Statistical Inference

https://people.bath.ac.uk/masss/APTS/lecture5.pdf

Simon Shaw

University of Bath

Lecture Five, 19 December 2019

Yesterday's lecture

- A decision rule δ_0 is admissible if there is no decision rule δ_1 which dominates it.
- Wald's Complete Class Theorem, CCT: a decision rule is admissible if and only if it is a Bayes rule for some prior distribution.
- Admissible decision rules respect the SLP.
- For quadratic loss, a point estimator for θ is admissible if and only if it is the conditional expectation with respect to some positive prior distribution $\pi(\theta)$.
- Level set property (LSP): a set $d \subset \Theta$ is a level set of the posterior distribution exactly when $d = \{\theta : \pi(\theta \,|\, x) \geq k\}$ for some k.
- If δ^* is a Bayes rule for $L(\theta, d) = |d| + \kappa(1 - 1_{\theta \in d})$ then it is a level set of the posterior distribution.

Confidence procedures and confidence sets

- We consider interval estimation, or more generally set estimation.
- Under the model $\mathcal{E} = \{\mathcal{X}, \Theta, f_X(x \mid \theta)\}$, for given data $X = x$, we wish to construct a set $C = C(x) \subset \Theta$ and the inference is the statement that $\theta \in C$.
- If $\theta \in \mathbb{R}$ then the set estimate is typically an interval.

Definition (Confidence procedure)

A random set $C(X)$ is a level-$(1-\alpha)$ confidence procedure exactly when

$$\mathbb{P}(\theta \in C(X) \mid \theta) \;\geq\; 1 - \alpha$$

for all $\theta \in \Theta$. C is an exact level-$(1-\alpha)$ confidence procedure if the probability equals $(1-\alpha)$ for all θ.

- The value $\mathbb{P}(\theta \in C(X) \,|\, \theta)$ is termed the coverage of C at θ.
- Exact is a special case: typically $\mathbb{P}(\theta \in C(X) \,|\, \theta)$ will depend upon θ.
- The procedure is thus conservative: for a given θ_0 the coverage may be much higher than $(1 - \alpha)$.

Uniform example

- Let $X_1, \ldots, X_n$ be independent and identically distributed $\text{Unif}(0, \theta)$ random variables where $\theta > 0$. Let $Y = \max\{X_1, \ldots, X_n\}$.
- We consider two possible sets: (aY, bY) where $1 \leq a < b$ and $(Y + c, Y + d)$ where $0 \leq c < d$.
  1. $\mathbb{P}(\theta \in (aY, bY) \,|\, \theta) = \left(\frac{1}{a}\right)^n - \left(\frac{1}{b}\right)^n$. Thus, the coverage probability of the interval does not depend upon θ.
  2. $\mathbb{P}(\theta \in (Y + c, Y + d) \,|\, \theta) = \left(1 - \frac{c}{\theta}\right)^n - \left(1 - \frac{d}{\theta}\right)^n$. In this case, the coverage probability of the interval does depend upon θ.

- We distinguish between the confidence procedure C, which is a random interval and so a function for each possible x, and the result when C is evaluated at the observation x, which is a set in Θ.

Definition (Confidence set)

The observed $C(x)$ is a level-$(1-\alpha)$ confidence set exactly when the random $C(X)$ is a level-$(1-\alpha)$ confidence procedure.

- If $\Theta \subset \mathbb{R}$ and $C(x)$ is convex, i.e. an interval, then a confidence set (interval) is represented by a lower and upper value.
- The challenge with confidence procedures is to construct one with a specified level: to do this we start with the level and then construct a C guaranteed to have this level.

Definition (Family of confidence procedures)

- $C(X; \alpha)$ is a family of confidence procedures exactly when $C(X; \alpha)$ is a level-$(1 - \alpha)$ confidence procedure for every $\alpha \in [0, 1]$.
- C is a nesting family exactly when $\alpha < \alpha'$ implies that $C(x; \alpha') \subset C(x; \alpha)$.

- If we start with a family of confidence procedures for a specified model, then we can compute a confidence set for any level we choose.

