STAT0008 Lecture 10
Bayesian Hypothesis Testing and Interval Estimation

Dr. Aidan O'Keeffe

Department of Statistical Science
University College London

10th December 2018

Outline

- Frequentist confidence intervals using a probability integral transform
- Bayesian approach to hypothesis testing
- Bayesian interval estimation
- Comparison between Bayesian and frequentist methods

Frequentist Confidence Intervals using a Probability Integral Transformation

From Lecture 9, recall that we can use a **pivotal quantity** to construct a $100(1-\alpha)\%$ confidence interval for an unknown parameter.

Last week, we identified common pivotal quantities for the scenario where $X_1, \ldots, X_n$ are iid $\mathcal{N}(\mu, \sigma^2)$. However, what about other distributions? How do we determine a pivotal quantity?

Once again, we make use of a **sufficient statistic** for a given parameter.

Frequentist Confidence Intervals using a Probability Integral Transformation

Suppose that we have $X_1, \ldots, X_n$ with $X_i \sim \mathcal{D}(\theta)$ and $T(\mathbf{X})$ is sufficient for θ where $T(\mathbf{X})$ has a continuous distribution with cumulative density function

$$F(t, \theta) = \mathbb{P}(T(\mathbf{X}) \leq t \mid \theta).$$

If we define the random variable

$$U(T(\mathbf{X}), \theta) = F(T(\mathbf{X}), \theta)$$

then $U(T(\mathbf{X}), \theta) \sim \mathcal{U}(0, 1)$ and, hence, $U(T(\mathbf{X}), \theta)$ is a pivotal quantity.

Proof

Given

$$U(T(\mathbf{X}), \theta) = F(T(\mathbf{X}), \theta)$$

It follows that

$$\begin{aligned}
\mathbb{P}(U(T(\mathbf{X}), \theta) \leq u) &= \mathbb{P}\left(F(T(\mathbf{X}), \theta) \leq u\right) \\
&= \mathbb{P}\left(F^{-1}(F(T(\mathbf{X}), \theta), \theta) \leq F^{-1}(u, \theta)\right) \\
&= \mathbb{P}\left(T(\mathbf{X}) \leq F^{-1}(u, \theta)\right).
\end{aligned}$$

The final line of the above is simply the cdf of $T(\mathbf{X})$ evaluated at $F^{-1}(u, \theta)$. Therefore, it follows that

$$\begin{aligned}
\mathbb{P}(U(T(\mathbf{X}), \theta) \leq u) &= F(F^{-1}(u, \theta), \theta) \\
&= u
\end{aligned}$$

Hence

$$U(T(\mathbf{X}), \theta) \sim \mathcal{U}(0, 1).$$

Notes

Since $T(\mathbf{X})$ is continuous and $F(t, \theta)$ is a strictly increasing function of t then F^{-1} exists.

Because $U(T(\mathbf{X}), \theta)$ is a pivotal quantity and the cumulative density function of U is known then we can re-arrange the cdf to obtain a confidence interval.

Example

Suppose that $X = x$ is a single observation on the lifetime of a component, where $X \sim \text{Exp}(\lambda)$. Construct a $100(1-\alpha)\%$ confidence interval for λ.

Example

Bayesian Approach to Hypothesis Testing

Suppose that $X_1, \ldots, X_n$ are iid random variables with some distribution, $\mathcal{D}(\theta)$, parameterised by θ. When using a Bayesian framework for inference, we know that we

1. Specify a prior distribution for θ.
2. Summarise the data, $\mathbf{x}$, using the likelihood function $\mathcal{L}(\theta \mid \mathbf{x})$.
3. Derive the **posterior distribution.**

We consider how to develop a Bayesian framework for hypothesis testing.

First, recall that the hypotheses that we've considered are **parametric** (i.e. specified in terms of the distributional parameter of interest θ).

For example

$$\mathsf{H}_0\colon \theta = \theta_0 \quad \text{versus} \quad \mathsf{H}_1\colon \theta = \theta_1.$$

Bayesian Approach to Hypothesis Testing

Since a prior distribution specifies our beliefs about θ, where the hypotheses are specified in terms of θ (i.e. are parametric) we can compute the probability of H_0 being true and the probability of H_1 being true, a priori.

Then, we can combine these probabilities with information from the data, using the likelihood function.

Finally we can obtain **posterior probabilities** for H_0 being true and for H_1 being true in light of our sample $\mathbf{x}$.

These posterior probabilities reflect our updated beliefs about our hypotheses, having observed the data, $\mathbf{x}$.

However, in the frequentist setting, we make a decision at the end of our test, using our data (test statistic) to answer the overall question 'do we reject H_0?' ('**Yes**' or '**No**'?).

How can we do this in a Bayesian way?

Bayesian Approach to Hypothesis Testing

Suppose we have the general hypotheses

$$\mathrm{H}_0 : \theta \in \Theta_0 \ \text{ versus } \ \mathrm{H}_1 : \theta \in \Theta_1.$$

We know that in the frequentist setting, we use Type I and Type II errors when deriving a test.

We can do something similar in Bayesian setting by assigning a **loss** to each of these possible errors.

