# Biost 517 / Biost 514 Applied Biostatistics I / Biostatistics I

Timothy A. Thornton, Ph.D.
Associate Professor of Biostatistics
University of Washington

## Lecture 8:

## Rules of Probability; Conditional Probability; Bayes Theorem; Diagnostic Tests: Sensitivity, Specificity, PPV, NPV

# Lecture Topics

- Rules of Probability
- Bayes Theorem
- Diagnostic Testing
  - Sensitivity
  - Specificity
  - Positive Predictive Value
  - Negative Predictive Value

# Probability

- Biostatisticians mostly use the long-run frequency definition of probability:
  - If we do lots of coin tosses, half of them will come up heads.
  - The long-run frequency of heads is ½, so P(heads)=1/2.
- There are other definitions of probability…
  - subjective probability
  - definitions based on symmetry
- … We care most about the properties of probability.

# Example: Coin Toss

In ther early 1700's, French naturalist George Louis Leclerc Buffon observed 2,048 heads in 4,040 tosses.

Relative frequency $= \frac{2048}{4040} = 0.5069.$

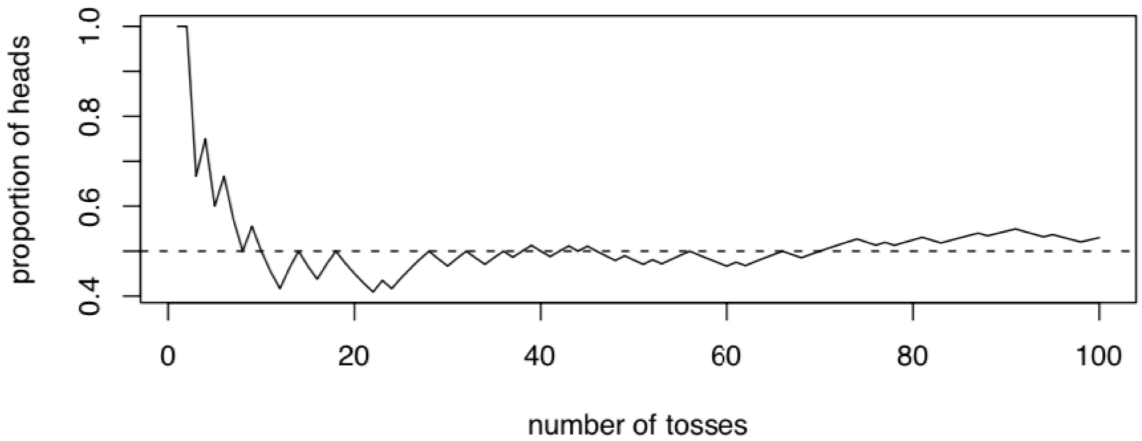
In the late 1800's, English statistician Karl Pearson observed 12,012 heads in 24,000 tosses.

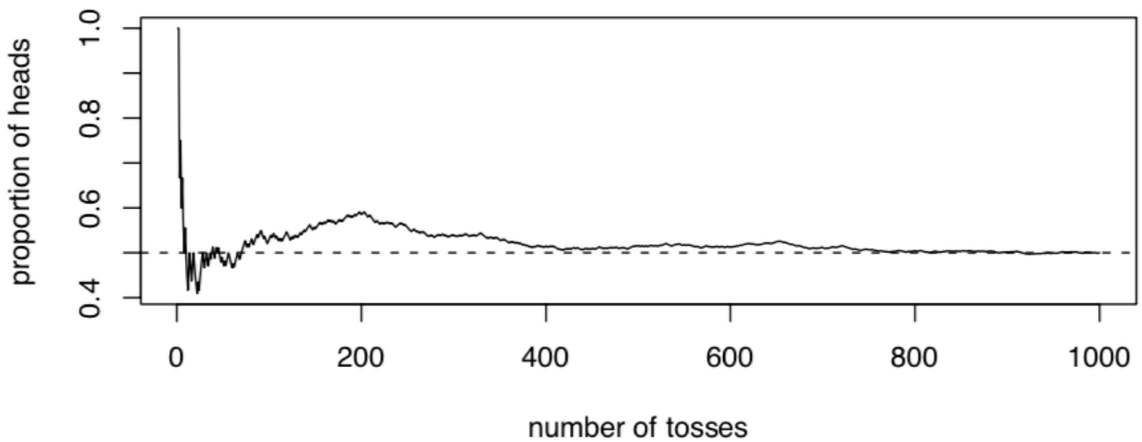
Relative frequency $= \frac{12012}{24000} = 0.5005.$

