Questions 2

Dr. Roman Belavkin

BIS4435

Question 1

Name and briefly describe three main sources of uncertainty.

Answer: *Here, we emphasise the following sources of uncertainty: Complexity, ignorance and randomness.*

Complexity *A system may be so complex that it is impossible to consider every possibility. The system can have many parts, and each part can have many states. The total number of states may be too large for any computer to process. Thus, it is necessary to come up with some heuristics and assumptions that lead to errors and uncertainty.*

Ignorance *Some information can be unknown. For example, some parts of the system can be hidden and unobservable.*

Randomness *It is possible (through not everyone is convinced) that the nature is random, and it makes it impossible to predict the future precisely. Thus, some part of uncertainty is irreducible.*

Question 2

What is probability? What is the probability of 'raining or not raining' tomorrow?

Answer: *Probability $P(E)$ of event E is a number between 0 and 1, such that 0 corresponds to an impossible event and 1 to a certain event.*

The statement 'raining or not raining' describes a certain event. Therefore, its probability is 1. Moreover, we can notice that the statement describes two events: Raining and NOT Raining. These events are disjoint (they cannot happen at the same time). If we denote the probability of the first event as $P(\mathit{Raining})$, then the probability of the second is

$$P(\mathit{NOT\ Raining}) = 1 - P(\mathit{Raining})$$

Because the events are disjoint, the probability of any of them happening is the sum of their probabilities:

$$\begin{aligned} P(\mathit{Raining\ OR\ NOT\ Raining}) &= P(\mathit{Raining}) + P(\mathit{NOT\ Raining}) \\ &= P(\mathit{Raining}) + 1 - P(\mathit{Raining}) = 1 \end{aligned}$$

Question 3

Suppose that a university database has a variable called 'Age' storing the age of a student, and it can have 100 values. What is your estimate of the prior probability of each value? What if the database contains records of a 20,000 of students, and all of them are older than 16?

Answer: *If we do not have any prior information, then the best we can do is to assign equal probabilities to all values. Thus, all values will have prior probability 1/100.*

The data suggests that it is very unlikely that a student is younger than 16, and therefore probability for values between 1-16 is almost zero.

Question 4

The probability of disjunction (logical OR) of several disjoint events is simply the sum of their probabilities. For example, for two independent events E_1 and E_2

$$P(E_1\, OR\, E_2) = P(E_1) + P(E_2)$$

If, however, the events are not disjoint (i.e. the events can happen together), then

$$P(E_1\, OR\, E_2) = P(E_1) + P(E_2) - P(E_1, E_2)$$

where $P(E_1, E_2)$ is the joint probability of E_1 and E_2 (logical AND).

a) What is the probability of disjunction of two events that are not disjoint but are independent of each other?

b) Consider two fair and independent coins tossed together. What is the probability that at least one of the coins will be 'Heads' (i.e. $P(\text{heads OR heads})$)?

Answer:

a) *Two events are independent if and only if*

$$P(E_1, E_2) = P(E_1)P(E_2)$$

Therefore, for independent events

$$P(E_1\, OR\, E_2) = P(E_1) + P(E_2) - P(E_1)P(E_2)$$

b) *For each fair coin $P(heads) = 1/2$. Because the coins are independent, using the above formula we have*

$$\begin{aligned} P(heads\ OR\ heads) &= P(heads) + P(heads) - P(heads)P(heads) \\ &= 1/2 + 1/2 - 1/2 \times 1/2 = 1 - 1/4 = 3/4 \end{aligned}$$

Question 5

Suppose you know $P(A, B)$ - the joint probability distribution of events A and B. Let also $P(A)$ and $P(B)$ be the probabilities of each event individually (i.e. the marginal probabilities, which can be computed from the joint distribution).

a) What these probabilities can tell you about the relation between events A and B?

b) How could you use this information to reduce the uncertainty about one event based on information about another?

c) Why is it better to use the information about event A to assess the probability of B, then simply using probability $P(B)$?

Answer:

a) *The most important thing that the joint probability can tell us is whether the events are independent or not. Indeed, if A and B are independent, then*

$$P(A, B) = P(A)P(B)$$

This, if the above equality does not hold, then there is statistical dependence between A and B.

b) *We can compute the conditional probabilities:*

$$P(A \mid B) = \frac{P(A, B)}{P(B)} \quad and \quad P(B \mid A) = \frac{P(A, B)}{P(A)}$$

Thus, given specific information about one event, we can compute the probability of another.

c) *If the events are not independent, then*

$$P(A \mid B) \neq P(A) \quad and \quad P(B \mid A) \neq P(B)$$

The use of the information about event A (conditional probability $P(B \mid A)$) will give us a more 'precise' judgement about the likelihood of B.

Question 6

Consider two systems: A bicycle and an airplane. Why is the uncertainty associated with an airplane higher than the uncertainty of a bicycle?

Answer: *An airplane is a more complex system than a bicycle. It has more parts and can have more states. Because the uncertainty is proportional to the number of states (more precisely* $\log M$*, where* M *is the number of states), an airplane is a more uncertain system than a bicycle.*

Question 7

Suppose that a database has recorded a very unusual case - its values are very different from other, more typical cases. Why is this case more interesting from the information theory point of view?

