Probability

CS 556

Probabilistic Models

- **Definition**: Probability is a branch of mathematics focusing on numerical descriptions of **how likely an event is to occur**.

- A probabilistic model is a **mathematical description of an uncertain situation**

Elements of a Probabilistic Model

- The sample space Ω, which is the set of all possible outcomes of an experiment.

- The probability law , which assigns a set A of possible outcomes (also called an event) a nonnegative number P(A) (called the probability of A) that specifies the “likelihood" of the elements of A.

The Axioms of Probability

- **Nonnegativity:** $0 \leq P(A) \leq 1$, for every event A

- **Additivity**: If A and B are two disjoint (i.e., *independent*) events, then the probability of their union satisfies:
$$P(A \wedge B) = P(A) + P(B)$$

- **Normalization:** Probability of the entire sample space is equal to 1, $P(\Omega) = 1$

- $P(A \mid B) = P(A) + P(B) - P(A \wedge B)$

Example 1

(Three Coin Tosses)

- Consider an experiment involving three coin tosses. The outcomes will be a 3 -long string of heads or tails.

Ω = {HHH, HHT, HTH,THH,THT,TTH,HTT,TTT}

- We assume that each possible outcome has the same probability $\frac{1}{8}$.

  What is the **probability of getting exactly two heads**?

Example 1

(Three Coin Tosses)

- Ω = {HHH, HHT, HTH,THH,THT,TTH,HTT,TTT}

- A = {exactly 2 heads occurs}
  = {HHT, HTH, THH}

- P(A) = P({HHT, HTH, THH})
  = P({HHT}) + P({HTH}) + P({THH})
  = 1/8 + 1/8 + 1/8
  = 3/8

Example 2

Consider the experiment of rolling a pair of 4 -sided fair dice. Compute the following probabilities:

- P({the sum of the rolls is even})
- P({the sum of the rolls is odd})
- P({the first roll is equal to the second})
- P({the first roll is larger than the second})
- P({at least one roll is equal to 4})

Example 2

- P({the sum of the rolls is even}) = $\frac{8}{16}$
- P({the sum of the rolls is odd}) = $\frac{8}{16}$
- P({the first roll is equal to the second}) = $\frac{4}{16}$
- P({the first roll is larger than the second}) = $\frac{6}{16}$
- P({at least one roll is equal to 4}) = $\frac{7}{16}$

Pairs of Events

- Consider two events A, B

- For each possible pair of values, we can define a *joint probability* :

$$p_{ij} = Pr[A = a_i, B = b_j]$$

Statistical Independence

- Two events A and B are said to be independent, if and only if

$$P(A \cap B) = P(A)P(B)$$

- That is, when knowing the value of A does not give us additional information for the value of B.

4-sided die example, probability of first roll = 1 and second roll = 1 ?

Conditional Probability

- When two events are not independent, knowing one allows better estimate of the other (e.g. outside temperature, season)

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Temperature

Season

- If independent $P(A|B)=P(A)$

Read: Probability of A given B

Example 3

- We toss a fair coin three successive times. We wish to find the condition probability P(A|B) where A and B are the events

- A={more heads than tails come up}

- B = {1st toss is a head}

Example 3

Ω = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

B = {HHH, HHT, HTH, HTT}

P(B) = $\frac{4}{8}$

(A ∩ B) = {HHH, HHT, HTH}

P(A∩B) = $\frac{3}{8}$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{3}{8}}{\frac{4}{8}} = \frac{3}{4}$$

Conditional Probability Example

- A jar contains black and white marbles. Two marbles are chosen without replacement. The probability of selecting a black marble and then a white marble is 0.34. The probability of selecting a black marble on the first draw is 0.47.

- What is the probability of selecting a white marble on the second draw, given that the first marble drawn was black?

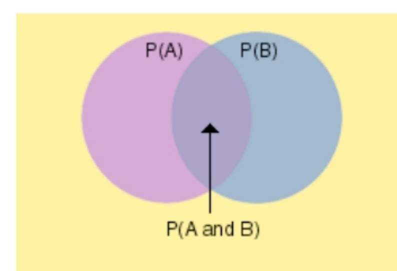
Conditional Probability Example

- A jar contains black and white marbles. Two marbles are chosen without replacement. The probability of selecting a black marble and then a white marble is 0.34. The probability of selecting a black marble on the first draw is 0.47.