While in a prison camp in Denmark during World War I, English mathematician John Kerrich observed 5,067 heads in 10,000 tosses.

Relative frequency $= \frac{5067}{10000} = 0.5067.$

# Example: Coin Toss



# Randomness and Probability

Randomness $\neq$ Complete Chaos!

A phenomenon is said to be **random** if individual outcomes are uncertain but there is a regular distribution of outcomes in a large number of repetitions.

The **probability** of any outcome of a random phenomenon is the proportion of times the outcome would occur in a very long series of repetitions.

# Sample Space

The **sample space**, denoted by $S$, is the set of all possible outcomes of a random phenomenon.

- Toss a coin and record the side facing up. Then $S = \{\text{Head}, \text{Tail}\} = \{H, T\}$.
- Toss a coin twice. Record the side facing up each time. Then $S = ?$
- Toss a coin twice. Record the number of heads in the two tosses. Then $S = ?$

An **event** is an outcome or a set of outcomes of a random phenomenon (i.e. a subset of the sample space).

Toss a coin three times. Then
$S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$.

- Let $A$ be the event that we get exactly two tails. Then $A = ?$
- Let $B$ be the event that we get at least one head. Then $B = ?$

# Probability Model

A **probability model** is a mathematical description of a random phenomenon consisting of two parts: a sample space $S$ and a way of assigning probabilities to events.

The **probability** of an event $A$, denoted by $P(A)$, can be considered the long run relative frequency of the event $A$.

# Set Notation

Suppose $A$ and $B$ are events in the sample space $S$. Then,

- $(A \cup B) \equiv (A \text{ or } B) \equiv$
  the set of all outcomes in $A$, or in $B$, or in both
- $(A \cap B) \equiv (A \text{ and } B) \equiv$
  the set of all outcomes that are in $A$ AND in $B$
- $(A \cap B = \emptyset) \equiv A$ and $B$ are disjoint $\equiv$
  $A$ and $B$ are mutually exclusive $\equiv$
  $A$ and $B$ have no outcomes in common
- $A^c \equiv$ the complement of $A \equiv$
  the event that $A$ does not occur.

# Example: Probability of Events

Toss a coin twice. Let $A$ be the event that we get 2 heads, $B$ the event that we get exactly 1 tail, and $C$ the event that we get at least one head. So,

$$A = \{HH\} \qquad B = \{TH, HT\} \qquad C = \{HH, HT, TH\}$$

- $A^c = ?$
- $B^c = ?$
- $A \cap B = ?$
- $A \cup B = ?$
- $A \cap C = ?$
- $B \cup C = ?$

# Rules of Probability

1. For any event $A$, $0 \leq P(A) \leq 1$.

2. $P(S) = 1$.

3. For any event $A$, $P(A^c) = 1 - P(A)$.

4. If $P(A \cap B) = \emptyset$, then

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

More generally,

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

# Probabilities in a Finite Sample Space

If the sample space is finite, each distinct event is assigned a probability. The probability of an event is the sum of the probabilities of the distinct outcomes making up the event.

If a random phenomenon has $k$ equally likely outcomes, each individual outcome has probability $\frac{1}{k}$. For any event $A$,

$$P(A) = \frac{\text{number of outcomes in } A}{\text{number of outcomes in } S}$$

# Example: Probabilities in a Finite Sample Space

Roll a fair die and looking at the face value.

