HW5, Theory of Inference 2015/6

Jonathan Rougier
School of Mathematics
University of Bristol UK

In these questions, we have $Y \sim \{f, \Omega\}$ as usual. The questions without marks are unlikely to turn up on an exam, but you should do them anyway, for practice.

1. Let X be a random quantity with distribution function F_X. Prove that if F_X is continuous then $Y := F_X(X)$ has a Uniform distribution. Note that this is a more general result than the easy proof in the case where F_X is strictly increasing. See Casella & Berger (2002, Thm 2.1.10).

   Answer. The difficulty is that F_X continuous does not imply that $F_X^{-1}(F_X(x)) = x$, because F_X may have flat segments indicating intervals for X for which $\Pr(a < X \leq b) = 0$.

2. This standard result is used in the Marginalisation theorem for confidence procedures. Let $q(y)$ and $r(y)$ be first-order sentences (i.e. statements about y which are either true or false). Prove that if

$$\forall(y)(q(y) \rightarrow r(y)),$$

   then $\Pr\{q(Y); \theta\} \leq \Pr\{r(Y); \theta\}$ for all $\theta \in \Omega$.

   Answer. There are lots of ways to answer this question. Here is mine. We have, by definition,

$$\Pr\{q(Y); \theta\} = \mathrm{E}\{\mathbb{1}_{q(Y)}; \theta\} = \sum\nolimits_y \mathbb{1}_{q(y)} \cdot f(y; \theta),$$

where $\mathbb{1}$ is the indicator function, as usual. Hence

$$\Pr\{r(Y);\theta\} - \Pr\{q(Y);\theta\} = \sum\nolimits_y (\mathbb{1}_{r(y)} - \mathbb{1}_{q(y)}) \cdot f(y;\theta).$$

If $q(y) \to r(y)$ then $\mathbb{1}_{q(y)} \leq \mathbb{1}_{r(y)}$, and $\mathbb{1}_{r(y)} - \mathbb{1}_{q(y)} \geq 0$. If this holds for all $y \in \mathcal{Y}$ then

$$\begin{aligned} \Pr\{r(Y);\theta\} - \Pr\{q(Y);\theta\} &= \sum\nolimits_y (\mathbb{1}_{r(y)} - \mathbb{1}_{q(y)}) \cdot f(y;\theta) \qquad \text{as above} \\ &\geq \sum\nolimits_y 0 \cdot f(y;\theta) = 0. \end{aligned}$$

As θ is arbitrary, this shows that $\Pr\{q(Y);\theta\} \leq \Pr\{r(Y);\theta\}$ for all θ, as required.

3. Consider functions $C : \mathcal{Y} \times [0,1] \times [0,1] \to 2^\Omega$ of the form

$$C(y,u,\alpha) = \big\{\theta_j \in \Omega : a \cdot u + b > \alpha\big\}.$$

Suppose that $U \sim \mathrm{U}[0,1]$, independently of Y for all $\theta \in \Omega$. State the model for (Y,U). Under what conditions on a and b is C a family of level $(1-\alpha)$ confidence procedures for θ? Interpret this result as a reflection on the usefulness of confidence procedures. [15 marks]

Answer. The model for (Y,U) is $\{f^*, \Omega\}$ where

$$f^*(y,u;\theta) = \Pr(U=u) \cdot \Pr(Y=y;\theta) = \mathbb{1}_{0 \leq u \leq 1} \cdot f(y;\theta),$$

by the independence of U and Y, and that U is Uniform. The Families of Confidence Procedures theorem states that C is a family of confidence procedures if (and only if) $a \cdot U + b$ is super-uniform for all θ. Let $G := a \cdot U + b$. Then we require $F_G(v) \leq v$ for all $v \in [0,1]$, where F_G is the distribution function of G. Now

$$\begin{aligned} F_G(v) &= \Pr(G \le v) \\ &= \Pr(a \cdot U + b \le v) \\ &= \Pr(U \le (v-b)/a) && \text{taking } a > 0 \\ &= (v-b)/a. \end{aligned}$$

So a and b must satisfy $(v-b)/a \le v$ for all $v \in [0,1]$. Taking $v = 0$ implies that $b \ge 0$. Taking $v = 1$ implies that $b \ge 1 - a$. So $a \ge 1$ and $b \ge 0$ suffice.

If we take $a \leftarrow 1$ and $b \leftarrow 0$, then C is a family of exact confidence procedures for θ. We also have, for any (a, b) satisfying the condition above, an uncountably infinite number of other families of confidence procedures for θ. That is a lot of confidence procedures. They are all completely useless, since they make no reference to y at all. So this example warns us that there is an uncountably infinite number of confidence procedures, for any level we care to select, which are useless. Therefore we have to make a special effort to select confidence procedures which are useful.

