P8130: Biostatistical Methods I

Lecture 3: Basic Probability Concepts

Cody Chiuzan, PhD
Department of Biostatistics
Mailman School of Public Health (MSPH)

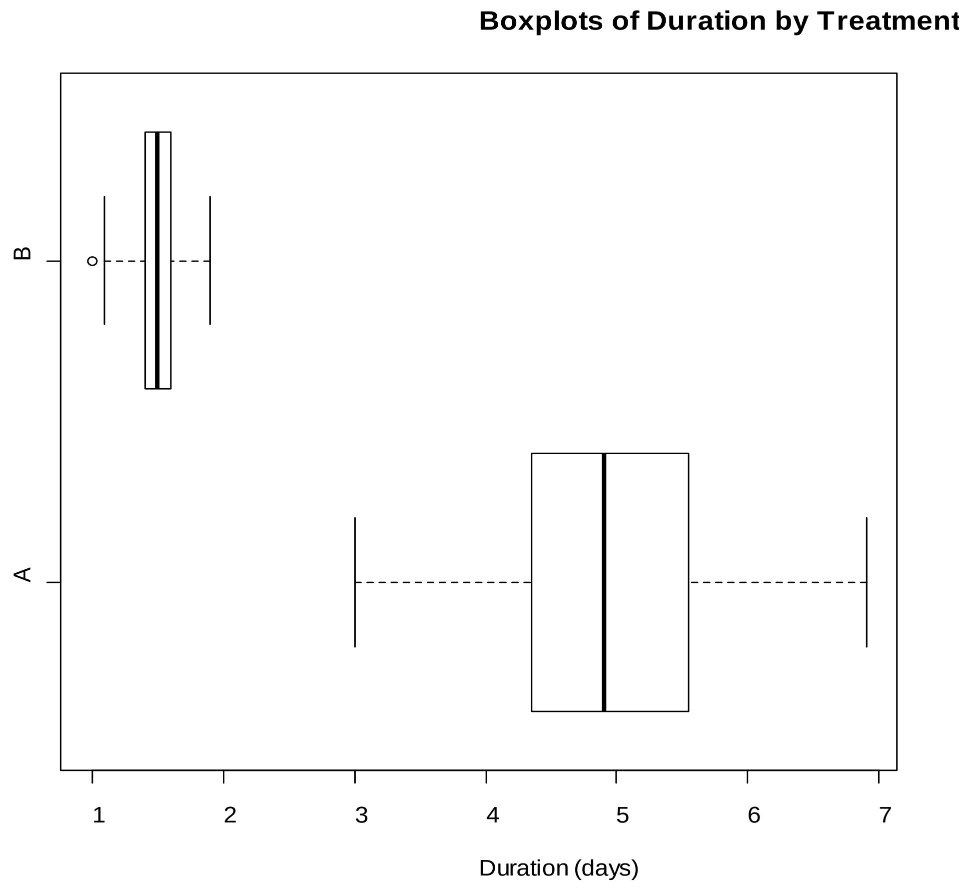
Lecture 2: Recap

- Descriptive statistics
  - Measures of location
  - Measures of dispersion
- Graphical displays
  - How to identify skewness?

Lecture 3: Outline

- Definitions and basic concepts
- Sets/Events/Rules
- Independent and conditional probability
- Bayes’ Theorem

Motivating Example



Setting: Two different treatments were randomly assigned to those diagnosed with the common cold. Left panel shows the box plots of duration of the cold for each treatment group.

Question: Which treatment worked better?

Answer: Treatment B seems to have a shorter cold duration, but…

How do we know if this difference is due to *chance alone* or because the treatments are *truly different*?

Probability Theory: Definitions

<u>Experiment:</u> A process which leads to a single outcome (or sample point) that cannot be predicted with certainty

<u>Sample Space (of an experiment):</u> The collection of all the possible outcomes (or sample points)

<u>Event:</u> An outcome or a set of outcomes from the sample space

<u>Probability of an event:</u> The relative frequency that an event occurs in a very large number of ‘trials’ (times it could have occurred)

> The probability of an event is a number between 0 and 1 that measures the likelihood that this event will occur when the experiment is performed.

