APTS Assessment on Statistical Inference

Simon Shaw, s.shaw@bath.ac.uk
University of Bath

Cambridge, 16-20 December 2019

Principles for Statistical Inference

1. Consider Birnbaum's Theorem, (WIP $\wedge$ WCP) $\leftrightarrow$ SLP. In lectures, we showed that (WIP $\wedge$ WCP) $\rightarrow$ SLP but not the converse. Hence, show that SLP $\rightarrow$ WIP and SLP $\rightarrow$ WCP.

2. Consider, given θ, a sequence of independent Bernoulli trials with parameter θ. We wish to make inferences about θ and consider two possible methods. In the first, we carry out n trials and let X denote the total number of successes in these trials. Thus, $X \,|\, \theta \sim Bin(n, \theta)$ with

$$f_X(x \,|\, \theta) \quad = \quad \binom{n}{x} \theta^x (1-\theta)^{n-x}, \quad x = 0, 1, \ldots, n.$$

In the second method, we count the total number Y of trials up to and including the rth success so that $Y \,|\, \theta \sim Nbin(r, \theta)$, the negative binomial distribution, with

$$f_Y(y \,|\, \theta) \quad = \quad \binom{y-1}{r-1} \theta^r (1-\theta)^{y-r}, \quad y = r, r+1, \ldots.$$

Suppose that we observe $x = r = 3$ and $y = n = 12$.

(a) For $\theta = 1/2$, calculate $\mathbb{P}(X \leq 3 \,|\, \theta = 1/2)$ and $\mathbb{P}(Y \geq 12 \,|\, \theta = 1/2)$. Consider the hypothesis test

$$H_0 : \theta = \frac{1}{2} \quad \text{versus} \quad H_1 : \theta < \frac{1}{2}.$$

For each method, what would you conclude for a test at significance level 5%? Interpret what this result says for the relationship between p-values and the Stopping Rule Principle (SRP).

(b) For a univariate parameter θ, a popular (default) noninformative prior distribution for θ in the model $\{\mathcal{X}, \Theta, f_X(x \,|\, \theta)\}$ is the Jeffreys prior,

$$\pi_X(\theta) \quad \propto \quad \sqrt{I_X(\theta)}$$

where the proportionality is with respect to θ and

$$I_X(\theta) \quad = \quad -\mathbb{E}\left(\frac{d^2}{d\theta^2} \log f_X(x \,|\, \theta) \,\middle|\, \theta \right)$$

is the Fisher information.

(i) Obtain the Jeffreys prior distribution for each of the two methods. You may find it useful to note that $E(Y \mid \theta) = \frac{r}{\theta}$. Are these prior distributions both proper?
[Hint: You may wish to consider the Beta distribution.]

(ii) For each method, calculate the posterior distribution for θ with the Jeffreys prior. Comment upon your answers.

(iii) What conclusions would a Bayesian statistician, using a prior distribution that reflected their prior knowledge about θ, do in this situation?

Statistical Decision Theory

3. Suppose we have a hypothesis test of two simple hypotheses

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

so that if H_i is true then X has distribution $f_i(x)$. It is proposed to choose between H_0 and H_1 using the following loss function.

		Decision	
		H_0	H_1
Outcome	H_0	c_{00}	c_{01}
	H_1	c_{10}	c_{11}

where $c_{00} < c_{01}$ and $c_{11} < c_{10}$. Thus, $c_{ij} = L(H_i, H_j)$ is the loss when the 'true' hypothesis is H_i and the decision H_j is taken. Show that a decision rule $\delta(x)$ for choosing between H_0 and H_1 is admissible if and only if

$$\delta(x) = \begin{cases} H_0 & \text{if } \frac{f_0(x)}{f_1(x)} > c, \\ H_1 & \text{if } \frac{f_0(x)}{f_1(x)} < c, \\ \text{either } H_0 \text{ or } H_1 & \text{if } \frac{f_0(x)}{f_1(x)} = c, \end{cases}$$

for some critical value $c > 0$.
[Hint: Consider Wald's Complete Class Theorem and a prior distribution $\pi = (\pi_0, \pi_1)$ where $\pi_i = \mathbb{P}(H_i) > 0$. You may assume that for all $x \in \mathcal{X}$, $f_i(x) > 0$.]

