# HW8

## Sam Olson

## Q1

Suppose there is one observation $X$ with pdf

$$f(x) = 2\theta(1 - 2x) + 2x, \quad \text{for } x \in [0, 1], \theta \in [0, 1].$$

Find the Bayes test for

$$H_0 : \theta \leq 0.4 \quad \text{vs.} \quad H_1 : \theta > 0.4$$

with respect to the uniform prior on $[0, 1]$.

### Answer

We are given the likelihood:

$$f(x \mid \theta) = 2\theta(1 - 2x) + 2x, \quad x \in [0, 1], \theta \in [0, 1]$$

and a uniform prior, corresponding to:

$$\pi(\theta) = 1, \quad \text{for } \theta \in [0, 1]$$

The posterior is then:

$$\pi(\theta \mid x) \propto f(x \mid \theta)\pi(\theta) = 2\theta(1 - 2x) + 2x$$

Normalizing the posterior (to ensure the function is proper):

$$\int_0^1 [2\theta(1 - 2x) + 2x] \, d\theta = (1 - 2x) \cdot 1 + 2x \cdot 1 = 1$$

Giving a posterior of the form:

$$\pi(\theta \mid x) = 2\theta(1 - 2x) + 2x$$

The posterior probabilities under the null is given by:

$$P(H_0 \mid x) = \int_0^{0.4} \pi(\theta \mid x) \, d\theta = (1 - 2x) \cdot (0.4)^2 + 2x \cdot 0.4 = 0.16(1 - 2x) + 0.8x$$

1

Simplifying:

$$P(H_0 \mid x) = 0.16 + 0.48x$$

We may then construct the posterior probability under the alternative:

$$P(H_1 \mid x) = 1 - P(H_0 \mid x) = 0.84 - 0.48x$$

By definition, the Bayes test rejects $H_0$ when $P(H_1 \mid x) > P(H_0 \mid x)$.

Using our now known values of these probabilities, we have:

$$0.84 - 0.48x > 0.16 + 0.48x \rightarrow 0.68 > 0.96x \rightarrow x < \frac{17}{24}$$

We may now define the (Bayes) test function $\varphi(x)$ as:

$$\varphi(x) = \begin{cases} 1, & \text{if } x < \dfrac{17}{24}, \\[2mm] 0, & \text{if } x \geq \dfrac{17}{24} \end{cases}$$

# Q2

Problem 9.13, Casella and Berger (2nd Edition)

Let $X$ be a single observation from the Beta$(\theta, 1)$ pdf.

## a)

Let $Y = -(\log X)^{-1}$. Evaluate the confidence coefficient of the set $[y/2, y]$.

### Answer

As given, $X \sim \text{Beta}(\theta, 1)$. So we know its pdf is of the form:

$$f_X(x) = \theta x^{\theta-1}, \quad 0 < x < 1$$

Define the continuous, monotonic, and one-to-one transformation (over the support):

$$Y = -\frac{1}{\log X} \Rightarrow X = e^{-1/Y}, \quad Y > 0$$

Via the change of variables, the pdf of Y has the form:

$$f_Y(y) = f_X(e^{-1/y}) \cdot \left| \frac{d}{dy} e^{-1/y} \right| = \theta \cdot e^{-\theta/y} \cdot \frac{1}{y^2} = \frac{\theta}{y^2} e^{-\theta/y}, \quad y > 0$$

It then follows:

$$P\left(\frac{Y}{2} \leq \theta \leq Y\right) = P\left(\theta \in \left[\frac{Y}{2}, Y\right]\right) = P\left(Y \in [\theta, 2\theta]\right)$$

Thus:

$$\text{CC} = \min_{\theta \in \Theta} P_\theta(\theta \in I(\underset{\sim}{Y})) = \int_\theta^{2\theta} \frac{\theta}{y^2} e^{-\theta/y} \, dy$$

Via u substitution to evaluate, where $u = \theta/y \Rightarrow y = \theta/u$, $dy = -\theta/u^2 du$:

$$\text{CC} = \int_{1/2}^1 e^{-u} \, du = e^{-1/2} - e^{-1} \approx 0.6065 - 0.3679 = 0.2386$$

## b)

Find a pivotal quantity and use it to set up a confidence interval having the same confidence coefficient as part a).

