning 2 are out)?

Seat One:
5 possible

Seat Two:
Only 4 possible

Seat Three:
only 3 possible

A
B
C
D
E

A
B
D
E

A
B
C
D

B
D
E

$$5x4x3 =$$

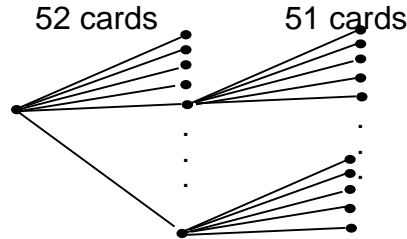$$\frac{5x4x3x2x1}{2x1} = \frac{5!}{2!} =$$

$$\frac{5!}{(5-3)!}$$

2. Permutation—without replacement

Note this also works for 5 people and 5 chairs:

$$\frac{5!}{(5-5)!} = \frac{5!}{0!} = 5!$$

2. Permutation—without replacement

How many two-card hands can I draw from a deck when order matters (e.g., ace of spades followed by ten of clubs is different than ten of clubs followed by ace of spades)

52 cards   51 cards

$$52 x 51 = \frac{52!}{50!} = \frac{52!}{(52-2)!}$$

2. Permutation—without replacement

How many ten-card hands can I draw from a deck when order matters...

$$52 \times 51 \times 50 \times 49 \times 48 \times 47 \times 46 \times 45 \times 44 \times 43 = \frac{52!}{42!} = \frac{52!}{(52-10)!}$$

Summary: permutation without replacement

"order matters" and "without replacement"→ use factorials→

$$\frac{(n \text{ people or cards})!}{(n \text{ people or cards} - r \text{ chairs or draws})!} = \frac{n!}{(n-r)!} = P(n,r) = {}^n P_r$$

$$\text{or } n(n-1)(n-2)\ldots(n-r+1)$$

Practice problem:

1. A wine taster claims that she can distinguish four different wines. What is the probability that she can do this by merely guessing (she is confronted with 4 unlabeled glasses)? (hint: without replacement)

Answer

P(success) = 1 (there's only way to get it right!) / total # of guesses she could make

Total # of guesses one could make randomly:

glass one: glass two: glass three: glass four:
4 choices   3 wine left   2 left   1 left

$$= 4 \times 3 \times 2 \times 1 = 4!$$

$$\therefore P(success) = 1 / 4! = 1/24 = .04167$$

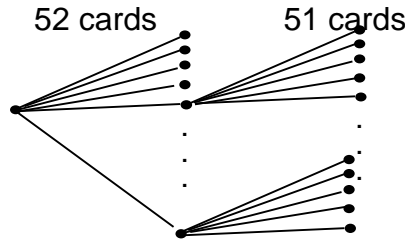# 3. Combinations—order doesn't matter (without replacement)

Introduction to combination function, or "choosing"

Written as: $_nC_r$ or $\begin{pmatrix} n \\ r \end{pmatrix}$

Spoken: "*n* choose *r*"

3. Combinations

How many two-card hands can I draw from a deck when order does <u>not</u> matter (e.g., ace of spades followed by ten of clubs is <u>the same</u> as ten of clubs followed by ace of spades)
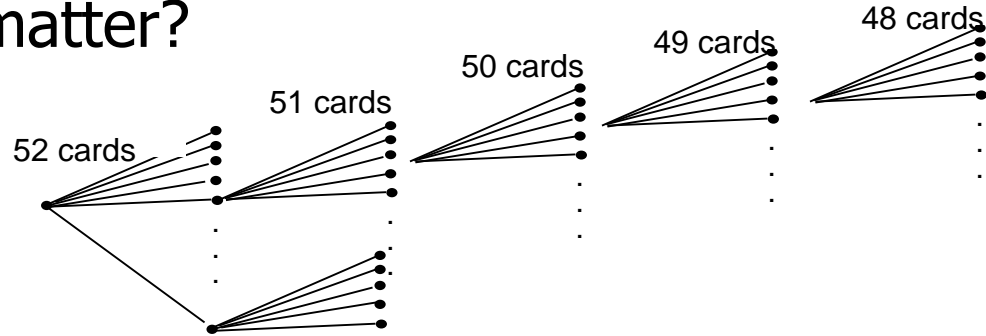
52 cards   51 cards

$$\frac{52x51}{2} = \frac{52!}{(52-2)!2}$$

3. Combinations

How many five-card hands can I draw from a deck when order does <u>not</u> matter?

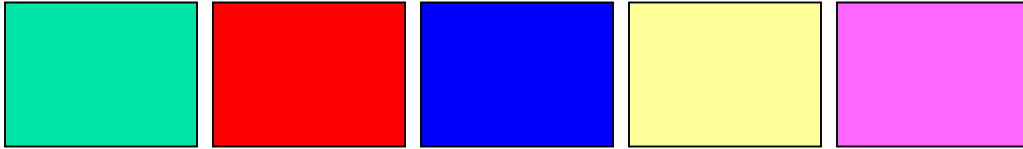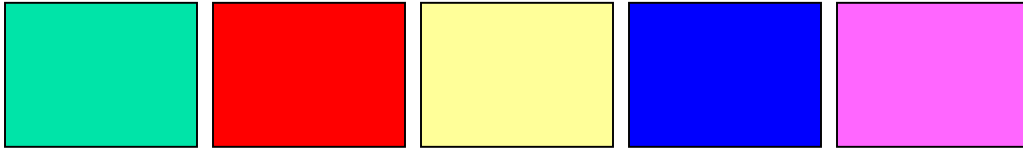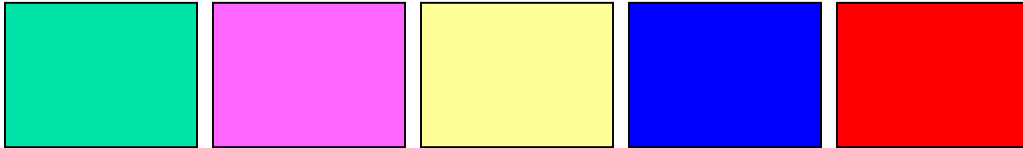$$\frac{52 \times 51 \times 50 \times 49 \times 48}{?}$$

3. Combinations

1.

2.

3.

….

How many repeats total??
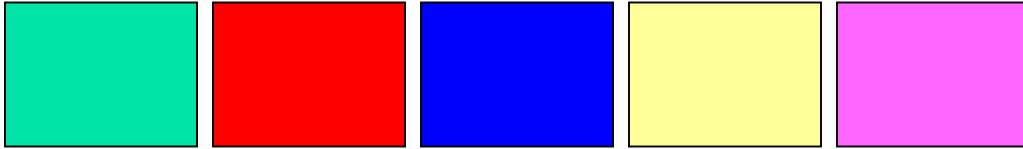
3. Combinations

1.

