Probability

Aleksandr Fisher

4/20/2020

Sample Space and events

- Probability formalizes chance variation or uncertainty in outcomes.
  - It might rain or be sunny today, we don't know which.
  - To formalize, we need to define the set of possible outcomes.
- **Sample space:** Ω the set of possible outcomes.
- **Event:** any subset of outcomes in the sample space

What is probability?

$$\text{P(event)} = \frac{\text{number of outcomes in the event}}{\text{total number of outcomes in the sample space}}$$

- Consider tossing a fair coin:
  - There are two possible outcomes - tossing a head or tossing a tail.
  - The sample space is the set of all possible outcomes, so the sample space is $\{\text{H, T}\}$.
  - Since the coin is fair, the outcomes are equally likely. The probability of the event toss a head is 0.5. In symbols, we can write this as $\text{P(H)} = 0.5$

Probability Distribution

- A **probability distribution** is a list of all of the possible outcomes of a random variable along with their corresponding probability values.

Events can be:

- **Independent** (each event is not affected by other events)
    - A coin does not "know" that it came up heads before . . . each toss of a coin is a perfect isolated thing.
- **Dependent** (also called "Conditional", where an event is affected by other events)
    - After taking one card from the deck there are less cards available, so the probabilities change
- **Mutually Exclusive** (events can't happen at the same time)
    - Heads and Tails are Mutually Exclusive
    - Kings and Hearts are not Mutually Exclusive, because we can have a King of Hearts!

Notation

- $P(A)$: Probability Of Event A
- $P(A^c)$: The probability that Event A will not occur
- $P(A \cap B)$: The probability that Events A and B both occur is the probability of the **intersection** of A and B.
- $P(A \cup B)$: The probability that Events A or B occur is the probability of the **union** of A and B
- $P(B \mid A)$: The **Conditional Probability** of B given A.

Probability Axioms

- Probability quantifies how likely or unlikely events are.
- We'll define the probability P(A) with three axioms:

1. (Nonnegativity) $P(A) \geq 0$ for every event A
2. (Normalization) $P(\Omega) = 1$
3. (Addition Rule) If two events A and B are mutually exclusive $P(A\text{or}B) = P(A) + P(AB)$

Complement of an Evenet

- Given an event A, and its complement A^c, the outcomes in Ω which are not in A, we have the complement rule:

$$P[A^c] = 1 - P[A]$$

-For instance, the probability of not throwing a 3 with a dice is:

$$P[A^c] = 1 - P[A] = 1 - 1/6 = 5/6$$

Union of two event

- For two events A and B we have the **addition rule**:

$$P[A \cup B] = P[A] + P[B] - P[A \cap B]$$

Union of two event

- Suppose that the probability of a fire breaking out in two houses in a given year is:
  - in house A: 60%, so $P(A) = 0.6$
  - in house B: 45%, so $P(B) = 0.45$
  - in at least one of the two houses: 80%, so $P(A \cup B) = 0.8$
- By summing $P(A)$ and $P(B)$, the intersection of A and B, i.e. $P(A \cap B)$, is counted twice. This is the reason we subtract it to count it only once.

Disjoint Events

- If A and B are *disjoint*:

$$P[A \cup B] = P[A] + P[B]$$

Intersection of two events

- If two events are independent, the probability of the intersection of the two events (i.e., the joint probability) is the probability of the two events occurring:

$$P(A \text{and} B) = P[A \cap B] = P[A] * P[B]$$

- For instance, if two coins are flipped, the probability of both coins being tails is:

$$P[Tail_1 \cap Tail_2] = 1/2 * 1/2 = 1/4$$

Conditional Probability

- Often, we would like to understand the probability of an event A, given some information about the outcome of event
- In that case, we have the **conditional probability rule**

$$P[A \mid B] = \frac{P[A \cap B]}{P[B]}$$

- Note that, in general, the probability of A given B is not equal to the probability of B given A, that is, $P(A|B) \neq P(B|A)$

Multiplication Rule

- Rearranging the conditional probability rule, we obtain the multiplication rule:

$$P[A \cap B] = P[B] \cdot P[A \mid B]$$

Bayes Theorem

From the formulas of the conditional probability and the multiplicative law, we can derive the Bayes' theorem:

$$P(B|A) = \frac{P(B \cap A)}{P(A)} \text{ (from the conditional probability)}$$

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

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)} \text{ (from the multiplicative law)}$$

which is equivalent to

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \text{ (Bayes' theorem)}$$

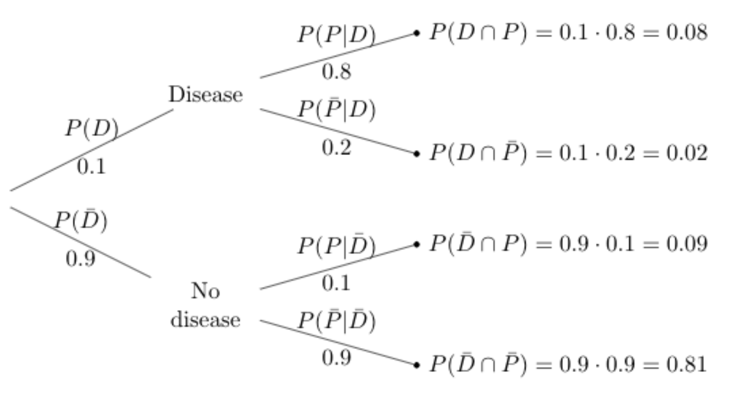
Example - Healthcare

- In order to determine the presence of a disease in a person, a blood test is performed.
- When a person has the disease, the test can reveal the disease in 80% of cases. When the disease is not present, the test is negative in 90% of cases.
- Experience has shown that the probability of the disease being present is 10%.

