Lecture 12: Uncertainty and Information

Dr. Roman V Belavkin

BIS4410

Contents

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)

1 What is probability?

What is Probability?

Definition 1. A measure $P(E)$ of uncertainty about event E ranging from impossible ($P(E) = 0$) to certain ($P(E) = 1$):

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

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

Set-Theoretic Intuition

- Events E are considered as subsets $E \subseteq U$ of the universal set U.
- Probability $P(U) = 1$, because the universe is certain.
- We can consider probabilities of negation (not E), disjunction (E_1 or E_2) and conjunction (E_1 and E_2) of events measures of the complement, union and intersection of 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 $E_1 \cap E_2 = \varnothing$.
- 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

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

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

Subjective and Objective Probability

- Two people may have different (subjective) experiences, sets of observations or measurements of the same phenomenon, and therefore they may believe in different (subjective) probability of the same event.
- Some scientists believe that there exists an objective probability law of the phenomenon (although it may be unknown to us).
- The weak law of large numbers states that empirical frequency of an event observed in independent and identically distributed (i.i.d.) experiments should converge to its probability:

$$\lim_{n\to\infty} \frac{n(E)}{n} = P(E)$$

2 Information

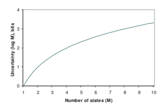
Complexity and Uncertainty

- A Boolean variable has two values, and probability 1/2 describes the maximum possible uncertainty about its values (e.g. a fair coin).
- A set of H Boolean variables has $M = 2\times2\times\cdots\times2$ possible configurations, and there are

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

- If a random phenomenon has M possible outcomes, then its *complexity* or maximal uncertainty is described by the *maximal entropy*:

$$H = \log_2 M$$



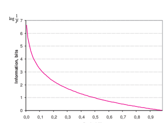
Surprise

- Notice that if $P(E) = \frac{1}{M}$, then

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

- *Random entropy* or *surprise* of event E with probability $P(E)$ is

$$H(E) = -\log_2 P(E)$$



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

Information

- Uncertainty can be reduced by obtaining unknown information.

- Thus, information is the difference of uncertainties:

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

- Information comes from measurements, tests or experiments.

Example 5 (Information). • Suppose you have 10 fair coins.

- Probability of each configuration is $1/2^{10}$, so that uncertainty is

$$H_{\text{before}}(10 \text{ coins}) = -\log_2 \frac{1}{2^{10}} = \log_2 2^{10} = 10$$

- How much information do we need to obtain to reduce the uncertainty to

$$H_{\text{after}}(10 \text{ coins}) = 2$$

- Answer: we need $8 = 10 - 2$ bits of information (i.e. state of 8 coins)

3 Decisions under Uncertainty

Payoff, Utility and Cost

- Our decisions (or choices) can have different *outcomes*:

$$E_1, E_2, \dots, E_m$$

- The main difference between the outcomes is that they carry different *payoffs* or *utilities*:

$$U(E_1) = £10, \quad U(E_2) = -£5, \ \dots$$

- The negative utilities are called *losses* or *costs*:

$$C(E_1) = -£10, \quad C(E_2) = £5, \ \dots$$

- Rational decisions should maximise utilities and minimise costs.

Expected Utility

- The outcomes of our decisions are often *uncertain*:

$$P(E_1), P(E_2), \dots, P(E_m)$$

Definition 6 (Expected utility)**.** The sum of products of outcomes' utilities $U(E_i)$ and their probabilities $P(E_i)$:

$$\mathbb{E}\{U\} = U(E_1)P(E_1) + U(E_2)P(E_2) + \cdots + U(E_m)P(E_m)$$

- If all probabilities are equal $P(E_i) = 1/m$, then expected utility is the same as the average utility:

$$\mathbb{E}\{U\} = \frac{U(E_1) + U(E_1) + \cdots + U(E_m)}{m}$$

- Rational decisions under uncertainty should maximise expected utility.