2.
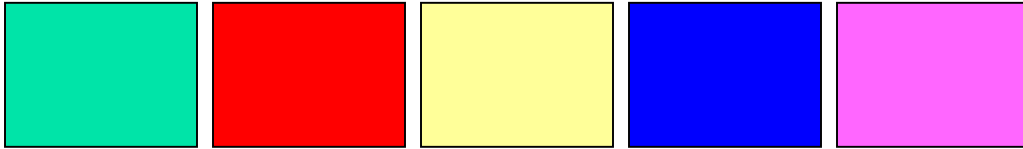
3.
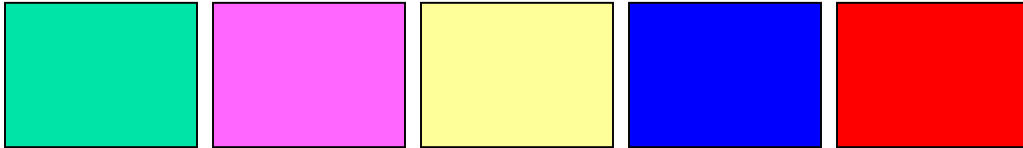
…. i.e., how many different ways can you arrange 5 cards…?

3. Combinations

That's a permutation without replacement!

5! = 120

$$\text{total \# of 5 - card hands} = \frac{52x51x50x49x48}{5!} = \frac{52!}{(52-5)!5!}$$

3. Combinations

- How many unique 2-card sets out of 52 cards?

$$\frac{52x51}{2} = \frac{52!}{(52-2)!2!}$$

- 5-card sets?

$$\frac{52x51x50x49x48}{5!} = \frac{52!}{(52-5)!5!}$$

- r-card sets?

$$\frac{52!}{(52-r)!r!}$$

- r-card sets out of n-cards?

$$\binom{n}{r} = \frac{n!}{(n-r)!r!}$$

Summary: combinations

"order doesn't matter" and "without replacement"→ use choosing function→

$$\binom{n}{r} = \frac{n!}{(n-r)!r!}$$

Examples—Combinations

A lottery works by picking 6 numbers from 1 to 49. How many combinations of 6 numbers could you choose?

$$\binom{49}{6} = \frac{49!}{43!6!} = 13{,}983{,}816$$

Which of course means that your probability of winning is 1/13,983,816!

Example—Combination!

How many ways can you get 3 heads in 5 coin tosses?

$$\binom{5}{3} = \frac{5!}{3!2!} = 10$$

Practice problems

- When you draw two cards from a deck, what's the probability that you will draw:

- 1. a pair of the same color

- 2. any pair

- 3. any two cards of the same color

Pair of the same color?

- P(pair of the same color) $= \dfrac{\#\,\text{pairs of same color}}{\text{total}\,\#\,\text{of two}-\text{card combinations}}$

$$\text{Denominator} =_{52}C_2 = \frac{52!}{50!\,x2!} = \frac{52x51}{2} = 1326$$

Numerator = red aces, black aces; red kings, black kings; etc....= 2x13 = 26

$$\text{So, P(pair of the same color)} = \frac{26}{1326} = 1.96\%\ \text{chance}$$

Any old pair?

- P(any pair) = $\dfrac{\text{\# pairs}}{\text{total \# of two} - \text{card combinations} = 1326}$

$$\text{number of different possible pairs of aces} =_4 C_2 = \frac{4!}{2!2!} = \frac{4x3}{2} = 6$$

$$\text{number of different possible pairs of kings} =_4 C_2 = \frac{4!}{2!2!} = \frac{4x3}{2} = 6$$

...

$$13x6 = 78 \text{ total possible pairs}$$

$$\therefore P(\text{any pair}) = \frac{78}{1326} = 5.9\% \text{ chance}$$

Two cards of same color?

Numerator: $_{26}C_2$ x 2 colors = 26!/(24!2!) = 325 x 2 = 650

Denominator = 1326

So, P (two cards of the same color) = 650/1326 = 49% chance

A little non-intuitive? Here's another way to look at it…

52 cards
26 red branches
26 black branches

From a Red branch: 26 black left, 25 red left

From a Black branch: 26 red left, 25 black left
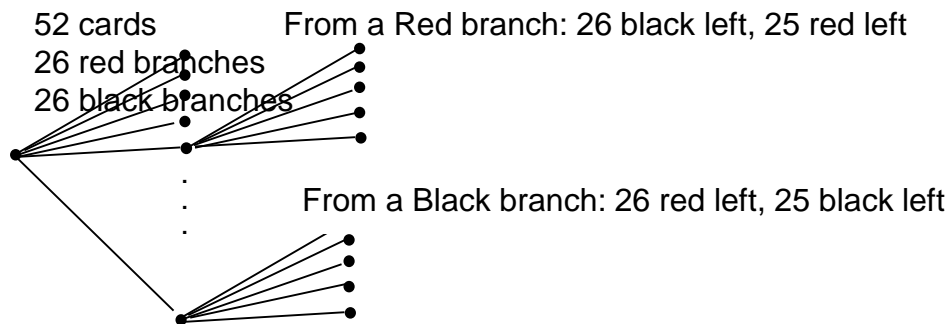
26x25 RR

26x26 RB

50/102

26x26 BR

Not quite 50/100

26x25 BB

The Birthday Problem

- In a room of 30 people what is the probability that at least two people share the same birthday?

  What would you guess is the probability? High or low?

Birthday problem answer

* *Trick!* 1- P(none) = P(at least one)

Use complement to calculate answer. It's easier to calculate 1- P(no matches) = the probability that at least one pair of people have the same birthday.

Birthday problem answer

What's the probability of no matches?

Denominator: how many sets of 30 birthdays are there?
--with replacement (permutation, order matters)
365^{30}

Numerator: how many different ways can you distribute 365 birthdays to 30 people without replacement?
--without replacement (permutation, order matters)
[365!/(365-30)!]= [365 x 364 x 363 x 364 x …..x(365-29)]

∴ P(no matches) = [365 x 364 x 363 x 364 x ….. x336] / 365^{30} =29.4%

∴ P(at least one match) = 1-29.4% = 70.6%

Combinations with replacement

"order doesn't matter" and "with replacement"→
use choosing function→

$$\binom{n+r-1}{r} = \frac{(n+r-1)!}{(n-1)!\,r!}$$

Statistics for Health Care

Module 2:

Rules of probability

Some useful rules of probability

- Mutually exclusive, exhaustive probabilities must sum to 1.
- Probability of at least one = 1 – P(none)
- AND means multiply
- OR means add
- If A and B are independent, then: P(A&B) = P(A)*P(B)

Mutually exclusive, exhaustive probabilities must sum to 1.

