Lecture 6: Uncertainty and Information

Dr. Roman V Belavkin

BIS3226

Contents

1 Introduction

Sources of Uncertainty

Complexity : the number of possible states of a system in question can be too large (e.g. predict how a chess game can develop after 10 moves?)

Ignorance : some important information about the system may not be available

Randomness : the system may be random by nature, and thus the uncertainty is irreducible.

Historical Background

1654 Blaise Pascal and Pierre Fermat

1657 Christian Huygens publishes *On Ratiocination in Dice Games*

1760 Thomas Bayes (conditional probability)

1812 Pierre–Simon Laplace

1933 Andrey Kolmogorov's axioms

1920–1940 Ronald Fisher, Abraham Wald (statistics)

1948 Claude Shannon (Information theory)

2 What is probability?

What is Probability?

Definition 1. The uncertainty about some event E can range from impossible to certain. Let us denote the *probability* of event E by $P(E)$ such that

$P(E) = 0$ means E is impossible;

$P(E) = 1$ means E is certain.

Thus, probability is a number between 0 and 1.

$$\text{(Impossible)} \qquad 0 \leq P(E) \leq 1 \qquad \text{(Certain)}$$

Example 2. For a fair coin, $P(heads) = \frac{1}{2} = 0.5$

Set-Theoretic Intuition

- Events E are considered as subsets of the universal set U:

$$E \subseteq U$$

- Probability is a *measure* of a subset $E \subseteq U$ such that

$$P(U) = 1$$

  (because the universe is certain).

- We can consider probabilities of negation, disjunction and conjunction of events:

$$P(\text{not } E)\,, \quad P(E_1 \text{ or } E_2)\,, \quad P(E_1 \text{ and } E_2)$$

- This corresponds to probability measure of the complement, union and intersection of the subsets:

$$P(\bar{E}) = P(U - E)\,, \quad P(E_1 \cup E_2)\,, \quad P(E_1 \cap E_2)$$

Additivity of Probabilities

- Events E_1 and E_2 are *disjoint* if $E_1 \cap E_2 = \emptyset$.
- For disjoint events

$$P(E_1 \text{ or } E_2) = P(E_1) + P(E_2)$$

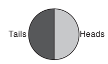
- For n disjoint events such that $E_1 \cup E_2 \cup \cdots \cup E_n = U$

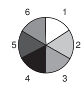
$$P(E_1) + P(E_2) + \cdots + P(E_n) = 1$$

  (because at least one of the events is certain)

Example 3. For a fair coin and a fair dice we have

$$\frac{1}{2} + \frac{1}{2} = 1 \qquad\qquad \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = 1$$





Probability of Negation

- Negation of an event 'not E' refers to its complement

$$\bar{E} = U - E$$

- E and $\bar{E}$ are disjoint and such that

$$E \cup \bar{E} = E \cup (U - E) = U$$

- Because $P(E \text{ or not } E) = P(E) + P(\text{not } E) = 1$ we have

$$P(\text{not } E) = 1 - P(E)$$

Question 1. *What is the probability of a void set $P(\emptyset)$?*

Joint Probability

- Co-occurrence of events E_1 and E_2 (e.g. clouds and rain) is their set intersection:

$$E_1 \cap E_2$$

- Probability of $E_1 \cap E_2$ is called *joint* probability and denoted $P(E_1 \cap E_2)$ or simply $P(E_1, E_2)$.
- For disjoint events $P(E_1 \cap E_2) = P(\emptyset) = 0$.

Example 4 (Two coins).

E_1	E_2
heads	heads
heads	tails
tails	heads
tails	tails

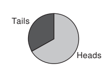
Where do Probabilities Come From?

