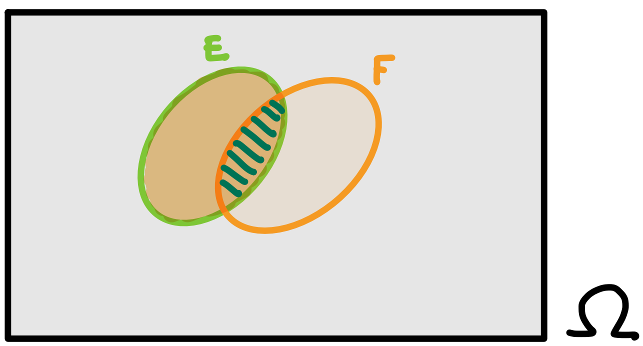
M378K: January 21st, 2026.

2.2. Conditional probability. In order to "build" more complicated (and useful!) random variables, it helps to review a bit more probability.

Definition 2.5. *Let E and F be two events on the same Ω such that $\mathbb{P}[E] > 0$. The conditional probability of F given E is defined as*

$$\mathbb{P}[F \mid E] = \frac{\boxed{\mathbb{P}[E \cap F]}}{\mathbb{P}[E]}.$$

Let's spend a moment with the geometric/informational perspective on this definition.



By far, the most popular problems relying on the notion of **conditional probability** are those to do with **specificity** and **sensitivity**[1] of medical tests.

Problem 2.5. *At any given time 2% of the population actually has a particular disease.*

A test indicates the presence of a particular disease 96% of the time in people who actually have the disease. The same test is positive 1% of the time when actually healthy people are tested.

Calculate the probability that a particular person actually has the disease **given** *that they tested positive.*

→ : E... the test was positive

F... the person has the disease

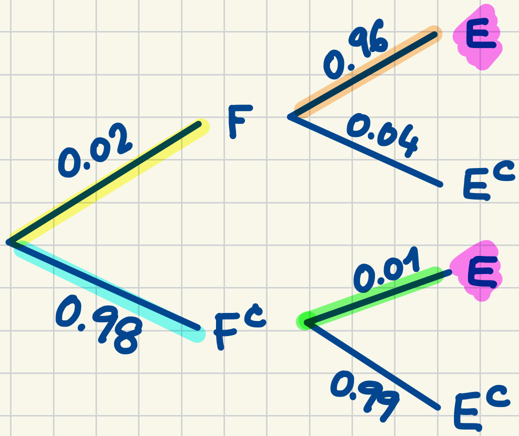
$\mathbb{P}[F] = 0.02$

$\mathbb{P}[E \mid F] = 0.96$

$\mathbb{P}[E \mid F^c] = 0.01$

$\mathbb{P}[F \mid E] = ?$

[1] https://en.wikipedia.org/wiki/Sensitivity_and_specificity



$$\mathbb{P}[F \mid E] = \frac{\mathbb{P}[F \cap E]}{\underbrace{\mathbb{P}[E]}_{\mathbb{P}[E \cap F] + \mathbb{P}[E \cap F^c]}} = \frac{\mathbb{P}[F] \cdot \mathbb{P}[E \mid F]}{\mathbb{P}[F] \cdot \mathbb{P}[E \mid F] + \mathbb{P}[F^c] \cdot \mathbb{P}[E \mid F^c]}$$

Bayes Theorem

$$\mathbb{P}[F \mid E] = \frac{0.02 \cdot 0.96}{0.02 \cdot 0.96 + 0.98 \cdot 0.01}$$

$$= \frac{2 \cdot 96}{2 \cdot 96 + 98} = \frac{96}{96+49} = \frac{96}{145} \quad \square$$

Moreover, now that we remember the definition of **conditional probability**, we can solve interesting problems such as this one:

Problem 2.6. *The number of pieces of gossip that break out in a particular high school in a week is modeled by a random variable Y with the following probability mass function:*

$$p_n := p_Y(n) = \frac{1}{(n+1)(n+2)} \quad \textit{for all } n \in \mathbb{N}_0.$$

Calculate the probability that at least one piece of gossip occurred in a week **given** *that at most four pieces of gossip occurred.*