Answer: *The fact that the case is unusual means that its probability is lower than the probabilities of more typical cases. The uncertainty is related to the probability as*

$$Uncertainty = \log \frac{1}{P}$$

and therefore there is a greater uncertainty associated with the unusual case. Information is measured by the reduction of uncertainty. Thus, because the unusual case reduces the uncertainty by a greater value, it brings more information than a more typical case.

Question 8

Let variable x can have values 1, 2 and 3 with probabilities $P(1) = 1/5$, $P(2) = 3/5$ and $P(3) = 1/5$. What is the expected value of x? Compare it with mean value of $(1, 2, 2, 2, 3)$?

Answer: *The expected value is*

$$E\{x\} = 1\frac{1}{5} + 2\frac{3}{5} + 3\frac{1}{5} = 2$$

The mean value of $(1, 2, 2, 2, 3)$ *is*

$$\frac{1 + 2 + 2 + 2 + 3}{5} = 2$$

The mean value here is the same as the expected value. In fact, we could interpret the mean value of $(1, 2, 2, 2, 3)$ *as the expected value of* x *with values 1, 2, 2, 2 and 3 with all probabilities equal 1/5.*

Question 9

Consider the following sets of values for variables x and y:

x	y
-1	-2
0	0
1	2

Compute the expected values and variances of x and y. You can compute them as the means and the mean square deviations. Compare the results. Which variable is more uncertain (risky)?

Answer: *First, we compute the expected values as the means:*

$$E\{x\} = Mean(x) = \frac{-1+0+1}{3} = 0$$
$$E\{y\} = Mean(y) = \frac{-2+0+2}{3} = 0$$

Thus, x and y have equal expected values. Next, we compute the deviations from the mean

$$\begin{array}{llll} \text{for } x: & -1-0=-1\,, & 0-0=0\,, & 1-0=1 \\ \text{for } y: & -2-0=-2\,, & 0-0=0\,, & 2-0=2 \end{array}$$

The variances we compute as the means of squared deviations

$$Var\{x\} = \frac{(-1)^2+0^2+1^2}{3} = \frac{2}{3}$$
$$Var\{y\} = \frac{(-2)^2+0^2+2^2}{3} = \frac{8}{3}$$

Thus, y has greater variance than x.

Although both variables have the same expected values (the means), their variances are different. Variable y is more uncertain (i.e. more risky) because it has a greater variance (higher dispersion).

Question 10

Compute the covariance and correlation between x and y from Question 9:

x	y
-1	-2
0	0
1	2

Are these variables correlated, uncorrelated or anticorrelated?

Answer: *From Question 9, the expected values are $E\{x\} = 0$ and $E\{y\} = 0$. The deviations are*

$$\begin{array}{llll} \text{for } x: & -1-0=-1\,, & 0-0=0\,, & 1-0=1 \\ \text{for } y: & -2-0=-2\,, & 0-0=0\,, & 2-0=2 \end{array}$$

We compute the covariance as the mean of deviations multiplied together:

$$Cov(x,y) = \frac{(-1)(-2) + 0 \times 0 + 1 \times 2}{3} = \frac{4}{3}$$

To find the correlation, we need simply to divide the covariance by the square root of the product of the variances (using results from Question 3.9, $Var(x) = \frac{2}{3}$ and $Var(y) = \frac{8}{3}$):

$$Corr(x,y) = \frac{4/3}{\sqrt{\frac{2}{3} \times \frac{8}{3}}} = \frac{4/3}{4/3} = 1$$

Thus, x and y are correlated.

Question 11

Suppose the database contains data for m independent variables. What should the covariance and the correlation matrices look like?

Answer: *Because all m variables are independent, their covariances as well as correlations should be zero. Thus, the covariance matrix will be*

$$\begin{pmatrix} Var(x_1) & 0 & \dots & 0 \\ 0 & Var(x_2) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & Var(x_m) \end{pmatrix}$$

and the correlation matrix will be

$$\begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

Thus, all the elements apart from the diagonal will be zero.

Question 12

Suppose of you have a choice between two lotteries A and B:

- Lottery A: The utility can have values -1, 0 or 1.
- Lottery B: The utility can have values -2, 0 or 2.

Suppose that all values have equal probabilities.

a) What choice does the maximum expected utility principle suggest?

b) Have you lost some information by using only the expected values?

Answer:

a) *Here, we note that the values of utilities are the same as the values of* x *and* y *in Question 3.8. So, we can use the results from Question 3.8 for the expected values:*

$$E\{A\} = 0, \quad E\{B\} = 0$$

Because both expected values are equal, there is no maximum, and therefore the maximum expected utility principle cannot suggest any choice (i.e. we are indifferent between A and B).

b) *Again, using the results of Question 3.9, we note that*

$$Var\{A\} = \frac{2}{3}, \quad Var\{B\} = \frac{8}{3}$$

So, lottery B is more risky. This information is lost if we rely solely on the expected values in our choice.