4. Let $X_1, \ldots, X_n$ be exchangeable random variables so that, conditional upon a parameter θ, the X_i are independent. Suppose that $X_i \mid \theta \sim N(\theta, \sigma^2)$ where the variance σ^2 is known, and that $\theta \sim N(\mu_0, \sigma_0^2)$ where the mean μ_0 and variance σ_0^2 are known. We wish to produce a point estimate d for θ, with loss function

$$L(\theta, d) = 1 - \exp\left\{-\frac{1}{2}(\theta - d)^2\right\}. \quad (1)$$

(a) Let $f(\theta)$ denote the probability density function of $\theta \sim N(\mu_0, \sigma_0^2)$. Show that $\rho(f, d)$, the risk of d under $f(\theta)$, can be expressed as

$$\rho(f, d) = 1 - \frac{1}{\sqrt{1 + \sigma_0^2}} \exp\left\{-\frac{1}{2(1 + \sigma_0^2)}(d - \mu_0)^2\right\}.$$

[Hint: You may use, without proof, the result that

$$(\theta - a)^2 + b(\theta - c)^2 \quad = \quad (1+b)\left(\theta - \frac{a+bc}{1+b}\right)^2 + \left(\frac{b}{1+b}\right)(a-c)^2$$

for any $a, b, c \in \mathbb{R}$ with $b \neq -1$.]

(b) Using part (a), show that the Bayes rule of an immediate decision is $d^* = \mu_0$ and find the corresponding Bayes risk.

(c) Find the Bayes rule and Bayes risk after observing $x = (x_1, \ldots, x_n)$. Express the Bayes rule as a weighted average of d^* and the maximum likelihood estimate of θ, $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, and interpret the weights.
[Hint: Consider conjugacy.]

(d) Suppose now, given data y, the parameter θ has the general posterior distribution $f(\theta \,|\, y)$. We wish to use the loss function $L(\theta, d)$, as given in equation (1), to find a point estimate d for θ. By considering an approximation of $L(\theta, d)$, or otherwise, what can you say about the corresponding Bayes rule?

Confidence sets and p-values

5. Show that if p is a family of significance procedures then

$$p(x; \Theta_0) \quad = \quad \sup_{\theta \in \Theta_0} p(x; \theta)$$

is a significance procedure for the null hypothesis $\Theta_0 \subset \Theta$, that is that $p(X; \Theta_0)$ is super-uniform for every $\theta \in \Theta_0$.

6. Suppose that, given θ, $X_1, \ldots, X_n$ are independent and identically distributed $N(\theta, 1)$ random variables so that, given θ, $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\theta, 1/n)$.

(a) Consider the test of the hypotheses

$$H_0 : \theta = 0 \quad \text{versus} \quad H_1 : \theta = 1$$

using the statistic $\overline{X}$ so that large observed values $\overline{x}$ support H_1. For a given n, the corresponding p-value is

$$p_n(\overline{x}; 0) \quad = \quad \mathbb{P}(\overline{X} \geq \overline{x} \,|\, \theta = 0).$$

We wish to investigate how, for a fixed p-value, the likelihood ratio for H_0 for versus H_1,

$$LR(H_0, H_1) \quad := \quad \frac{f(\overline{x} \,|\, \theta = 0)}{f(\overline{x} \,|\, \theta = 1)}$$

changes as n increases.

(i) Use R to create a plot of $LR(H_0, H_1)$ for each $n \in \{1, \ldots, 20\}$ where, for each n, $\overline{x}$ is the value which corresponds to a p-value of 0.05.
[Hint: You may need to utilise the `qnorm` and `dnorm` functions. The look of the plot may be improved by using a log-scale on the axes.]

(ii) Comment on your plot, in particular on what happens to the likelihood ratio as n increases. What is the implication for hypothesis testing and the corresponding (fixed) p-value?

(b) Consider the test of the hypotheses

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

using once again $\overline{X}$ as the test statistic.

(i) Suppose that $\overline{x} > 0$. Show that

$$lr(H_0, H_1) \;:=\; \min_{\theta>0} \frac{f(\overline{x} \mid \theta = 0)}{f(\overline{x} \mid \theta)} \;=\; \exp\left\{-\frac{n}{2}\overline{x}^2\right\}.$$

(ii) Use R to create a plot of $lr(H_0, H_0)$ for a range of p-values for H_0 from 0.001 to 0.1.[1] 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.[2]

[1] The plot doesn't depend upon the actual choice of n and so you may choose $n = 1$. Once again, the look of the plot may be improved by using a log-scale on the axes.

[2] For the origins of the use of 0.05 see Cowles, M. and C. Davis (1982). On the origins of the .05 level of statistical significance. *American Psychologist 37(5)*, 553-558.