CSE 400: Fundamentals of Probability in Computing

Lecture 5 — Bayes' Theorem, Random Variables, and Probability Mass Function

1 Bayes' Theorem

1.1 Weighted Average of Conditional Probabilities

Let A and B be events.

We may express the event A as

$$A = AB \cup AB^c$$

since, for an outcome to be in A, it must either be in both A and B, or be in A but not in B.

The events AB and AB^c are mutually exclusive. By Axiom 3,

$$\Pr(A) = \Pr(AB) + \Pr(AB^c)$$

Using the definition of conditional probability,

$$\Pr(AB) = \Pr(A \mid B)\Pr(B)$$

$$\Pr(AB^c) = \Pr(A \mid B^c)\Pr(B^c)$$

Hence,

$$\Pr(A) = \Pr(A \mid B)\Pr(B) + \Pr(A \mid B^c)\Pr(B^c)$$

$$= \Pr(A \mid B)\Pr(B) + \Pr(A \mid B^c)[1 - \Pr(B)]$$

Interpretation: The probability of event A is a weighted average of conditional probabilities, where the weights are the probabilities of the events on which the conditioning is done.

1.2 Learning by Example — Example 3.1 (Part 1)

Problem Statement

An insurance company classifies people into two categories:

Accident prone

Not accident prone

Given:

$$\Pr(\text{Accident within 1 year} \mid \text{Accident prone}) = 0.4$$

$$\Pr(\text{Accident within 1 year} \mid \text{Not accident prone}) = 0.2$$

$$\Pr(\text{Accident prone}) = 0.3$$

Find the probability that a new policyholder will have an accident within one year.

Definitions

Let:

A_1: policyholder has an accident within one year

A: policyholder is accident prone

Solution

Condition on whether the policyholder is accident prone:

$$\Pr(A_1) = \Pr(A_1 \mid A)\Pr(A) + \Pr(A_1 \mid A^c)\Pr(A^c)$$

Substitute values:

$$\Pr(A_1) = (0.4)(0.3) + (0.2)(0.7)$$

$$= 0.12 + 0.14$$

$$= 0.26$$

1.3 Learning by Example — Example 3.1 (Part 2)

Problem Statement

Suppose a new policyholder has had an accident within one year. What is the probability that the policyholder is accident prone?

Solution

We seek:

$$\Pr(A \mid A_1)$$

By definition of conditional probability:

$$\Pr(A \mid A_1) = \frac{\Pr(A \cap A_1)}{\Pr(A_1)}$$

Using multiplication rule:

$$\Pr(A \cap A_1) = \Pr(A)\Pr(A_1 \mid A)$$

Thus,

$$\Pr(A \mid A_1) = \frac{(0.3)(0.4)}{0.26}$$

$$= \frac{0.12}{0.26}$$

$$= \frac{6}{13}$$

1.4 Formal Introduction: Law of Total Probability

Suppose:

$$B_1, B_2, \dots, B_n$$

are mutually exclusive events

$$\bigcup_{i=1}^{n} B_i = B$$

Exactly one of the events $B_1, \dots, B_n$ must occur.
Writing:

$$A = \bigcup_{i=1}^{n} AB_i$$

and noting that the events AB_i are mutually exclusive, we have:

$$\Pr(A) = \sum_{i=1}^{n} \Pr(AB_i)$$

Using conditional probability:

$$\Pr(AB_i) = \Pr(A \mid B_i) \Pr(B_i)$$

Hence,

$$\Pr(A) = \sum_{i=1}^{n} \Pr(A \mid B_i) \Pr(B_i)$$

This is the Law of Total Probability (Formula 3.4).

1.5 Bayes' Formula (Proposition 3.1)

Using:

$$\Pr(AB_i) = \Pr(B_i \mid A) \Pr(A)$$

and substituting into the law of total probability, we obtain:

$$\Pr(B_i \mid A) = \frac{\Pr(A \mid B_i) \Pr(B_i)}{\sum_{j=1}^{n} \Pr(A \mid B_j) \Pr(B_j)}$$

Where:

$$\Pr(B_i)$$

is the a priori probability

$$\Pr(B_i \mid A)$$

is the posterior probability

1.6 Learning by Example — Example 3.2 (Cards Problem)

Setup

Three cards:
One card: red–red (RR)
One card: black–black (BB)
One card: red–black (RB)
A card is randomly selected and placed down. The upper side is observed to be red.
Find the probability that the other side is black.

