Chapter 2 - Probability and Distributions

Read sections 2.1, 2.2.1 - 2.2.4, 2.3, 2.5

In Chapter 1 we reviewed how to sample; how to conduct an experiment; and how to describe the resulting data using numerical and graphical summaries. In the rest if the course, we turn our attention to **Inferential Statistics**, the process of taking results about a sample and generalizing them to the population from which the sample was taken.

We already know that we can infer things about a population from **random samples**.

In order to fully understand inferential statistics, we need the language of probability, which is the topic of this chapter.

Probability

(2.1, 2.2.1 - 2.2.4)

Terminology:

1. **Outcome** - one possible value of a variable
2. **Sample Space** - the set of all possible outcomes of a variable.
3. **Event** - a group of one or more outcomes; An event 'occurs' when one of the outcomes in the event occurs.

EXAMPLES:

- Flip a coin once and measure the face of the coin. The sample space of all possible outcomes is

$$S = \{\_\_\_\_\_\}.$$

  The event that a head occurs is

$$A = \{\_\_\_\_\_\}.$$

- Consider drawing a single card from a deck of 52 playing cards and recording the rank and suit of the card. The sample space of all 52 possible outcomes is

$$S = \{\_\_\_\_\_\}.$$

  The event that the card is a heart is

$$A = \{\_\_\_\_\_\}.$$

- Flip a coin twice. The sample space of all possible outcomes is

$$S = \{\_\_\_\_\_\}.$$

  The event that at least one head occurs is

$$A = \{\_\_\_\_\_\}.$$

- Count the number of buffalo observed at a certain location in YNP. The sample space of all possible outcomes is

$$S = \{\_\_\_\_\_\}.$$

  The event that more than ten buffalo are observed is

$$A = \{\_\_\_\_\_\}.$$

- Measure the time it takes for some randomly chosen human being to run a mile. The sample space of all outcomes is

$$S = [\_\_\_\_\_).$$

  Consider the event that the mile is completed under 5 minutes,

$$S = [\_\_\_\_\_).$$

LAW OF LARGE NUMBERS:

Probability is a "long-term" relative frequency or proportion. The probability of an event A, written $P(A)$, is the proportion of times an outcome in the event occurs in many (i.e., an infinite number) of independent and identical trials.

QUESTIONS:

1. Toss a fair coin. How do you know $P(\text{Head})=0.5$?

Toss	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	etc.
Proportion of heads						

2. Net worth is defined as the current value of one's assets less liabilities. If the threshold level of being wealthy is having a net worth of \$1 million or more, then 3.5 million of the 100 million households in America are considered wealthy (from "The Millionaire Next Door: The Surprising Secrets of American's Wealthy", *The New York Times on the Web*, 1996). Randomly draw a household from the US. How do we know that $P(\text{getting millionaire}) = .035$?

The Rules of Probability

1. $\boxed{\mathbf{0 \leq P(A) \leq 1}}$

   How often an event A occurs must be somewhere between "never" (probability = 0) and "always" (probability =1).

2. **Addition Rule for Disjoint Events** $\longrightarrow$ $\boxed{\mathbf{P(A \text{ or } B) = P(A) + P(B)}}$

   - If the events A and B are NOT *disjoint* then $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$.
   - Two events are **disjoint** (or **mutually exclusive**) if they cannot possibly occur simultaneously.

     EXAMPLE: Suppose we are drawing a single card from a deck. Let $A = \{2\heartsuit, ...,\text{Ace } \heartsuit\}$ and $B = \{2\diamondsuit, ...,\text{Ace } \diamondsuit\}$ Then A and B are disjoint and the probability of getting a heart or a diamond on a draw from a deck of 52 cards is

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

     .

3. **Complement Rule** $\longrightarrow$ $\boxed{P(A^c) = 1 - P(A)}$

   - The notation "A^c" is read the "complement of A".
   - A^c contains all of the outcomes not in A.
   - Either A occurred or A did not occur and they are disjoint, so their probabilities must sum to 1. $P(A) + P(A^c) = 1$.

