HW3, Theory of Inference 2016/7

Jonathan Rougier
School of Mathematics
University of Bristol UK

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. Prove that LP $\rightarrow$ WIP, and that LP $\rightarrow$ WCP. [10 marks]

   You should use this question to practice writing really clear and compelling proofs.

   Answer. The LP asserts that if $\mathcal{E}$ and $\mathcal{E}'$ are two experiments with the same parameter, and if $f(x;\bullet) = c(x,x') \cdot f'(x';\bullet)$, then $\mathrm{Ev}(\mathcal{E},x) = \mathrm{Ev}(\mathcal{E}',x')$.

   (a) LP $\rightarrow$ WIP. The WIP states that if $f(x;\bullet) = f(x';\bullet)$ in an experiment $\mathcal{E}$, then $\mathrm{Ev}(\mathcal{E},x) = \mathrm{Ev}(\mathcal{E},x')$. Letting $\mathcal{E}' = \mathcal{E}$, the condition of the WIP satisfies the condition of the LP with $c(x,x') = 1$, and hence the LP implies the WIP.

   (b) LP $\rightarrow$ WCP. The WCP states that if $\mathcal{E}_1$ and $\mathcal{E}_2$ are two experiments with the same parameter, and $\mathcal{E}^*$ is a mixture experiment with known probabilities (p_1,p_2) where $p_2 = (1-p_1)$, then $\mathrm{Ev}(\mathcal{E}^*,(i,x_i)) = \mathrm{Ev}(\mathcal{E}_i,x_i)$. For the mixture experiment we have

   $$f^*((i,x_i);\theta) = p_i \cdot f_i(x_i;\theta) \qquad \text{for all } \theta \in \Omega,$$

   which satisfies the condition of the LP with $c((i,x_i),x_i) = p_i$. Hence the LP implies that $\mathrm{Ev}(\mathcal{E}^*,(i,x_i)) = \mathrm{Ev}(\mathcal{E}_i,x_i)$, which is the WCP.

2. A Traffic Inspector sits outside the Bristol Royal Infirmary, on a road known to have a very high level of pollution. In one hour she records 423 passing cars (excluding vans and lorries), noting for each car the number of empty seats.

(a) Write down a model for this experiment, bearing in mind that the number of cars seen in an hour (N) is itself a random quantity. [10 marks]

Answer. There can be from 0 to 4 empty seats in a car, so let X_i be the number of cars with i empty seats, $i = 0, \ldots, 4$. Let N be the number of cars. So if $y = (n, x_0, \ldots, x_4)$, the sample space for the experiment is

$$\mathcal{Y} = \left\{ (n, x_0, \ldots, x_4) \in \mathbb{N}^6 : \sum_{i=0}^{4} x_i = n \right\},$$

where $\mathbb{N} = \{0, 1, \ldots\}$. We can let N be Poisson, with rate $\lambda > 0$, so

$$\mathrm{p}(n; \lambda) = e^{-\lambda} \lambda^n / n!.$$

Then using a Multinomial model with probabilities $\theta = (\theta_0, \ldots, \theta_4)$,

$$p(x \mid n; \theta) = \frac{n!}{x_0! \cdots x_4!} \prod_{i=0}^{4} (\theta_i)^{x_i}.$$

where the first term is the multinomial coefficient. These two choices would be standard. Hence the parameter space is

$$\Omega = \mathbb{R}_{++} \times \left\{ \theta \in \mathbb{R}^5 : \theta_i \geq 0, \sum\nolimits_i \theta_i = 1 \right\},$$

and the model is

$$f_{N,X}(n, x; \lambda, \theta) = \mathrm{p}(n; \lambda) \cdot \mathrm{p}(x \mid n; \theta).$$

(b) State what is meant by an 'ancillary' random quantity, and discuss whether N is ancillary. [10 marks]

Answer. N would be ancillary for the model with parameter (λ, θ) if the model factorised as

$$\mathrm{p}(n, x; \lambda, \theta) = \mathrm{p}(n) \cdot \mathrm{p}(x \mid n; \lambda, \theta),$$

so that all the parameters were in the conditional probability distribution (the fact that $X \mid N$ does not depend on λ is immaterial). So N is *not* ancillary, because its distribution depends on λ, which is part of the parameter. However, if λ were known then N would be ancillary, and if we accept the Sure Thing Principle (STP, which we do accept as being self-evident), then we can treat N as though it were ancillary. In this case we can base our

inference about θ around the conditional distribution $\mathrm{p}(x \mid n; \theta)$, treating N as though it were known and not random. Technically, this is the Strong Conditionality Principle (SCP).

(c) State the Conditionality Principle (CP), and show that it is implied by the Likelihood Principle (LP). [10 marks]

Answer. I will state the CP in the context of the question. The CP asserts that if N is ancillary (i.e. λ is known), then the evidence about Θ from the full experiment with (N, X) is the same as the evidence about Θ from the conditional experiment with $X \mid (N = n)$. To prove this, suppose that N is ancillary, in which case

$$f_{N,X}(n, x; \theta) = f_N(n) \cdot f_{X|N}(x \mid n; \theta) \quad \text{for all } \theta.$$

This satisfies the condition of the LP with $c((n, x), x) = f_N(n)$, where the lefthand side is the full experiment, and the righthand side is the conditional experiment.

