APTS Statistical Inference, assessment questions, 2016

To students and supervisors. These questions are formative rather than summative. It may be necessary to delete some parts of some questions for a summative assessment, or to provide additional questions.

1. In the notes and lecture I oversimplified the statement and proof of the Stopping Rule Theorem, i.e. that LP $\rightarrow$ SRP. Here is a better introductory paragraph, to replace the first and second paragraphs in section 2.5.

   "Consider a sequence of random quantities $X_1, X_2, \ldots$ with marginal PMFs

   $$f_n(x_1, \ldots, x_n; \theta) \qquad n = 1, 2, \ldots,$$

   where consistency requires that

   $$f_n(x_1, \ldots, x_n; \theta) = \sum_{y_1} \cdots \sum_{y_m} f_{n+m}(x_1, \ldots, x_n, y_1, \ldots y_m; \theta)$$

   for each $n, m \in 1, 2, \ldots$. In a sequential experiment, the number of X's that are observed is not fixed in advanced but depends on the values seen so far. That is, at time j, the decision to observe X_{j+1} can be modelled by a probability $p_j(x_1, \ldots, x_j)$. We can assume, resources being finite, that the experiment must stop at specified time m, if it has not stopped already, hence $p_m(x_1, \ldots, x_m) = 0$. Denote the stopping rule as $\tau := (p_1, \ldots, p_m)$."

   Now prove Theorem 2.6, with this more general definition of a stopping rule (which allows the rule to be stochastic). Hint: consider the case where the outcome is (x_1, x_2), to show that $f((x_1, x_2); \theta) = c(x_1, x_2) \cdot f_2((x_1, x_2); \theta)$ for some $c > 0$, and then generalize.

2. This is a question about Lindley's paradox (Lindley, 1957), P-values, and the conventional 0.05 threshold.

   (a) Suppose you are sitting in a bar talking to an experimental psychologist about significance levels. An informal statement of *Lindley's paradox* is that a P-value for H_0 smaller than α can correspond to a likelihood ratio $f_0(y)/f_1(y)$ greater than $1/\alpha$. Provide a proof of this statement which you can sketch on a napkin (by all means include the napkin in your answer.) Hint: see DeGroot and Schervish (2002, sec. 8.9), or work backwards from the next question.

(b) Study Figure 1, and see if you can replicate it, either showing your workings or including your code. What is this graph showing? For example, how would you feel about rejecting H_0 in favour of H_1 at a P-value of 0.0499, when $n = 20$?

Some hints:

$$\bar{X} \sim \mathrm{N}(\mu, \sigma^2/n) \tag{1a}$$

with σ^2 known (take $\sigma = 1$). The competing hypotheses are

$$H_0 : \mu = 0 \quad \text{versus} \quad H_1 : \mu = 1, \tag{1b}$$

i.e. a separation of σ. In this case, a sensible statistic for forming a P-value for H_0 is $g(x) = \bar{x}/\sigma = \bar{x}$, which will be larger under H_1 than H_0. This gives

$$p(\bar{x}; H_0) = \Pr(\bar{X} \geq \bar{x}; H_0) = 1 - \Phi(\bar{x}; 0, 1/n)$$

where Φ is the Normal distribution function with specified expectation and variance. You can then find the value of $\bar{x}$ which implies a P-value of 0.05 (use the `qnorm` function), and at this value compute the likelihood ratio $\phi(\bar{x}; H_0)/\phi(\bar{x}; H_1)$ where ϕ is the Normal PDF (use the `dnorm` function).

(c) Consider

$$H_0 : \mu = 0 \quad \text{versus} \quad H_1 : \mu > 0$$

in the case where σ^2 is known and n is fixed. Produce a graph showing the value of the minimum likelihood ratio over H_1 for a range of P-values for H_0 from 0.001 to 0.1. Check your graph against the minimum shown in Figure 1. Hint: you should be able to compute this graph directly. You might find Edwards *et al.* (1963) or Goodman (1999a,b) helpful.

Comment on whether the conventional choice of 0.05 is a suitable threshold for choosing between hypotheses, or whether some other choice might be better. You may also like to reflect on the origin of the value 0.05, see Cowles and Davis (1982).

3. Here's a question about confidence sets and P-values for the Poisson distribution. Let the model be

$$Y_1, \dots, Y_n \overset{\text{iid}}{\sim} \text{Poisson}(\lambda), \quad \lambda > 0.$$