- For example, P(A) + P(~A) = 1

- If the probability of rain tomorrow is 20%, then the probability of it not raining tomorrow is 80%.
- If the chance of a woman NOT developing moderate to severe cellulite is 50%, then the chance that she develops the condition is 50%.

Genetics example:

- In genetics, if both the mother and father carry one copy of a recessive disease-causing mutation (d), there are three possible outcomes:

  - child is not a carrier (DD)
  - child is a carrier (Dd)
  - child has the disease (dd)

  - P(genotype=DD)=.25
  - P(genotype=Dd)=.50
  - P(genotype=dd)=.25

mutually exclusive, exhaustive probabilities sum to 1.

P(at least one)= 1 − P(none)

What's the probability of rolling at least 1 six when you roll 4 dice?

It's hard to calculate the probability of *at least one* (=1 OR 2 OR 3 OR 4 sixes). It's easier to calculate the probability of getting 0 sixes.

P(no sixes) = $(5/6)^4$ = 48.2%

P(at least one six) = 1-P(no sixes) = 51.8%

P(at least one)= 1 – P(none)

- If the probability of transmitting HIV is 1/500 per single act of unprotected sex, what is the probability that the uninfected partner of a discordant couple seroconverts by the end of 100 acts?

Seroconversion requires <u>at least one</u> transmission
P(at least one) = 1- P(no transmissions)

P(no transmissions in 100 acts) = $(\frac{499}{500})^{100} = .819$

$\therefore$**P(at least one) = 1 -.819 = .181**

AND means multiply

- If A and B are independent, then:
- P(A&B) = P(A)*P(B)

- In a genetic cross involving a recessive disease gene (d), what's the chance that two heterozygous parents (Dd) pass on the disease to their child.
- 0.5 chance that dad passes on d
- 0.5 chance that mom passes on d

- 0.5 x 0.5 = 25% chance the child will have the disease (dd)

AND means multiply

- What's the probability that you win the lottery (pick 6 of 49 numbers) twice in a row?

1/13,983,816* 1/13,983,816 = .000000000000005

AND means multiply

- (From Week 1) What's the probability that a woman is BOTH a super-high user of lipstick (218 mg/day) AND that she uses the lipstick with the highest detected lead levels (7.19 ppm)?

1/30,000 * 1/400 = 1 in 12 million women

OR means add

- If A and B are mutually exclusive events, then the probability of A or B is:

- $P(A) + P(B)$

- What is the chance of getting a 5 or a 6 when you roll a die?

- $P(5 \text{ OR } 6) = 1/6 + 1/6 = 2/6$

OR means add

In a genetic cross involving a recessive disease gene (d), what's the chance that two heterozygous parents (Dd) will have a heterozygote child (Dd)?

- 0.5 x 0.5 = 25% chance the child will inherit D from dad and d from mom
- 0.5 x 0.5 = 25% chance the child will inherit d from dad and D from mom

$\therefore$ P(child is Dd) = 25% + 25% = 50%

Blood type probabilities:

Out of 100 donors . . . . .

84 donors are RH+	16 donors are RH-
38 are O+	7 are O-
34 are A+	6 are A-
9 are B+	2 are B-
3 are AB+	1 is AB-

Source: AABB.ORG

What's the probability that a random donor will be AB?

3% + 1% = 4%

Statistical Independence

Formal definition: A and B are independent if and only if
P(A&B)=P(A)*P(B)

Joint Probability: The probability of two events happening simultaneously.

Marginal probability: This is the probability that an event happens at all, ignoring all other outcomes.

Is RH-status independent of blood type?

Out of 100 donors . . . . .

84 donors are RH+	16 donors are RH-
38 are O+	7 are O-
34 are A+	6 are A-
9 are B+	2 are B-
3 are AB+	1 is AB-

P(RH +) = 84%; P(O) = 45%; 45%*84% =37.8%
In fact, 38% are O+; so, yes, independent!

Independent ≠ mutually exclusive

- Events A and ~A are mutually exclusive, but they are NOT independent.

- P(A&~A)= 0

- P(A)*P(~A) ≠ 0

  Conceptually, once A has happened, ~A is impossible; thus, they are completely dependent.

Statistics for Health Care

Module 3:
Probability trees and conditional probability

Recall simple genetics example:

- In genetics, if both the mother and father carry one copy of a recessive disease-causing mutation (d), there are three possible outcomes:

  - P(genotype=DD)=.25
  - P(genotype=Dd)=.50
  - P(genotype=dd)=.25

Using a probability tree

What's the chance of having a heterozygote child (Dd) if both parents are heterozygote (Dd)?

Mother's allele   Father's allele   Child's outcome

$P(\male D=.5)$

$P(DD)=.5*.5=.25$

$P(\female D=.5)$

$P(\male d=.5)$

$P(Dd)=.5*.5=.25$

$P(\male D=.5)$

$P(dD)=.5*.5=.25$

50%!

$P(\female d=.5)$

$P(\male d=.5)$

$P(dd)=.5*.5=.25$

1.0
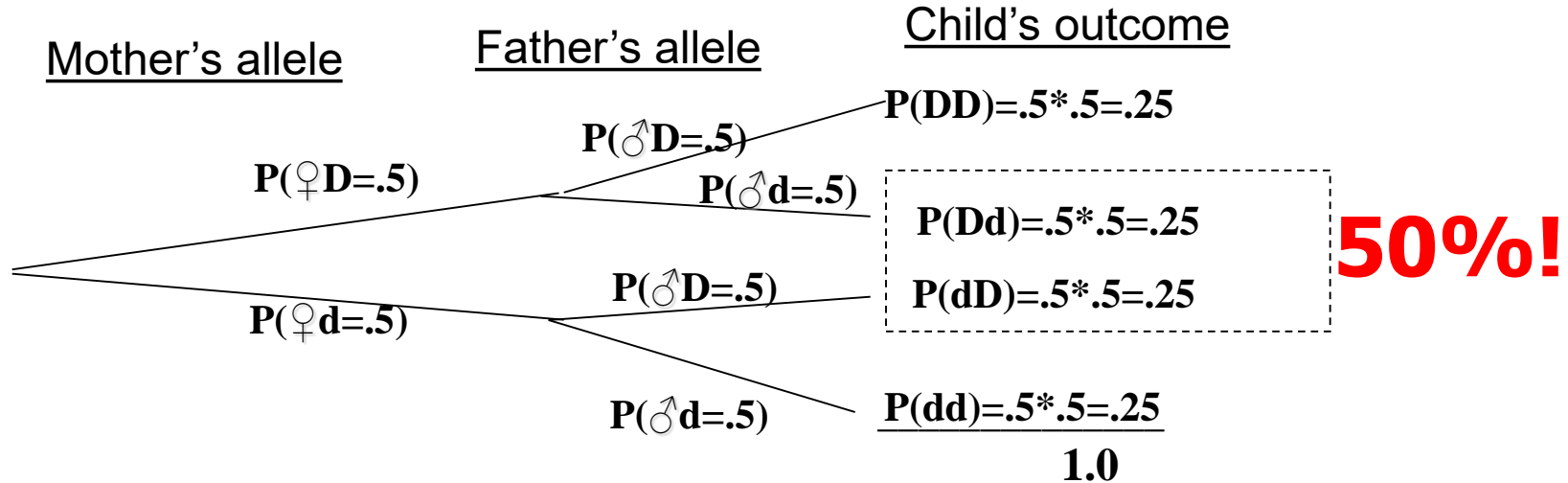
Recall: "and" means multiply; "or" means add

