M378K: August 29^{th}, 2025.

M378K Introduction to Mathematical Statistics
Problem Set #2
Discrete random variables.

2.1. **Probability mass function.** Recall the following definition from the last class:

Definition 2.1. *Given a set B, we say that a random variable Y is $B-$valued if*

$$\mathbb{P}[Y \in B] = 1.$$

We reserve special terminology for random variables Y depending on the cardinality of the set B from the above definition. In particular, we have the following definition:

Definition 2.2. *A random variable Y is said to be discrete if there exists a set S such that :*

- *Y is $S-$valued, and*
- *S is either **finite** or **countable**.*

Problem 2.1. *Provide an example of a **discrete** random variable.*

- roll of a die: $S = \{1, 2, 3, 4, 5, 6\}$
- coin toss: $S = \{H, T\}$
- finite uniform
- Bernoulli r.v. $S = \{\underset{\text{failure}}{0}, \underset{\text{success}}{1}\}$

Our next task is to try to keep track of the probabilities that Y takes specific values from S. In order to be more "economical", we introduce the following concept:

"**Definition 2.3.**" *The support S_Y of a random variable Y is the "**smallest**" set S such that Y is $S-$valued.*

Problem 2.2. *What is the **support** of the random variable you provided as an example in the above problem?*

✓

Y is still discrete.

Problem 2.3. *Let $y \in S_Y$. Is it possible to have $\mathbb{P}[Y = y] = 0$?*

$\boxed{\text{No!}}$ Assume, to the contrary, that such a y exists.

Set $\tilde{S}_Y = S_Y \setminus \{y\}$

Then, $\mathbb{P}[Y \in \tilde{S}_Y] = \underbrace{\mathbb{P}[Y \in S_Y]}_{1} - \underbrace{\mathbb{P}[Y = y]}_{0} = 1$

and $\tilde{S}_Y$ is "smaller". $\square$

Usually, we are interested in calculating and modeling probabilities that look like this

$$\mathbb{P}[Y \in A] \quad \text{for some } A \subset S_Y.$$

Note that, if we know the probabilities of the form

$$\mathbb{P}[Y = y] \quad \text{for all } y \in S_Y,$$

then we can calculate any probability of the above form. *How?*

$$\boxed{\mathbb{P}[Y \in A] = \sum_{y \in A} \mathbb{P}[Y = y]}$$

So, if we "tabulate" the probabilities of the form $\mathbb{P}[Y = y]$ for all $y \in S_Y$, we have sufficient information to calculate any probability of interest to do with the random variable Y. This observation motivates the following definition:

Definition 2.4. *The probability mass function (pmf) of a* **discrete** *random variable Y is the function $p_Y : S_Y \to \mathbb{R}$ defined as*

$$p_Y(y) = \mathbb{P}[Y = y] \quad \textit{for all } y \in S_Y.$$

Can you think of different ways in which to display the pmf?

- formula: e.g., geometric $p_Y(k) = (1-p)^{k-1} p \quad k = 1, \dots, \dots$
- roll of a die: $p_Y(k) = \frac{1}{6} \quad k = 1, \dots, 6$
- a distribution table.

y	y_1	y_2	....	y_k	.....
$p_Y(y)$	p_1	p_2	....	p_k	....

What are the immediate properties of every pmf? Does the "reverse" hold, i.e., if a function p_Y satisfies you stated, is it always a pmf of **some** random variable?

$$\begin{cases} \bullet\ p_Y(y) > 0 \quad \text{for all } y \in S_Y \\ \bullet\ \sum_{y \in S_Y} p_Y(y) = 1 \end{cases}$$

What is the pmf of the random variable which you provided as an example above?

- coin toss: $S_Y = \{H, T\}$

  e.g., fair coin

  $p_Y(H) = \frac{1}{2}$

  $p_Y(T) = \frac{1}{2}$

- Bernoulli r.v.: $S_Y = \{0, 1\}$

  $p_Y(0) = 1 - p,$

  $p_Y(1) = p$ for $p \in (0,1)$

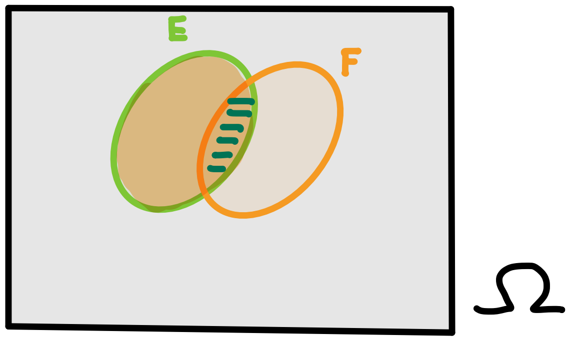
  parameter.

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 2.4. *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

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

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

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

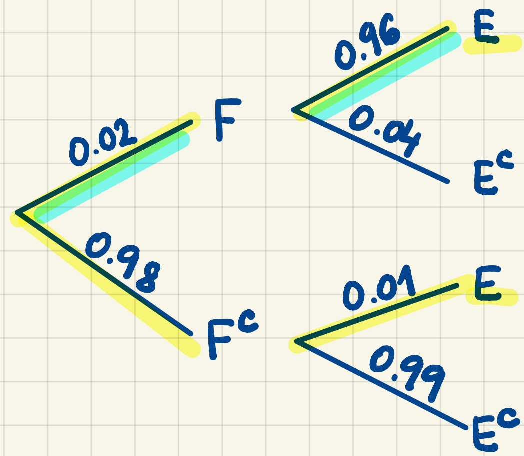
$\mathbb{P}[F \mid E] = ?$

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

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

$$= \frac{\mathbb{P}[E|F]\cdot\mathbb{P}[F]}{\mathbb{P}[E|F]\cdot\mathbb{P}[F] + \mathbb{P}[E|F^c]\cdot\mathbb{P}[F^c]}$$

Bayes Theorem



$$\mathbb{P}[F|E] = \frac{0.02\cdot 0.96}{0.02\cdot 0.96 + 0.98\cdot 0.01} = \frac{2\cdot 96}{2\cdot 96 + 98}$$

$$= \frac{96}{96+49} = \frac{96}{145}$$

$\square$