APTS Assessment on Statistical Inference

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

Warwick, 13-16 December 2022

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. Suppose that we have two discrete experiments $\mathcal{E}_1 = \{\mathcal{X}_1, \Theta, f_{X_1}(x_1 \,|\, \theta)\}$ and $\mathcal{E}_2 = \{\mathcal{X}_2, \Theta, f_{X_2}(x_2 \,|\, \theta)\}$ and that, for $x_1' \in \mathcal{X}_1$ and $x_2' \in \mathcal{X}_2$,

$$f_{X_1}(x_1' \,|\, \theta) \quad = \quad c f_{X_2}(x_2' \,|\, \theta) \tag{1}$$

for all θ where c is a positive constant not depending upon θ (but which may depend on x_1', x_2') and $f_{X_1}(x_1' \,|\, \theta) > 0$. We wish to consider estimation of θ under a loss function $L(\theta, d)$ which is strictly convex in d for each θ. Thus, for all $d_1 \neq d_2 \in \mathcal{D}$, the decision space, and $\alpha \in (0, 1)$,

$$L(\theta, \alpha d_1 + (1 - \alpha) d_2) \quad < \quad \alpha L(\theta, d_1) + (1 - \alpha) L(\theta, d_2).$$

For the experiment $\mathcal{E}_j$, $j = 1, 2$, for the observation x_j we will use the decision rule $\delta_j(x_j)$ as our estimate of θ so that

$$\text{Ev}(\mathcal{E}_j, x_j) \quad = \quad \delta_j(x_j).$$

Suppose that the inference violates the strong likelihood principle so that, whilst equation (1) holds, $\delta_1(x_1') \neq \delta_2(x_2')$.

(a) Let $\mathcal{E}^*$ be the mixture of the experiments $\mathcal{E}_1$ and $\mathcal{E}_2$ according to mixture probabilities $1/2$ and $1/2$. For the outcome (j, x_j) the decision rule is $\delta(j, x_j)$. If the Weak Conditionality Principle (WCP) applies to $\mathcal{E}^*$ show that

$$\delta(1, x_1') \quad \neq \quad \delta(2, x_2').$$

(b) An alternative decision rule for $\mathcal{E}^*$ is

$$\delta^*(j, x_j) \quad = \quad \begin{cases} \frac{c}{c+1}\delta(1, x_1') + \frac{1}{c+1}\delta(2, x_2') & \text{if } x_j = x_j' \text{ for } j = 1, 2, \\ \delta(j, x_j) & \text{otherwise.} \end{cases}$$

Show that if the WCP applies to $\mathcal{E}^*$ then δ^* dominates δ so that δ is inadmissible. [Hint: First show that $R(\theta, \delta^*) = \frac{1}{2}\mathbb{E}[L(\theta, \delta^*(1, X_1)) \,|\, \theta] + \frac{1}{2}\mathbb{E}[L(\theta, \delta^*(2, X_2)) \,|\, \theta]$.]

(c) Comment on the result of part (b).

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	Decision 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

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

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 $X_i \,|\, \theta \sim N(\theta, \sigma^2)$ 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

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

   (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

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

   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 (2), 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 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.