Independence

Formal definition: A and B are independent if and only if P(A&B)=P(A)*P(B)

The mother's and father's alleles are segregating independently.

P(♂D/♀D)=.5 and P(♂D/♀d)=.5

Conditional Probability: Read as "the probability that the father passes a D allele given that the mother passes a d allele."

Joint Probability: The probability of two events happening simultaneously.

What father's gamete looks like is not dependent on the mother's – doesn't depend which branch you start on

Marginal probability: This is the probability that an event happens at all, ignoring all other outcomes.

Formally, P(DD)=.25=P(D♂)*P(D♀)

On the tree

Mom and dad are *independent* because the fathers' probabilities are the same regardless of the mother's outcome.

Mother's allele

Father's allele

Child's outcome

P(♀D=.5)

P(♂D)=.5

P(DD)=.5*.5=.25

P(♂d=.5)

P(Dd)=.5*.5=.25

P(♀d=.5)

P(♂D=.5)

P(dD)=.5*.5=.25

P(♂d=.5)

P(dd)=.5*.5=.25

1.0
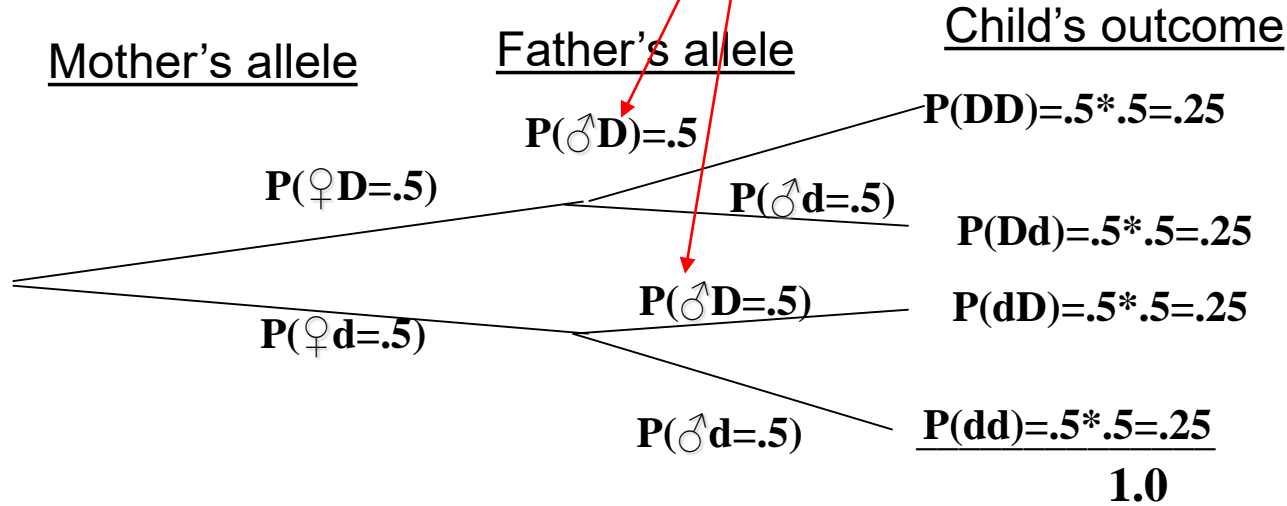
Conditional probabilities:

P(♂D/♀D)=.5

When A and B are independent, probability of A does not depend on B:

