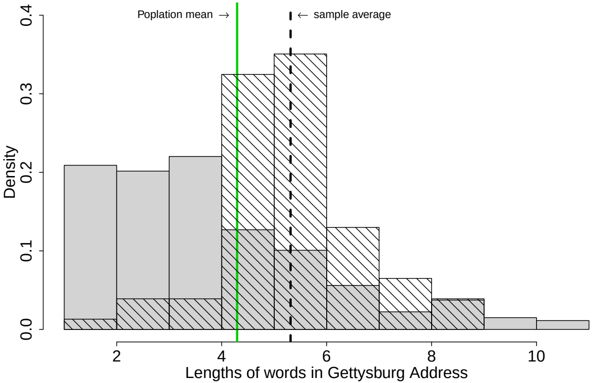
Why random samples? I

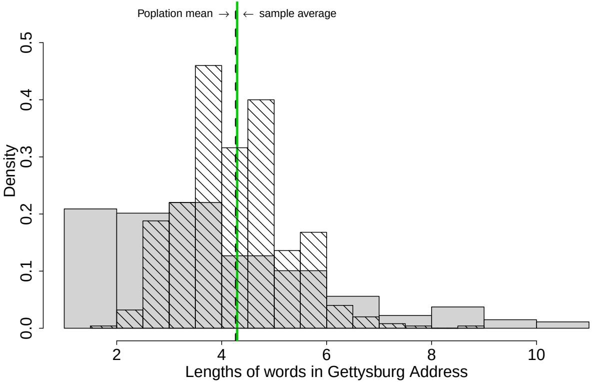
Last time we saw that samples you created by looking at the document were **biased**.



On average your sample averages were higher than the true population mean.

Why random samples? II

Random samples produced by R did not have that problem.



Probability theory tries to provide a mathematical explanation for this.

Probability vs. Statistics

- **Population:** A collection of all *units* of interest (people, households, items produced, etc.).
- **Sample:** A subset of a population that is actually observed.
- **Random Sample:** A sample that gives an equal pre-assigned chance to every unit in population to enter sample.
  - Frees the sample from bias (so that the sample is representative of population)
  - Effectively neutralize all confounding factors at once
- **Probability:** If we know the population (or at least a reasonable model for it), then we can determine what a random sample is "likely" to look like.
- **Statistics (Inference):** If we only have the sample, what can we reasonably conclude about the population?

Assigning Probabilities

Symmetry of outcomes: all outcomes of an experiment are assumed equally likely

Assume N outcomes: Probability of an outcome: $1/N$.
Event is defined as a subset of possible outcomes.
Probability of event containing n outcomes: n/N

- requires finitely many and equally likely outcomes
- can be determined by counting outcomes

The Axioms of Probability

Sample space S: set of possible outcomes.

Events: collection $\mathcal{A}$ of subsets of S.

A **probability** on a sample space S (and a set $\mathcal{A}$ of events) is a function which assigns each event A (in $\mathcal{A}$) a value in $[0, 1]$ and satisfies the following rules:

- **Axiom 1:** All probabilities are nonnegative:

$$P(A) \geq 0 \qquad \text{for all events } A.$$

- **Axiom 2:** The probability of the whole sample space is 1:

$$P(S) = 1.$$

- **Axiom 3 (Addition Rule):** If two events A and B are disjoint then

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

The Axioms of Probablity

Suppose a random experiment has N different outcomes, such that ith outcomes occurs with probability p_i, $i = 1, \ldots, N$. It is natural to define the probability of an event as the sum of the probabilities of the distinct outcomes making up the event.

