HW1, Theory of Inference 2016/7

Jonathan Rougier
School of Mathematics
University of Bristol UK

A general statement on homeworks. These homeworks are an opportunity for you to develop your understanding, and to practice your maths and communication skills. If you hand-in your homeworks, you will get feedback on how well you are doing. You are *strongly encouraged* to hand-in your homeworks.

It is crucial that you express your answers clearly, in well-structured sentences. Where you are writing maths, your writing must be tidy enough that there can be no ambiguity about symbols and the names of variables. The way you lay-out your maths must be logical and clear, using indentation, alignment, and other standard conventions. This is a skill you must master before you leave the university. Future employers and colleagues will rightly be critical of sloppy thinking and sloppy communicating.

I am told by students that my marking criteria are very strict. Please be absolutely clear that I am marking according to the the criteria that you will be judged by when you leave the university. Do not be put off by low marks. Come to an Office Hour to discuss how you could have done better, and study the solutions.

In the following questions, I show marks in square brackets, to give you an idea of the approximate tariff the question would carry in an exam.

1. The usual convention in statistics is to write $\mathrm{p}(x)$ for $\Pr(X = x)$, $\mathrm{p}(x \mid y)$ for $\Pr(X = x \mid Y = y)$, and so on. Under this convention, the definition of the conditional probability $\mathrm{p}(x \mid y)$ is any function that satisfies

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

   This definition implies that $\mathrm{p}(x \mid y)$ is arbitrary when $\mathrm{p}(y) = 0$, because in this case the relation reads $0 = \mathrm{p}(x \mid y) \cdot 0$.

(a) Consider the model $\{\mathcal{X} \times \mathcal{Y}, \Omega, f_{X,Y}\}$, where I will write $p(x, y \mid \theta)$ for $f_{X,Y}(x, y; \theta)$, and so on. I will treat Ω as uncountable. Prove that if $\mathrm{p}(y) > 0$, then

$$\mathrm{p}(x \mid y) = \int_\Omega \mathrm{p}(x \mid y, \theta)\, \mathrm{p}(\theta \mid y)\, \mathrm{d}\theta.$$

This is eq. (1.11) in the handout. [10 marks]

If you would like to read more about conditional probabilities, then see sections 2.1 to 2.3 of the handout *Statistical Modelling*, available at `https://people.maths.bris.ac.uk/~mazjcr/BMB/2016/BMBmodelling.pdf`.

2. Consider the case where $\mathcal{Y} = \Omega = \{-, +\}$, and where

$f_Y(y; \theta) =$

	$\theta = -$	$\theta = +$
$y = -$	0.80	0.05
$y = +$	0.20	0.95

Suppose that Θ represents whether or not a person has a disease, and Y represents whether or not that person tests positive for the disease. Think of Y as an algorithm for predicting Θ. Such as algorithm is certified in terms of its *sensitivity*, which is is $f_Y(+; +)$, and its *specificity*, which is $f_Y(-; -)$.

(a) Identify the sensitivity and specificity of the algorithm from the values of f_Y. Suppose you have tested positive. Compute the *likelihood ratio*

$$f_Y(+; +)/f_Y(+; -).$$

What does this tell you about whether you have the disease? [10 marks]

(b) Suppose that you now find out that the base rate of the disease in people similar to you is 5 people in every 1000. What is the probability that you have the disease, given that you have tested positive? [10 marks]

If you are planning to participate in any form of routine screening, or you have been to the doctor who has arranged for you to have some tests, then you should read *Reckoning with Risk* by Gerd Gigerenzer.