Sample space: $S = \{1, 2, 3, 4, 5, 6\}$

This is a finite sample space, and each outcome is equally likely. That is,

$$P(X = j) = 1/6, \ \forall j \in S$$

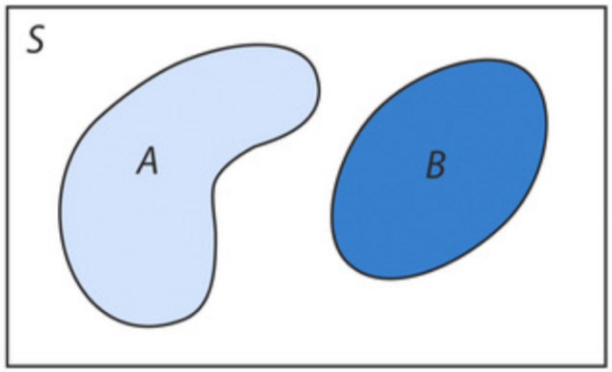
where $X$ is the face value of the die after rolling.

$P(X \geq 5) = P(X = 5) + P(X = 6) = 1/6 + 1/6 = 1/3$

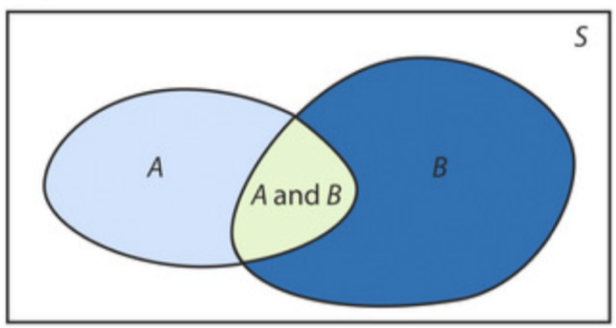
$P(X \leq 2) = ?$

# Venn Diagrams

A **Venn diagram** is a graphical representation of events in a sample space. The sample space $S$ is represented as the rectangle and the events are areas within $S$. Below is a Venn diagram with two disjoint events.




Below is a Venn diagram with events $A$ and $B$ overlapping.



# Independent Events and Disjoint Events

Two events $A$ and $B$ are **independent** if knowing that one occurs does not change the probability that the other occurs.

If $A$ and $B$ are independent, then

$$P(A \cap B) = P(A)\,P(B)$$

Events $A$ and $B$ are **disjoint** or **mutually exclusive** if they have no outcomes in common, so that

