

Statistics

Introduction to Data Science - 2020 Semester A Final Project

Ori Darshan: 212458244

1 Section 1: Probability and Bayes Theorem

1.1 Question 1

1.1.1 Section a:

Question:

Around $1/125$ of all births are nonidentical twins, and $1/300$ are identical twins. Elvis had a twin brother who died in birth. What is the probability that Elvis had an identical twin?

(you can assume that the probability for boys/girls is $1/2$)

Answer:

We search for the probability of Elvis to have an identical twin. From all the population, the probability is $1/300$ (given). How the information we have changes that number?

We know that Elvis had a twin, from all possibilities, $1/300$ is the chance for an identical and $1/125$ are the chances for a nonidentical twins.

combining these two option we get

$$\frac{1}{300} + \frac{1}{125} = \frac{17}{1500}$$

We won't actually use this number, since with the fact that Elvis had a twin we will look only on the 'twin related' statistics.

From all possibilities for twins, what is the probability for an identical twins?

identical:

$$\frac{1}{300} / \frac{17}{1500} = \frac{5}{17}$$

nonidentical:

$$\frac{1}{125} / \frac{17}{1500} = \frac{12}{17}$$

We know that the twin was a boy, all of the identical twins are the same gender while only $1/2$ of the nonidentical twins will have the same gender (given).

Elvis can be part of the $1/300$ births of identical twins or part of the $(1/125)/2 = 1/250$ births of nonidentical, same-gender births.

The probability for a same gender twin from all possible twins is:

$$\begin{aligned} P(\text{same gender twins}) &= \frac{5}{17} + 0.5 \cdot \frac{12}{17} \\ &= \frac{11}{17} \end{aligned}$$

We will use Bayes theorem to calculate the probability for an identical twin given there was a twin from the same gender, as a base we will use the probabilities shown above for identical twins given there was a twin.

$$P(\text{identical} \mid \text{same gender}) = \frac{P(\text{same gender} \mid \text{identical}) \cdot P(\text{identical})}{P(\text{same gender})}$$

Inserting previous calculations:

$$P(\text{identical} \mid \text{same gender}) = \frac{1 \cdot \frac{5}{17}}{\frac{11}{17}} = \frac{5}{11}$$

The probability Elvis's twin was an identical twin is:

$$\boxed{P = \frac{5}{11}}$$

1.1.2 Section b

Question:

There are two cookie bowls, in bowl 1 there are 10 almond cookies and 30 chocolate cookies. in bowl 2 there are 20 almond cookies and 20 chocolate cookies.

Eric chose a random bowl and took 1 random cookie from it. the cookie was a chocolate cookie.

What is the probability that Eric chose bowl 1?

Answer:

The probability to pick bowl 1 is $1/2$ (bowl 1 or bowl 2), The probability to pick a chocolate cookie is $50/80=5/8$.

Given the cookie was chocolate cookie, the probability for choosing bowl 1 changes according to Bayes theorem:

$$P(\text{bowl 1} \mid \text{chocolate cookie}) = \frac{P(\text{chocolate cookie} \mid \text{bowl 1}) \cdot P(\text{bowl 1})}{P(\text{chocolate cookie})}$$

Inserting known probabilities:

$$P(\text{bowl 1} \mid \text{chocolate cookie}) = \frac{\frac{3}{4} \cdot \frac{1}{2}}{\frac{5}{8}} = \frac{3}{5}$$

The probability Eric chose bowl 1 is:

$$P = \frac{3}{5}$$

1.2 Question 2

Question:

In 1995, M&M company added the blue color. before this year, the color distribution was:

30% Brown, 20% Yellow, 20% Red, 10% Green, 10% Orange, 10% Tan

Since 1995, the distribution looked like that:

24% Blue, 20% Green, 16% Orange, 14% Yellow, 13% Red, 13% Brown

Your friend have 2 M&M bags, one from 1994 and one from 1996 and he is not willing to expose which one belongs to which year. But he gives you one candy from each bag. one candy is yellow and the other one is green.

