HW6, Theory of Inference 2015/6

Jonathan Rougier
School of Mathematics
University of Bristol UK

With these questions, you can get a feeling for how much detail is required by the tariff; the last question, for example, requires quite a lot of detail for 15 marks. Remember this very general advice about answering exam questions in maths: (i) make sure you have defined all of the terms in the question, and (ii) if you are asked to prove something, make sure you state exactly what it is you need to show.

1. Suppose that $Y \sim \text{Binomial}(p, n)$ where $p \in [0, 1]$ and n is specified. Give an example of a composite hypothesis with zero measure. [5 marks]

   Answer. A composite hypothesis contains more than one element of the parameter space. In this case the parameter space $[0, 1]$ has the cardinality of the continuum, and thus any countable subset has measure zero. So $H_0 : p \in \{0.1, 0.2\}$ is an example of a composite hypothesis with zero measure.

2. Apparently, 71 people out of 103 experienced 'enhanced gastric transport' after eating a serving of probiotic yogurt each day for two weeks. Starting from the model $Y \sim \text{Binomial}(p, n = 103)$, with $p \in [0, 1]$, sketch a Wilks 95% confidence set for p. You should provide as much information as you can without using a calculator. [10 marks]

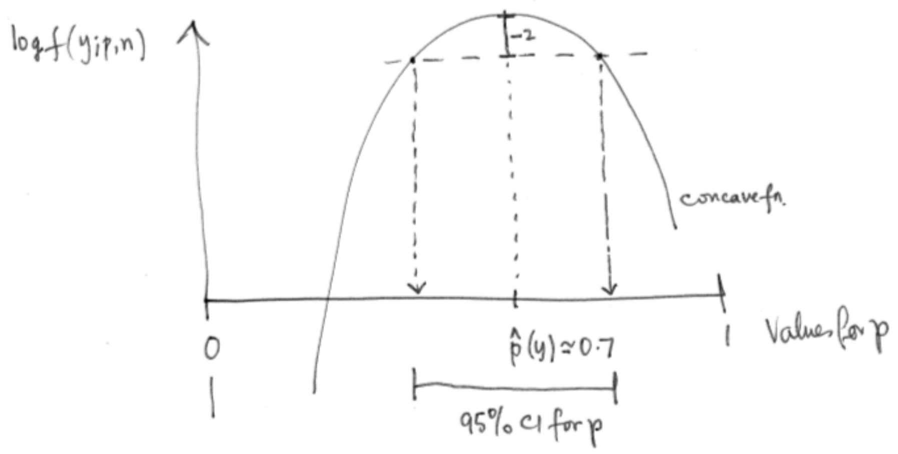
   Answer. This is a 1D problem. In this case, the Wilks 95% confidence procedure has, by Allan's Rule of Thumb,

$$C(y; 0.05) = \{p \in [0, 1] \mid \log f(y; p, n) > \log f(y; \hat{p}(y), n) - 2\}$$

where n is known and the '2' comes from the more general formula $F^{-1}_{\chi^2_1}(1 - \alpha)/2$. For the Binomial model,

$$\log f(y; p, n) = c + y \log p + (n - y) \log(1 - p)$$

where c is a constant that depends on (y, n) but not on p. It is easy to see that $\hat{p}(y) = y/n$. The second derivative of $\log f(y; p, n)$ with respect to p is negative, so $\log f(71; p, n = 103)$ is concave for $p \in (0, 1)$, peaking at $\hat{p}(y) = 71/103 \approx 0.7$, and asymptoting at $p = 0$ and $p = 1$. So here is my sketch:



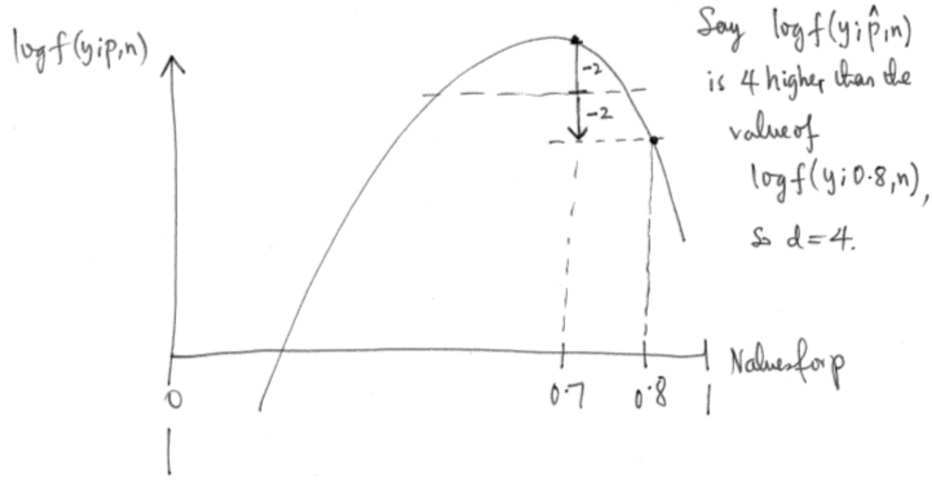
3. Suppose you wanted to test the null hypothesis $H_0 : p = 0.8$, and you had found out that 0.8 was outside the Wilks 95% confident set. How would you go about computing the P-value for H_0; illustrate with a sketch. [10 marks]

Answer. The P-value is the smallest value for α for which $p = 0.8$ is outside the level $(1-\alpha)$ confidence set. Since the parameter space is continuous, this is the value of α for which 0.8 is on the boundary of the level $(1-\alpha)$ confidence set.

