

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

1 Bayes' Theorem

1.1 Weighted Average of Conditional Probabilities

Definitions and Notation

Let A and B be events in a probability space.

The event A can be expressed as:

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

because any outcome in A must either:

1. Belong to both A and B , or
2. Belong to A but not B .

Assumptions

- AB and AB^c are mutually exclusive events.
- Probability axioms hold (especially additivity for mutually exclusive events).

Derivation

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

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

Using conditional probability:

$$Pr(AB) = Pr(A|B)Pr(B), \quad Pr(AB^c) = Pr(A|B^c)Pr(B^c)$$

Hence:

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

Result

The probability of event A is a weighted average of conditional probabilities, where the weights are the probabilities of the conditioning events.

1.2 Learning by Example — Insurance Example

Problem Statement

Population divided into:

- Accident-prone
- Not accident-prone

Given:

$$Pr(\text{accident} \mid \text{accident-prone}) = 0.4$$

$$Pr(\text{accident} \mid \text{not accident-prone}) = 0.2$$

$$Pr(\text{accident-prone}) = 0.3$$

Find probability a new policyholder has an accident in one year.

Solution

Let:

- A_1 : accident occurs within one year
- A : person is accident-prone

Using total probability:

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

Follow-up: Posterior Probability

Given accident occurred, find probability person is accident-prone.

$$Pr(A|A_1) = \frac{Pr(AA_1)}{Pr(A_1)} = \frac{Pr(A)Pr(A_1|A)}{Pr(A_1)} = \frac{(0.3)(0.4)}{0.26} = \frac{13}{6}$$

This illustrates Bayes' reasoning: updating probability after observing an event.

1.3 Law of Total Probability

Assumptions

Let B_1, B_2, \dots, B_n be:

- Mutually exclusive
- Exactly one of these events occurs.

Representation

Since B_i are mutually exclusive:

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

Using conditional probability:

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

This is the Law of Total Probability.

1.4 Bayes Formula (Proposition)

Using:

$$Pr(AB_i) = Pr(B_i | A)Pr(A)$$

We obtain:

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

Interpretation

- $Pr(B_i)$: a priori probability
- $Pr(B_i | A)$: posteriori probability after observing A .

1.5 Example — Card Problem

Setup

Three cards:

1. RR (both sides red)
2. BB (both sides black)
3. RB (one red, one black)

A card selected randomly and placed face up. Upper side is red.
Find probability other side is black.

Notation

- RR, BB, RB: events representing card types
- R : upturned side is red

Computation

Using conditional probabilities:

$$Pr(RB|R) = \frac{Pr(RB \cap R)}{Pr(R)}$$

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

Thus probability other side is black is $\frac{1}{3}$.

2 Random Variables

2.1 Motivation

In experiments, interest often lies in a function of outcomes:

- Dice \rightarrow sum of values
- Coins \rightarrow number of heads

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

2.2 Definition

A random variable on sample space Ω is a function:

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

assigning each sample point ω a real number $X(\omega)$.

Restriction

Focus is on discrete random variables:

- Finite or countably infinite values.

2.3 Distribution of a Random Variable

Important components:

1. Set of values X can take.
2. Probabilities of those values.

For value a :

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

is an event; hence:

$$Pr(X = a)$$

Collection of these probabilities forms the distribution of X .

2.4 Visualization

Distribution represented as bar diagram:

- x-axis: possible values
- bar height at a : computed from probabilities of corresponding events.

2.5 Discrete vs Continuous

Discrete

- Countable support
- Probability mass function
- Probabilities assigned to single values

Continuous

- Uncountable support
- Probability density function
- Probabilities assigned to intervals
- Individual points have probability zero.

2.6 Example — Tossing 3 Fair Coins

Let Y : number of heads.

Possible values:

$$0, 1, 2, 3$$

Probabilities:

$$P(Y = 0) = \frac{1}{8}, \quad P(Y = 1) = \frac{3}{8}, \quad P(Y = 2) = \frac{3}{8}, \quad P(Y = 3) = \frac{1}{8}$$

Since Y must take one of these values:

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

3 Probability Mass Function (PMF)

3.1 Definition

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

Let discrete random variable have range:

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

Function:

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

is called the Probability Mass Function (PMF).

Since X must take one of the values:

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

3.2 Example — Two Tosses of a Fair Coin

Sample space:

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

Let X : number of heads.

PMF:

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

Example computation:

$$P(X > 0) = P(X = 1) + P(X = 2) = 1/2 + 1/4 = 3/4$$

3.3 Example — Given PMF

Given:

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

Step 1: Use normalization

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

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

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

$$ce^\lambda = 1 \Rightarrow c = \frac{1}{e^\lambda}$$

Step 2: Find probabilities

$$p(i) = \frac{1}{e^\lambda} \frac{\lambda^i}{i!}$$

$$P(X = 0) = c = \frac{1}{e^\lambda}$$

$$P(X > 2) = 1 - [p(0) + p(1) + p(2)]$$

4 End of Lecture 5 Key Takeaways

1. Probability of an event can be written as a weighted sum of conditional probabilities.
2. Law of Total Probability connects partitioned events to overall probability.
3. Bayes' theorem updates prior probabilities using observed evidence.
4. Random variables map outcomes to real numbers; distributions assign probabilities to their values.
5. PMF describes probabilities for discrete random variables and must sum to 1.