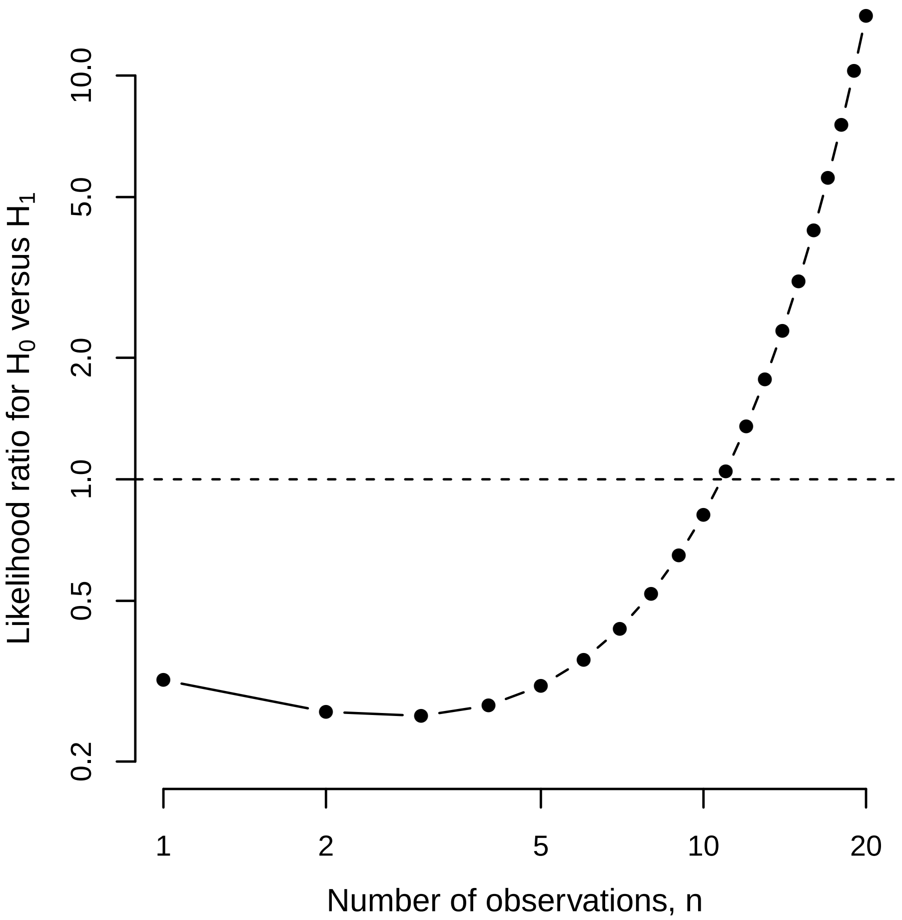

Figure 1: The likelihood ratio corresponding to a P-value for H_0 of 0.05, for the model and hypotheses given in (1).

(a) Show that the log-likelihood function has the form

$$\begin{aligned}\ell(\lambda; \boldsymbol{y}) &= \sum_{i=1}^{n} \Big(-\lambda + y_i \log(\lambda) - \log(y_i!) \Big) \\ &= c - n\lambda + n\bar{y}\log(\lambda)\end{aligned}$$

where $\bar{y} := n^{-1}(y_1 + \cdots + y_n)$, and c is some additive constant which can be ignored. Check that the Maximum Likelihood (ML) estimator is $\hat{\lambda}(\boldsymbol{y}) = \bar{y}$ (don't forget to check the second-order condition).

(b) Now suppose that $\bar{y} = 0.20$ from $n = 20$ observations. Draw the log-likelihood function, and indicate the Wilks's Theorem 95% approximately exact confidence set.

(c) Compute the P-value for the hypothesis $H_0 : \lambda = 0.1$. Hint: although you could do this by trial and error, you can also do it exactly (i.e. to within computer precision) once you know the value of the log-likelihood at $\lambda = 0.1$.

4. Here is a question on the Neyman-Pearson approach to Null Hypothesis Significance Testing (NHST). Suppose there are two simple hypotheses,

$$H_0 : Y \sim f_0 \quad \text{versus} \quad H_1 : Y \sim f_1.$$

   (a) You must choose between H_0 and H_1 according to the loss function

		'True' H_0	'True' H_1
Choose	H_0	c_{00}	c_{01}
	H_1	c_{10}	c_{11}

   where $c_{00} < c_{10}$ and $c_{11} < c_{01}$. Show that a decision rule for choosing between H_0 and H_1 is admissible if and only if it has the form

$$\frac{f_0(y)}{f_1(y)} \begin{cases} < c & \text{choose } H_1 \\ = c & \text{toss a coin} \\ > c & \text{choose } H_0 \end{cases} \tag{2}$$

   for some *critical value* $c > 0$. Hint: find the Bayes Rule for prior probabilities (π_0, π_1) where $\pi_1 = 1 - \pi_0$, and then apply the Complete Class Theorem (CCT).

   (b) Neyman-Pearson proposed a different approach. Let $\mathcal{R} \subset \mathcal{Y}$ be the region in which H_0 is rejected. Define the Type 1 (α) and Type 2 (β) error rates as

$$\alpha := \Pr(Y \in \mathcal{R}; H_0) \tag{Type 1}$$
$$\beta := \Pr(Y \notin \mathcal{R}; H_1) \tag{Type 2}$$

   Provide a verbal description of each of these errors. Now show that $\alpha + c \cdot \beta$ is minimised when

$$\mathcal{R} = \left\{ y \in \mathcal{Y} : f_0(y) / f_1(y) < c \right\} \tag{3}$$

   for some $c > 0$. Conclude by explaining why $\mathcal{R}$ defined in (3) minimises the β for any specified α.