Constructing confidence procedures

- The general approach to construct a confidence procedure is to invert a test statistic.
- In the Uniform example, the coverage of the procedure (aY, bY) does not depend upon θ because the coverage probability could be expressed in terms of $T = Y/\theta$ where the distribution of T did not depend upon θ.
  - T is an example of a pivot and confidence procedures are straightforward to compute from a pivot.
  - However, a drawback to this approach in general is that there is no hard and fast method for finding a pivot.
- An alternate method which does work generally is to exploit the property that *every* confidence procedure corresponds to a hypothesis test and vice versa.

Consider a hypothesis test where we have to decide either to accept that an hypothesis H_0 is true or to reject H_0 in favour of an alternative hypothesis H_1 based on a sample $x \in \mathcal{X}$.

- The set of x for which H_0 is rejected is called the rejection region.
- The complement, where H_0 is accepted, is the acceptance region.
- A hypothesis test can be constructed from any statistic $T = T(X)$.

Definition (Likelihood Ratio Test, LRT)

The likelihood ratio test (LRT) statistic for testing $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_0^{\mathrm{c}}$, where $\Theta_0 \cup \Theta_0^{\mathrm{c}} = \Theta$, is

$$\lambda(x) \; = \; \frac{\sup_{\theta \in \Theta_0} L_X(\theta; x)}{\sup_{\theta \in \Theta} L_X(\theta; x)}.$$

A LRT at significance level α has a rejection region of the form $\{x \, : \, \lambda(x) \leq c\}$ where $0 \leq c \leq 1$ is chosen so that $\mathbb{P}(\text{Reject } H_0 \,|\, \theta) \leq \alpha$ for all $\theta \in \Theta_0$.

Example

- Let $X = (X_1, \ldots, X_n)$ and suppose that the X_i are independent and identically distributed $N(\theta, \sigma^2)$ random variables where σ^2 is known.
- Consider the likelihood ratio test for $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. Then, as the maximum likelihood estimate (mle) of θ is $\overline{x}$,

$$\lambda(x) = \frac{L_X(\theta_0; x)}{L_X(\overline{x}; x)} = \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left((x_i - \theta_0)^2 - (x_i - \overline{x})^2\right)\right\}$$
$$= \exp\left\{-\frac{1}{2\sigma^2} n(\overline{x} - \theta_0)^2\right\}.$$

Notice that, under H_0, $\frac{\sqrt{n}(\overline{X} - \theta_0)}{\sigma} \sim N(0,1)$ so that

$$-2\log\lambda(X) = \frac{n(\overline{X} - \theta_0)^2}{\sigma^2} \sim \chi^2_1,$$

the chi-squared distribution with one degree of freedom.

Example continued

- The rejection region is $\{x : \lambda(x) \leq c\} = \{x : -2\log\lambda(x) \geq k\}$.
- Setting $k = \chi^2_{1,\alpha}$, where $\mathbb{P}(\chi^2_1 \geq \chi^2_{1,\alpha}) = \alpha$, gives a test at the exact significance level α.

The acceptance region of this test is $\{x : -2\log\lambda(x) < \chi^2_{1,\alpha}\}$ where

$$\mathbb{P}\left(\frac{n(\overline{X}-\theta_0)^2}{\sigma^2} < \chi^2_{1,\alpha} \,\middle|\, \theta = \theta_0\right) \;=\; 1-\alpha.$$

This holds for all θ_0 and so, additionally rearranging,

$$\mathbb{P}\left(\overline{X} - \sqrt{\chi^2_{1,\alpha}}\frac{\sigma}{\sqrt{n}} < \theta < \overline{X} + \sqrt{\chi^2_{1,\alpha}}\frac{\sigma}{\sqrt{n}} \,\middle|\, \theta\right) \;=\; 1-\alpha.$$

Thus, $C(X) = (\overline{X} - \sqrt{\chi^2_{1,\alpha}}\frac{\sigma}{\sqrt{n}}, \overline{X} + \sqrt{\chi^2_{1,\alpha}}\frac{\sigma}{\sqrt{n}})$ is an exact level-$(1-\alpha)$ confidence procedure with $C(x)$ the corresponding confidence set.

- Note that we obtained the level-$(1-\alpha)$ confidence procedure by inverting the acceptance region of the level α significance test.
- This correspondence, or duality, between acceptance regions of tests and confidence sets is a general property.

