Lies, Damned Lies, and Statistics

This phrase was popularized in the U.S. by Mark Twain, who attributed it to Benjamin Disraeli — "There are three kinds of lies: lies, damned lies, and statistics." This course will focus on a practical understanding of probability and statistics.

Study Group Leader: John Blankinship ej_blank@msn.com

- How to calculate the probability of simple and compound events
- Practical applications to poker hands, Blackjack, the Arizona Lottery and other games of chance
- Statistics can be tricky: conditional probabilities, correlation vs. causation, Simpson's Paradox, deceptive statistics
- Why are so many natural phenomena in the real world "normally distributed" as a bell-shaped curve?
- How do statistical polls work, and what is a margin of error?
- What does statistically significant mean?

All Statistics Are Man-Made

Any statistical number is the result of a human counting, measuring or calculating something

From Joel Best's book, *More Damned Lies and Statistics*

- Americans have a widespread, naïve faith in the power of numbers. We seem to believe that any number is better than no number. This faith rests on some dubious assumptions. The first is a belief that numbers are by nature factual. This ignores a basic truth — that all numbers are products of human efforts.
- Numbers do not exist independent of people. Statistics are social constructions, and the process by which they are created inevitably shapes the resulting numbers.
- The question to ask about any number is not "Is it true?" — but rather "How was it produced?" Who counted what, how, and why? Also, who paid to have it counted? Some healthy skepticism is in order. Philosophical Question: What is a fact?

Why does Statistics have a reputation for lies?

- Statistics and statistical polls are often used to persuade people about hot button issues about which they have strong opinions
- People may not trust the statistics based on personal experience
- People may think the statistics are biased or manipulated with a vested interest, and may think they are being misled
- People have seen instances where the presentation of statistical numbers and graphs has misled them on issues
- People are overwhelmed by a barrage of statistics and numbers, and don’t know what to believe
- Statistics involves some mathematics and concepts that some people may not understand or trust

How to Calculate the Probability of Simple Events

We intuitively understand that:

- Probability is the likelihood that an event will occur
- The probability of a *Heads* on a coin toss is ½
- There is a "1 in 2 chance" of getting a *Heads* on a coin toss
- If we were to toss a coin 100 times, we would expect (on average) around 50 *Heads* — but we could get 47 or 56
- For 1000 tosses, proportion of *Heads* probably be closer to 50%

The probability of an event is calculated as:

$$\text{Probability} = \frac{\text{number of elementary events favorable to the event}}{\text{total number of equally likely elementary events}}$$

The **elementary events** of an experiment are the set of (equally likely) possible outcomes of that experiment

<u>Example #1</u> — A fair coin is tossed. What is the probability of getting a *Heads*?

- There are two equally likely elementary events: {H, T}
- One of these events (*Heads*) is favorable to "getting a *Heads*"
- Therefore, the probability of getting a *Heads* $= \frac{1}{2} = 0.5$
- The expected number of *Heads* in 60 tosses is: $\frac{1}{2} \times 60 = 30$
- Pr (*Heads*) + Pr (*Tails*) = 1 Pr (*Tails*) = 1 – Pr (*Heads*) = $\frac{1}{2}$

<u>Example #2</u> — A single die is tossed. What is the probability of getting a "2 or less?"

- There are six equally likely elementary events: {1, 2, 3, 4, 5, 6}
- Two of these elementary events {1, 2} are favorable to the event "getting a 2 or less"
- Therefore, the probability of "getting a 2 or less" $= \frac{2}{6} = \frac{1}{3}$
- The expected number of "getting a 2 or less" in 60 tosses of a die is $\frac{1}{3} \times 60 = 20$
- Pr (NOT getting a 2 or less) = 1 – Pr (2 or less) = $1 - \frac{1}{3} = \frac{2}{3}$

Example #3 — A card is drawn from a shuffled standard deck of 52 cards. What is the probability of drawing a spade?

