3 Topics: Bayes' rule, law of total probability

3.1 Prerequisites: Lecture 7

Exercise 3- 1: (Suggested for personal/peer tutorial) *The Prisoner's Dilemma*: Three prisoners A, B, C are in solitary confinement under sentence of death, but each knows that one of them, chosen at random with equal probability, is to be pardoned. Prisoner A begs the governor to tell him whether he, A, is to be pardoned or executed. The governor refuses to answer this, but he does say that B is to be executed. The governor thinks that he isn't giving useful information, as A knows that at least one of B and C must die.

(a) A suddenly feels much happier, as he believes his chances of being pardoned have *risen* from 1/3 to 1/2. The governor, who, if A were actually to be pardoned, would be equally likely to give C's name rather than B's, is mystified by A's euphoria. Who is correct?

[Hint: Let A, B, C be the events that A, B or C respectively are pardoned. Then A, B, C partition Ω. Now let G_{AB} be the event that the governor tells A that B is to be executed. You want $\mathrm{P}(A|G_{AB})$, so consider the three conditional probabilities of G_{AB} given A, B and C respectively, and then use Bayes Theorem.]

(b) What should C feel if he overhears the governor's reply, but assumes that the question was asked by one of the warders? (consider the event G_{WB} that the governor tells a warder that B is to be executed).

Solution:

(a) By the general Bayes' rule, we have

$$\mathrm{P}(A|G_{AB}) = \frac{\mathrm{P}(A \cap G_{AB})}{\mathrm{P}(G_{AB})} = \frac{\mathrm{P}(G_{AB}|A)\mathrm{P}(A)}{\mathrm{P}(G_{AB})},$$

where

$$\mathrm{P}(G_{AB}) = \mathrm{P}(G_{AB}|A)\mathrm{P}(A) + \mathrm{P}(G_{AB}|B)\mathrm{P}(B) + \mathrm{P}(G_{AB}|C)\mathrm{P}(C).$$

We know that, for each prisoner, the probability of being pardoned is given by $\mathrm{P}(A) = \mathrm{P}(B) = \mathrm{P}(C) = 1/3$. Also,

- given that A is being pardoned, the governor can tell A that B (or C) will be executed with equal probability, so and $\mathrm{P}(G_{AB}|A) = 1/2$.
- However, conditional on B being pardoned, the governor cannot tell A that B will be executed so $\mathrm{P}(G_{AB}|B) = 0$ and,
- conditional on C being pardoned, the governor can now only choose to say that B will be executed, hence $\mathrm{P}(G_{AB}|C) = 1$,

and hence by Bayes theorem using an identical calculation to above, we have

$$\mathrm{P}(A|G_{AB}) = \frac{1/2 \cdot 1/3}{1/2 \cdot 1/3 + 0 \cdot 1/3 + 1 \cdot 1/3} = 1/3,$$

and hence governor is correct.

(b) Now $\mathrm{P}(G_{WB}|A) = 1/2$, $\mathrm{P}(G_{WB}|B) = 0$ but $\mathrm{P}(G_{WB}|C) = 1/2$, so by Bayes theorem

$$\begin{aligned}\mathrm{P}(C|G_{WB}) &= \frac{\mathrm{P}(G_{WB}|C)\mathrm{P}(C)}{\mathrm{P}(G_{WB}|A)\mathrm{P}(A) + \mathrm{P}(G_{WB}|B)\mathrm{P}(B) + \mathrm{P}(G_{WB}|C)\mathrm{P}(C)} \\ &= \frac{1/2 \cdot 1/3}{1/2 \cdot 1/3 + 0 \cdot 1/3 + 1/2 \cdot 1/3} = \frac{1}{2},\end{aligned}$$

and hence C is right to feel happier.

Exercise 3- 2: A diagnostic test has a probability 0.95 of giving a positive result when applied to a person suffering from a certain disease, and a probability 0.10 of giving a (false) positive when applied to a non-sufferer. It is estimated that 0.5 % of the population are sufferers. Suppose that the test is now administered to a person about whom we have no relevant information relating to the disease (apart from the fact that he/she comes from this population). Calculate the following probabilities:

(a) that the test result will be positive;

(b) that, given a positive result, the person is a sufferer;

(c) that, given a negative result, the person is a non-sufferer;

(d) that the person will be misclassified.

Solution: Let $T \equiv$ "Test positive", $S \equiv$ "Sufferer". Then $\mathrm{P}(T|S) = 0.95$, $\mathrm{P}(T|S^c) = 0.10$, $\mathrm{P}(S) = 0.005$.

