HW2, Bayesian Modelling B 2016/17

Jonathan Rougier
School of Mathematics
University of Bristol UK

In the homeworks, questions with marks are officially 'exam-style', although you can expect any homework question to appear as an exam question, unless it is explicitly 'not examinable'.

Hand in Q4 and Q5 to be marked. Try the others now or later!

1. Here are some questions about the probability calculus. They test your ability to do mathematics; that is, to present logical and compelling arguments, which combine clarity of expression with elegance and absence of superfluity. Just do a couple of these, to check that you can. I am taking it for granted that you know the propositional calculus (also known as 'zeroth order logic'). If not, you might find the wikipedia page `https://en.wikipedia.org/wiki/Propositional_calculus` helpful.

   Let $A, A_1, A_2, \dots, B$ be a set of propositions.

   (a) State the three axioms of probability. Distinguish between 'finite' and 'countable' additivity, and show that the latter implies the former.

   Answer.

   i. $\Pr(A) \geq 0$.
   ii. If A is certainly true, then $\Pr(A) = 1$.
   iii. If A and B are mutually exclusive, then $\Pr(A \vee B) = \Pr(A) + \Pr(B)$.

   The final axiom extends by induction to: if $A_1, \dots, A_n$ are mutually exclusive, then $\Pr(A_1 \vee \cdots \vee A_n) = \Pr(A_1) + \cdots + \Pr(A_n)$; see below. This is *finite additivity.* A stronger axiom is: if $A_1, A_2, \dots$ are mutually exclusive, then $\Pr(A_1 \vee A_2 \vee \cdots) = \Pr(A_1) + \Pr(A_2) + \cdots$. This is countable additivity. It implies finite additivity if we take $A_{n+1} = A_{n+2} = \cdots = \mathsf{FALSE}$, because if $\{A_1, \dots, A_n\}$ are mutually exclusive, then $\{A_1, A_2, \cdots\}$ are mutually exclusive, and

   $$\begin{aligned}
   \Pr(A_1 \vee \cdots \vee A_n) &= \Pr(A_1 \vee \cdots \vee A_n \vee \mathsf{FALSE}) \\
   &= \Pr(A_1 \vee \cdots \vee A_n \vee A_{n+1} \vee \cdots) \\
   &= \Pr(A_1) + \cdots + \Pr(A_n) + \Pr(A_{n+1}) + \cdots \quad \text{by countable additivity} \\
   &= \Pr(A_1) + \cdots + \Pr(A_n),
   \end{aligned}$$

   because $\Pr(A_{n+1}) = \Pr(A_{n+2}) = \cdots = \Pr(\mathsf{FALSE}) = 0$, as shown below.

   (b) Prove each of the following:

i. $\Pr(\neg A) = 1 - \Pr(A)$.

Answer. A and $\neg A$ are mutually exclusive, and $A \vee \neg A = \mathsf{TRUE}$. Hence

$$1 = \Pr(A \vee \neg A) = \Pr(A) + \Pr(\neg A),$$

from which the result follows. Note that if $A = \mathsf{FALSE}$, then $\Pr(\neg A) = 1$ and $\Pr(A) = 0$.

ii. If $\{A_1, \dots, A_n\}$ are mutually exclusive, then

$$\Pr(A_1 \vee \cdots \vee A_n) = \sum_{i=1}^{n} \Pr(A_i).$$

Answer. '$\vee$' is commutative, and hence

$$A_1 \vee \cdots \vee A_n = A_1 \vee (A_2 \vee \cdots \vee A_n) = A_1 \vee A_{-1},$$

say. If $\{A_1, \dots, A_n\}$ are mutually exclusive, then A_1 and A_{-1} are mutually exclusive, and so $\Pr(A_1 \vee \cdots \vee A_n) = \Pr(A_1) + \Pr(A_{-1})$. The result then follows, by applying the same argument to A_{-1}.

iii. If $A \to B$, then $\Pr(A) \le \Pr(B)$.

Answer. If $A \to B$, then $B = A \vee (B \wedge \neg A)$. A and $B \wedge \neg A$ are mutally exclusive, and therefore

$$\Pr(B) = \Pr(A) + \Pr(B \wedge \neg A) \ge \Pr(A),$$

as required.

iv. If $\Pr(A) = 0$, then $\Pr(A \wedge B) = 0$.

Answer. Follows from Frechét's inequality.

v. $\Pr(A \wedge B) \le \min\{\Pr(A), \Pr(B)\}$. This is known as *Fréchet's inequality.*

Answer. As $(A \wedge B) \to A$, so $\Pr(A \wedge B) \le \Pr(A)$. Likewise, $(A \wedge B) \to B$, so $\Pr(A \wedge B) \le \Pr(B)$, as required.

