Statistical Independence

- In the last example, we had several probabilities of the form $P(A|B) = P(A)$

Statistical Independence

- In the last example, we had several probabilities of the form $P(A|B) = P(A)$
- In this case, we say that A is *statistically independent (s.i.)* of B, since the probabilities of A are not affected by knowledge of A having occurred

- By the chain rule, $P(A \cap B) = P(A|B)P(B)$. So if $P(A) > 0$, $P(B > 0)$, and $P(A|B) = P(A)$, then

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

- By the chain rule, $P(A \cap B) = P(A|B)P(B)$. So if $P(A) > 0$, $P(B > 0)$, and $P(A|B) = P(A)$, then

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

- By the chain rule, $P(A \cap B) = P(A|B)P(B)$. So if $P(A) > 0$, $P(B > 0)$, and $P(A|B) = P(A)$, then

$$
\begin{aligned}
P(B|A) &= \frac{P(A \cap B)}{P(A)} \\
&= \frac{P(A|B)P(B)}{P(A)} \\
&= \frac{P(A)P(B)}{P(A)}
\end{aligned}
$$

- By the chain rule, $P(A \cap B) = P(A|B)P(B)$. So if $P(A) > 0$, $P(B > 0)$, and $P(A|B) = P(A)$, then

$$
\begin{aligned}
P(B|A) &= \frac{P(A \cap B)}{P(A)} \\
&= \frac{P(A|B)P(B)}{P(A)} \\
&= \frac{P(A)P(B)}{P(A)} \\
&= P(B)
\end{aligned}
$$

- So, if A is statistically independent of B, then B is statistically independent of A

- So, if A is statistically independent of B, then B is statistically independent of A

- Thus, we can write $P(A \cap B) = P(A)P(B)$, and we use this for our definition of statistical independence because it works even when one of $P(A) = 0$ or $P(B) = 0$

- So, if A is statistically independent of B, then B is statistically independent of A

- Thus, we can write $P(A \cap B) = P(A)P(B)$, and we use this for our definition of statistical independence because it works even when one of $P(A) = 0$ or $P(B) = 0$

Events A and B are *statistically independent (s.i.)* if and only if (iff)

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

- Events that arise from completely separate random phenomena are statistically independent.

EX **Take 2:** XOR of Two Independent Binary Values

Flip a fair coin with sides labeled '0' and '1' two times. Let E_i denote a '1' on the top face on flip i. Let F denote the event that the XOR of the values observed on the top faces on the two flips is '1'.

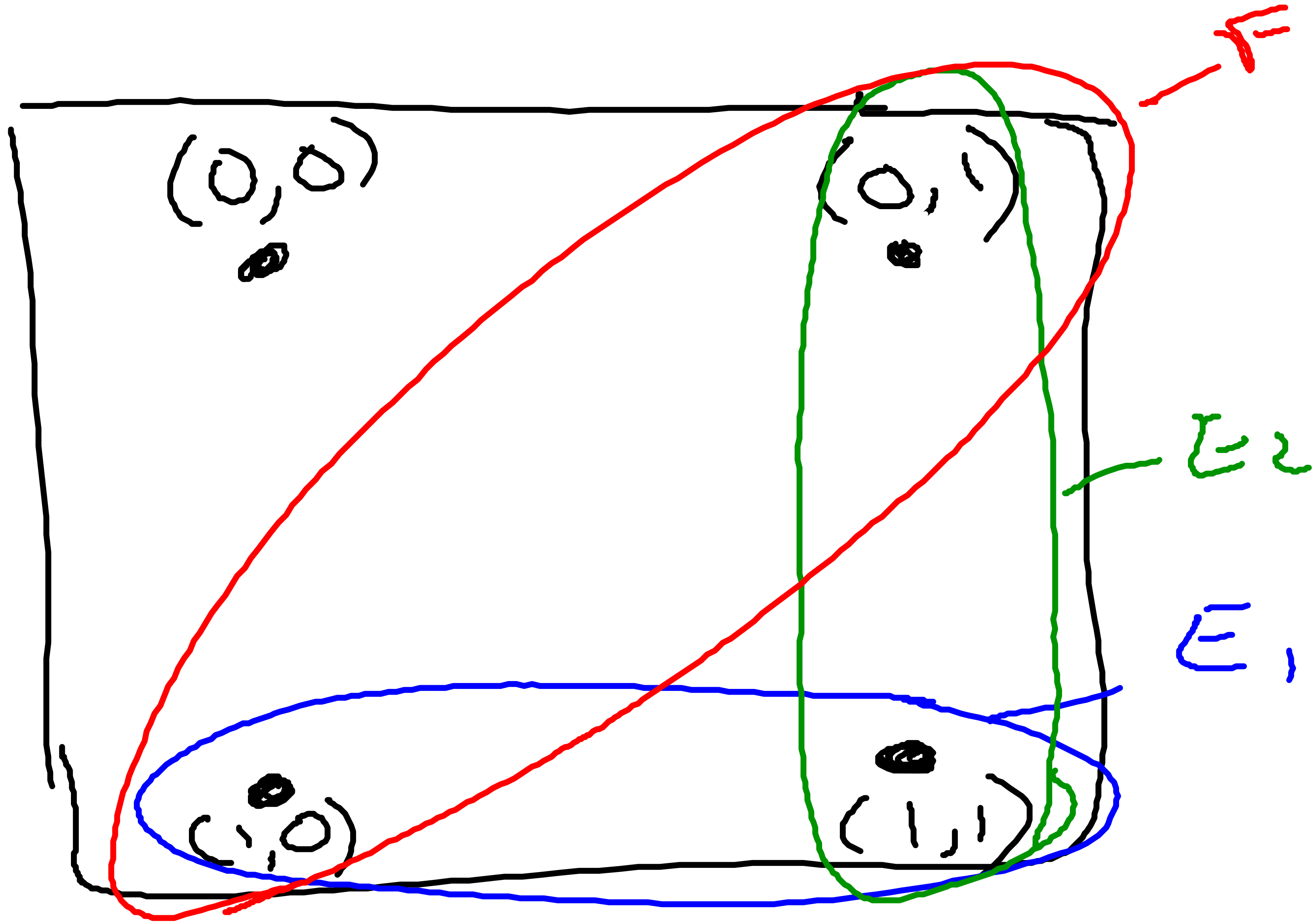
Flip 1	Flip 2	XOR	Prob
0	0	0	1/4
1	0	1	1/4
0	1	1	1/4
1	1	0	1/4