(c) Define

$$r(y) := f_0(y)/f_1(y).$$

Show that $\mathrm{E}\{r(Y); H_1\} = 1$ and $\mathrm{E}\{r(Y); H_0\} \geq 1$ (hint: Jensen's inequality). Sketch, on the same axis, the PMFs of $r(Y)$ under H_0 and H_1. Add a critical value c, and shade in the areas representing the α and β values.

(d) The *operating characteristics curve (OCC)* of $\mathcal{R}$ in (3) plots α (x-axis) against β (y-axis) for varying c. Sketch this curve: you will need to show that it goes through $(0, 1)$ and $(1, 0)$, and also that it is convex. Hint for the last part: consider tossing a coin to choose between two different values of c: what would α and β be for this mixed decision rule?

(e) (Harder.) Explain why the gradient of the OCC is $\Delta\beta/\Delta\alpha \approx -1/c$. Add a few different specific values of c to your sketch: you should make sure that the aspect ratio of your sketch is approximately 1.

(f) Now consider

$$H_0 : Y_1, \dots, Y_n \overset{\text{iid}}{\sim} f_0 \quad \text{versus} \quad H_1 : Y_1, \dots, Y_n \overset{\text{iid}}{\sim} f_1.$$

Draw a new sketch showing the two cases where n is small and n is large (i.e. two OCCs on the same sketch). On the sketch draw the line $\alpha = 0.05$ and comment on the value of β in the two cases.

(g) Finally, explain how a sketch such as the one you have drawn in Q4f can be used to choose the sample size and the critical value for target values such as $\alpha = 0.05$ and $\beta = 0.2$. (These are common values in practice.)

5. Here is a brief exercise on one- and two-sided hypothesis tests. Suppose that $Y \sim f(\cdot\,; \mu, \sigma^2)$ where μ and σ^2 are respectively the expectation and variance of Y. Consider two different Null Hypothesis Significance Tests (NHSTs):

Test A	Test B
$H_0 : \kappa = c$	$H_0 : \kappa \geq c$
$H_1 : \kappa \neq c$	$H_1 : \kappa < c$

where $\kappa := \sigma/\mu \in \mathbb{R}_{++}$. Sketch the parameter space with μ on the x-axis, and σ on the y-axis. In this parameter space draw the set $\kappa = c$ for some $c > 0$. Now add three different 95% confidence sets for (μ, σ), corresponding to observations y, y', and y'', which satisfy the following tableau:

	y	y'	y''
Test A	Reject H_0	Fail to reject H_0	Reject H_0
Test B	Accept H_0 & reject H_1	Undecided	Reject H_0 & accept H_1

all at a significance level of 5%.

Brief comment. Test A is known as a 'one-sided' test, and Test B as a 'two-sided' test (also known as 'one-tailed' and 'two-tailed'). I find talk of one-sided and two-sided tests arcane and unhelpful. The key difference, it seems to me, is that sometimes one of the hypotheses, usually H_0, is reduced to a tiny set, so that it is never possible to 'Accept H_0': this seems to be the characteristic of a two-sided test. But if H_0 is a decent size, then it becomes possible to 'Accept H_0': a one-sided test.

In applications you will usually want to do a one-sided test. For example, if μ is the performance of a new treatment relative to a control, then you can be fairly sure *a priori* that $\mu = 0$ is false: different treatments seldom have identical effects. What you want to know is whether the new treatment is worse or better than the control: i.e. you want $H_0 : \mu \leq 0$ versus $H_1 : \mu > 0$. In this case you can find in favour of H_0, or in favour of H_1, or be undecided.

Jonathan Rougier
University of Bristol
Dec 2015

References

M.S. Bartlett, 1957. A comment on D.V. Lindley's statistical paradox. *Biometrika*, **44**, 533–534.

M. Cowles and C. Davis, 1982. On the origins of the .05 level of statistical significance. *American Psychologist*, **37**(5), 553–558.

M.H. DeGroot and M.J. Schervish, 2002. *Probability and Statistics*. Addison-Wesley Publishing Co., Reading MA, 3rd edition.

W. Edwards, H. Lindman, and L.J. Savage, 1963. Bayesian statistical inference for psychological research. *Psychological Review*, **70**(3), 193–242.

S. Goodman, 1999a. Toward evidence-based medical statistics. 1: The p-value fallacy. *Annals of Internal Medicine*, **130**, 995–1004.

S. Goodman, 1999b. Toward evidence-based medical statistics. 2: The Bayes factor. *Annals of Internal Medicine*, **130**, 1005–1013.

D.V. Lindley, 1957. A statistical paradox. *Biometrika*, **44**, 187–192. See also Bartlett (1957).