Probability of Events: Examples

Example 1:

Disease Lung Cancer	Smoking Status	
	Never Smoker	Previous/Current Smoker
Yes	0.40	0.45
No	0.05	0.10

Probability of cancer with no smoking history =

Probability of cancer with smoking history =

Probability of no cancer with no smoking history =

Probability of no cancer with smoking history =

Sum of all probabilities: 0.40 + 0.45 + 0.05 + 0.10 = 1.00

Probability of Events: Examples

<u>Example 2:</u> Rolling one die (one time):

1. What is the sample space?

2. What is the probability of getting a 3?

3. What if we roll the die twice? What is the probability of getting no 3s in two succesive throws?

<u>Example 3:</u> Tossing two fair coins (one time):

1. What is the sample space?

2. What is the probability of getting a HH?

3. What is the probability of getting at least one head (H) in one toss?

Theoretical vs Empirical Probabilities

Motivating Example: In class, 4 out of 10 people are allergic to peanuts.

If you perform an experiment choosing 4 random students from this group, without replacement, what is the probability that they will all have peanuts allergy?

Let's first compute the **theoretical probability**:

P(1st pers. allergic) · P(2nd pers. allergic) · P(3rd pers. allergic) · P(4th pers. allergic) =

Theoretical vs Empirical Probabilities

What if it’s not feasible to conduct the experiment?

We need to simulate it and compute the **experimental (empirical) probability** instead.

<u>One way</u> is to conduct (generate) 10 trials for each person and record the the number of cases where all 4 have a positive response (i.e., allergic to peanuts):

1st pers.	Y	N	Y	N	Y	N	N	Y	N	N
2nd pers.	Y	N	N	Y	Y	N	N	Y	N	N
3rd pers.	Y	N	N	N	Y	N	N	Y	N	Y
4th pers.	N	Y	N	N	Y	Y	N	N	Y	N
All 4 pers.	**N**	**N**	**N**	**N**	**Y**	**N**	**N**	**N**	**N**	**N**

Empirically, there is a 1/10 =0.10 chance that all 4 subjects chosen at random out of 10 people will have peanuts allergies. What if we would perform 50 ‘trials’ for each person, would the empirical probability change?

Theoretical vs Empirical Probabilities

In general, the ***theoretical probability*** (assuming that each of the possible outcomes has an equal chance of occurring) is defined by:

$$P(\text{event}) = \frac{Number\ of\ sample\ points\ in\ the\ event\ of\ interest}{Total\ number\ of\ points\ in\ the\ sample\ space}$$

Theoretical probabilities cannot always be computed, and thus we use the empirical approximation instead.

- Note that experimental probabilities will differ; thus, more trials you perform, the closer the calculations will be for experimental and theoretical probabilities (>25 trials) – <u>Law of Large Numbers</u>

Operations on Events

Intersection ($A \cap B$): the event that both event A and event B occur when the experiment is conducted.

Union ($A \cup B$): the event that either event A or event B (or both) occurs when the experiment is conducted.

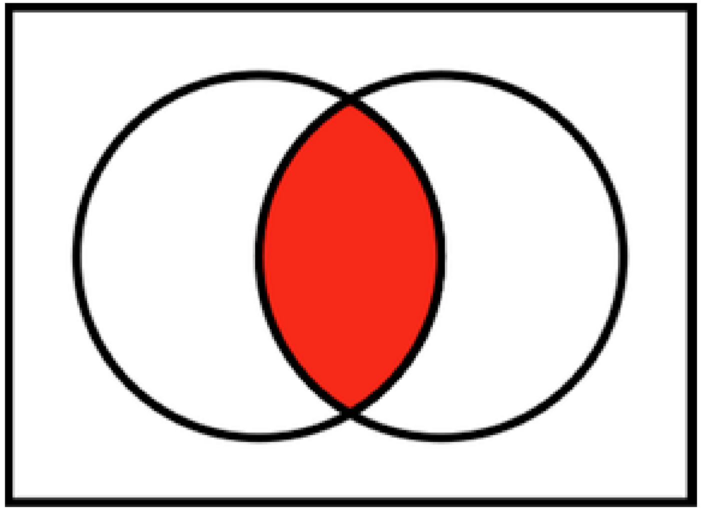
Complement (A^c or $\bar{A}$): the collection of outcomes that do not belong to event A.