(a) Using the law of total probability, we have

$$\mathrm{P}(T) = \mathrm{P}(T|S)\mathrm{P}(S) + \mathrm{P}(T|S^c)\mathrm{P}(S^c) = (0.95 \times 0.005) + (0.1 \times 0.995) = 0.10425.$$

(b) By Bayes' formula, we have

$$\mathrm{P}(S|T) = \frac{\mathrm{P}(T|S)\mathrm{P}(S)}{\mathrm{P}(T|S)\mathrm{P}(S) + \mathrm{P}(T|S^c)\mathrm{P}(S^c)} = \frac{0.95 \times 0.005}{(0.95 \times 0.005) + (0.1 \times 0.995)} = 0.0456.$$

(c) Again, by Bayes' formula, we have

$$\mathrm{P}(S^c|T^c) = \frac{\mathrm{P}(T^c|S^c)\mathrm{P}(S^c)}{\mathrm{P}(T^c)} = \frac{0.9 \times 0.995}{1 - 0.10425} = 0.9997.$$

(d) Finally, we have

$$\mathrm{P}(M) = \mathrm{P}(T \cap S^c) + \mathrm{P}(T^c \cap S) = \mathrm{P}(T|S^c)\mathrm{P}(S^c) + \mathrm{P}(T^c|S)\mathrm{P}(S) = 0.09975.$$

Exercise 3- 3: Show that the Bayes' rule and the law of total probability also hold with extra conditioning:

(a) Bayes' rule with extra conditioning: For events A, B, E with $\mathrm{P}(A \cap E) > 0, \mathrm{P}(B \cap E) > 0$, we have

$$\mathrm{P}(A|B \cap E) = \frac{\mathrm{P}(B|A \cap E)\mathrm{P}(A|E)}{\mathrm{P}(B|E)}$$

(b) Consider events A, E with $\mathrm{P}(E) > 0$ and let $\{B_i : i \in \mathcal{I}\}$ denote a partition of Ω, with $\mathrm{P}(B_i \cap E) > 0$ for all $i \in \mathcal{I}$. Then,

$$\mathrm{P}(A|E) = \sum_{i \in \mathcal{I}} \frac{\mathrm{P}(A \cap B_i \cap E)}{\mathrm{P}(E)} = \sum_{i \in \mathcal{I}} \mathrm{P}(A|B_i \cap E)\mathrm{P}(B_i|E).$$

Solution:

(a) Using the definition of the conditional probability, we can write

$$\mathrm{P}(A|B \cap E)\mathrm{P}(B|E) = \frac{\mathrm{P}(A \cap B \cap E)}{\mathrm{P}(B \cap E)} \frac{\mathrm{P}(B \cap E)}{\mathrm{P}(E)} = \frac{\mathrm{P}(A \cap B \cap E)}{\mathrm{P}(E)}$$

$$= \frac{P(B|A \cap E)P(A \cap E)}{P(E)} = P(B|A \cap E)P(A|E),$$

which implies the Bayes' rule with extra conditioning.

(b) From the definition of the conditional probability, we have

$$P(A|E) = \frac{P(A \cap E)}{P(E)}.$$

We can now apply the law of total probability to the event $A \cap E$ and get

$$P(A \cap E) = \sum_{i \in \mathcal{I}} P(A \cap E \cap B_i) = \sum_{i \in \mathcal{I}} P(A|E \cap B_i)P(E \cap B_i).$$

Dividing by $P(E)$ leads to

$$P(A|E) = \sum_{i \in \mathcal{I}} \frac{P(A \cap E \cap B_i)}{P(E)} = \sum_{i \in \mathcal{I}} P(A|E \cap B_i)\frac{P(E \cap B_i)}{P(E)} = \sum_{i \in \mathcal{I}} P(A|E \cap B_i)P(B_i|E).$$

3.2 Prerequisites: Lecture 8

Exercise 3- 4: Athletes are routinely tested for the use of performance-enhancing drugs. When a test is to be carried out, the athlete provides two blood samples, the first of which is then tested. The test used is quite accurate, in that it correctly indicates the *presence* of drugs in 99.5% of tests, and correctly indicates the *absence* of drugs in 98% of tests. If this test is positive, indicating that drugs are present, the second sample is tested, using the same test, and if the second test is also positive, then the athlete is deemed to be using drugs.