(d) Treating N as ancillary, how does the Traffic Inspector analyse her results, if she adopts the CP? [10 marks]

Answer. Her model becomes the conditional model

$$\mathcal{E} = \{x \in \mathcal{X}, \theta \in \Omega, f_X\}, \tag{1}$$

where $f_X(x; \theta) = f_{X|N}(x \,|\, 423; \theta)$. She will need to specify the set of possible inferences about Θ, namely $\mathcal{A}$, and the consequences of choosing a when $\Theta = \theta$, represented as the loss function $L(a, \theta)$. She will want her Ev to obey the LP, because otherwise it would be difficult for her to justify the use of the CP. If she wants her Ev to obey the LP then she can let Ev be the Bayes rule for some prior distribution π: we have proved that such decision rules always obey the LP. So she will need a prior distribution over Ω. The Dirichlet distribution is the standard distribution in this case, for which she will have to choose concentration parameters, see `https://en.wikipedia.org/wiki/Dirichlet_distribution`. These concentration parameters can reflect her beliefs about Θ, or they can be 'vague'. Unless she has very strong beliefs about Θ, with $N = 423$ she can choose vague concentration parameters, because her choice is unlikely to make any difference to her inference.

In this answer, I have said rather more than is needed in an exam. You do

not need to know about the Dirichlet distribution. But the need to make sure that Ev respects the LP is important, given that the CP has been used to simplify the inference.

3. This question on stopping rules is from David MacKay's *Information Theory, Inference, and Learning Algorithms* (CUP, 2003), sec. 37.2. It is not the kind of question I would set in an exam, except for the last part. You will need to use R to compute the probabilities.

   In an expensive laboratory, Dr Bloggs tosses a coin twelve times and the result is HHHTHHHHTHHT. He is interested to know whether Θ, the probability of tails (T) is not equal to 0.5.

   (a) Dr Bloggs consults a Frequentist statistician, who tells him that he needs to compute a P-value, namely $\Pr(R \leq 3; n = 12, \theta = 0.5)$, where R is the number of tails. Give the formula for this P-value, and show that it is equal to 0.07.

   Answer. R is a Binomial random quantity, with

   $$\Pr(R = r; n, \theta) = \binom{n}{r} \theta^r \cdot (1 - \theta)^{n-r}.$$

   So

   $$\Pr(R \leq 3; n = 12, \theta = 0.5) = \sum_{r=0}^{3} \binom{12}{r} \left(\frac{1}{2}\right)^n = 0.07299805$$

   using the R command `pbinom(3, 12, 0.5)`.

   (b) Dr Bloggs is vexed; the P-value is not below the threshold of 0.05, which is what he needs to publish in the prestigious Journal of Experimental Coin Psychology. But when he talks to his statistician he discovers that the statistician failed to account for his (Dr Bloggs's) stopping rule. Dr Bloggs had decided to toss the coin until 3 tails had appeared. In that case, says the statistician, N, the sample size, is the random quantity, and we need to compute $\Pr(N \geq 12; r = 3, \theta = 0.5)$. Give the formula for $\Pr(N = n; r, \theta)$, and show that the P-value is equal to 0.03.

   Hint: To get r tails on the nth toss, we need to get exactly $r - 1$ tails in $n - 1$ tosses, and then a tail on the nth toss.

Answer. Taking the hint,

$$\Pr(N = n; r, \theta) = \Pr(R = r - 1; n - 1, \theta) \cdot \theta.$$

Hence

$$\begin{aligned}
\Pr(N \geq 12; r = 3, \theta = 0.5) &= 1 - \sum_{n=3}^{11} \Pr(N = n; r = 3, \theta = 0.5) \\
&= 1 - \sum_{n=3}^{11} \Pr(R = 2; n - 1, 0.5) \cdot 0.5 \\
&= 1 - \sum_{n=3}^{11} \binom{n-1}{2} \left(\frac{1}{2}\right)^n \\
&= 0.03271484
\end{aligned}$$

using the R command `1 - sum(choose(3:11 - 1, 2) * (1/2)^(3:11))`.

(c) Dr Bloggs is delighted, and writes up his result without delay, with the respectably low P-value of 0.03. What does the Stopping Rule Principle (SRP) say about this experiment? What do we infer about the Frequentist practice of using P-values for inference? What would a Bayesian statistician do in this situation? [15 marks]

Answer. The SRP states that the Dr Bloggs's inference about Θ should be invariant to his stopping rule; i.e. he should have made the same inference about Θ whether he did 12 trials regardless of the outcome, or whether he kept tossing the coin until he got 3 tails.

Clearly, though, Dr Bloggs's inference about Θ, presented in the form of a P-value for the simple hypothesis $\Theta = 0.5$, depends on the stopping rule, because he got one P-value for one stopping rule, and another P-value for another stopping rule (what he claims was his actual stopping rule). So Dr Bloggs's inference has violated the SRP.

The SRP is implied by the LP, and so by violating the SRP, Dr Bloggs's inference about Θ violates the LP. In general, all inference based on P-values violates the LP, because a P-value is a Frequentist construction, certified according to its sampling distribution (we will cover this in the lectures).

A Bayesian statistician would base her inference about Θ on the posterior distribution, which requires her to specify a prior distribution for Θ. Usually, this would be a Beta distribution (for convenience). Depending on the

inference she wanted to do, she would select a Bayes rule for her loss function. I reckon she might want a set estimator for Θ (to see whether 0.5 is inside it), in which case she might choose the 95% high posterior density region for Θ, which satisfies the necessary condition for set estimators to be Bayes rules according to the standard loss function.

I've said a bit more in this last paragraph than would be expected in an exam.