If $p = 0.8$ is outside the 95% confidence set, then it must be above the upper limit of the confidence set, because the confidence set contains $p = 0.7$ (the ML estimate), and it is convex because the log-likelihood is concave. So I need to increase the level of the confidence set until 0.8 is at its upper limit. This involves finding the vertical distance from $\log f(71; 71/103, 103)$ to the level set that just touches $p = 0.8$; call this distance d. Then solve

$$d = F^{-1}_{\chi^2_1}(1-\alpha)$$

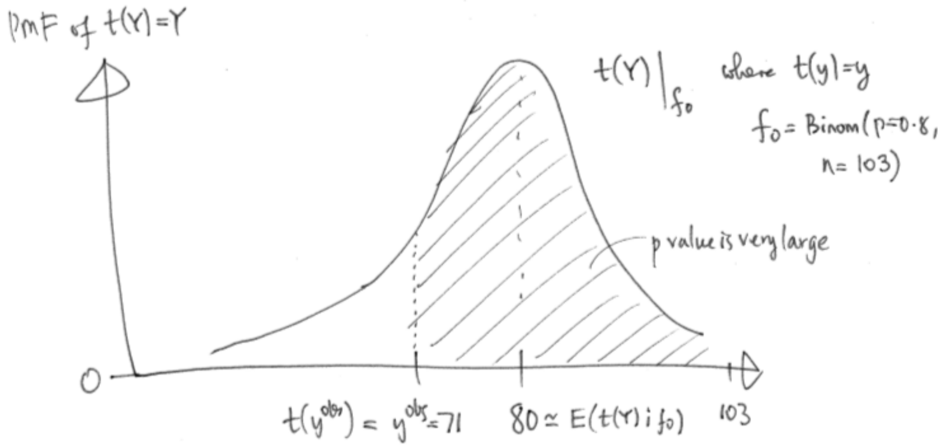
for α, which is the P-value. This value of α will be smaller than 0.05. Here is my sketch:



4. Suppose instead that you were asked to compute the P-value of $f_0 = \text{Binomial}(p = 0.8, n = 103)$ directly, using the test statistic $t(y) = y$. Provide a sketch showing the value that you would compute. Again, provide as much detail as you can without using a calculator. [10 marks]

Answer. As is well known, $\text{E}(Y; p, n) = np$, and $\text{Var}(Y; p, n) = np(1-p)$. So under the null model f_0, Y has expectation $103 \times 0.8 \approx 80$, variance

$103 \times 0.8 \times 0.2 \approx 1.6$, and standard deviation ≈ 1.3. This allows us to sketch the distribution of $t(Y) = Y$ under f_0, and show the P-value as an area. Here is my sketch:



5. Outline Fisher's dichotomy for interpreting P-values, and explain the main difficulty with this interpretation. [15 marks]

Answer. Suppose we have computed a small P-value such as $p(y^{\text{obs}}; t) = 0.015$. We know, from the definition of a significance procedure, that

$$\Pr\{p(Y; t) \leq 0.015; f_0\} \leq 0.015,$$

because $p(Y; t)$ is super-uniform under f_0. Fisher's dichotomy is, for a small p-value:

> EITHER f_0 is 'true' and something improbable has occurred, OR f_0 is false.

If we accept this dichotomy at face value, then we might conclude, on the basis that improbable events happen infrequently, that f_0 is probably false. This is technically illogical (we have not computed $\Pr\{f_0 = \textsf{true} | Y = y^{\text{obs}}\}$), yet it seems natural.

The main difficulty is that we know, a priori, that f_0 is false. We cannot read the book of nature, so all hypotheses are products of our minds (or, as a Martian would say, "our puny minds"), and formulated according to the very simple toolbox of statistical models which we have developed. Nature is much richer in her complexity than we can imagine, or express. So what we must really acknowledge is the much less interesting dichotomy

> f_0 is false. EITHER our significance procedure is not powerful enough to reveal this to us, OR it is.

Assuming a sensible choice for t, the power of a significance procedure is determined mainly by the sample size, which we denote as n. When n is small, we expect $p(Y; t)$ to be approximately uniform, which equates to approximately meaningless. When n is large, we expect $p(Y; t)$ to be tiny, because f_0 is false.

People who use significance procedures to 'reject' the null model f_0 must belive that somewhere between $n = 1$ and $n =$ large there is a 'sweet spot' in which a large P-value indicates that f_0 is 'good enough', and a small P-value indicates that f_0 should be rejected. Unfortunately, it is very difficult to determine either where the sweet-spot is, or what we mean by 'good enough'.

The prevailing view among professional statisticians is that the P-value is an indicator of the power of the significance procedure to reject f_0 (e.g, the sample size), rather than the 'truth' of f_0.