Suppose that an athlete is selected at random, and two blood samples (regarded as identical) are obtained. Let events T_1 and T_2 correspond respectively to the events that first and second samples test positive, and let C be the event that drugs are actually present in the samples.

It is estimated that only 1 athlete in 1000 gives samples in which drugs are actually present

If it is assumed the results of the two tests are *conditionally independent* given the presence or absence of drugs in the samples, give expressions for, and evaluate

(a) the probability that the first test is positive

(b) the conditional probability that drugs are actually present in the sample, given that the first test is positive.

(c) the probability that both tests are positive, so that the athlete fails the test

(d) the conditional probability that drugs are actually present in the sample, given that both tests are positive.

Solution: Let $T_1 \equiv$ "first test positive", $T_2 \equiv$ "second test positive", $C \equiv$ "drugs present in sample". Then we read off from the question that

$$P(T_1|C) = P(T_2|C) = 0.995 \quad P(T_1^c|C^c) = P(T_2^c|C^c) = 0.98,$$

and

$$\mathrm{P}(C) = 0.001, \qquad \mathrm{P}(C^c) = 0.999.$$

(a) By the Theorem of Total Probability

$$\mathrm{P}(T_1) = \mathrm{P}(T_1|C)\mathrm{P}(C) + \mathrm{P}(T_1|C^c)\mathrm{P}(C^c) = 0.995 \times 0.001 + (1 - 0.98) \times 0.999 = 0.021.$$

(b) By Bayes Theorem

$$\mathrm{P}(C|T_1) = \frac{\mathrm{P}(T_1|C)\mathrm{P}(C)}{\mathrm{P}(T_1|C)\mathrm{P}(C) + \mathrm{P}(T_1|C^c)\mathrm{P}(C^c)} = \frac{0.995 \times 0.001}{0.995 \times 0.001 + (1 - 0.98) \times 0.999} = 0.047.$$

(c) By the Theorem of Total Probability and conditional independence

$$\begin{aligned}\mathrm{P}(T_1 \cap T_2) &= \mathrm{P}(T_1 \cap T_2|C)\mathrm{P}(C) + \mathrm{P}(T_1 \cap T_2|C^c)\mathrm{P}(C^c)\\ &= \mathrm{P}(T_1|C)\mathrm{P}(T_2|C)\mathrm{P}(C) + \mathrm{P}(T_1|C^c)\mathrm{P}(T_2|C^c)\mathrm{P}(C^c)\\ \\ &= 0.995^2 \times 0.001 + (1 - 0.98)^2 \times 0.999 = 0.0014.\end{aligned}$$

(d) By Bayes Theorem

$$\begin{aligned}\mathrm{P}(C|T_1 \cap T_2) &= \frac{\mathrm{P}(T_1 \cap T_2|C)\mathrm{P}(C)}{\mathrm{P}(T_1 \cap T_2|C)\mathrm{P}(C) + \mathrm{P}(T_1 \cap T_2|C^c)\mathrm{P}(C^c)}\\ &= \frac{\mathrm{P}(T_1|C)\mathrm{P}(T_2|C)\mathrm{P}(C)}{\mathrm{P}(T_1|C)\mathrm{P}(T_2|C)\mathrm{P}(C) + \mathrm{P}(T_1|C^c)\mathrm{P}(T_2|C^c)\mathrm{P}(C^c)}\\ \\ &= \frac{0.995^2 \times 0.001}{0.995^2 \times 0.001 + (1 - 0.98)^2 \times 0.999} = 0.712\end{aligned}$$

Exercise 3- 5: Prove Lemma 6.1.11: Any countable union can be written as a countable union of disjoint sets. I.e. let $A_1, A_2, \cdots \in \mathcal{F}$ and define $D_1 = A_1, D_2 = A_2 \setminus A_1, D_3 = A_3 \setminus (A_1 \cup A_2), \ldots$. Then $\{D_i\}$ is a collection of disjoint sets and $\cup_{i=1}^n A_i = \cup_{i=1}^n D_i$ for n being any positive integer or ∞.

Solution: By construction, the D_is are disjoint.

(a) We show that $\cup_{i=1}^n A_i \subseteq \cup_{i=1}^n D_i$: Suppose that $\omega \in \cup_{i=1}^n A_i$. Then $\omega \in A_i$ for at least one (finite) $i \leq n$. Let us denote by I the smallest index $i \leq n$ such that $\omega \in A_i$, i.e. $\omega \in A_I$, but $\omega \notin A_i$ for $i < I$. Hence, $\omega \in D_I$ which implies that also $\omega \in \cup_{i=1}^n D_i$.