$P(E_1)$

$1/2$

E_2

F



$$P(E_1) = P(E_2) = P(F) = 1/2$$

$P(E_1|E_2) = 1/2$ $\quad$ $P(E_2|E_1) = 1/2$ $\quad$ $P(F|E_1) = \frac{1}{2}$

$P(F|E_1 \cap E_2) = 0$, F and $E_1 \cap E_2$ are m.e.

- Events that arise from completely separate random phenomena are statistically independent.

EX **Take 3:** A fair six-sided die is rolled twice. What is the probability of observing a 1 or a 2 on the top face on either roll of the die?

$$P(E_1 \cup E_2)$$

$$= P(E_1) + P(E_2) - P(\underbrace{E_1 \cap E_2}_{\text{s.i.}})$$

$$= 1/3 + 1/3 - \underbrace{P(E_1)}_{(1/3)}\underbrace{P(E_2)}_{(1/3)}$$

$$= 5/9$$

~~Take~~

$$P(E_1 \cup E_2) = 1 - P\left(\overline{E_1 \cup E_2}\right)$$

$$= 1 - P(\underbrace{\overline{E_1} \cap \overline{E_2}}_{s.i.})$$

$$= 1 - P(\overline{E_1})P(\overline{E_2})$$

$$= 1 - (2/3)(2/3)$$

$$= 1 - [1 - P(E_1)][1 - P(E_2)]$$

$$= 1 - [1 - p]^2$$

$$p = P(E_1) = P(E_2)$$

EEL 4930 Lecture 9:

One-dimensional Statistics

Populations and Sampling

POPULATIONS AND SAMPLING

A population is a group of people, objects, events, observations, etc. that is being studied.

Populations and Sampling

A population is a group of people, objects, events, observations, etc. that is being studied.

Often we are trying to assess some qualities or properties of that population. We call these parameters.

POPULATIONS AND SAMPLING

A population is a group of people, objects, events, observations, etc. that is being studied.

Often we are trying to assess some qualities or properties of that population. We call these parameters.

When the population is too large to directly measure the parameters of interest, then we try to draw inferences from a subset of the population

A sample from a population is a subset of the population that can be used to draw inferences about the parameters of interest.

A sample from a population is a subset of the population that can be used to draw inferences about the parameters of interest.

- A sample is usually drawn randomly from the population

A sample from a population is a subset of the population that can be used to draw inferences about the parameters of interest.

- A sample is usually drawn randomly from the population
- We usually require that each member of the sample is chosen independently from other members

A sample from a population is a subset of the population that can be used to draw inferences about the parameters of interest.

- A sample is usually drawn randomly from the population
- We usually require that each member of the sample is chosen independently from other members
- Often, but not always, each member in the population is equally likely to be included in the sample

A statistic is a measurement of a quality or property on a sample that is used to assess a parameter of the whole population.

A statistic is a measurement of a quality or property on a sample that is used to assess a parameter of the whole population.

When samples are small, the statics often provide little or no information about the parameters.

A statistic is a measurement of a quality or property on a sample that is used to assess a parameter of the whole population.

When samples are small, the statics often provide little or no information about the parameters.

- For example, consider the problem of determining whether a coin is fair or two-sided. The result of flipping a coin one time provides no useful information for determining that

When samples are larger, they generally more accurately represent the population

When samples are larger, they generally more accurately represent the population

There are generally two cases that we will encounter

1. When designing an experiment, the statistician can choose the ~~population~~ sample size to balance between being able to generate a useful statistic and the cost of taking more samples

1. When designing an experiment, the statistician can choose the sample ~~population~~ size to balance between being able to generate a useful statistic and the cost of taking more samples

2. Sometimes the experiment has already been carried out or is not under the control of the statistician. For instance, the statistician wants to assess something based on an existing survey or compare effects of a change in laws on a set of states. In this case, the ~~population~~ sample size is fixed

We will be using Python to compute statistics on samples and determine whether and how well these statistics represent the parameters of the populations being studied