Definitions

Let:
RR, BB, RB: type of selected card
R: upturned side is red

Solution

$$\Pr(RB \mid R) = \frac{\Pr(R \mid RB)\Pr(RB)}{\Pr(R \mid RR)\Pr(RR) + \Pr(R \mid RB)\Pr(RB) + \Pr(R \mid BB)\Pr(BB)}$$

Substitute values:

$$= \frac{(1/2)(1/3)}{(1)(1/3) + (1/2)(1/3) + (0)(1/3)}$$

Simplify:

$$= \frac{1/6}{1/3 + 1/6}$$

$$= \frac{1/6}{1/2}$$

$$= \frac{1}{3}$$

2 Random Variables

2.1 Motivation and Concept

When an experiment is performed, interest often lies in a function of the outcome rather than the outcome itself.

Examples:
Tossing dice: sum of values

Tossing coins: number of heads

These real-valued functions defined on the sample space are called random variables.

A random variable assigns:

A real number to each outcome

Probabilities to possible values

2.2 Definition

A random variable X on a sample space Ω is a function:

$$X : \Omega \to \mathbb{R}$$

that assigns to each sample point $\omega \in \Omega$ a real number $X(\omega)$.

In this lecture, attention is restricted to discrete random variables, whose values form a finite or countably infinite set.

2.3 Distribution of a Random Variable

Two key components:

The set of values the random variable can take

The probabilities with which it takes those values

For a value a,

$$\{\omega \in \Omega : X(\omega) = a\}$$

is an event, denoted $\{X = a\}$.

The probability:

$$\Pr[X = a]$$

is defined via the probability of the corresponding event.

The collection of these probabilities over all possible values constitutes the distribution of X.

2.4 Example — Tossing 3 Fair Coins

Let Y be the number of heads.

Possible values:

$$Y \in \{0, 1, 2, 3\}$$

Probabilities:

$$\Pr(Y = 0) = \Pr(t, t, t) = \frac{1}{8}$$

$$\Pr(Y = 1) = \Pr(t, t, h), (t, h, t), (h, t, t) = \frac{3}{8}$$

$$\Pr(Y = 2) = \Pr(t,h,h),(h,t,h),(h,h,t) = \frac{3}{8}$$

$$\Pr(Y = 3) = \Pr(h,h,h) = \frac{1}{8}$$

Since Y must take one of these values:

$$\sum_{i=0}^{3} \Pr(Y = i) = 1$$

3 Probability Mass Function (PMF)

3.1 Concept

A random variable that takes at most a countable number of possible values is called discrete.

Let X be a discrete random variable with range:

$$R_X = \{x_1, x_2, x_3, \ldots\}$$

The function:

$$p_X(x_k) = \Pr(X = x_k)$$

is called the Probability Mass Function (PMF) of X.

Since X must take one of its possible values:

$$\sum_k p_X(x_k) = 1$$

3.2 Example — Two Independent Tosses of a Fair Coin

Sample space:

$$\Omega = \{(H,H),(H,T),(T,H),(T,T)\}$$

Let:

$$X = \text{number of heads}$$

PMF:

$$p_X(x) = \begin{cases} \frac{1}{4}, & x = 0 \text{ or } x = 2 \\ \frac{1}{2}, & x = 1 \\ 0, & \text{otherwise} \end{cases}$$

3.3 Example — Given PMF

Given:

$$p(i) = c\frac{\lambda^i}{i!}, \quad i = 0, 1, 2, \ldots$$

Since:

$$\sum_{i=0}^{\infty} p(i) = 1$$

we have:

$$c\sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = 1$$

Using:

$$\sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = e^{\lambda}$$

Thus:

$$c = e^{-\lambda}$$

Hence:

$$\Pr(X = 0) = c = e^{-\lambda}$$

$$\Pr(X > 2) = 1 - [\Pr(X = 0) + \Pr(X = 1) + \Pr(X = 2)]$$

End of Lecture 5 Scribe