P(♂D/♀D)=.5

P(♂D/♀d)=.5

P(♂D)=.5

On the tree

Mother's allele

Father's allele

Child's outcome

$P(\text{♂}D/\text{♀}D)=.5$

$P(\text{♀}D=.5)$

$P(DD)=.5*.5=.25$

$P(\text{♂}d=.5)$

$P(Dd)=.5*.5=.25$

$P(\text{♂}D=.5)$

$P(dD)=.5*.5=.25$

$P(\text{♀}d=.5)$

$P(\text{♂}d=.5)$

$\underline{P(dd)=.5*.5=.25}$

1.0
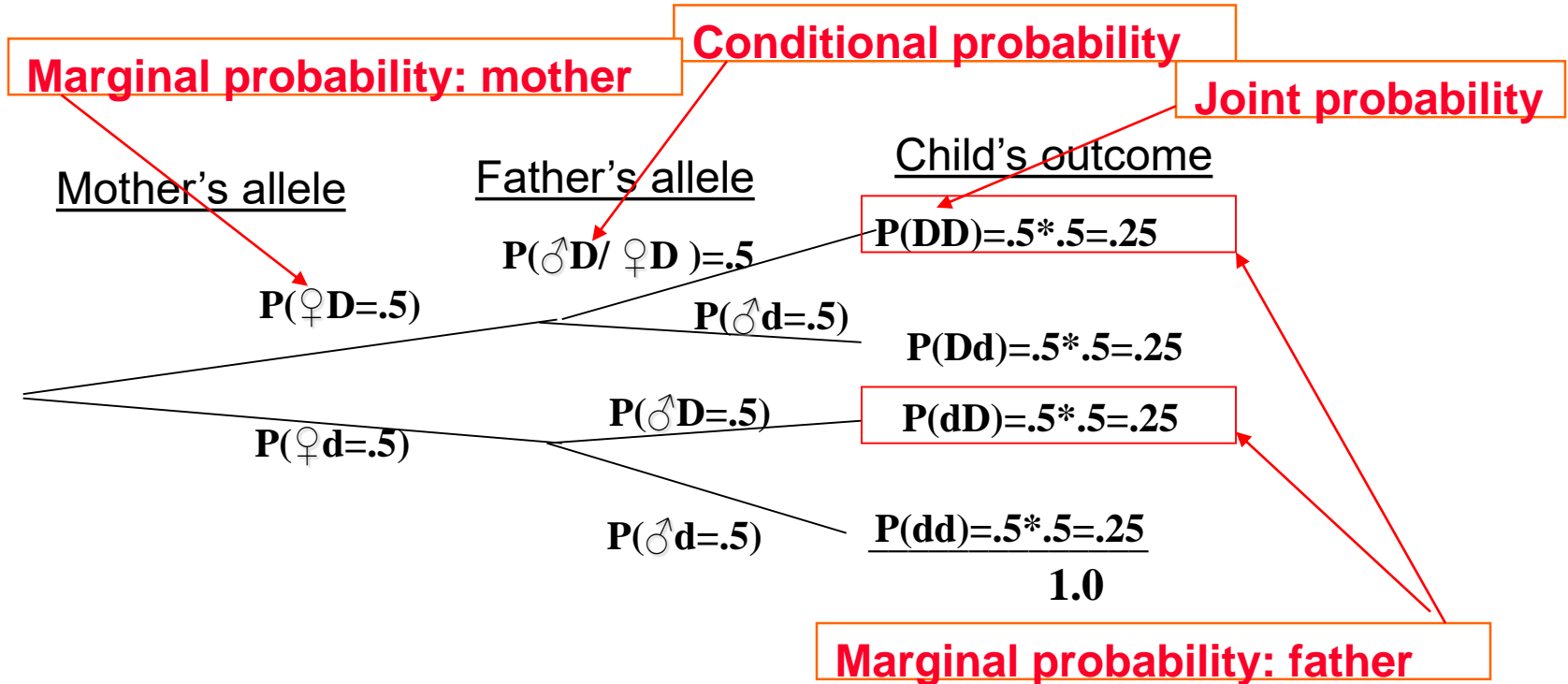
Example with *dependent* events

If HIV has a prevalence of 3% in San Francisco, and a particular HIV test has a false positive rate of .001 and a false negative rate of .01, what is the probability that a random person selected off the street will test positive?

Joint probability of being + and testing +

Marginal probability of carrying the virus.

P(test +/+)=.99   P (+, test +)=.0297

P(+)=.03

P(test - /+)= .01

P(+, test -)=.003

P(test +/-) = .001   P(-, test +)=.00097

P(-)=.97

P(test -/-) = .999   P(-, test -) = .96903
                                              1.0

$\therefore P(\text{test } +)=.0297+.00097=.03067$

Marginal probability of testing positive
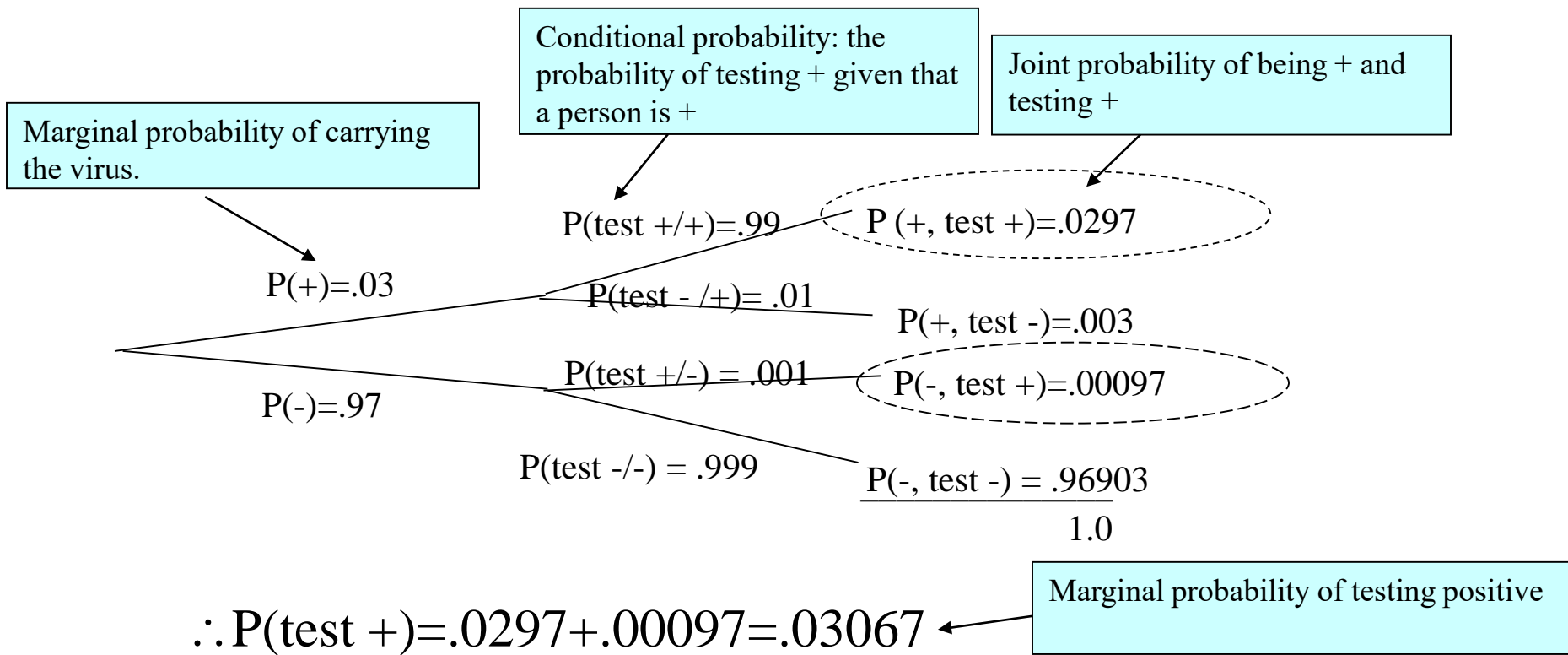
Formal evaluation of independence: P(+&test+)=.0297

P(+)*P(test+) = 03*.03067=.00092 $\therefore$ Dependent!