$\rightarrow$:

$$\mathbb{P}[Y \geq 1 \mid Y \leq 4] = \frac{\mathbb{P}[1 \leq Y \leq 4]}{\mathbb{P}[Y \leq 4]} =$$

$$= \frac{p_1 + p_2 + p_3 + p_4}{p_0 + p_1 + p_2 + p_3 + p_4} =$$

$$= \frac{\left(\frac{1}{2} - \cancel{\frac{1}{3}}\right) + \cdots + \left(\cancel{\frac{1}{5}} - \frac{1}{6}\right)}{\left(1 - \cancel{\frac{1}{2}}\right) + \left(\cancel{\frac{1}{2}} - \cancel{\frac{1}{3}}\right) + \cdots + \left(\cancel{\frac{1}{5}} - \frac{1}{6}\right)}$$

$$= \frac{\frac{1}{2} - \frac{1}{6}}{1 - \frac{1}{6}} = \frac{\frac{2}{6}}{\frac{5}{6}} = \frac{2}{5} \quad \square$$

3. Independent events

What if knowing that an event happened in fact does **not** give any information about the probability of another event?

Definition 3.1. *We say that events E and F on Ω are* independent *if*

$$\mathbb{P}[E \cap F] = \mathbb{P}[E]\mathbb{P}[F].$$

In the case when E or F have a positive probability, it's possible to rewrite the above condition in a different (illustrative!) way. *How?*

Assume that $\mathbb{P}[E] > 0$.

$$\mathbb{P}[F \mid E] = \frac{\mathbb{P}[E \cap F]}{\mathbb{P}[E]} = \frac{\cancel{\mathbb{P}[E]} \cdot \mathbb{P}[F]}{\cancel{\mathbb{P}[E]}} = \mathbb{P}[F]$$

Now that we know the notion of **independence**, we can construct random variables in many creative ways.

Example 3.2. *A fair coin is tossed repeatedly and* **independently** *until the first* Heads. *Let the random variable Y represent the total number of* Tails *observed by the end of the procedure.*

What is the support of the random variable Y?

$$S_Y = \{0, 1, 2, \ldots\} = \mathbb{N}_0$$

What is the **probability mass function** *of the random variable Y?*

for $y \in S_Y$:

$$p_Y(y) = \left(\frac{1}{2}\right)^y \left(\frac{1}{2}\right) = \left(\frac{1}{2}\right)^{y+1}$$

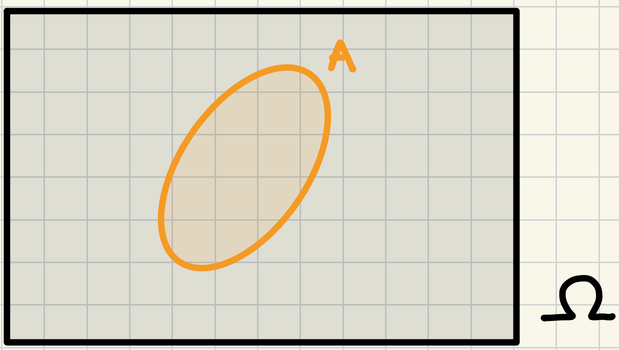
the # of failures until 1st success

Named Discrete Distributions.

Def'n. **Bernoulli Trials** have two possible outcomes.
They are also known as **indicators**
(or indicator random variables).

Usually, the outcomes are interpreted as

$$\begin{cases} 1 \text{ for "success"} \\ 0 \text{ for "failure"} \end{cases}$$



$$\mathbb{I}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}$$

$$\mathbb{I}_A = \begin{cases} 1 & \text{if } A \text{ happened} \\ 0 & \text{if } A \text{ did not happen} \end{cases}$$

Example. Y_i, $i = 1, 2$ ... result of a throw of a regular die

$$S_{Y_1} = S_{Y_2} = \{1, 2, \dots, 6\}$$

We win if the result on the 1st die is even
and
the result on the 2nd die is prime.

$$\begin{cases} I_1 = \mathbb{I}_{\{Y_1 \in \{2,4,6\}\}} \\ I_2 = \mathbb{I}_{\{Y_2 \in \{2,3,5\}\}} \end{cases}$$

Then, our indicator of a win is $\boxed{I_1 \cdot I_2}$ □