$$P(A) + P(A^c) = 1$$

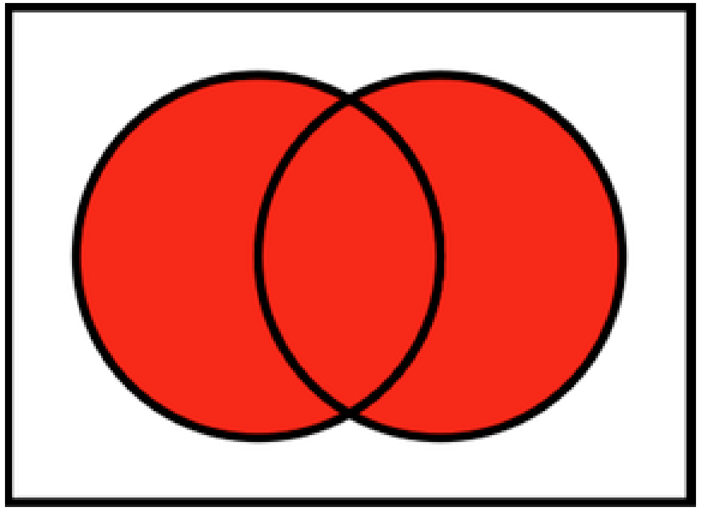
Additive Rule of Probability: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Operations on Events

<u>Venn Diagrams</u>: graphical representation of which sample points make up which events.



$A \cap B$



$A \cup B$

Draw the complement of event A.

Mutually Exclusive Events

Two (or more events) are mutually exclusive (disjoint) if the following is true:
If one event occurs in an experiment, the other event cannot occur.

Mutually exclusive events implies: $P(A \cap B) = 0$

Additive Rule of Probability becomes: $P(A \cup B) = P(A) + P(B)$

Examples:

Mutually Exclusive: Getting a head or getting a tail when tossing a coin

Non Mutually Exclusive: Listening to the lecture and checking Facebook (can happen simultaneously)

Conditional Probability

<u>Notation:</u> $P(A|B)$ denotes the probability that event A occurs given that event B has already occurred

Formula: $P(A|B) = \frac{P(A \cap B)}{P(B)}, given\ that\ P(B) \neq 0.$

Multiplicative Rule of Probability: $P(A \cap B) = P(B)P(A|B)$ or
$P(A \cap B) = P(A)P(B|A)$

Question: Is $P(A|B)$ equal to $P(B|A)$?

Independent Events

Two events are independent if the occurrence of event B does not affect the probability that event A occurs (and vice versa).

A and B are independent events <u>if and only if</u>:

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

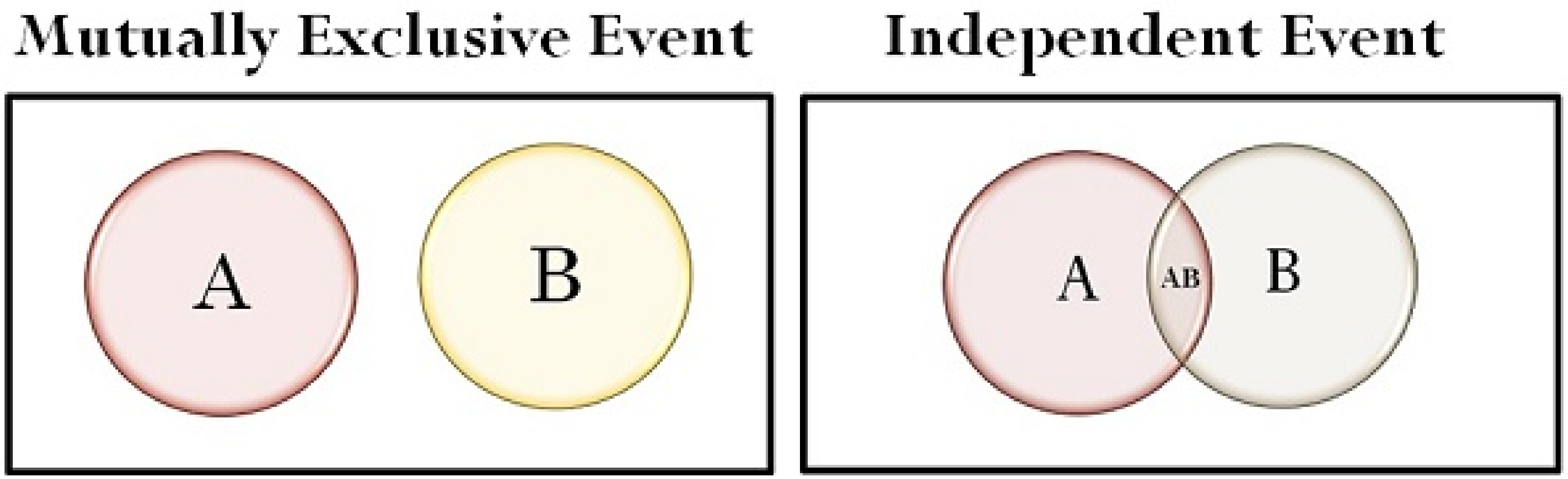
Two events that are not independent are said <u>to be dependent</u>.