"Law of total probability"

$$P(\text{test} +) = P(\text{test} + /HIV+)P(HIV+) + P(\text{test} + /HIV-)P(HIV\text{-})$$

One of these has to be true (mutually exclusive, collectively exhaustive). They sum to 1.0.

$$P(\text{test} +) = .99(.03) + .001(.97)$$

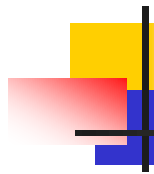$$P(B) = P(B / A)P(A) + P(B / \sim A)P(\sim A)$$

Example

- A 54-year old woman has an abnormal mammogram; what is the chance that she has breast cancer?
  - Sensitivity = 90%
  - Specificity = 89%
  - Prevalence of BC in 54-year old women = .3%

		Condition (as determined by "gold standard") Biopsy		P(true+/test+)
		Condition positive	Condition negative	
Test outcome Mammogram	Test outcome positive	True positive	False positive (Type I error)	Positive predictive value = Σ True positive/ Σ Test outcome positive
	Test outcome negative	False negative (Type II error)	True negative	Negative predictive value = Σ True negative/ Σ Test outcome negative
		Sensitivity = Σ True positive/ Σ Condition positive	Specificity = Σ True negative/ Σ Condition negative	P(true-/test-)

True positive rate
P(test+/true+)

True negative rate
P(test-/true-)

False negative rate = 1- True positive rate

False positive rate = 1- True negative rate

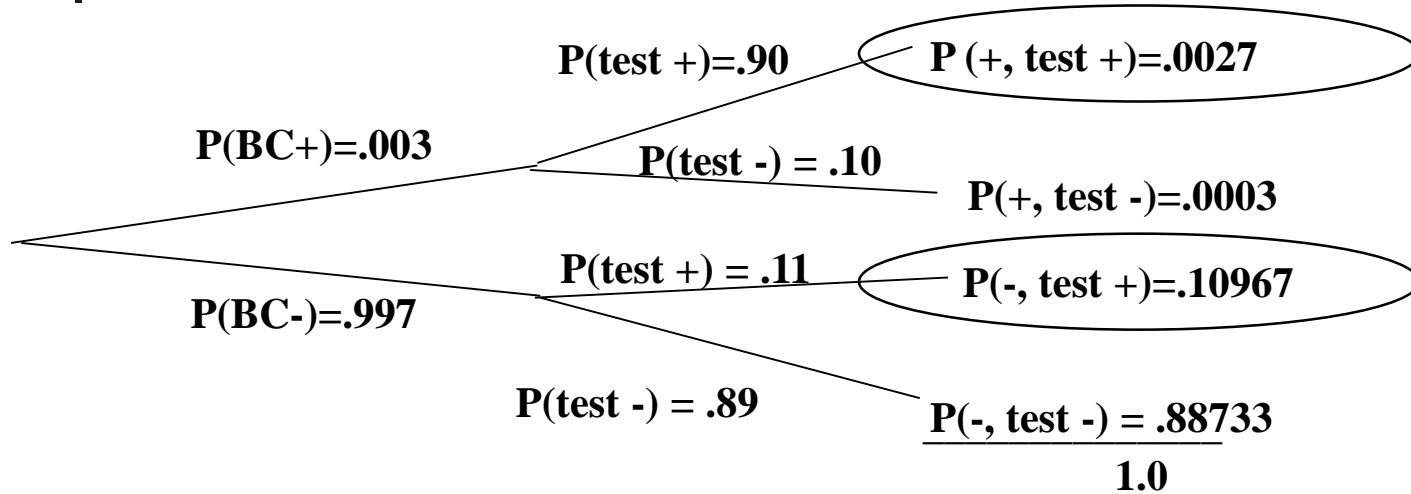|		Condition (as determined by "Gold standard")		
		Positive	Negative	
Test outcome	Positive	True Positive	False Positive (Type I error, P-value)	→ Positive predictive value
	Negative	False Negative (Type II error)	True Negative	→ Negative predictive value
		↓ Sensitivity	↓ Specificity	

		Patients with bowel cancer (as confirmed on endoscopy)		
		Positive	Negative	?
FOB test	Positive	TP = 2	FP = 18	= TP / (TP + FP) = 2 / (2 + 18) = 2 / 20 ≡ **10%**
	Negative	FN = 1	TN = 182	= TN / (TN + FN) 182 / (1 + 182) = 182 / 183 ≡ **99.5%**
		↓ = TP / (TP + FN) = 2 / (2 + 1) = 2 / 3 ≡ **66.67%**	↓ = TN / (FP + TN) = 182 / (18 + 182) = 182 / 200 ≡ **91%**	

Example: Mammography

P(test +)=.90   P (+, test +)=.0027

P(BC+)=.003

P(test -) = .10   P(+, test -)=.0003

P(test +) = .11   P(-, test +)=.10967

P(BC-)=.997

P(test -) = .89   P(-, test -) = .88733

                 1.0

P(BC/test+)=.0027/(.0027+.10967)=2.4%

Statistics for Health Care

Module 4:

Bayes' Rule

Bayes' Rule: derivation

- Definition:

Let A and B be two events with P(B) ≠ 0. The conditional probability of A given B is:

$$P(A/B) = \frac{P(A \& B)}{P(B)}$$

Bayes' Rule: derivation

$$P(A/B) = \frac{P(A \& B)}{P(B)}$$

can be re-arranged to:

$$P(A \& B) = P(A/B)P(B)$$

and, since also:

$$P(B/A) = \frac{P(A \& B)}{P(A)} \qquad \therefore P(A \& B) = P(B/A)P(A)$$

$$P(A/B)P(B) = P(A \& B) = P(B/A)P(A)$$

$$\therefore P(A/B) = \frac{P(B/A)P(A)}{P(B)}$$

Bayes' Rule:

$$P(A \, / \, B) = \frac{P(B \, / \, A)P(A)}{P(B)}$$

OR

$$P(A \, / \, B) = \frac{P(B \, / \, A)P(A)}{P(B \, / \, A)P(A) + P(B \, / \sim A)P(\sim A)}$$

From the "Law of Total Probability"

Bayes' Rule:

- Why do we care??

- Why is Bayes' Rule useful??

- It turns out that sometimes it is very useful to be able to "flip" conditional probabilities. That is, we may know the probability of A given B, but the probability of B given A may not be obvious. An example will help…
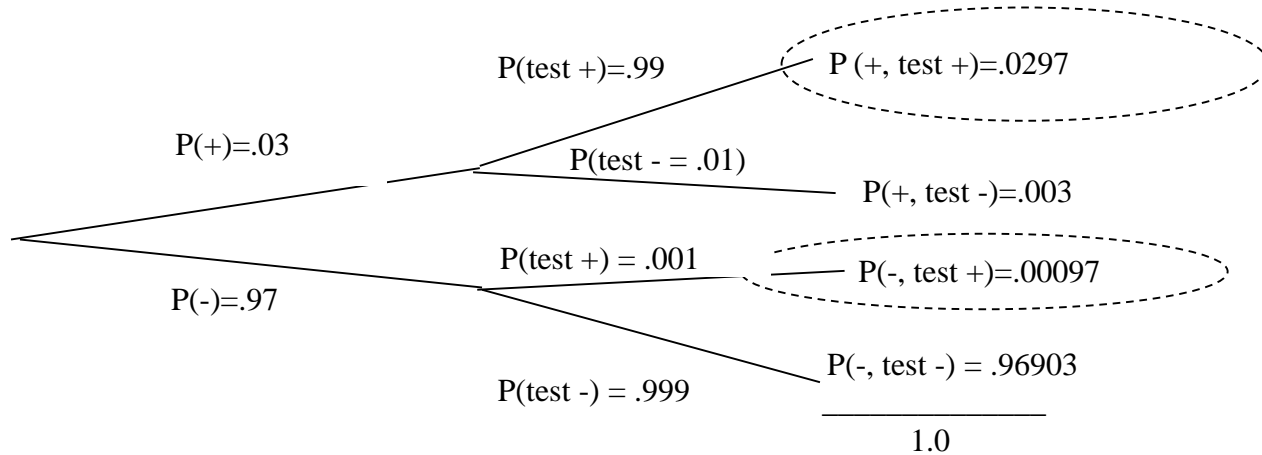
Positive predictive value

If HIV has a prevalence of 3% in San Francisco, and a particular HIV test has a false positive rate of .001 and a false negative rate of .01, what is the probability that a random person who tests positive is actually infected (also known as "positive predictive value")?