- There are 52 elementary events: { A♠, 2♠, 3♠ … J♣, Q♣, K♣}
- 13 of these elementary events {A♠, 2♠, 3♠ … J♠, Q♠, K♠} are favorable to the event "drawing a spade"
- Therefore, the probability of "drawing a spade" = $\frac{13}{52} = \frac{1}{4}$
- Pr (NOT drawing a spade) = 1 – Pr (drawing a spade) = 1-¼ = ¾

The probability of an event has a value between 0 and 1

- The probability of an "impossible event" is 0 — e.g. tossing a die and getting a 7 or greater
- The probability of a "certain event" is 1 — e.g. tossing a die and getting a 10 or less
- The probability of an event can be expressed as:
  a fraction — e.g. ¼
  a decimal — e.g. 0.25
  a percentage — e.g. 25%
  a phrase like "1 in 4 chance"
  odds — e.g. 3 to 1 against (or 1 to 3 in favor)

How to Calculate the Probability of Compound Events

Example #4 — Two fair coins are tossed. What is the probability that both coin tosses are *Heads*?

- The probability that second coin is *Heads* does not depend on whether first coin is *Heads*. The two events are **independent**.
- If the occurrence or non-occurrence of event E_1 does not affect the probability of event E_2, then E_1 and E_2 are independent events; otherwise, they are dependent events.
- If two events E_1 and E_2 are independent events, then the probability that they will **both** occur is:

Multiplicative Probability Rule for Independent Events E_1 and E_2:

$$\Pr(E_1 \text{ and } E_2) = \Pr(E_1) \times \Pr(E_2)$$

The Multiplicative Probability Rule for Independent Events can be extended to more than two events as:

$$\Pr(E_1 \text{ and } E_2 \text{ and } E_3 \ldots) = \Pr(E_1) \times \Pr(E_2) \times \Pr(E_3) \times \ldots$$

- Using the Multiplicative Probability Rule for Independent Events:

Pr (getting H on first coin AND getting H on second coin)
= Pr (getting H on first coin) × Pr (getting H on second coin)

$$= \frac{1}{2} \times \frac{1}{2} = \frac{1}{4} = 0.25$$

Probability can also be figured by counting the elementary events:

- There are 2 × 2 = 4 equally likely "compound" elementary events for the tossing of two coins: {HH, HT, TH, TT}
- One of these elementary events (HH) is favorable to the event "both coin tosses are *Heads*"
- Therefore, probability that both coin tosses are *Heads* = $\frac{1}{4}$
- Also, probability that both coin tosses are *Tails* = $\frac{1}{4}$

  Probability that the two coin tosses do not match = $\frac{1}{2}$

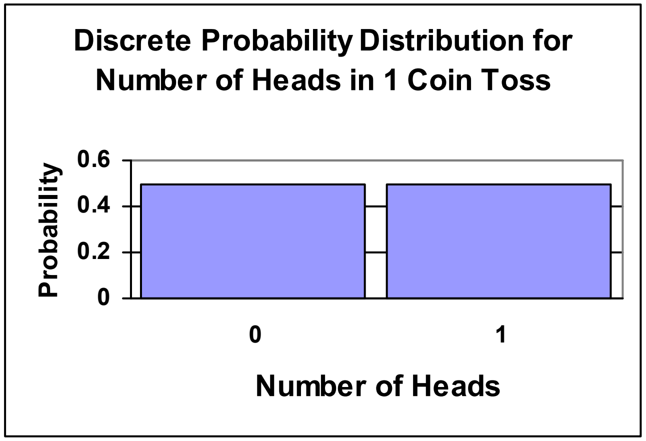
  Pr (at least one of the two coin tosses is *Heads*) = $\frac{3}{4}$