     EXAMPLE: Suppose we are drawing a single card from a deck. Let $A = \{2\heartsuit, ...,\text{Ace } \heartsuit\}$. Then $A^c = \{2\clubsuit, 3\clubsuit, \text{ ... },\text{Ace } \clubsuit, 2\diamondsuit, 3\diamondsuit, \text{ ... }, \text{ Ace } \diamondsuit, 2\spadesuit, 3\spadesuit, \text{ ... }, \text{ Ace } \spadesuit\}$.

     $$P(A) = \_\_\_\_\_$$
     $$P(A^c) = \_\_\_\_\_$$

4. **Multiplication Rule** $\longrightarrow$ $\boxed{\mathbf{P(A \text{ and } B) = P(A|B) \times P(B)}}$

   - $P(A|B)$ is the probability of A occurring given you know B has occurred.

   EXAMPLES:

(a) Suppose we are drawing a single card from a deck. Let $A = \{Q\spadesuit\}$ and $B = \{2\spadesuit, ...,$Ace $\spadesuit\}$.

$$P(A) = \_\_\_\_\_$$

$$P(B) = \_\_\_\_\_$$

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

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

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

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

(b) Suppose that there are 5 wealthy households (having a net worth of over a million dollars) out of the 21 households in some rural county in Montana. From the tax rolls, randomly choose two households from this county. Let B be the event that the first household is wealthy. Let A be the event that the second household is wealthy.

$$P(B) = \_\_\_\_\_$$

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

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

(c) On gradation day at a large university, one graduate is selected at random. Let A represent the event that the student is an engineering major, and let B represent the event that the student took a calculus course in college. Which probability is greater, $P(A|B)$ or $P(B|A)$?

- The events A and B are **independent** if the knowledge that one of the events occurred (or did not occur) does not change the probability of the other event occurring (or not occurring),

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

- **Multiplication Rule for Independent Events** $\longrightarrow$ $\boxed{P(A \text{ and } B) = P(A) \times P(B)}$ if A and B are **independent**.

EXAMPLES:

(a) Toss a coin twice.

i. Let B be the event of a head on the first toss. Let A be the event of a head on the second toss.

$P(A) =$ _____

$P(B) =$ _____

$P(A|B) =$ _____

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

ii. Are the coin tosses independent?

(b) Suppose that there are 3.5 million millionaire households out of the 100 million households in the US. From the rax rolls, randomly choose 2 households from the US. Let B be the event that the first household is a millionaire. Let A be the event that the second household is a millionaire.

i.

$P(B) =$ _____

$P(A|B) =$ _____

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

ii. Are A and B independent?

iii. When choosing two households from the total of 21 households in some Montanan county, are A and B independent?

An Example using all of the rules of probability:

Suppose that the sample space of human blood types is $S = \{O, A, B, AB\}$. For the American population, the following table gives the probabilities for each:

American blood type distribution:

Blood type	O	A	B	AB
U.S. probability	0.45	0.40	0.11	?

Assume that the following table gives the human blood type probabilities for the population in China:

Chinese blood type distribution:

Blood type	O	A	B	AB
China probability	0.35	0.27	0.26	0.12

1. In the United States, what is the P(AB)?

2. Maria has type B blood. She can safely receive blood transfusions from people with blood types O or B. What is the probability that a randomly chosen American can donate blood to Maria?

3. Bozeman Deaconess Hospital needs a donor with type A blood. Ten American donors come in that day. What is the probability that the first nine people do not have type A blood but that the 10th person does have type A blood.

4. What is the probability that at least one of the ten people has type A blood?

5. Choose an American and a Chinese at random, independently of each other. What is the probability that both have type O blood?

6. What is the probability that both have the same blood type?

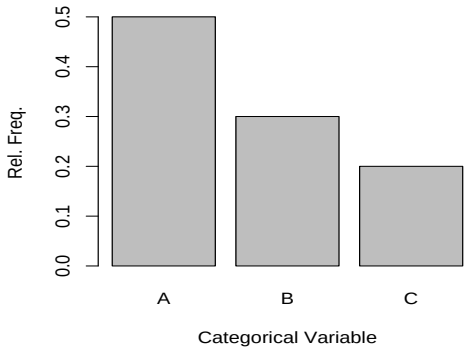
Distributions

(2.5)

A **Population Distribution** gives all the values of a variable for a population, and the probabilities (or relative frequencies) with which the variable takes on these values.