Answer: using probability tree

P(test +)=.99   P (+, test +)=.0297

P(+)=.03

P(test - = .01)   P(+, test -)=.003

P(test +) = .001   P(-, test +)=.00097

P(-)=.97

P(test -) = .999   P(-, test -) = .96903

1.0

$$P(+/test+) = \frac{.0297}{.0297 + .00097} = 96.8\%$$

Answer: using Bayes' rule

$$P(true + / test+) = \frac{P(test + / true+)P(true+)}{P(test+)}$$

$$= \frac{P(test + / true+)P(true+)}{P(test + / true+)P(true+) + P(test + / true-)P(true-)}$$

$$= \frac{.99(.03)}{.99(.03) + .001(.97)} = 96.8\%$$

		Condition (as determined by "gold standard")		
		Condition positive	Condition negative	P(true+/test+)
Test outcome	Test outcome positive	True positive	False positive (Type I error)	Positive predictive value = Σ True positive/ Σ Test outcome positive
	Test outcome negative	False negative (Type II error)	True negative	Negative predictive value = Σ True negative/ Σ Test outcome negative
		Sensitivity = Σ True positive/ Σ Condition positive	Specificity = Σ True negative/ Σ Condition negative	P(true-/test-)

P(true+/test+)

P(true-/test-)

True positive rate
P(test+/true+)

True negative rate
P(test-/true-)

False negative rate = 1- True positive rate   False positive rate = 1- True negative rate

Practice problem

- Suppose that 15% of the country was exposed to a dangerous chemical Z due to a particular corporation's negligence. You are a lawyer who is suing the corporation on behalf of a pancreatic cancer patient. You want to provide some evidence that your client's cancer was likely to be due, at least in part, to exposure to Z. However, your client has no idea if she was exposed or not. If exposure to Z quadruples one's lifetime risk of pancreatic cancer, from .0001 to .0004 (RR=4.0), what's the probability that your client was exposed to Z?

- We have: P(cancer/exposure)
- We want: P(exposure/cancer)

- Use Bayes' Rule!

Answer, using Bayes' Rule

$$P(Z/PC) = \frac{P(PC/Z)P(Z)}{P(PC)}$$

P(*Z*) = .15
P(*PC*/*Z*) = .0004
P(*PC*/~*Z*) = .0001
P(*PC*) =P(*PC*/*Z*) P(*Z*) + P(*PC*/~*Z*)P(~*Z*) =.0004(.15) + .0001(.85) = .000145

$$P(Z/PC) = \frac{P(PC/Z)P(Z)}{P(PC)} = \frac{.0004(.15)}{.000145} = 41.4\%$$

Answer, using a tree

P(*Z*) = .15
P(*PC/Z*) = .0004
P(*PC/~Z*) = .0001

Use Bayes' rule to "update" probabilities

- A box has three coins. One has two heads, one has two tails, and the other is a fair coin with one head and one tail.
  - A coin is chosen at random, is flipped, and comes up heads. What is the probability that the coin chosen is the two-headed coin?
  - Suppose that the coin is thrown a second time and comes up heads again. What is the probability that the chosen coin is the two-headed coin?

Use Bayes' rule to "update" probabilities

- Prior P (two-headed coin) = 1/3

- After flipping the coin and obtaining a head, how does this probability change?

Use Bayes' rule to "update" probabilities

- Updated P (two-headed coin) = 2/3
- After flipping the coin again and obtaining a head, how does this probability change?

Thought problem…

- You are on the Monty Hall show. You are presented with 3 doors (A, B, C), one of which has a valuable prize behind it (the other two have nothing). You choose door A; Monty Hall opens door B and shows you that there is nothing behind it. Then he gives you the choice of sticking with A or switching to C. Do you stay or switch; or does it make no difference?

Try this link:

- http://www.nytimes.com/2008/04/08/science/08monty.html#

Statistics for Health Care

Conditional probability, Bayes' rule, and the odds ratio

The Risk Ratio

	Exposure (E)	No Exposure (~E)
Disease (D)	a	b
No Disease (~D)	c	d
	a+c	b+d

$$RR = \frac{P(D/E)}{P(D/\sim E)} = \frac{a/(a+c)}{b/(b+d)}$$

risk to the exposed

risk to the unexposed

Hypothetical Data

	High Systolic BP	Normal BP
Congestive Heart Failure	400	400
No CHF	1100	2600
	1500	3000

$$RR = \frac{400/1500}{400/3000} = 2.0$$

Case-control study example:

- You sample 50 stroke patients and 50 controls without stroke and ask about their smoking in the past.

Hypothetical results:

	Smoker (E)	Non-smoker (~E)	
Stroke (D)	**15**	**35**	50
No Stroke (~D)	**8**	**42**	50

What's the risk ratio here?

	Smoker (E)	Non-smoker (~E)	
Stroke (D)	15	35	50
No Stroke (~D)	8	42	50

Tricky: There is no risk ratio, because we cannot calculate the risk (or rate) of disease!!

The odds ratio…

- <u>We cannot calculate a risk ratio from a case-control study.</u>

- BUT, we can calculate an <u>odds ratio</u>…

The Odds Ratio (OR)

	Smoker (E)	Smoker (~E)	
Stroke (D)	15	35	50
No Stroke (~D)	8	42	50

These data give: P(E/D) and P(E/~D).

Luckily, you can flip the conditional probabilities using Bayes' Rule:

$$P(D/E) = \frac{P(E/D)P(D)}{P(E)}$$

Unfortunately, our sampling scheme precludes calculation of the marginals: P(E) and P(D), but turns out we don't need these if we use an odds ratio because the marginals cancel out!