$$P(A \cap B) = \emptyset$$

If $A$ and $B$ are disjoint (or mutually exclusive), then

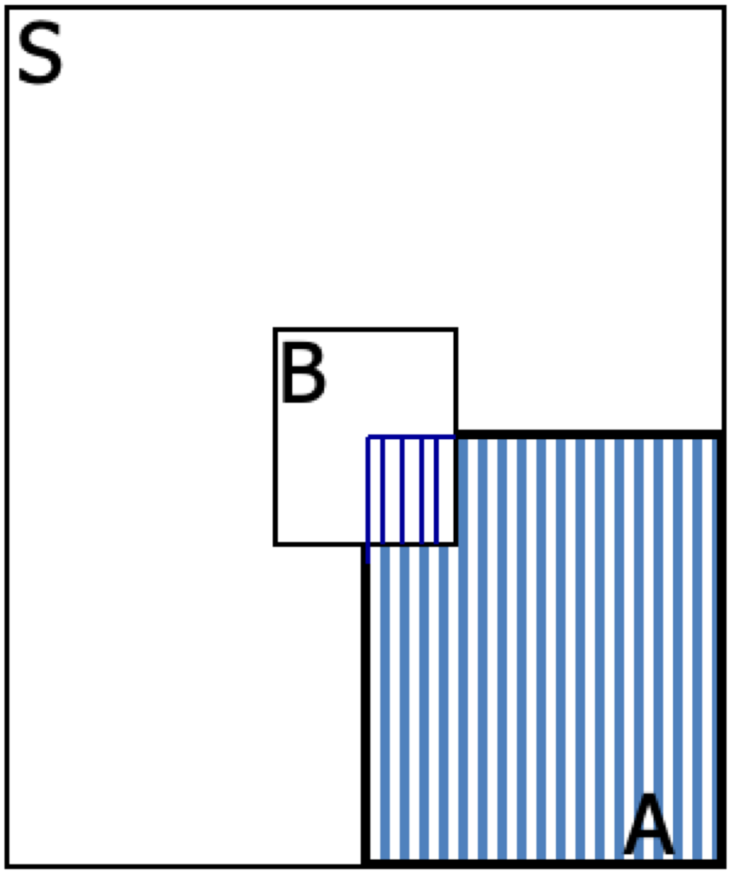
$$P(A \cup B) = P(A) + P(B)$$

# Caution: Independent vs. Mutually Exclusive

"Independent" and "mutually exclusive" sound similar in everyday language. But in probability they mean VERY different things.

If A and B are disjoint, then the fact A occurs tells us that B can not occur. So disjoint events are not independent.

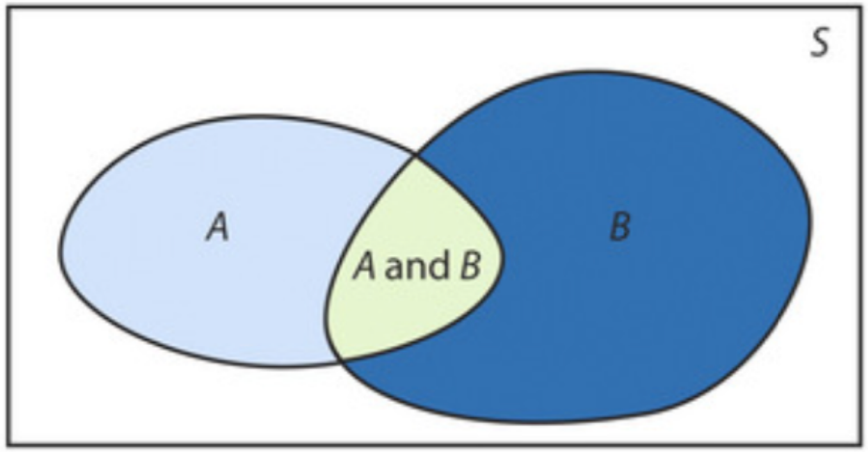
# Independent Events




- Independence is more difficult to visualize in a Venn diagram because it involves the probabilities of the events rather than just the outcomes that make up the events.
- Independence means that $P(A|B) = P(A)$, so in a Venn diagram, the area of $A$ in relation to the sample space $S$ is the same as the area of $A \cap B$ in relation to $B$

# Conditional Probability

**Conditional Probability:** $P(B|A)$




Idea of $P(B|A)$: Given that $A$ occurs, what is the probability that $B$ also occurs?

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

# Example: Deck of Cards

A deck of cards has 4 suits: $\heartsuit, \clubsuit, \spadesuit, \diamondsuit$. There are 13 cards in each suit: 2 through 10, jack, queen, king, and ace. So, there are $4 \times 13 = 52$ cards in the deck.

# Example: Conditional Probability

- A deck of cards is shuffled and the top two cards are put on a table, face down. You win \$10 if the second card is the $Q\heartsuit$.
  - What is your chance of winning?
  - You turn over the first card. It is $7\clubsuit$. Now, what is your chance of winning?
- Five cards are dealt off the top of a well-shuffled deck.
  - What is the probability that the third card is a $\spadesuit$?
  - You turn over the first four cards and none of them are a $\spadesuit$. What is the probability that the fifth card is a $\spadesuit$?