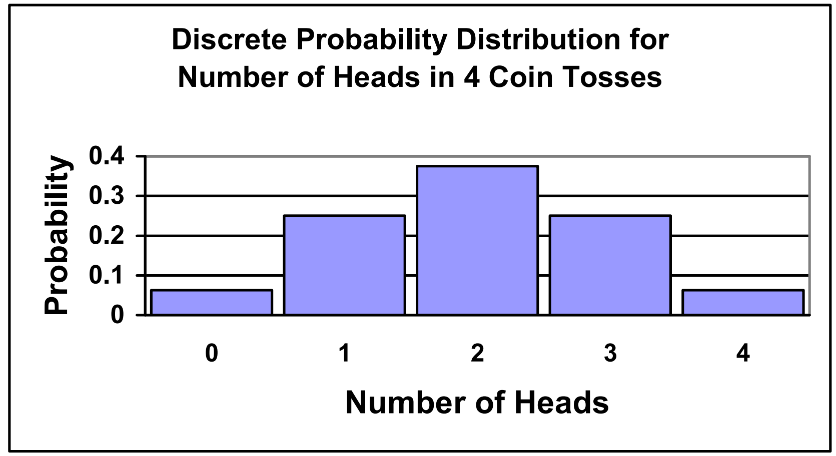
<u>Example #5</u> — Four fair coins are tossed. What is the probability of getting no *Heads*? 1 *Head*? 2 *Heads*? 3 *Heads*? 4 *Heads*?

- There are 2 × 2 × 2 × 2 = 16 equally likely compound elementary events:

HHHH HHHT HHTH HHTT
HTHH HTHT HTTH HTTT
THHH THHT THTH THTT
TTHH TTHT TTTH TTTT

Event	**# of Elementary Events Favorable to Event**	**Prob (Event)**
getting 0 *Heads*	1	$1/16 = .0625$
getting 1 *Head*	4	$4/16 = 1/4 = .25$
getting 2 *Heads*	6	$6/16 = 3/8 = .375$
getting 3 *Heads*	4	$4/16 = 1/4 = .25$
getting 4 *Heads*	1	$1/16 = .0625$





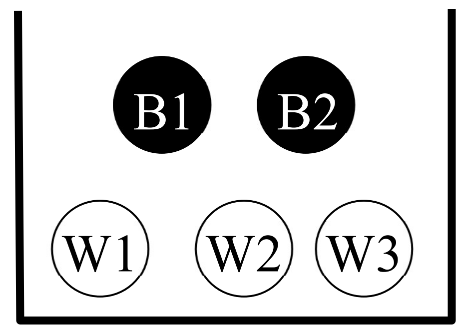
Discrete probability distributions specify the probability associated with each of a set of discrete outcomes

Example #6 — Consider all the families in the United States with exactly four children. Assuming that births of boys and girls are equally likely, random and independent, what proportion of these families would you expect to have exactly two boys and two girls?

- This problem is analogous to a fair coin being flipped four times.
- Therefore, you would expect 37.5% of the families to have exactly two boys and two girls.

<u>Example #7</u> — A ball is drawn at random from a box containing 3 white and 2 black balls. The ball is replaced back in box, and a second ball is drawn. What is the probability both balls are white?

This is called <u>sampling with replacement</u>. Because the first ball is put back into the box before the second ball is drawn, the probability that the second ball is white does not depend on whether the first ball is white. Therefore, the two events are independent.



- Using the Multiplicative Probability Rule for Independent Events:

Pr (first ball is white AND second ball is white) =

Pr (first ball is white) × Pr (second ball is white) = $\frac{3}{5} \times \frac{3}{5} = \frac{9}{25} = 0.36$

- Probability can also be figured by counting elementary events

There are 5 × 5 = 25 equally likely <u>compound</u> elementary events for the drawing of 2 balls <u>with replacement</u> from the box:

W1-W1	**W1-W2**	**W1-W3**	W1-B1	W1-B2
W2-W1	**W2-W2**	**W2-W3**	W2-B1	W2-B2
W3-W1	**W3-W2**	**W3-W3**	W3-B1	W3-B2
B1-W1	B1-W2	B1-W3	B1-B1	B1-B2
B2-W1	B2-W2	B2-W3	B2-B1	B2-B2

Nine of these elementary events (in **bold**) are favorable to the event that both balls are white. As a result:

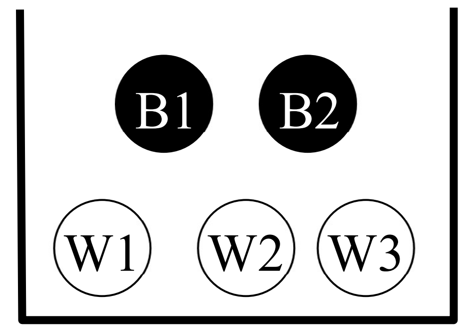
Probability that both balls are white = $\frac{9}{25}$ = 0.36

Also: Probability that both balls are black = $\frac{4}{25}$ = 0.16

Probability that the balls are different = $\frac{12}{25}$ = 0.48

Example #8 — Two balls are randomly drawn from box containing 3 white and 2 black balls — i.e. a ball is drawn from the box, and then without replacing it back into the box, a second ball is drawn.

This is called sampling without replacement. In this case, the drawing of the two balls are dependent events. As shown below, the probability that the second ball is white does depend on whether the first ball is white.



- If first ball is white, then 2 white and 2 black balls are left in box, and conditional probability that second ball is white = $^2/_4$ = $^1/_2$
- But if first ball is black, then 3 white and 1 black ball will be left in box and conditional probability that second ball is white = $^3/_4$

Multiplicative Probability Rule for Dependent Events E_1 and E_2:

Pr (E_1 and E_2) = Pr (E_1) × Pr (E_2 given that E_1 has occurred)

- Using Multiplicative Rule of Probability for <u>Dependent Events:</u>

Pr (first ball is white AND second ball is white) =

Pr (1st ball is white) × Pr (2nd ball is white given 1st ball is white)

$$= \frac{3}{5} \times \frac{2}{4} = \frac{6}{20} = \frac{3}{10} = 0.3$$

- Probability can also be figured by counting elementary events

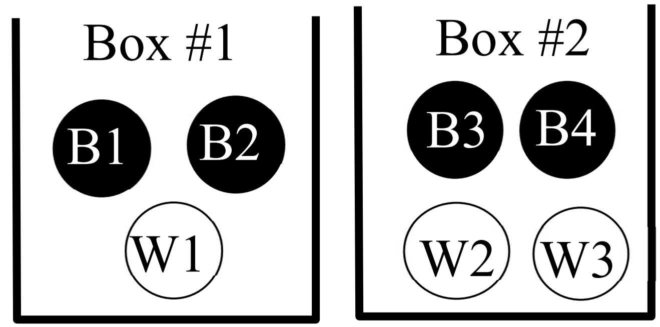
There are 5 × 4 = 20 equally likely compound elementary events for the drawing of 2 balls <u>without replacement</u> from the box:

W1-W2	**W2-W1**	**W3-W1**	B1-W1	B2-W1
W1-W3	**W2-W3**	**W3-W2**	B1-W2	B2-W2
W1-B1	W2-B1	W3-B1	B1-W3	B2-W3
W1-B2	W2-B2	W3-B2	B1-B2	B2-B1

Probability that both balls are white = $\frac{6}{20} = \frac{3}{10}$

Pr (both black) = $\frac{2}{20} = \frac{1}{10}$ Pr (both different) = $\frac{12}{20} = \frac{3}{5}$

<u>Example #9</u> - Box #1 contains 1 white and 2 black balls. Box #2 contains 2 white and 2 black balls. A ball is drawn at random from each box. What is the probability that the ball drawn from Box #1 OR Box #2 (or both) is white?