- What is the probability of selecting a white marble on the second draw, given that the first marble drawn was black?

$$P(\text{White} \mid \text{Black}) = \frac{P(\text{Black} \cap \text{White})}{P(\text{Black})} = \frac{0.34}{0.47} = 0.72$$

A is black in first draw, B is white in second draw

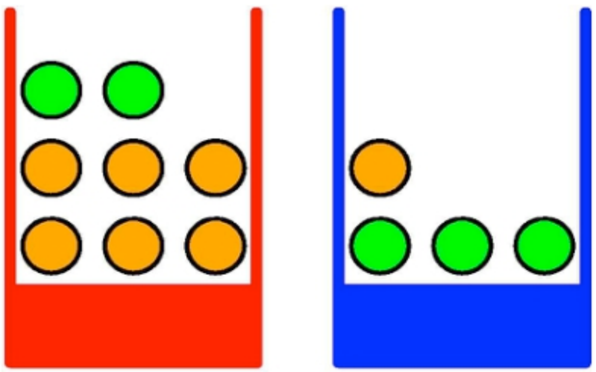


Sum and Product Rules (1/7)

Example:

- We have two boxes: one red and one blue
- Red box: 2 apples and 6 oranges
- Blue box: 3 apples and 1 orange
- Pick red box 40% of the time and blue box 60% of the time, then pick one item of fruit

What is the probability of picking an apple?



C.M. Bishop, "Pattern Recognition and Machine Learning", 2006

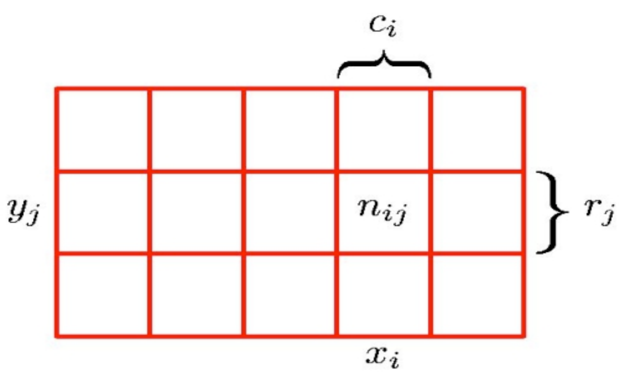
Sum and Product Rules (2/7)

- Define:
  - B: *random variable* for box picked (r or b)
  - F: *random variable* for identity of fruit (a or o)

- p(B=r)= $\frac{4}{10}$ & p(B=b)= $\frac{6}{10}$

- Events are mutually exclusive and include all possible outcomes => their probabilities must sum to 1

Sum and Product Rules (3/7)



$$p(X = x_i) = \frac{c_i}{N}.$$

Marginal Probability

$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$

Joint Probability

$$p(Y = y_j | X = x_i) = \frac{n_{ij}}{c_i}$$

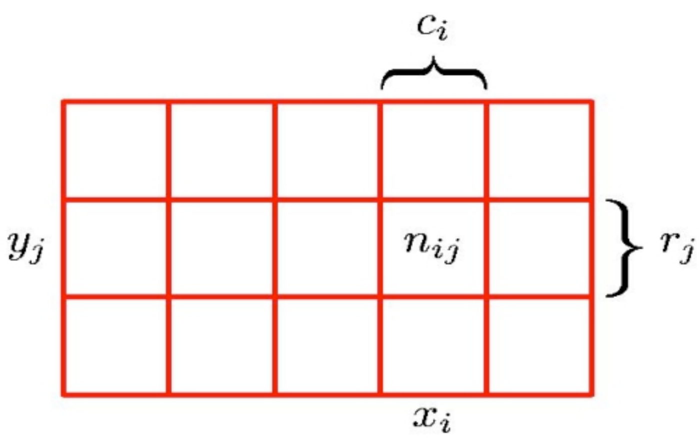
Conditional Probability

N = Size of Sample Space

n_{ij} is the number of occurrences of $X = x_i$ & $Y = y_j$

C.M. Bishop, “Pattern Recognition and Machine Learning”, 2006

Sum and Product Rules (4/7)



Sum Rule

$$p(X = x_i) = \frac{c_i}{N} = \frac{1}{N}\sum_{j=1}^{L} n_{ij}$$

$$= \sum_{j=1}^{L} p(X = x_i, Y = y_j)$$

Product Rule

$$\begin{aligned} p(X = x_i, Y = y_j) &= \frac{n_{ij}}{N} = \frac{n_{ij}}{c_i} \cdot \frac{c_i}{N} \\ &= p(Y = y_j | X = x_i) p(X = x_i) \end{aligned}$$

Sum and Product Rules (5/7)

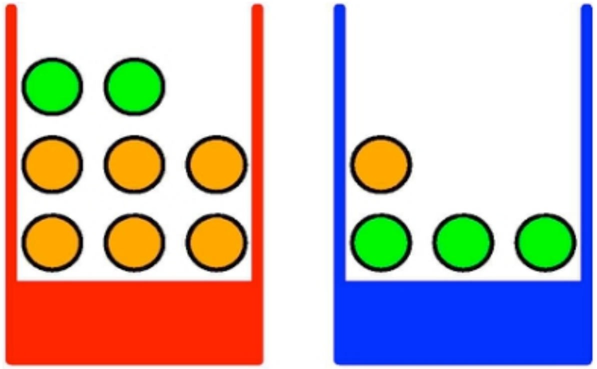
- Sum Rule $$p(X) = \sum_{Y} p(X, Y)$$
- Product Rule $$p(X, Y) = p(Y \mid X)\, p(X)$$