Example - Healthcare

- A researcher would like to know the probability that an individual has the disease given that the result of the test is positive.
- To answer this question, the following events are defined:
  - P: the test result is positive
  - D: the person has the disease

Example - Healthcare



Applying Bayes Rule

$$P[A \mid B] = \frac{P[A \cap B]}{P[B]}$$

$$P[Dis \mid Pos] = \frac{P[Dis \cap Pos]}{P[Pos]}$$

- From the tree diagram, we can see that $P[Dis \cap Pos] = 0.08$
- A positive test result is possible under two scenarios:
  - (i) when a person has the disease
  - (ii) when the person does not actually have the disease
- In order to find the probability of a positive test result, P(P), we need to sum up those two scenarios:

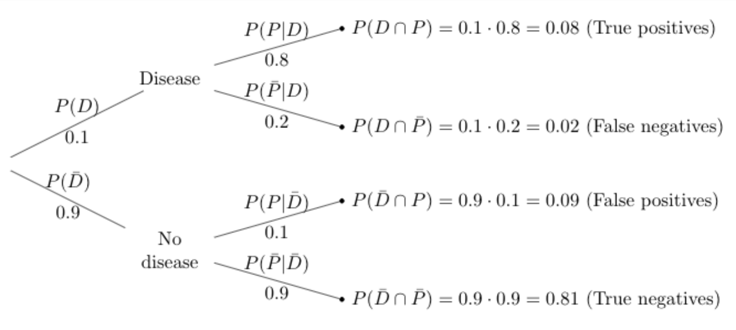
Applying Bayes Rule

$$P(P) = P[Dis \cap Pos] + P[Dis^c \cap Pos] = 0.08 + 0.09 = 0.17$$

$$P[Dis \mid Pos] = \frac{0.08}{0.17} = 0.47$$

- The probability of having the disease given that the result of the test is positive is only 47%.
- This means that in this specific case (with the same percentages), an individual has less than 1 chance out of 2 of having the disease knowing that his test is positive!

Accuracy Measures



False Negatives

- The false negatives (FN) are the number of people incorrectly labeled as not having the disease or the condition, when in reality it is present.
- It is like telling a women who is 7 months pregnant that she is not pregnant.

$$FN = P(D \cap P^c) = 0.02$$

False positives

- The false positives (FP) are the number of people incorrectly labeled as having the disease or the condition, when in reality it is not present.
- It is like telling a man he is pregnant.

$$FP = P(D^c \cap P) = 0.09$$

Sensitivity

- The **sensitivity** of a test, also referred as the **recall**, measures the ability of a test to detect the condition when the condition is present (the percentage of sick people who are correctly identified as having the disease):

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Sensitivity = \frac{0.8}{0.8 + 0.2} = 0.8$$

Specificity

- The specificity of a test measures the ability of a test to correctly exclude the condition when the condition is absent (the percentage of healthy people who are correctly identified as not having the disease):

$$Specificity = \frac{TN}{TN + FP}$$

$$Sensitivity = \frac{0.9}{0.9 + 0.1} = 0.9$$

Positive Predicted Value

- The positive predictive value, also referred as the **precision**, is the proportion of positives that correspond to the presence of the condition, so the proportions of positive results that are true positive results:

$$PPV = \frac{TP}{TP + FP} = 0.476$$

Negative Predicted Value

- The negative predictive value is the proportion of negatives that correspond to the absence of the condition, so the proportions of negative results that are true negative results:

$$NPV = \frac{TN}{TN + FN} = 0.476$$

- From the tree diagram we have:

$$NPV = \frac{TN}{TN + FN} = \frac{P[Dis^c \cap Pos^c]}{P(P^c)} = \frac{0.81}{0.81 + 0.02} = 0.98$$

Expected Values and Variance

- Most of the time we want to know what the expected value of a distribution is and its variance.
- The expected value of the binomial is the **mean** of the distribution.

$$E(X) = size * p$$

- The **variance** is defined as: The average of the squared differences from the Mean.
- To calculate the variance follow these steps: Work out the Mean (the simple average of the numbers) Then for each number: subtract the Mean and square the result (the squared difference).

Frequentist vs. Bayesian

- **Frequentist**: probabilities reflect relative frequency in a large number of trials
- **Bayesian**: probabilities are subjective beliefs about outcomes
- Don't worry about it ... we'll be sticking to frequentism in this class

See also:

- https://tophat.com/marketplace/science-&-math/statistics/full-course/statistics-for-social-science-stephen-hayward/211/34407/
- https://towardsdatascience.com/the-9-concepts-and-formulas-in-probability-that-every-data-scientist-should-know-a0eb8d3ea8c4