**Answer**

The pdf of X is:

$$f_X(x) = \theta x^{\theta-1}, \quad x \in (0, 1)$$

Consider the transformation:

$$T = X^\theta$$

The cdf of the statistic $T$ is:

$$P(X^\theta \leq t) = P(X \leq t^{1/\theta}) = \int_0^{t^{1/\theta}} \theta x^{\theta-1} dx = t$$

This is a recognizable form! Specifically, $T = X^\theta \sim \text{Uniform}(0, 1)$, and is a pivotal quantity (it does not depend on the parameter $\theta$).

To construct a confidence interval, we want to find values $a, b \in (0, 1)$ such that:

$$P(a \leq T \leq b) = b - a = 0.239$$

Substituting known relations:

$$P(a \leq X^\theta \leq b) = 0.239$$

Solving for $\theta$:

$$a \leq X^\theta \leq b \rightarrow \frac{\log a}{\log X} \leq \theta \leq \frac{\log b}{\log X}$$

Noting: $0 < X < 1 \rightarrow \log X < 0$

**c)**

Compare the two confidence intervals.

**Answer**

The interval in part a) is a special case of part b), but doesn't always produce the shortest possible confidence interval. The pivotal method in part b) is more optimal as it minimizes the interval length by choosing $a = \alpha$ and $b = 1$, so the method from part b) produces the shortest $1 - \alpha$ confidence interval via:

$$\left\{ \theta : 0 \leq \theta \leq \frac{\log \alpha}{\log X} \right\}$$

For $\alpha = 0.239$, the coverage from part a), part b) gives a shorter interval. However, the methods from part a) and b) agree in some instances (when $a = e^{-1}$ and $b = e^{-1/2}$).

I'd recommend using part a) for simplicity, but part b) for minimal length/optimality, which can be dependent upon the context of application (for example, Dr. Nordman gave an example of economists sometimes being interested in one-sided confidence intervals for the maximum/upper bounds, which would be an instance of using the method from part b).

# Q3

Problem 9.16, Casella and Berger (2nd Edition)

Let $X_1, \ldots, X_n$ be i.i.d. $N(\theta, \sigma^2)$, where $\sigma^2$ is known. For each of the following hypotheses, write out the acceptance region of a level $\alpha$ test and the $1 - \alpha$ confidence interval that results from inverting the test.

## a)

$H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$

**Answer**

As given, we have n iid Normal random variables. We may normalize these as follows (under $H_0$):

$$Z = \frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}} \sim N(0, 1)$$

For the null hypothesis as given,

Reject $H_0$ if:

$$|Z| > z_{\alpha/2} \quad \Longleftrightarrow \quad \left|\bar{X} - \theta_0\right| > z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

And accept $H_0$ if:

$$\theta_0 \in \left[\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \ \bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right]$$

Inverting the test yields the $(1 - \alpha)$ confidence interval for $\theta$:

$$I(\underset{\sim}{X}) = \left[\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \ \bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right]$$

## b)

$H_0 : \theta \geq \theta_0$ versus $H_1 : \theta < \theta_0$

**Answer**

We again have our standard normal iid random variable(s):

$$Z = \frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}}$$

As we now are dealing with a one-sided alternative hypothesis, we have a UMP test of the form (taken for granted, not proven):

Reject $H_0$ if:

$$Z < -z_\alpha \quad \Longleftrightarrow \quad \bar{X} - \theta_0 < -z_\alpha \cdot \frac{\sigma}{\sqrt{n}}$$

And accept $H_0$ if:

$$\bar{X} \geq \theta_0 - z_\alpha \cdot \frac{\sigma}{\sqrt{n}}$$

Inverting the test yields the one-sided confidence interval:

$$I(\underset{\sim}{X}) = \left(-\infty, \bar{X} + z_\alpha \cdot \frac{\sigma}{\sqrt{n}}\right]$$

**c)**

$H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$

**Answer**

Once more, we have our standard normal iid random variables of the form:

$$Z = \frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}}$$

And given the direction of our one-sided alternative hypothesis, we have a UMP test of the form (again, taken for granted, not proven):