Games and Lotteries

- Games and lotteries are typical examples of situations when the outcomes of decisions are uncertain.
- A *lottery* is a set of outcomes $\{E_1, \dots, E_m\}$ with their utilities $U(E_i)$ and probabilities $P(E_i)$.
- Economic theory states that lottery B is preferred to lottery A if and only if lottery B has greater expected utility:

$$A \lesssim B \qquad \Longleftrightarrow \qquad \mathbb{E}_A\{U\} \leq \mathbb{E}_B\{U\}$$

Example 7.
- Consider two lotteries A and B with outcomes E_1 and E_2
- Let $U(E_1) = -£100$ and $U(E_2) = £100\}$
- Let $P_A(E_1) = 1/2$ and $P_B(E_1) = 1/4$ $(P_{A,B}(E_2) = 1 - P_{A,B}(E_1))$
- Then

$$\mathbb{E}_A\{U\} = £0 \quad \leq \quad \mathbb{E}_B\{U\} = £50$$

Risk

- What should we choose when two lotteries have the same expected utility?
- Different lotteries may have different *risk*, which can be measured as standard deviation or variance of utility from its expected value.

Example 8.
- Consider two lotteries A and B with outcomes E_1, E_2, E_3, E_4
- Let $U(E_i) \in \{-£1000, -£1, £1, £1000\}$
- Let $P_A(E_i) \in \{0, \frac{1}{2}, \frac{1}{2}, 0\}$
- Let $P_B(E_i) \in \{\frac{1}{2}, 0, 0, \frac{1}{2}\}$
- Both lotteries have zero expected utility $\mathbb{E}_A\{U\} = \mathbb{E}_B\{U\} = 0$, but the risk is different.

4 Human Perception of Risk

The Allais paradox

Consider two lotteries:

A : $p(£300) = \frac{1}{3}$ (and $p(£0) = \frac{2}{3}$)

B : $p(£100) = 1$

- Most of the people seem to prefer $A \lesssim B$

- Note that

$$\begin{aligned}\mathbb{E}_A\{x\} &= 300 \cdot \frac{1}{3} + 100 \cdot 0 + 0 \cdot \frac{2}{3} = 100 \\ \mathbb{E}_B\{x\} &= 300 \cdot 0 + 100 \cdot 1 + 0 \cdot 0 = 100\end{aligned}$$

- Lottery B has no risk.
- Both lotteries are about *gaining* utility.

Remark 1. *It has been noticed that people tend to be risk averse, when choosing between gains.*

The Allais paradox (2)

Consider two lotteries:

C : $p(-\pounds 300) = \frac{1}{3}$ (and $p(\pounds 0) = \frac{2}{3}$)

D : $p(-\pounds 100) = 1$

- Most of the people seem to prefer $C \gtrsim D$
- Note that

$$\begin{aligned}\mathbb{E}_C\{x\} &= -300 \cdot \frac{1}{3} - 100 \cdot 0 - 0 \cdot \frac{2}{3} = -100 \\ \mathbb{E}_D\{x\} &= -300 \cdot 0 - 100 \cdot 1 - 0 \cdot 0 = -100\end{aligned}$$

- Lottery C is risky.
- Both lotteries are about *loosing* utility.

Remark 2. *It has been noticed that people tend to be risk taking, when choosing between losses.*

Decision Theories

Normative

- How decisions should be made: *rational* agents acting according to their *preferences* and *utilities*.
- If the choices are made under uncertainty, then one maximises the *expected utility* with respect to some probability measures.

Descriptive

- How people make decisions.
- People do not always make decisions according to normative theory.

- There is a discord between the *normative* and *descriptive* theories of choice (due to many paradoxes).

  How wonderful that we have met with a paradox. Now we have some hope of making progress.

  Niels Bohr

Why is it a paradox?

- Normative theory states that one person should either be risk averse or risk taking, but *not both*.
- Tversky and Kahneman (1981) proposed *prospect theory* to explain human decision-making
- Prospect theory is not normative (i.e. not very satisfactory for mathematicians).

Remark 3 (Risk and Information Utility)**.**

- *Recall that uncertainty is related to information (i.e. as reduction of uncertainty).*
- *Information also carries some utility (e.g. for future decisions), but its value is not the same for gains and losses.*
- *Lotteries with no risk cannot give any information.*

Additional Reading

1. Tversky and Kahneman (1974):

   Judgment under Uncertainty: Heuristics and Biases

2. Tversky and Kahneman (1981):

   The framing of decisions and the psychology of choice

3. Both articles are available at Daniel Kahneman's webpage:

   `http://www.princeton.edu/~kahneman/`

References

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*, 453–458.