Why do we need to learn techniques for counting? If each outcome equally likely, $p_1 = \ldots = p_N = 1/N$, then for any event A,

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

This setup satisfies the 3 axioms

- $P(A) \geq 0$
- $P(S) = \dfrac{\#(S)}{\#(S)} = 1$
- If A and B are disjoint then

$$P(A \cup B) = \frac{\#(A \cup B)}{\#(S)} = \frac{\#(A)}{\#(S)} + \frac{\#(B)}{\#(S)} = P(A) + P(B).$$

A Set Theory Primer

A set is *"a collection of definite, well distinguished objects of our perception or of our thought"*. (GEORG CANTOR, 1845-1918)

Some important sets:

- $N = \{1, 2, 3, \ldots\}$, the set of *natural numbers*
- $Z = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$, the set of *integers*
- $R = (-\infty, \infty)$, the set of *real numbers*

Intervals are denoted as follows:

$[0, 1]$ the interval from 0 to 1 including 0 and 1

$[0, 1)$ the interval from 0 to 1 including 0 but not 1

$(0, 1)$ the interval from 0 to 1 not including 0 and 1

If a is an element of the set A then we write $a \in A$.

If a is not an element of the set A then we write $a \notin A$.

A Set Theory Primer

Suppose that A and B are subsets of S (denoted as $A, B \subseteq S$).

- The *empty set* is denoted by $\emptyset$
  (Note: $\emptyset \subseteq A$ for all subsets A of S).
- *Complement of A (A^c or A'):*
  Set of all elements in S that are not in A.
- *Intersection of A and B ($A \cap B^c$):*
  Set of all elements in S which are both in A and in B.

A Set Theory Primer

Suppose that A and B are subsets of S (denoted as $A, B \subseteq S$).

- *Difference of A and B $(A \setminus B) = (A \cap B^c)$:*
  Set of all elements in A which are not in B.
- *Union of A and B $(A \cup B)$:*
  Set of all elements in S that are in A or in B or in both. Note that $A \cap A^c = \emptyset$ and $A \cup A^c = S$
- A and B are *disjoint* if A and B have no common elements, that is $A \cap B = \emptyset$. Two events A and B with this property are said to be *mutually exclusive*.

The Mathematics of Probability

We assume the collection of events $\mathcal{A}$ satisfies:

- $\phi, S \in \mathcal{A}$
- if $A \in \mathcal{A}$ then $A^c \in \mathcal{A}$
- If $A, B \in \mathcal{A}$ then $A \cap B \in \mathcal{A}$
- If $A, B \in \mathcal{A}$ then $A \cup B \in \mathcal{A}$

The Mathematics of Probability

Let A and B be events in an outcome set S.

Partition rule: $\mathrm{P}(A) = \mathrm{P}(A \cap B) + \mathrm{P}(A \cap B^c)$

Example: Roll a pair of fair dice

$$\begin{aligned}&\mathrm{P}(\text{Total of 10})\\&\quad = \mathrm{P}(\text{Total of 10 and double}) + \mathrm{P}(\text{Total of 10 and no double})\\&\quad = \frac{1}{36} + \frac{2}{36} = \frac{3}{36} = \frac{1}{12}\end{aligned}$$

Complement rule: $\mathrm{P}(A^c) = 1 - \mathrm{P}(A)$

Example: Often useful for events of the type "at least one" or "at least as large as some small number"

$$\mathrm{P}(\text{Total is at least 4}) = 1 - \mathrm{P}(\text{Total is less than 4}) = 1 - \frac{3}{36} = \frac{33}{36}$$

The Mathematics of Probability

Let A and B be events in an outcome set S.

Containment rule: $\mathrm{P}(A) \leq \mathrm{P}(B)$ for all $A \subseteq B$

Example: Compare "two ones" with "any double",

$$\frac{1}{36} = \mathrm{P}(\text{two ones}) \leq \mathrm{P}(\text{any double}) = \frac{6}{36} = \frac{1}{6}$$

The Mathematics of Probability

Inclusion exclusion formula: $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$

Example: Roll a pair of fair dice

$$\begin{aligned}
&\mathrm{P}(\text{Total of 10 or double}) \\
&\quad = \mathrm{P}(\text{Total of 10}) + \mathrm{P}(\text{Double}) - \mathrm{P}(\text{Total of 10 and double}) \\
&\quad = \frac{3}{36} + \frac{6}{36} - \frac{1}{36} = \frac{8}{36} = \frac{2}{9}
\end{aligned}$$

The two events are Total of $10 = \{(4,6),(6,4),(5,5)\}$ and

$$\text{Double} = \{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}$$

The intersection is Total of 10 and double $= \{(5,5)\}$.
Adding the probabilities for the two events, the probability for the event $\{(5,5)\}$ is added twice, so we need to subtract this probability back out.

