

# ML - Probability Practice

Sarah Dominguez

2024-08-12

**Part A.** Visitors to your website are asked to answer a single survey question before they get access to the content on the page. Among all of the users, there are two categories: Random Clicker (RC), and Truthful Clicker (TC). There are two possible answers to the survey: yes and no. Random clickers would click either one with equal probability. You are also giving the information that the expected fraction of random clickers is 0.3. After a trial period, you get the following survey results: 65% said Yes and 35% said No. What fraction of people who are truthful clickers answered yes? Hint: use the rule of total probability.

## Solution:

### Step 1: Define Initial Probabilities

**Random Clicker (RC):** Clicks “Yes” or “No” with equal probability:

- $P(Yes|RC) = 0.5$
- $P(No|RC) = 0.5$

**Truthful Clickers (TC):** Someone who always clicks the correct answer:

- **Probability of a user being a Random Clicker (RC):**  
 $P(RC) = 0.3$
- **Probability of a user being a Truthful Clicker (TC):**  
 $P(TC) = 1 - P(RC) = 0.7$

**Step 2: Survey Results** After the trial period, the survey results indicate:

- $P(Yes) = 0.65$  -  $P(No) = 0.35$

**Step 3: Calculate the Fraction of Truthful Clickers Who Answered “Yes” Using the Rule of Total Probability:**

- $P(Yes) = P(Yes|TC) \times P(TC) + P(Yes|RC) \times P(RC)$

```
# Calculate the probability of a truthful clicker answering "Yes"
prob_yes_truth <- (0.65 - (0.5 * 0.3)) / 0.7
```

```
# Convert to percentage
fraction <- prob_yes_truth * 100
```

```
# Print the result
```

```
print(paste("The fraction of people who are truthful clickers answered 'Yes' is", round(fraction, 2), "%"))
```

```
## [1] "The fraction of people who are truthful clickers answered 'Yes' is 71.43 %"
```

## Part B

Imagine a medical test for a disease with the following two attributes:

- The sensitivity is about 0.993. That is, if someone has the disease, there is a probability of 0.993 that they will test positive.
- The specificity is about 0.9999. This means that if someone doesn't have the disease, there is probability of 0.9999 that they will test negative.
- In the general population, incidence of the disease is reasonably rare: about 0.0025% of all people have it (or 0.000025 as a decimal probability).

Suppose someone tests positive. What is the probability that they have the disease?

## Solution:

### Step 1: Define Initial Probabilities:

- $P(D) = 0.000025$  (Probability of having the disease)
- $P(\text{NoD}) = 1 - 0.000025 = 0.999975$  (Probability of not having the disease)

### Sensitivity of the test is 0.993, leading to:

- $P(+Pos D) = 0.993$  (Probability of testing positive given that the individual has the disease)
- $P(-Neg D) = 1 - 0.993 = 0.007$  (Probability of testing negative given that the individual has the disease)

### Specificity of the test is 0.9999, leading to:

- $P(-Neg \text{NoD}) = 0.9999$  (Probability of testing negative given that the individual does not have the disease)
- $P(+Pos \text{NoD}) = 1 - 0.9999 = 0.0001$  (Probability of testing positive given that the individual does not have the disease)

**Question: “Suppose someone tests positive. What is the probability that they actually have the disease?”**

We calculate the following joint probabilities using Bayes' Theorem:

- $P(+Pos \text{ 'and' } D) = P(+Pos D) \times P(D)$
- $P(+Pos \text{ 'and' } \text{NoD}) = P(+Pos \text{NoD}) \times P(\text{NoD})$

```
# Calculating joint probabilities
joint_prob_pos_disease <- 0.993 * 0.000025
joint_prob_pos_no_disease <- 0.0001 * 0.9999

# Calculate the total probability of testing positive
total_prob_positive <- joint_prob_pos_disease + joint_prob_pos_no_disease
```

```
prob_disease_given_pos <- joint_prob_pos_disease / total_prob_positive
prob_no_disease_given_pos <- joint_prob_pos_no_disease / total_prob_positive

# Display the result
prob_disease_given_pos
```

```
## [1] 0.1988944
```

```
print(paste("The probability of someone having the disease given that they tested positive is", round(p
```

```
## [1] "The probability of someone having the disease given that they tested positive is 19.89 %"
```