# Multiplication Rule

From the definition of conditional probability, we have

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

We obtain **The Multiplication Rule** from conditional probability:

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

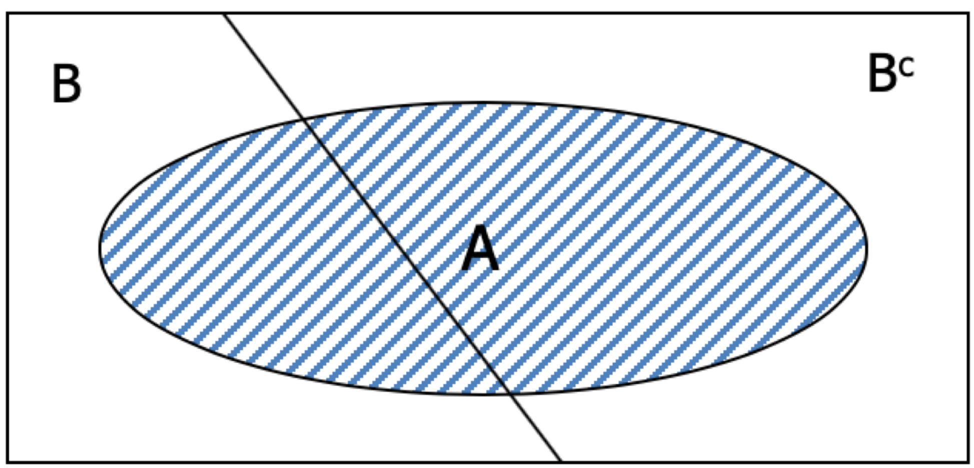
# Example: Multiplication Rule

Two cards are dealt off the top of a well-shuffled deck. What is the probability that the first card is a $\diamondsuit$ and the second card is a $\clubsuit$?

# Partitioning Events into Disjoint Events

- When calculating the probability of an event $A$, it is often helpful to partition $A$ based on disjoint events where the union of these events is the entire sample space $S$, e.g., $B$ and $B^c$
- Note that $A = (A \cap B) \cup (A \cap B^c)$. So we have the following:

$$P(A) = P(A \cap B) + P(A \cap B^c) = P(A|B)P(B) + P(A|B^c)P(B^c)$$



# Partitioning example

John has a disease that is known to be caused by three different bacteria. The only available treatment is 95% effective in eliminating the infection if John has the common A strain, 20% effective if John has subtype B, and never effective if John has a third, mutant subtype. Assume that 70% of infections are from the common A strain, a quarter of new infections are from subtype B, and the remainder of new infections are from the new mutant subtype.

John takes the treatment. What are the chances John is cured?

# Partitioning example

P(John is cured)=P(John is Cured and has Strain A)+ P(John is Cured and has Subtype B)+ P(John is Cured and has Mutant Subtype)

= P(John is Cured|Strain A)P(Strain A)+ P(John is Cured|Subtype B)P(Subtype B)+ P(John is Cured|Mutant Subtype)P(Mutant Subtype)

= (0.95)(0.70)+(0.20)(0.25)+(0)(0.05)=0.715

So the probability that John is cured is 0.715 (or 71.5% chance of being cured)

# Bayes' Theorem

- **Bayes' theorem/Bayes' rule**

Multiplication Rule

$$P(B \mid A) = \frac{P(A\, and\, B)}{P(A)} = \frac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid not\, B)P(not\, B)}$$

Definition of conditional probability

Partitioning

# Bayes' rule… Not just for Bayesians!

- "The term 'controversial theorem' sounds like an oxymoron, but Bayes' theorem has played this part for two-and-a-half centuries. Twice it has soared to scientific celebrity, twice it has crashed, and it is currently enjoying another boom." -Bradley Efron, *Science* 2013

# Diagnostic Testing: Sensitivity and Specificity

- We characterize the "accuracy" of a diagnostic test based on the *sensitivity* and *specificity* of the test.
- In this setting a person is either diseased (D) or healthy (H); the test is either postive (+) or negative (-)

- Sensitivity of test: Positivity in diseased
  – P( + | D )

- Specificity of test: Negativity in healthy (non-diseased)
  – P( - | H )

# Predictive Values

- Neither sensitivity nor specificity directly answers the patient's question: I got a positive test result. Am I diseased? Or: I got a negative test result. Am I healthy?