4. Using Wilks's Theorem, construct a family of approximately exact confidence procedures which does not satisfy the Level Set Property. Sketch a level 95% confidence set from this family in the case where $\Omega \subset \mathbb{R}$. [10 marks]

Answer. Let $w(y, \theta_j)$ be the Wilks's Theorem test statistic,

$$w(y, \theta_j) := -2 \log \frac{f(y; \theta_j)}{f(y; \hat{\theta}(y))},$$

where $\hat{\theta}$ is the Maximum Likelihood Estimator. Wilks's Theorem asserts that, under certain regularity conditions, $w(Y, \theta_j)\big|_{\theta=\theta_j}$ has an approximately χ^2_d distribution for all $\theta_j \in \Omega$, where $d := \dim \Omega$. According to the Probability Integral Transform,

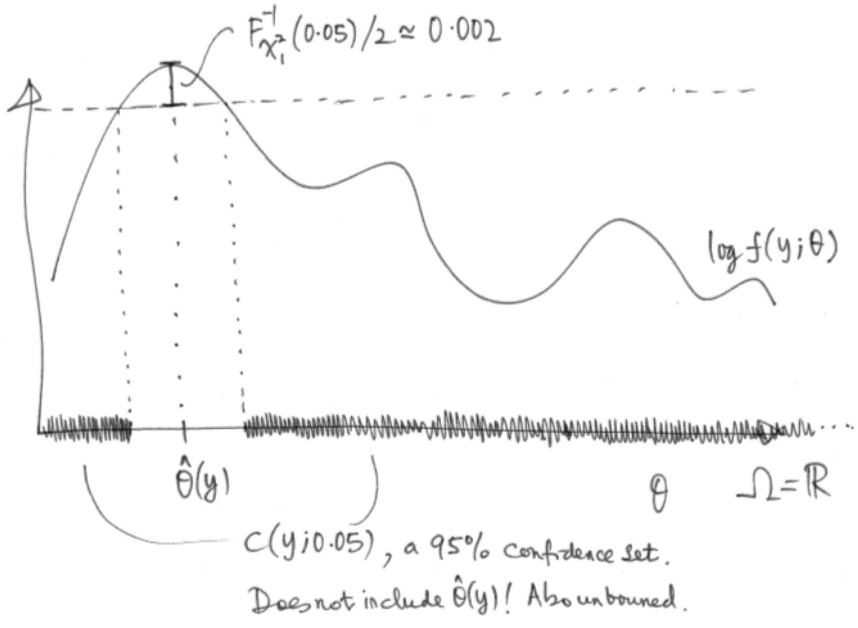
$$U := F_{\chi^2_d}\big(w(Y, \theta_j)\big)$$

has an approximately Uniform distribution for all $\theta_j \in \Omega$, where $F_{\chi^2_d}$ is the distribution function of the χ^2_d distribution.

In the lecture we used that if U is uniform, then $1 - U$ is also uniform. But what if we had continued with U instead of $1 - U$? From the Family of Confidence Procedures theorem, the acceptance region of the level $(1 - \alpha)$ confidence procedure would have been

$$
\begin{aligned}
&F_{\chi^2_d}\big(w(y, \theta_j)\big) > \alpha \\
&\qquad \iff w(y, \theta_j) > F^{-1}_{\chi^2_d}(\alpha) \\
&\qquad \iff -2 \log \frac{f(y; \theta_j)}{f(y; \hat{\theta}(y))} > F^{-1}_{\chi^2_d}(\alpha) \\
&\qquad \iff \log f(y; \theta) < \log f(y; \hat{\theta}(y)) - F^{-1}_{\chi^2_d}(\alpha)/2
\end{aligned}
$$

which is the opposite of the Level Set Property. Here is my sketch. This is a legitimate but alarming confidence set, and definitely not a confidence interval.



5. Prove that confidence sets which satisfy the Level Set Property are 'transformation invariant'. Start by exploring what this might mean

(hint: think about expressing the same model with a different parameter ϕ, where $g : \theta \mapsto \phi$ is bijective). [15 marks]

Answer. Suppose we had expressed our model in terms of ϕ instead of θ, so that

$$Y \sim f_\phi(y; \phi) \qquad \phi \in \Phi,$$

where $f_\phi(y; \phi) = f(y; g^{-1}(\phi))$ and $\Phi = g\Omega$. For C to be transformation invariant, we would want $C_\phi(y) = gC(y)$ for all y. That is, we would get the same set in Ω whether we specified our model with θ itself, or with ϕ and then transformed. We have already shown that the MLE is transformation invariant.

Assume that our confidence procedures both have the Level Set Property, i.e.

$$C(y) = \left\{\theta \in \Omega \,\middle|\, f(y; \theta) \geq c\right\}$$
$$C_\phi(y) = \left\{\phi \in \Phi \,\middle|\, f_\phi(y; \phi) \geq c\right\}$$

where, for reasons that will become apparent, I have used the same threshold c in both cases. We have to show that $\theta \in C(y) \iff \phi \in C_\phi(y)$. So:

$$\begin{aligned}
\theta \in C(y) &\iff f(y; \theta) \geq c \\
&\iff f(y; g^{-1}g(\theta)) \geq c \\
&\iff f(y; g^{-1}(\phi)) \geq c \qquad\qquad \text{as } \phi = g(\theta) \\
&\iff f_\phi(y; \phi) \geq c \\
&\iff \phi \in C_\phi(y),
\end{aligned}$$

as needed to be shown.