Reject $H_0$ if:

$$Z > z_\alpha \quad \Longleftrightarrow \quad \bar{X} - \theta_0 > z_\alpha \cdot \frac{\sigma}{\sqrt{n}}$$

And accept $H_0$ if:

$$\bar{X} \leq \theta_0 + z_\alpha \cdot \frac{\sigma}{\sqrt{n}}$$

Inverting the test gives the one-sided interval:

$$I(\underset{\sim}{X}) = \left[\bar{X} - z_\alpha \cdot \frac{\sigma}{\sqrt{n}}, \infty\right)$$

# Q4

Problem 9.11, Casella and Berger (2nd Edition)

If $T$ is a continuous random variable with cdf $F_T(t \mid \theta)$ and $\alpha_1 + \alpha_2 = \alpha$, show that an $\alpha$-level acceptance region of the hypothesis $H_0 : \theta = \theta_0$ is

$$\{t : \alpha_1 \leq F_T(t \mid \theta_0) \leq 1 - \alpha_2\},$$

with associated confidence $1 - \alpha$ set

$$\{\theta : \alpha_1 \leq F_T(t \mid \theta) \leq 1 - \alpha_2\}.$$

## Answer

As given:

Let $T$ be a continuous test statistic with cumulative distribution function $F_T(t \mid \theta)$.

Under the null hypothesis $H_0 : \theta = \theta_0$, we define the transformed variable:

$$U = F_T(T \mid \theta_0)$$

Since $T$ is continuous and $F_T(\cdot \mid \theta_0)$ is strictly increasing, via PIT, we know that under $H_0$:

$$U \sim \text{Uniform}(0, 1)$$

So we then construct an acceptance region that excludes the lower $\alpha_1$ and upper $\alpha_2$ tails of this uniform distribution.

More descriptively, we accept $H_0$ if:

$$\alpha_1 \leq F_T(t \mid \theta_0) \leq 1 - \alpha_2$$

Meaning the probability of rejecting $H_0$ is the sum of the tail probabilities:

$$P_{\theta_0}\left(F_T(T \mid \theta_0) < \alpha_1\right) = \alpha_1$$

And:

$$P_{\theta_0}\left(F_T(T \mid \theta_0) > 1 - \alpha_2\right) = \alpha_2$$

Taken together, the total probability of rejection under $H_0$ is:

$$P_{\theta_0}(\text{Reject } H_0) = \alpha_1 + \alpha_2 = \alpha$$

So this acceptance region defines a level $\alpha$ test.

We then construct a $1 - \alpha$ confidence set by inverting the acceptance region:

We then fix $t_{\text{obs}}$, and define the set of all parameter values $\theta$ for which $t_{\text{obs}}$ lies within the acceptance region:

$$C(t_{\text{obs}}) = \{\theta : \alpha_1 \leq F_T(t_{\text{obs}} \mid \theta) \leq 1 - \alpha_2\}$$

Since $F_T(T \mid \theta) \sim \text{Uniform}(0, 1)$ (noting PIT) under the true parameter $\theta$, we have:

$$P_\theta \left(\alpha_1 \leq F_T(T \mid \theta) \leq 1 - \alpha_2\right) = 1 - \alpha_1 - \alpha_2 = 1 - \alpha$$

Thus, the random set

$$\{\theta : \alpha_1 \leq F_T(t \mid \theta) \leq 1 - \alpha_2\}$$

is a confidence set for $\theta$ with coverage probability $1 - \alpha$.

So we have an acceptance region (level $\alpha$):

$$\{t : \alpha_1 \leq F_T(t \mid \theta_0) \leq 1 - \alpha_2\}$$

With confidence set (level $1 - \alpha$):

$$\{\theta : \alpha_1 \leq F_T(t \mid \theta) \leq 1 - \alpha_2\}$$

We may take advantage of the PIT further (knowing $F_T(t \mid \theta) \sim U(0, 1)$) to note:

$$\alpha_1 = \alpha_2 \rightarrow C_{\underset{\sim}{X}} = \left\{\theta : \frac{\alpha}{2} \leq F_T(t \mid \theta) \leq 1 - \frac{\alpha}{2}\right\}$$