The Odds Ratio (OR)

	Exposure (E)	No Exposure (~E)
Disease (D)	a	b
No Disease (~D)	c	d

Odds of exposure in the cases

Odds of exposure in the controls

$$OR = \dfrac{\dfrac{P(E/D)}{P(\sim E/D)}}{\dfrac{P(E/\sim D)}{P(\sim E/\sim D)}} = \dfrac{\dfrac{a}{b}}{\dfrac{c}{d}} = \dfrac{ad}{bc}$$

The Odds Ratio (OR)

Odds of exposure in the cases

Odds of disease in the exposed

$$OR = \dfrac{\dfrac{P(E/D)}{P(\sim E/D)}}{\dfrac{P(E/\sim D)}{P(\sim E/\sim D)}}$$

But, this expression is mathematically equivalent to:

$$\dfrac{\dfrac{P(D/E)}{P(\sim D/E)}}{\dfrac{P(D/\sim E)}{P(\sim D/\sim E)}}$$

Odds of exposure in the controls

Odds of disease in the unexposed

Backward from what we want…

The direction of interest!

Proof via Bayes' Rule

$$\frac{P(E/D)}{P(\sim E/D)}$$ Odds of exposure in the cases

$$\frac{P(E/\sim D)}{P(\sim E/\sim D)}$$ Odds of exposure in the controls

Bayes' Rule

$$\frac{\dfrac{P(D/E)P(E)}{P(D)}}{\dfrac{P(D/\sim E)P(\sim E)}{P(D)}}$$

$$\frac{\dfrac{P(\sim D/E)P(E)}{P(\sim D)}}{\dfrac{P(\sim D/\sim E)P(\sim E)}{P(\sim D)}}$$

$=$

$$\frac{P(D/E)}{P(\sim D/E)}$$ Odds of disease in the exposed

$$\frac{P(D/\sim E)}{P(\sim D/\sim E)}$$ Odds of disease in the unexposed

What we want!

The odds ratio here:

	Smoker (E)	Non-smoker (~E)	
Stroke (D)	15	35	50
No Stroke (~D)	8	42	50

$$OR = \frac{\dfrac{15}{35}}{\dfrac{8}{42}} = \frac{15 * 42}{35 * 8} = 2.25$$

■Interpretation: there is a 2.25-fold higher <u>odds</u> of stroke in smokers vs. non-smokers.

The rare disease assumption

$$OR = \frac{\dfrac{P(D \,/\, E)}{P(\sim D \,/\, E)}}{\dfrac{P(D \,/\, \sim E)}{P(\sim D \,/\, \sim E)}} \approx \frac{P(D \,/\, E)}{P(D \,/\, \sim E)} = RR$$

When a disease is rare:
P(~D) = 1 - P(D) $\cong$ 1

Statistics for Health Care

Module 5:

Diagnostic testing

Measures of diagnostic testing

- Sensitivity

- Specificity

- Positive predictive value (PPV)

- Negative predictive value (NPV)

Characteristics of a diagnostic test

Sensitivity= Probability that, *if you truly have the disease*, the diagnostic test will catch it.

Specificity=Probability that, *if you truly do not have the disease*, the test will register negative.
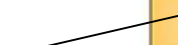
Positive or negative predictive value (PPV or NPV)

- PPV = the probability that if you test positive for the disease, you actually have the disease.

- NPV = the probability that if you test negative for a disease, you actually do not have the disease.

- Depends on the characteristics of the test (sensitivity, specificity) <u>and the prevalence of disease</u>.

Condition
(as determined by "**gold standard**")

		Condition positive	Condition negative	
Test outcome	Test outcome positive	**True positive**	**False positive** (Type I error)	**Positive predictive value** = Σ True positive/ Σ Test outcome positive
	Test outcome negative	**False negative** (Type II error)	**True negative**	**Negative predictive value** = Σ True negative/ Σ Test outcome negative
		Sensitivity = Σ True positive/ Σ Condition positive	**Specificity** = Σ True negative/ Σ Condition negative	

True positive rate

True negative rate

False negative rate = 1- True positive rate

False positive rate = 1- True negative rate

Fun example/bad investment

- http://www.cellulitedx.com/en-us/
  - "A patient who tests positive has approximately a 70% chance of developing moderate to severe cellulite." (PPV)
  - "A patient who tests negative has approximately a 50% chance of not developing moderate to severe cellulite." (NPV)

CelluliteDX Data

Researchers compared the ACE genotype of 200 women with moderate to severe cellulite (cases) and 200 cellulite-free controls.

They found that a significantly higher proportion of cases carried at least one D allele.

	Cellulite		
ACE genotype	**+**	**-**	
DD/DI	168	148	316
II	32	52	84
	200	200	

Sensitivity and Specificity

Cellulite

Cellulite DX Test

	+	-	
+	**168**	**148**	**316**
-	**32**	**52**	**84**
	200	**200**	

Sensitivity=168/200=.84

32 false negatives out of 200 cases

Specificity= 52/200 =.26

148 false positives out of 200 controls

What is the positive predictive value? Tricky!

Cellulite

Cellulite DX Test

	+	-	
+	168	148	316
-	32	52	84
	200	200	

PPV and NPV cannot be calculated directly because the prevalence of cellulite in the sample differs from the prevalence in the general population.

Use Bayes' rule!

Prevalence of disease in the population = 65% (according to the company).

Or use a tree:

Alternatively, calculate with a simple 2x2 table:

Cellulite

Cellulite DX Test

	+	-	
+			
-			
	65	35	100

To solve for PPV and NPV, must incorporate the true prevalence. Trick: start with 100 people…

PPV and NPV calculation

Cellulite

Cellulite DX Test

	+	-	
+	55		
-		9	
	65	35	100

Then apply sensitivity and specificity….

True positives=65*.84= 55

True negatives = 35*.26= 9

PPV and NPV calculation

Cellulite

Cellulite DX Test

	+	-	
+	55	26	81
-	10	9	19
	65	35	100

Then apply sensitivity and specificity….

True positives=65*.84= 55

True negatives = 35*.26= 9

PPV and NPV calculation

Cellulite

Cellulite DX Test

	+	-	
+	55	26	81
-	10	9	19
	65	35	100

PPV=55/81=68%

68% of those who test positive will get cellulite.

NPV=9/19=47%

47% of those who test negative will NOT get cellulite. I.e, 53% will get cellulite.

Conclusions

- A patient who does not take the test has a 65% chance of getting moderate to severe cellulite (=the prevalence of the disease in the general population)

- Four out of five women will test positive (81%). A positive test moves a woman's risk from 65% to 68%.

- One out of five women will test negative (19%). A negative tests moves a woman's risk from 65% to 53%.

- Conclusion: the test is basically useless.

Conclusions

- Sensitivity and specificity do not depend on the prevalence of disease.

- PPV and NPV (what patients care about!) depend on the prevalence of disease.

- Just because a risk factor is significantly associated with a disease does not make it a useful screening test!