- Positive predictive value: Probability of disease when test is positive
  – P( D | + )

- Negative predictive value: Probability of healthy when test is negative
  – P( H | - )

# Computing Predictive Values

- Apply Bayes' Rule to relate PPV and NPV to Sensitivity and Specificity

$$\Pr\left(D \mid +\right) = \frac{\Pr\left(+ \mid D\right)\Pr\left(D\right)}{\Pr\left(+ \mid D\right)\Pr\left(D\right) + \Pr\left(+ \mid H\right)\Pr\left(H\right)}$$

$$\Pr\left(H \mid -\right) = \frac{\Pr\left(- \mid H\right)\Pr\left(H\right)}{\Pr\left(- \mid H\right)\Pr\left(H\right) + \Pr\left(- \mid D\right)\Pr\left(D\right)}$$

# PPV: Depends on Prevalence

- Need to know sensitivity, specificity, AND prevalence of disease

$$\Pr(D \mid +) = \frac{\Pr(+ \mid D)\Pr(D)}{\Pr(+ \mid D)\Pr(D) + \Pr(+ \mid H)\Pr(H)}$$

$$PPV = \frac{Sens \times Prev}{Sens \times Prev + (1 - Spec) \times (1 - Prev)}$$

# NPV: Depends on Prevalence

- Need to know sensitivity, specificity, AND prevalence of disease

$$\Pr(H \mid -) = \frac{\Pr(- \mid H)\Pr(H)}{\Pr(- \mid H)\Pr(H) + \Pr(- \mid D)\Pr(D)}$$

$$NPV = \frac{Spec \times (1 - Prev)}{Spec \times (1 - Prev) + (1 - Sens) \times Prev}$$

# Ex: Syphilis and VDRL

- Sensitivity of test: Probability of positive in diseased
  - 90% of subjects with syphilis test positive
  - (Actually depends on duration of infection)

- Specificity of test: Probability of negative in healthy
  - 98% of subjects without syphilis test negative
  - (Actually depends on age, and presence/absence of certain other diseases)

# Ex: PPV, NPV at STD Clinic

- Ex: 1000 patients at STD clinic
  - Prevalence of syphilis 30%
  - PPV: 95% with positive VDRL have syphilis (270/284)

| | | Syphilis | | |
|---|---|---|---|---|
| | | Yes | No | Tot |
| VDRL | Pos | 270 | 14 | 284 |
| | Neg | 30 | 686 | 716 |
| Total | | 300 | 700 | 1000 |

# Ex: PPV, NPV in Marriage License

- Ex: Screening for marriage license
  - Prevalence of syphilis 2%
  - PPV: 47% with positive VDRL have syphilis (18/38)

| | | Syphilis Yes | No | Tot |
|---|---|---|---|---|
| VDRL | Pos | 18 | 20 | 38 |
| | Neg | 2 | 960 | 962 |
| Total | | 20 | 980 | 1000 |

# Bottom Line

- The Positive and Negative Predictive Values are clinically important, but the Sensitivity and Specificity are more likely to be generalizable.
- Sensitivity and Specificity are properties of a test, whereas Predictive Values are properties of a test *in a population*.
- The relationship between (Sensitivity, Specificity) and (PPV, NPV) depends heavily on the prevalence.
- When a disease or condition  is rare, even a test with excellent sensitivity can have extremely poor PPV.

# Rare Disease

- Consider testing 10,000 people, ten of whom have the disease, using a test with 95% sensitivity and 95% specificity.
  - 95% sensitivity means that we expect 9 -10 of the people with the disease to test positive.
  - 95% specificity means that about 9500 of the 9990 people without the disease will test negative.
- So roughly 500 people test positive, roughly 10 of them have the disease, so PPV≈10/500=2%.