Law of Total Probability

- If an event A can occur in m different ways and if these m different ways are mutually exclusive, then the probability of A occurring is the sum of the probabilities of the sub-events

$$P(X = x_i) = \sum_j P(X = x_i \mid Y = y_j)\, P(Y = y_j)$$

Sum and Product Rules (6/7)



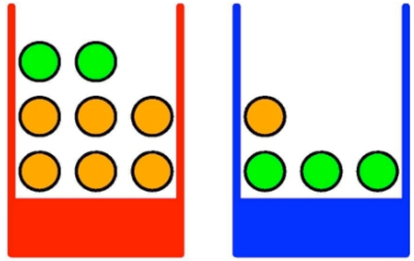
- Back to the fruit baskets
  - $p(B=r)= \dfrac{4}{10}$ and $p(B=b)= \dfrac{6}{10}$
  - $p(B=r) + p(B=b) = 1$
- Conditional probabilities

$$(1)\ p(F=a \mid B = r) = \frac{1}{4} \qquad (2)\ p(F=o \mid B = r) = \frac{3}{4}$$

$$(3)\ p(F=a \mid B = b) = \frac{3}{4} \qquad (4)\ p(F=o \mid B = b) = \frac{1}{4}$$

Sum and Product Rules (7/7)

Note: $p(F = a \mid B = r) + p(F = o \mid B = r) = 1$

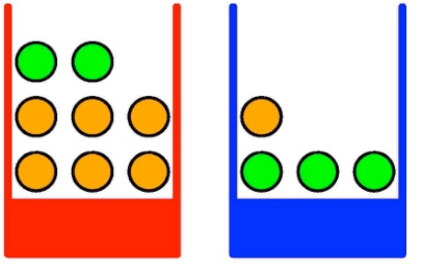


What is the probability of picking an apple?

$p(F=a) = ?$

Sum and Product Rules (7/7)

Note: $p(F= a \mid B = r) + p(F=o \mid B = r) = 1$



What is the probability of picking an apple?

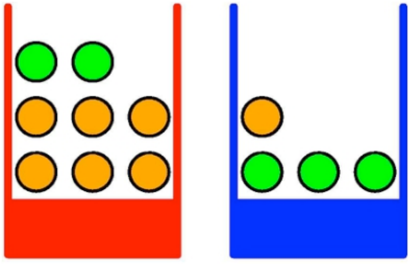
$$p(F=a) = p(F=a \mid B = r)\ p(B=r) + p(F=a \mid B = b)\ p(B=b)$$
$$= 1/4 * 4/10 + 3/4 * 6/10 = 11/20$$

What is the probability of picking an orange?

$p(F=o) = ?$

Sum and Product Rules (7/7)

Note: $p(F = a \mid B = r) + p(F = o \mid B = r) = 1$



What is the probability of picking an apple?

$$p(F=a) = p(F=a \mid B = r)\, p(B=r) + p(F=a \mid B = b)\, p(B=b)$$
$$= 1/4 * 4/10 + 3/4 * 6/10 = 11/20$$

What is the probability of picking an orange?

$$p(F=o) = p(F=o \mid B = r)\, p(B=r) + p(F=o \mid B = b)\, p(B=b)$$
$$= 3/4*4/10 + 1/4*6/10 = 3/10 + 3/20 = 9/20$$

C.M. Bishop, "Pattern Recognition and Machine Learning", 2006

Law of Total Probability

$$P_X(x) = \sum_{y \in Y} P(x, y)$$

$$P(x \mid y) = \frac{P(x, y)}{P(y)}$$

Bayes Rule

Product Rule

$$P(x \mid y) = \frac{P(x,y)}{P(y)} = \frac{P(y \mid x)P(x)}{\sum_{x \in X} P(y \mid x)P(x)}$$

$$\text{posterior} = \frac{\text{likelihood * prior}}{\text{evidence}}$$

- x is the unknown cause
- y is the observed evidence
- Bayes rule shows how probability of x changes after we have observed y

Bayes Rule on the Fruit Example

- Suppose we have selected an orange.
  Which box did it come from?

- What is the probability it came from the red box?

$$p(B = r \,|\, F = o) = \frac{p(F = o \,|\, B = r)P(B = r)}{p(F = o)} = \frac{\frac{3}{4} \times \frac{4}{10}}{\frac{9}{20}} = \frac{2}{3}$$