Intersection of two independent events:

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

Why is this possible?

Mutually Exclusive vs Independent



Let’s consider the following situations:

1. Two events that are mutually exclusive and not independent

2. Two events that are independent and not mutually exclusive

3. Can two events that are mutually exclusive be independent too, and viceversa?

4. Can two events be non-mutually exclusive and dependent?

Law of Total Probability (LTP)

<u>Problem:</u> Sometimes you do not have direct information on the probability of an event (A)

<u>Solution:</u> Split the sample space into a set of disjoint (mutually exclusive) events B_i.

The unconditional probability of event A is given by:

$$P(A) = \textstyle\sum_{i=1}^{k} P(A \cap B_i)$$

OR

$$P(A) = \sum_{i=1}^{k} P(B_i)P(A|B_i)$$

<u>Idea:</u> Probability of event A is unknown, but you know its occurrence under different disjoint scenarios and the probabilities associated with each one of these scenarios.

Example: Law of Total Probability (LTP)

Find the total probability of developing glaucoma in the next 5 years.

Age group	Population proportion in the age group	Age specific probability to develop glaucoma
60-64	0.45	0.02
65-69	0.28	0.05
70-74	0.20	0.09
>74	0.07	0.15

P(glaucoma) = ?

Interpretation:

What if there are 5,000 subjects in our population; how many people you would expect to develop glaucoma?

Bayes' Theorem

Nothing more than the definition of conditional probabilities combined with thebLaw of Total Probability (LTP).

Let us assume that we have a partition of B_k events. We know $P(A|B_i)$ and we want to find the $P\left(B_j\middle|A\right)$ (flipped conditional probability). It follows that:

$$P\left(B_j\middle|A\right) = \frac{P(B_j \cap A)}{P(A)} = \frac{P\left(A\middle|B_j\right)P(B_j)}{P(A)}$$

Now, we use the LTP to expand $P(A)$ in the denominator and obtain:

$$P\left(B_j\middle|A\right) = \frac{P\left(A\middle|B_j\right)P(B_j)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots P(A|B_k)P(B_k)}$$

Example: Bayes' Theorem

A new biomarker is being tested for prostate cancer.

Let D - probability of cancer positive and T - probability of test positive, with the following given probabilities:

Sensitivity: $P(T|D) = 0.89$

Specificity: $P(T^c|D^c) = 0.75$

Prevalence: $P(D) = 0.01$

Calculate:

1. The positive predictive value (PPV), i.e., the probability that someone who tests positive actually has the disease:

$P(D|T) =?$

Example

2. The negative predictive value (NPV), i.e., probability that someone who tests negative using this new biomarker does not have prostate cancer:

$P(D^c|T^c) = ?$

Readings

Rosner, *Fundamentals of Biostatistics*, Chapter 3

- Sections: 3.1 – 3.7
- Sections: 3.9 (covers ROC curve); read for next time!