(b) Next we show that $\cup_{i=1}^n A_i \supseteq \cup_{i=1}^n D_i$: Let $\omega \in \cup_{i=1}^n D_i$. Then $\omega \in D_i$ for some finite $i \leq n$. But then $\omega \in A_i$ and also $\omega \in \cup_{i=1}^n A_i$.

We remark that our arguments are valid both for finite n and for the case when $n = \infty$.

3.3 Prerequisites: Lecture 9

Exercise 3- 6: Consider a function with domain $\mathcal{X}$ and co-domain $\mathcal{Y}$, i.e. $f : \mathcal{X} \to \mathcal{Y}$. For any collection of subsets $B_i \subseteq \mathcal{Y}$, $i \in \mathcal{I}$ where $\mathcal{I}$ denotes an (arbitrary) index set, show that

$$f^{-1}\left(\bigcap_{i\in\mathcal{I}} B_i\right) = \bigcap_{i\in\mathcal{I}} f^{-1}(B_i).$$

Solution: We have that

$$\begin{aligned} x \in f^{-1}\left(\bigcap_{i\in\mathcal{I}} B_i\right) &\Leftrightarrow f(x) \in \bigcap_{i\in\mathcal{I}} B_i \\ &\Leftrightarrow \forall\, i \in \mathcal{I} : f(x) \in B_i \\ &\Leftrightarrow \forall\, i \in \mathcal{I} : x \in f^{-1}(B_i) \\ &\Leftrightarrow x \in \bigcap_{i\in\mathcal{I}} f^{-1}(B_i). \end{aligned}$$

Exercise 3- 7: Let X denote a discrete random variable on $(\Omega, \mathcal{F}, \mathrm{P})$ and let $g : \mathbb{R} \to \mathbb{R}$ be a deterministic function. Show that $Y = g(X)$ is also a discrete random variable and find the probability mass function of Y.

Solution: We will discuss this in Chapter 9 in the lecture notes: For $Y = g(X)$, we have that $\mathrm{Im}(Y) = \{g(X(\omega)) : \omega \in \Omega\}$ is countable since $\mathrm{Im}(X) = \{X(\omega) : \omega \in \Omega\}$ is countable. Moreover, for all $y \in \mathbb{R}$, we have

$$\begin{aligned} Y^{-1}(\{y\}) &= \{\omega \in \Omega : Y(\omega) = y\} = \{\omega \in \Omega : g(X(\omega)) = y\} \\ &= \{\omega \in \Omega : X(\omega) \in \{x \in \mathrm{Im}X : g(x) = y\}\} \\ &= \{\omega \in \Omega : X(\omega) \in \bigcup_{x\in\mathrm{Im}X:g(x)=y} \{x\}\} \\ &= X^{-1}\left(\bigcup_{x\in\mathrm{Im}X:g(x)=y} \{x\}\right) \\ &\overset{\text{Lemma 7.1.2}}{=} \bigcup_{x\in\mathrm{Im}X:g(x)=y} X^{-1}(\{x\}) \\ &= \bigcup_{x\in\mathrm{Im}X:g(x)=y} \{\omega \in \Omega : X(\omega) = x\} \in \mathcal{F}, \end{aligned} \tag{3.1}$$

since each event $\{\omega \in \Omega : X(\omega) = x\} \in \mathcal{F}$ for all $x \in \mathbb{R}$ and, by the definition of a σ-algebra, a countable union of elements of $\mathcal{F}$ is in $\mathcal{F}$, too. Since X is discrete, we indeed have that $\{x \in \mathrm{Im}X : g(x) = y\} \subseteq \mathrm{Im}X$ is (at most) countably infinite. Hence we can conclude that $Y = g(X)$ is indeed a discrete random variable.

Recall:

$$Y^{-1}(\{y\}) = \bigcup_{x\in\mathrm{Im}X:g(x)=y} \{\omega \in \Omega : X(\omega) = x\}$$

We can compute the p.m.f. of Y as follows:

$$p_Y(y) = \mathrm{P}(Y = y) = \mathrm{P}(g(X) = y) = \sum_{x \in \mathrm{Im} X : g(x) = y} \mathrm{P}(X = x),$$

so we are just summing up the probabilities for all values x for which $g(x) = y$. Here we used the fact that the union in (3.1) is a countable union of disjoint events, hence Axiom (iii) of the definition of the probability measure applies.