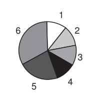
- If there are n disjoint events, then we could assume that all

$$P(E_1) = P(E_2) = \cdots = P(E_n) = \frac{1}{n}$$

- It would be much better to use the *empirical frequency* function

$$P(E_i) \approx \frac{n(E_i)}{n} = \frac{\text{no. of times event } E_i \text{ occurs}}{\text{no. of independent tests}}$$

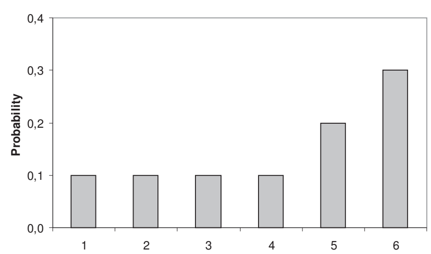




Example 5. Flip a coin or roll a dice several times to estimate the probabilities.

Probability Distributions

We can plot probabilities of all events on a graph, which shows probability **distribution**



3 Conditional Probability and Independence

Conditional Probability

Question 2. *How likely is it to rain if you see clouds?*

- Probability of E_1 *conditional* to event E_2 (i.e. if E_2 has happened):

$$P(E_1 \mid E_2)$$

- Joint probability $P(E_1 \cap E_2)$ can be computed using conditional by multiplication:

$$\begin{aligned} P(E_1 \cap E_2) &= P(E_1 \mid E_2)P(E_2) \\ P(E_1 \cap E_2) &= P(E_2 \mid E_1)P(E_1) \end{aligned}$$

- If $P(E_2) > 0$, then the *Bayes formula* holds

$$P(E_1 \mid E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)}$$

Independence

Question 3. • *How likely is it to rain?*

- *How likely is it to rain if you see clouds?*
- *Are rain and clouds independent?*

Definition 6. Event E_1 is *independent* of E_2 if

$$P(E_1 \mid E_2) = P(E_1)$$

(i.e. knowledge of E_2 does not change the probability of E_1)

Because $P(E_1 \cap E_2) = P(E_1 \mid E_2)P(E_2)$, the joint probability of indepednent events is

$$P(E_1 \cap E_2) = P(E_1)P(E_2)$$

A Simple Credit Score Example

- You selected at random records of 20 customers and divided them based on *homeowner* and *credit score*.

		Credit score	
		Low	High
Homeowner	No	7	3
	Yes	2	8

- Can you tell from this data whether credit score and homeownership depend on each other?

A Simple Credit Score Example (sol.)

		Credit score		
		Low	High	
Homeowner	No	7	3	**10**
	Yes	2	8	**10**
		9	**11**	**20**

$$P(\mathrm{H},\mathrm{C}) = \left\{ \frac{7}{20}, \frac{2}{20}, \frac{3}{20}, \frac{8}{20} \right\}$$

$$P(\mathrm{H}) = \left\{ \frac{10}{20}, \frac{10}{20} \right\}, \qquad P(\mathrm{C}) = \left\{ \frac{9}{20}, \frac{11}{20} \right\}$$

Our simple test for independence $P(\mathrm{H},\mathrm{C}) \stackrel{?}{=} P(\mathrm{H})P(\mathrm{C})$

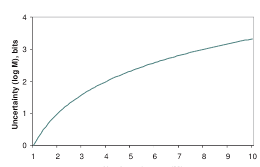
4 Uncertainty and Information

Uncertainty and Information

- If each event can have 2 states (i.e. True, False), then for a system that consists of H number of events there are

$$M = 2^H \text{ possible states}$$

- To measure uncertainty we can use $H = \log_2 M$ (aka *entropy*)



- Information reduces uncertainty

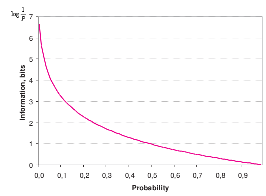
$$\text{Information} = \text{Uncertainty before} - \text{Uncertainty after}$$

- Information can be seen as negative entropy $I = -H$

Information and Probability

- What if probabilities are not equal? (i.e. $P(E) \neq \frac{1}{M}$)
- We can express the uncertainty as

$$H(E) = \log_2 \frac{1}{P(E)} = -\log_2 P(E)$$



Example 7. Compare information from observing events with probabilities $\frac{1}{8}$ and $\frac{1}{2}$