The Complement Rule

The axioms are the fundamental building blocks of probability. Any other probability relationships can be derived from the axioms.

Show that $P(A^c) = 1 - P(A)$

This proof asks us to confirm an equation
mathematical expression A = mathematical expression B

General form of a proof:

- First, write down any existing definitions or previously proven facts you can think of that are related to any formulas/symbols appearing in expressions A and B
- Start the proof with the left side (expression A) or with the most complex of the two expressions.
- Use algebra and established statistical facts to re-write this right-side expression until it equals the left-side

The Complement Rule

The axioms are the fundamental building blocks of probability.
Any other probability relationships can be derived from the axioms.

[Some knowledge of sets will be needed too.]

Show that $P(A^c) = 1 - P(A)$

- A & A^c are disjoint (mutually exclusive, don't overlap).
  So, $P(A \cup A^c) = P(A) + P(A^c)$ (Axiom 3).
- Also, $A \cup A^c = S$.
  So, $P(A \cup A^c) = P(S) = 1$ (Axiom 2).
- Therefore, $P(A) + P(A^c) = 1$.
- That is, $P(A^c) = 1 - P(A)$.

We only needed 1st step of proof algorithm this time - gather info.

Isn't $0 \leq P(A) \leq 1$?
...but Axiom 1 is just $P(A) \geq 0$.

The axioms are the fundamental building blocks of probability.

Any other probability relationships can be derived from the axioms.

Let's prove that $P(A) \leq 1$.

- By the complement rule, $1 - P(A) = P(A^c)$
- and $P(A^c) \geq 0$ (Axiom 1).
- So, $1 - P(A) \geq 0 \implies 1 \geq P(A)$.

Some more probability facts

We can also prove ...

- The Law of Total Probability = Partition Rule

$$P(A) = P(A \cap B) + P(A \cap B^c)$$
$$\text{or} \;\; P(A) = P(A \cap B) + P(\text{“}A - B\text{''})$$

- The Inclusion-Exclusion Formula

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

- Probability for subsets If $A \subseteq B$, then $P(A) \leq P(B)$

  Let's try to prove the last one.

Proving a Conditional Statement

The axioms are the fundamental building blocks of probability.
Any other probability relationships can be derived from the axioms.

[Some knowledge of sets will be needed too.]

Show that if $A \subseteq B$, then $P(A) \leq P(B)$

Suppose $A \subseteq B$.

- Then, $A \cap B = A$
- Always true: $P(A \cap B) + P(A^c \cap B) = P(B)$ (Axiom 3)
- So, $P(A) + P(A^c \cap B) = P(B)$
- and $P(A) \leq P(A) + P(A^c \cap B)$ since $P(A^c \cap B) \geq 0$ (Axiom 1)
- Putting everything together... $P(A) \leq P(B)$

Conditional Probability

Probability gives chances for events in outcome set S.

Example: Number of Deaths in the U.S. in 1996

Cause	All ages	1-4	5-14	15-24	25-44	45-64	$\geq$ 65
Heart	733,125	207	341	920	16,261	102,510	612,886
Cancer	544,161	440	1,035	1,642	22,147	132,805	386,092
HIV	32,003	149	174	420	22,795	8,443	22
Accidents[1]	92,998	2,155	3,521	13,872	26,554	16,332	30,564
Homicide[2]	24,486	395	513	6,548	9,261	7,717	52
All causes	2,171,935	5,947	8,465	32,699	148,904	380,396	1,717,218

[1] Accidents and adverse effects, [2] Homicide and legal intervention

Probabilities and conditional probabilities for causes of death:

- $\mathrm{P}(\text{accident}) = 92,998/2,171,935 = 0.04282$
- $\mathrm{P}(5 \leq \text{age} \leq 14) = 8,465/2,171,935 = 0.00390$
- $\mathrm{P}(\text{accident and } 5 \leq \text{age} \leq 14) = 3,521/2,171,935 = 0.00162$
- $\mathrm{P}(\text{accident} \mid 5 \leq \text{age} \leq 14) = 3,521/8,465 = 0.41595$

Conditional Probability

$$
\begin{aligned}
P(\text{accident}|5 \leq \text{age} \leq 14) &= \frac{3,521}{8,465} = \frac{3,521/2,171,935}{8,465/2,171,935} \\
&= \frac{P(\text{accident and } 5 \leq \text{age} \leq 14)}{P(5 \leq \text{age} \leq 14)}
\end{aligned}
$$

Conditional probability of A given B

$$
\mathrm{P}(A \mid B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}, \qquad \text{if } \mathrm{P}(B) > 0
$$

Conditional Probability

$\rightsquigarrow$ measure conditional probability with respect to a subset of S

Conditional probability of A given B

$$\mathrm{P}(A \mid B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}, \qquad \text{if } \mathrm{P}(B) > 0$$

If $\mathrm{P}(B) = 0$ then $\mathrm{P}(A|B)$ is undefined.

$\rightsquigarrow$ **Multiplication rule:** $\mathrm{P}(A \cap B) = \mathrm{P}(A \mid B) \times \mathrm{P}(B)$.

Independence

Example: Roll two fair dice

What is probability that 2nd die shows ⚀?

$\mathrm{P}(\text{2nd die} = ⚀) = \frac{1}{6}$

What is probability 2nd die shows ⚀ if 1st die showed ⚀?

$$\mathrm{P}(\text{2nd die} = ⚀ \mid \text{1st die} = ⚀) = \frac{1}{6}$$

. . . and if the 1st die did not show ⚀?

$$\mathrm{P}(\text{2nd die} = ⚀ \mid \text{1st die} \neq ⚀) = \frac{1}{6}$$

Independence

The event A is **independent** of the event B if its chances are not affected by the occurrence of B,

$$\mathrm{P}(A|B) = \mathrm{P}(A).$$

You can show that the following definitions of independence are equivalent.

$$\mathrm{P}(A|B) = \mathrm{P}(A)$$

$$\mathrm{P}(B|A) = \mathrm{P}(B)$$

$$\mathrm{P}(A \cap B) = \mathrm{P}(A) \times \mathrm{P}(B)$$

Independence

Suppose event A is independent of event B.
Then, knowing that B has occurred does not effect the probability of event A occurring: $\mathrm{P}(A|B) = \mathrm{P}(A)$. Now,

$$\mathrm{P}(B|A) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(A)} = \frac{\mathrm{P}(A|B)\mathrm{P}(B)}{\mathrm{P}(A)} = \frac{\mathrm{P}(A)\mathrm{P}(B)}{\mathrm{P}(A)} = \mathrm{P}(B)$$

Thus, event B is independent of event A.
The argument in the other direction is exactly the same.
So, the following two statements are equivalent:

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

and we simply state that events A and B are independent.

Independence

Also, if A and B are independent events, then

$$\mathrm{P}(A \cap B) = \mathrm{P}(A|B)\mathrm{P}(B) = \mathrm{P}(A)\mathrm{P}(B)$$

and in the other direction... If $\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B)$, then

$$P(A|B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} = \frac{\mathrm{P}(A)\mathrm{P}(B)}{\mathrm{P}(B)} = \mathrm{P}(A).$$

Thus, the following three statements are equivalent definitions of independence of events A and B :

$$\mathrm{P}(A|B) = \mathrm{P}(A)$$

$$\mathrm{P}(B|A) = \mathrm{P}(B)$$

$$\mathrm{P}(A \cap B) = \mathrm{P}(A) \times \mathrm{P}(B)$$