What is the probability that the yellow candy came from the bag of 1994?

Answer:

At first glance on the distributions, we can see that the green candy is more likely to come from the 1996's bag while the yellow candy is more likely to come from the 1994's bag. Let's calculate that.

Assuming Eric chose the candies at random (otherwise Eric decided what option is true), we need to calculate these parameters in order to use Bayes theorem:

$$P(1 \text{ yellow, } 1 \text{ green} \mid \text{yellow from 1994}), \quad P(\text{yellow from 1994}), \quad P(1 \text{ green, } 1 \text{ yellow})$$

$$P(1 \text{ yellow, } 1 \text{ green} \mid \text{yellow from 1994}) = P(\text{green in 1996}) = 0.2$$

$$P(\text{yellow from 1994}) = 0.2$$

$$\begin{aligned} P(1 \text{ green, } 1 \text{ yellow}) &= P(\text{green 1994, yellow 1996}) + P(\text{green 1996, yellow 1994}) \\ &= 0.1 \cdot 0.14 + 0.2 \cdot 0.2 = 0.054 \end{aligned}$$

Using Bayes theorem, we get:

$$P(\text{yellow from 1994} \mid 1 \text{ green, } 1 \text{ yellow}) = \frac{P(1 \text{ green, } 1 \text{ yellow} \mid \text{yellow from 1994}) \cdot P(\text{yellow from 1994})}{P(1 \text{ green, } 1 \text{ yellow})}$$

With the statistics discovered earlier:

$$P(\text{yellow from 1994} \mid 1 \text{ green, } 1 \text{ yellow}) = \frac{0.2 \cdot 0.2}{0.054} = \frac{20}{27}$$

This result align with our assumptions.

The probability the yellow candy is from the 1994's bag is:

$$P = \frac{20}{27}$$

1.3 Question 3

Question:

You went to the doctor due to a penetrating nail. The doctor chose you at random to make a Swine influenza test. It is known that 1 out of 10,000 are infected by Swine influenza. The test is 99% accurate in the sense that the probability for a false positive is 1%. which means that the probability a healthy man was diagnosed as ill is 1%. The probability for false negative (ill man as healthy) is 0%. You was found positive to Swine influenza.

a: What is the probability you have Swine influenza?

b: Let's say you returned from Thailand lately, and you know that 1 out of 200 people who returned from Thailand lately return with Swine influenza.

Given the same situation like section a. What is the (updated) probability you have Swine influenza?

Answer a:

To use Bayes theorem, we need to find what is the probability you will get a positive test.

$$P(\text{positive})=P(\text{ill})+P(\text{healthy but positive}) = \frac{1}{10000} + \frac{9999}{10000} \cdot \frac{1}{100} = \frac{10099}{1000000} = 0.010099$$

Now using Bayes theorem we get:

$$P(\text{ill} \mid \text{positive}) = \frac{P(\text{positive} \mid \text{ill}) \cdot P(\text{ill})}{P(\text{positive})}$$

$$P(\text{ill} \mid \text{positive}) = \frac{1 \cdot 0.0001}{0.010099} = \frac{100}{10099}$$

Given a positive test, the probability you are ill is:

$$P = \frac{100}{10099}$$

Answer b:

We will update the parameters for $P(\text{ill})$ and $P(\text{positive})$ according to the new information.

The probability for getting a positive result will be:

$$P(\text{positive})=P(\text{ill})+P(\text{healthy but positive}) = \frac{1}{200} + \frac{199}{200} \cdot \frac{1}{100} = \frac{100}{20000} + \frac{199}{20000} = \frac{299}{20000}$$

Now back to Bayes:

$$P(\text{ill} \mid \text{positive}) = \frac{P(\text{positive} \mid \text{ill}) \cdot P(\text{ill})}{P(\text{positive})}$$

$$P(\text{ill} \mid \text{positive}) = \frac{1 \cdot \frac{1}{200}}{\frac{299}{20000}} = \frac{100}{299}$$

Given positive test and that you came back from Thailand, the probability you are ill is:

$$P = \frac{100}{299}$$