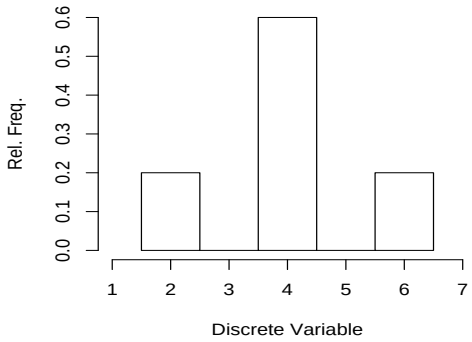
1. Categorical Distribution

Categories	Probability
A	0.5
B	0.3
C	0.2



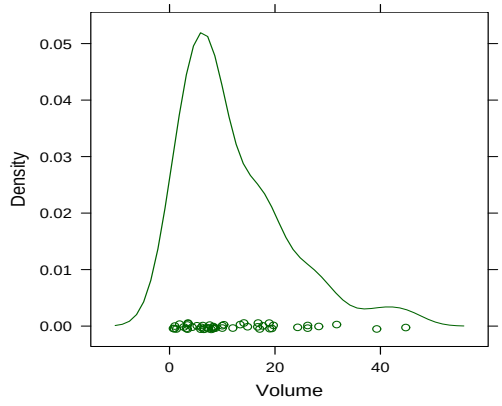
2. Discrete Numerical Distribution

Discrete Values	Probability
2	0.2
4	0.6
6	0.2



3. Continuous Numerical Distribution (also known as a **density curve**).

```
> library(lattice)
> trellis.par.set(col.whitebg())
> densityplot(Volume)
```



Properties of Density Curves

- The probability of obtaining a value in an interval is the area under the curve above the interval.
- A density curve is always 0 or larger.
- The total area under a density curve is 1.
- $P(X = c) = 0$ for any constant c.

EXAMPLES:

1. Witch's Hat Distribution

2. Uniform Distribution

Population Parameters: numbers that describe a population distribution; a numerical value calculated from all individuals in a population.

For numerical (discrete or continuous) population distributions.

- CENTER
  - The **population mean** is μ ("mew")
  - $\mu = \begin{cases} \sum_x xP(x) & \text{if x is discrete} \\ \int_{-\infty}^{\infty} xf(x)dx & \text{if x is continuous} \end{cases}$
- SPREAD
  - The **population variance** is σ^2 ("sigma squared")
  - The **population standard deviation** is σ
  - $\sigma^2 = \begin{cases} \sum_x (x-\mu)^2 P(x) & \text{if x is discrete} \\ \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx & \text{if x is continuous} \end{cases}$

For categorical population distributions:

- The **population proportion** with which some categories occurs over all individuals in the population is given by p.

STATISTICS vs. PARAMETERS:

Recall the table of population parameters and the statistics that estimate them:

Statistics	Parameters
$\bar{x}$	μ
$\tilde{x}$	$\tilde{\mu}$
s	σ
s^2	σ^2
$\hat{p}$	p

Statistic: A numerical value calculated from a sample of individuals.

- Sample Mean: $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
- Sample Variance: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
- Sample Proportion: $p = \frac{\text{number of } 1'\text{s}}{n}$ is the proportion of 1's in the sample

TRANSFORMATIONS: Shifting and Rescaling:

Shifting - adding (or subtracting) a constant c to (or from) every data value will shift the distribution of data values upward (or downward) by c. The measures of center (mean and median) shift by c, but the measures of spread (standard deviation and IQR) remain unchanged.

Rescaling - multiplying (or dividing) every data value by a constant b will multiple (or divide) both the measures of center (mean and median) and the measures of spread (standard deviation and IQR) by b.

Mathematically, if

$$Y = bX + c$$

then

$$\mu_Y = b\mu_X + c \qquad \text{and} \qquad \sigma_Y = |b|\sigma_X.$$

Exercises

Probability basics on p. 116: 2.1, 2.5, 2.7, 2.9ab, 2.11, 2.13
Conditional probability on p. 119: 2.15, 2.17, 2.19
Small populations on p. 122: 2.27, 2.29, 2.31
Distributions on p. 125: 2.43