Theorem (Duality of Acceptance Regions and Confidence Sets)

1. For each $\theta_0 \in \Theta$, let $A(\theta_0)$ be the acceptance region of a test of $H_0 : \theta = \theta_0$ at significance level α. For each $x \in \mathcal{X}$, define $C(x) = \{\theta_0 \, : \, x \in A(\theta_0)\}$. Then $C(X)$ is a level-$(1-\alpha)$ confidence procedure.
2. Let $C(X)$ be a level-$(1-\alpha)$ confidence procedure and, for any $\theta_0 \in \Theta$, define $A(\theta_0) = \{x \, : \, \theta_0 \in C(x)\}$. Then $A(\theta_0)$ is the acceptance region of a test of $H_0 : \theta = \theta_0$ at significance level α.

Proof

1. As we have a level α test for each $\theta_0 \in \Theta$ then $\mathbb{P}(X \in A(\theta_0) \,|\, \theta = \theta_0) \geq 1 - \alpha$. Since θ_0 is arbitrary we may write θ instead of θ_0 and so, for all $\theta \in \Theta$,

$$\mathbb{P}(\theta \in C(X) \,|\, \theta) \;=\; \mathbb{P}(X \in A(\theta) \,|\, \theta) \;\geq 1 - \alpha.$$

   Hence, $C(X)$ is a level-$(1 - \alpha)$ confidence procedure.

2. For a test of $H_0 : \theta = \theta_0$, the probability of a Type I error (rejecting H_0 when it is true) is

$$\mathbb{P}(X \notin A(\theta_0) \,|\, \theta = \theta_0) \;=\; \mathbb{P}(\theta_0 \notin C(X), \,|\, \theta = \theta_0) \;\leq\; \alpha$$

   since $C(X)$ is a level-$(1 - \alpha)$ confidence procedure. Hence, we have a test at significance level α. □

A possibly easier way to understand the relationship between significance tests and confidence sets is by defining the set $\{(x,\theta) : (x,\theta) \in \tilde{C}\}$ in the space $\mathcal{X} \times \Theta$ where $\tilde{C}$ is also a set in $\mathcal{X} \times \Theta$.

- For fixed x, define the confidence set as $C(x) = \{\theta : (x,\theta) \in \tilde{C}\}$.
- For fixed θ, define the acceptance region as $A(\theta) = \{x : (x,\theta) \in \tilde{C}\}$.

Example revisited

Letting $x = (x_1, \ldots, x_n)$, with $z^2_{\alpha/2} = \chi^2_{1,\alpha}$, define the set

$$\{(x,\theta) : (x,\theta) \in \tilde{C}\} = \left\{(x,\theta) : -z_{\alpha/2}\sigma/\sqrt{n} < \overline{x} - \theta < z_{\alpha/2}\sigma/\sqrt{n}\right\}.$$

The confidence set is then

$$C(x) = \left\{\theta : \overline{x} - z_{\alpha/2}\sigma/\sqrt{n} < \theta < \overline{x} + z_{\alpha/2}\sigma/\sqrt{n}\right\}$$

and acceptance region

$$A(\theta) = \left\{x : \theta - z_{\alpha/2}\sigma/\sqrt{n} < \overline{x} < \theta + z_{\alpha/2}\sigma/\sqrt{n}\right\}.$$

Good choices of confidence procedures

- In the previous chapter, we showed that, under the generic loss $L(\theta, d) = |d| + \kappa(1 - 1_{\theta \in d})$, a necessary condition for admissibility was that d was a level set of the posterior distribution.
- We now proceed by consider confidence procedures that satisfy a level set property for the likelihood $L_X(\theta; x) = f_X(x \,|\, \theta)$.

Definition (Level set property, LSP)

A confidence procedure C has the level set property exactly when

$$C(x) \;=\; \{\theta \,:\, f_X(x \,|\, \theta) > g(x)\}$$

for some $g : \mathcal{X} \to \mathbb{R}$.

We now show that we can construct a family of confidence procedures with the LSP. The result has pedagogic value, because it can be used to generate an uncountable number of families of confidence procedures, each with the level set property.