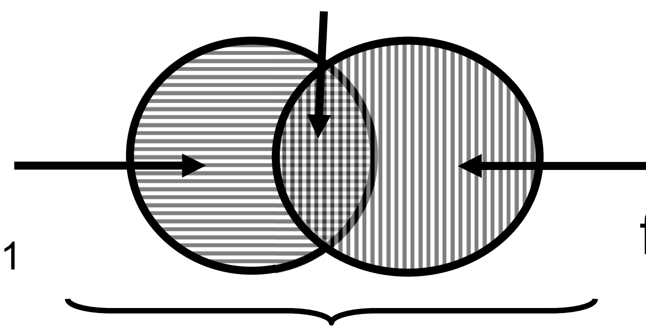
Additive Rule of Probability for Events E_1 OR E_2

$$Pr(E_1 \text{ OR } E_2) = Pr(E_1) + Pr(E_2) - Pr(E_1 \text{ AND } E_2)$$

This rule is illustrated by the following <u>Venn diagram</u>:

Cross-hatched area contains events favorable to event (E_1 AND E_2)

Left circle contains elementary events favorable to event E_1



Right circle contains elementary events favorable to event E_2

Total shaded area contains elementary events favorable to event (E_1 OR E_2)

- Using the Additive Rule of Probability for events E_1 or E_2:

Pr (ball drawn from Box 1 OR Box 2 is white) =

Pr (ball drawn from Box #1 is white)

\+ Pr (ball drawn from Box #2 is white)

– Pr (balls drawn from Box 1 and 2 are both white)

$$= \frac{1}{3} + \frac{1}{2} - \left(\frac{1}{3} \times \frac{1}{2}\right) = \frac{2}{6} + \frac{3}{6} - \frac{1}{6} = \frac{4}{6} = \frac{2}{3} = 0.667$$

- Probability can also be figured by counting elementary events

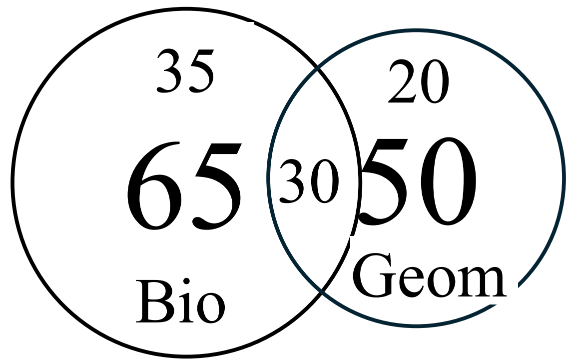
There are 3 × 4 = 12 equally likely compound elementary events for the drawing of 1 ball from Box #1 and 1 ball from Box #2:

B1-B3	B1-B4	**B1-W2**	**B1-W3**
B2-B3	B2-B4	**B2-W2**	**B2-W3**
W1-B3	**W1-B4**	**W1-W2**	**W1-W3**

Probability that Ball #1 OR Ball #2 is white = $^{8}/_{12} = {}^{2}/_{3}$

Pr (both W) = $^{2}/_{12}$ Pr (both B) = $^{4}/_{12}$ Pr (different) = $^{6}/_{12}$

<u>Example #10</u> — Out of 100 10th-graders at Central High School, 65 are taking Biology, 50 are taking Geometry, and 30 are taking both. How many are taking Biology or Geometry (or both)?



N(Bio or Geom) = N(Bio) + N(Geom) – N(Bio and Geom)
= 65 + 50 – 30 = 85

Pr (student taking Bio) = 0.65   Pr (student taking Geom) = 0.5
Pr (student taking both Bio and Geom) = 0.3

Pr (Bio or Geom) = Pr (Bio) + Pr (Geom) – Pr (Bio and Geom)
= 0.65 + 0.5 – 0.3 = 0.85
Pr (student taking neither Bio nor Geom) = 1 – 0.85 = 0.15 = 15%