2. Let C be a proposition. Prove that if $\Pr(C) > 0$, then $\Pr(\cdot \mid C)$ obeys the three axioms of probability (finite additivity suffices). [10 marks]

Answer. It is always true that

$$\Pr(A \wedge C) = \Pr(A \mid C) \cdot \Pr(C).$$

If $\Pr(C) > 0$, then

$$\Pr(A \mid C) = \frac{\Pr(A \wedge C)}{\Pr(C)}.$$

It follows that $\Pr(A \mid C) \ge 0$, which is axiom i. If $A = \mathsf{TRUE}$, then $A \wedge C = C$, and

$$\Pr(\mathsf{TRUE} \mid C) = \Pr(A \mid C) = \Pr(C) / \Pr(C) = 1,$$

which is axiom ii. If $A = A_1 \vee A_2$, where A_1 and A_2 are mutually exclusive, then

$$\begin{aligned}
\Pr(A_1 \vee A_2 \mid C) &= \frac{\Pr\{(A_1 \vee A_2) \wedge C\}}{\Pr(C)} \\
&= \frac{\Pr\{(A_1 \wedge C) \vee (A_2 \wedge C)\}}{\Pr(C)} \\
&= \frac{\Pr(A_1 \wedge C) + \Pr(A_2 \wedge C)}{\Pr(C)} \\
&= \Pr(A_1 \mid C) + \Pr(A_2 \mid C)
\end{aligned}$$

using that '$\wedge$' is distributive over '$\vee$', and that $A_1 \wedge C$ and $A_2 \wedge C$ are mutually exclusive. This is axiom iii.

3. Let X, Y, Z be collections of random quantities.

   (a) Prove that $X \perp\!\!\!\perp Y \mid Z$ if and only if:

   $$\mathrm{p}(x, y \mid z) = g(x, z) \cdot h(y, z)$$

   for some g, h, whenever $\mathrm{p}(z) > 0$. [10 marks]

   Answer. Take as given throughout that $\mathrm{p}(z) > 0$.

   'Only if' follows from the equivalence between $X \perp\!\!\!\perp Y \,|\, Z$ and $\mathrm{p}(x, y|z) = \mathrm{p}(x|z) \cdot \mathrm{p}(y|z)$, where in this case $g(x, z) = \mathrm{p}(x \mid z)$ and $h(y, z) = \mathrm{p}(y \mid z)$.

   'If' is slightly trickier. We assume that

   $$\mathrm{p}(x, y \mid z) = g(x, z) \cdot h(y, z) \tag{†}$$

   for some g, h. Note that

   $$1 = \sum\nolimits_x \sum\nolimits_y \mathrm{p}(x, y \mid z) = \sum\nolimits_x \sum\nolimits_y g(x, z) \cdot h(y, z) = \sum\nolimits_x g(x, z) \cdot \sum\nolimits_y h(y, z).$$

   Hence, starting from (†),

   $$\mathrm{p}(x, y \mid z) = g(x, z) \cdot h(y, z) \cdot \sum\nolimits_x g(x, z) \cdot \sum\nolimits_y h(y, z), \tag{‡}$$

   after multiplying by 1. But also note that

   $$\mathrm{p}(x \mid z) = \sum\nolimits_y \mathrm{p}(x, y \mid z) = g(x, z) \sum\nolimits_y h(y, z),$$

   and similarly for $\mathrm{p}(y \mid z)$. Substituting both of these results into (‡) gives

   $$\mathrm{p}(x, y \mid z) = \mathrm{p}(x \mid z) \cdot \mathrm{p}(y \mid z),$$

   which is equivalent to $X \perp\!\!\!\perp Y \mid Z$.

   (b) Express the statement after the colon in the previous question using predicate logic (i.e., with $\exists$ and $\forall$ quantifiers). [5 marks]

   Answer.

   $$\exists g \, \exists h \, \forall x \, \forall y \, \forall z \, \big\{[\mathrm{p}(z) = 0] \vee [\mathrm{p}(x, y \mid z) = g(x, z) \cdot h(y, z)]\big\}.$$

4. Prove that

$$\vDash X \mid Y \iff f_{X|Y}(x \mid y) = \prod_{i=1}^{m} f_{X_i|Y}(x_i \mid y).$$

[10 marks]

Answer. Let $X = (X_1, \dots, X_m)$. $\vDash X \mid Y$ exactly when $X_A \perp\!\!\!\perp X_B \mid Y$ for every disjoint A and B. For simplicity, write the righthand expression as $\mathrm{p}(x \mid y) = \prod_{i=1}^{m} \mathrm{p}(x_i \mid y)$.

