CSE400: Fundamentals of Probability in Computing

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

Name: Masoom Choksi
AUID: AU2440263

Bayes' Theorem

Weighted Average of Conditional Probabilities

Let A and B be events. We may express the event A as

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

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

Since AB and AB^c are mutually exclusive, by Axiom 3 of probability,

$$\begin{aligned}
\Pr(A) &= \Pr(AB) + \Pr(AB^c) \\
&= \Pr(A \mid B)\Pr(B) + \Pr(A \mid B^c)\Pr(B^c) \\
&= \Pr(A \mid B)\Pr(B) + \Pr(A \mid B^c)\left[1 - \Pr(B)\right].
\end{aligned}$$

Thus, the probability of event A is a weighted average of the conditional probabilities, with weights given by the probabilities of the events on which it is conditioned.

Learning by Example

Example 3.1 (Part 1) An insurance company classifies people into two categories:

- Accident-prone
- Not accident-prone

The probability that an accident-prone person has an accident within a 1-year period is 0.4, while this probability is 0.2 for a person who is not accident-prone. It is assumed that 30% of the population is accident-prone.

Question: What is the probability that a new policyholder will have an accident within a year?

Solution:

Let

- A_1 denote the event that the policyholder has an accident within a year,
- A denote the event that the policyholder is accident-prone.

Then,

$$\begin{aligned}
\Pr(A_1) &= \Pr(A_1 \mid A)\Pr(A) + \Pr(A_1 \mid A^c)\Pr(A^c) \\
&= (0.4)(0.3) + (0.2)(0.7) \\
&= 0.12 + 0.14 \\
&= 0.26.
\end{aligned}$$

Example 3.1 (Part 2) Suppose that a new policyholder has an accident within a year. What is the probability that the policyholder is accident-prone?

Solution:

We compute

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

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

$$\begin{aligned}
\Pr(A \mid A_1) &= \frac{\Pr(A)\Pr(A_1 \mid A)}{\Pr(A_1)} \\
&= \frac{(0.3)(0.4)}{0.26} \\
&= \frac{6}{13}.
\end{aligned}$$

Formal Introduction

Law of Total Probability

Let $B_1, B_2, \ldots, B_n$ be mutually exclusive and exhaustive events such that

$$\bigcup_{i=1}^{n} B_i = S.$$

Then for any event A,

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

This is known as the **Law of Total Probability**.

Bayes' Formula

Using

$$\Pr(A \cap B_i) = \Pr(B_i \mid A)\Pr(A),$$

we obtain

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

This is known as **Bayes' Formula** (Proposition 3.1), where:

- $\Pr(B_i)$ is the *a priori* probability,
- $\Pr(B_i \mid A)$ is the *a posteriori* probability after observing event A.

Example 3.2

Three cards are identical in form:

- One card has both sides red (RR),
- One card has both sides black (BB),
- One card has one red and one black side (RB).

One card is randomly selected and placed on the ground. If the upper side is red, what is the probability that the other side is black?

Solution:

Let

- RR, RB, BB denote the events that the selected card is all-red, red-black, or all-black respectively,
- R denote the event that the upturned side is red.

We compute

$$\Pr(RB \mid R) = \frac{\Pr(R \mid RB)\Pr(RB)}{\Pr(R)}.$$

Now,

$$\Pr(R) = \Pr(R \mid RR)\Pr(RR) + \Pr(R \mid RB)\Pr(RB) + \Pr(R \mid BB)\Pr(BB).$$

Substituting values,

$$\begin{aligned}\Pr(RB \mid R) &= \frac{(1/2)(1/3)}{(1)(1/3) + (1/2)(1/3) + (0)(1/3)} \\ &= \frac{1/6}{1/2} \\ &= \frac{1}{3}.\end{aligned}$$

Random Variables

Motivation and Concept

In many experiments, we are interested in a function of the outcome rather than the outcome itself.

- In tossing two dice, the sum is often of interest.
- In flipping coins, the number of heads may be of interest.

A **random variable** is a real-valued function defined on the sample space. Its value is determined by the outcome of an experiment, and probabilities are assigned to its possible values.

Example

Suppose three fair coins are tossed. Let Y denote the number of heads observed.

Then Y takes values $\{0, 1, 2, 3\}$ with probabilities:

$$\begin{aligned}
\Pr(Y = 0) &= \Pr(t, t, t) = \frac{1}{8}, \\
\Pr(Y = 1) &= \Pr(t, t, h), (t, h, t), (h, t, t) = \frac{3}{8}, \\
\Pr(Y = 2) &= \Pr(t, h, h), (h, t, h), (h, h, t) = \frac{3}{8}, \\
\Pr(Y = 3) &= \Pr(h, h, h) = \frac{1}{8}.
\end{aligned}$$

Since Y must take one of these values,

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

Probability Mass Function

Concept

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

Let X be a discrete random variable with range

$$R_X = \{x_1, x_2, x_3, \dots\}.$$

The function

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

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

Since X must take one of the values in R_X,

$$\sum_k p(x_k) = 1.$$

Example

The PMF of a random variable X is given by

$$p(i) = c\lambda^i, \quad i = 0, 1, 2, \dots,$$

where $\lambda > 0$.

Since probabilities must sum to 1,

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

This gives

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

hence

$$c = 1 - \lambda.$$

Therefore,

$$\Pr(X = 0) = p(0) = 1 - \lambda,$$

and

$$\begin{aligned}\Pr(X > 2) &= 1 - \Pr(X \leq 2) \\ &= 1 - \sum_{i=0}^{2} (1-\lambda)\lambda^i.\end{aligned}$$