'Only if'. Use the Telescope theorem to write

$$\mathrm{p}(x \mid y) = \mathrm{p}(x_1 \mid y) \prod_{i=2}^{m} \mathrm{p}(x_i \mid x_{1:(i-1)}, y).$$

But if $\vDash X \mid Y$, then $X_i \perp\!\!\!\perp X_{1:(i-1)} \mid Y$, setting $A = \{i\}$ and $B = \{1, \dots i-1\}$. Hence

$$\mathrm{p}(x \mid y) = \mathrm{p}(x_1 \mid y) \prod_{i=2}^{m} \mathrm{p}(x_i \mid y),$$

as required.

'If'. Let A and B be arbitrary disjoint sets, and set $C = \{1, \dots, m\} \setminus \{A \cup B\}$. Then $\mathrm{p}(x \mid y) = \mathrm{p}(x_A, x_B, x_C \mid y)$, and

$$\begin{aligned}
\mathrm{p}(x_A, x_B \mid y) &= \sum_{x_C} \mathrm{p}(x_A, x_B, x_C \mid y) \\
&= \sum_{x_C} \prod_{i \in A} \mathrm{p}(x_i \mid y) \cdot \prod_{j \in B} \mathrm{p}(x_j \mid y) \cdot \prod_{k \in C} \mathrm{p}(x_k \mid y) \\
&= \prod_{i \in A} \mathrm{p}(x_i \mid y) \cdot \prod_{j \in B} \mathrm{p}(x_j \mid y).
\end{aligned}$$

This has the form

$$\mathrm{p}(x_A, x_B \mid y) = g(x_A, y) \cdot h(x_B, y)$$

which is equivalent to $X_A \perp\!\!\!\perp X_B \mid Y$, as required (see Q3).

Note: in the question I should have added "If $\mathrm{p}(y) > 0$, then …".

5. Let $X = (X_1, \dots, X_m)$. The PMF f_X is *exchangeable* exactly when

$$\mathrm{p}(x_1, \dots, x_m) = \mathrm{p}(x_{\pi_1}, \dots, x_{\pi_m})$$

whenever $(\pi_1, \dots, \pi_m)$ is a permutation of $(1, \dots, m)$.

(a) Show that if f_X is exchangeable, then f_{X_A} is exchangeable, where X_A is any subset of X. [10 marks]

Answer. Let $X = (X_1, \dots, X_m)$, and set $A = (1, \dots, k)$, without loss of generality, with $B = (k+1, \dots, m)$. Let π' be an arbitrary permutation of A, so that (π', B) is a permutation of $(1, \dots, m)$. Then by exchangeability, $\mathrm{p}(x_A, x_B) = \mathrm{p}(x_{\pi'}, x_B)$. Marginalizing over all but the first k arguments gives $\mathrm{p}(x_A) = \pi(x_{\pi'})$, showing that f_{X_A} is exchangeable.

(b) Show that if X is conditionally IID given Θ, then X is exchangeable. [10 marks]

Answer. If X is conditionally IID given Θ, then $\models X \mid \Theta$ and $f_{X_i|\Theta} = f_{X_1|\Theta}$. Hence, by the previous result (Q4),

$$\mathrm{p}(x \mid \theta) = \prod_{i=1}^{m} f_{X_1|\Theta}(x_i \mid \theta).$$

From the definition of conditional probability,

$$\mathrm{p}(\theta, x) = \mathrm{p}(\theta)\,\mathrm{p}(x \mid \theta) = \mathrm{p}(\theta) \prod_{i=1}^{m} f_{X_1}(x_i \mid \theta).$$

Now marginalize over Θ to give

$$\mathrm{p}(x) = \int_{\Omega} \mathrm{p}(\theta)\,\mathrm{p}(x \mid \theta)\,\mathrm{d}\theta = \int_{\Omega} \mathrm{p}(\theta) \prod_{i=1}^{m} f_{X_1}(x_i \mid \theta)\,\mathrm{d}\theta.$$

This is a symmetric function of x, and therefore X is exchangeable.

(c) Explain why exchangeability can be interpreted as 'similar but not identitical'. [10 marks]

Answer. The essence of exchangeability is that the labels, $i = 1, \ldots, m$ do not matter. So while it is *not* the case that X_i and X_j are identical, it *is* the case that marginal distributions of X_i and X_j are identical, and in this sense X_i and X_j are similar. So thus we have 'similar but not identical'. This argument can be extended to consider pairs, triples, and all tuples; for example, the marginal distribution of any pair of X's is identical. So 'similar' in the context of exchangeability is very strong. If you believe that X is exchangeable, and you are offered a sample of size n, then you will be indifferent about the sample that you get.