Assume that the losses owing to Type I and Type II errors are a and b, respectively. Then we have the following table of losses

	True state of nature	
Decision	$\theta \in \Theta_0$ (H_0 true)	$\theta \in \Theta_1$ (H_1 true)
Retain H_0	0	b (Type II Error)
Reject H_0	a (Type I Error)	0

Bayesian Approach to Hypothesis Testing

What we have done here, is to specify a set of decisions (i.e. a 'decision space')

$$\mathcal{D} = \{\text{Reject } \mathsf{H}_0, \text{Retain } \mathsf{H}_0\}$$

and an associated loss function $L(\theta, d)$ (for $d \in \mathcal{D}$) such that

$$L(\theta, \text{Retain } \mathsf{H}_0) = \begin{cases} 0 & \text{if } \theta \in \Theta_0; \\ b & \text{if } \theta \notin \Theta_0. \end{cases}$$

$$L(\theta, \text{Reject } \mathsf{H}_0) = \begin{cases} a & \text{if } \theta \in \Theta_0; \\ 0 & \text{if } \theta \notin \Theta_0. \end{cases}$$

Then, conditional on observing the data $\mathbf{x}$, we can evaluate

$$\mathbb{E}\left[L(\theta, \text{Retain } \mathsf{H}_0 \mid \mathbf{x})\right] \quad \text{and} \quad \mathbb{E}\left[L(\theta, \text{Reject } \mathsf{H}_0 \mid \mathbf{x})\right]$$

and reject H_0 if

$$\mathbb{E}\left[L(\theta, \text{Reject } \mathsf{H}_0 \mid \mathbf{x})\right] < \mathbb{E}\left[L(\theta, \text{Retain } \mathsf{H}_0 \mid \mathbf{x})\right].$$

Bayesian Approach to Hypothesis Testing

In words, we reject the null hypothesis if, having observed the data, we expect a smaller loss under the alternative hypothesis than under the null.

Note that if $a = b$ then we'd choose to accept the hypothesis with the highest posterior probability.

We note that losses do not have to be constants, a and b. We could specify the loss as a function of θ if desired.

Bayesian Hypothesis Testing: Example

Electrical components are produced by two factories, A and B. The proportions of defective components produced at factories A and B are known to be 0.2 and 0.3, respectively. A sample of 110 components is sent to an external company for analysis and 28 of these components are found to be defective. We know that all 110 components were made at the same factory but, unfortunately, en route to analysis, the information regarding the factory at which the components were produced is lost and cannot be recovered.

Under the a priori assumption that each factory is equally likely to have made the sample of components, construct a Bayesian hypothesis test of the null hypothesis that the components were made at Factory A.

You should assume that the loss on choosing Factory A as the producer where Factory B actually produced is $\pounds b$ and, similarly, the loss on choosing Factory B as the producer where Factory A actually produced is $\pounds a$.

Bayesian Hypothesis Testing: Example

Bayesian Hypothesis Testing: Example

Bayesian Hypothesis Testing: Example

Bayesian Hypothesis Testing

For a pair of simple hypotheses, i.e.

$$\mathsf{H}_0\colon \theta = \theta_0 \;\; \text{versus} \;\; \mathsf{H}_1\colon \theta = \theta_1$$

if the loss is constant, we see that the Bayesian hypothesis test procedure results in a likelihood ratio test. This is because

$$\begin{aligned}\mathbb{E}(L(\theta, \text{Retain } \mathsf{H}_0) \mid \mathbf{x}) &= L(\theta_1, \text{Retain } \mathsf{H}_0)\mathbb{P}(\theta = \theta_1 \mid \mathbf{x}) \\ &= b[\mathbb{P}(\theta = \theta_1 \mid \mathbf{x})]\end{aligned}$$

$$\begin{aligned}\mathbb{E}(L(\theta, \text{Reject } \mathsf{H}_0) \mid \mathbf{x}) &= L(\theta_0, \text{Reject } \mathsf{H}_0)\mathbb{P}(\theta = \theta_0 \mid \mathbf{x}) \\ &= a[\mathbb{P}(\theta = \theta_0 \mid \mathbf{x})]\end{aligned}$$

Then, if we reject the null hypothesis where

$$\begin{aligned}&\mathbb{E}(L(\theta, \text{Reject } \mathsf{H}_0) \mid \mathbf{x}) < \mathbb{E}(L(\theta, \text{Retain } \mathsf{H}_0) \mid \mathbf{x}) \\ \implies &a[\mathbb{P}(\theta = \theta_0 \mid \mathbf{x})] < b[\mathbb{P}(\theta = \theta_1 \mid \mathbf{x})] \qquad (1)\end{aligned}$$

Bayesian Hypothesis Testing

re-arranging (1), we reject H_0 if

$$\frac{\mathbb{P}(\theta = \theta_1 \mid \mathbf{x})}{\mathbb{P}(\theta = \theta_0 \mid \mathbf{x})} > \frac{a}{b} \tag{2}$$

We know that, for $i \in \{0, 1\}$

$$\mathbb{P}(\theta = \theta_i \mid \mathbf{x}) = \frac{\pi(\theta_i)\mathcal{L}(\theta_i \mid \mathbf{x})}{C(\mathbf{x})} \quad \text{(Bayes' theorem)}$$

with $C(\mathbf{x})$ some normalising constant.

If π_i is the prior probability that H_i is true we have $\pi_1 = 1 - \pi_0$ and (2) is written 'Reject H_0 if'

$$\frac{(1 - \pi_0)\mathcal{L}(\theta_1 \mid \mathbf{x})}{\pi_0 \mathcal{L}(\theta_0 \mid \mathbf{x})} > \frac{a}{b}$$

$$\implies \frac{\mathcal{L}(\theta_1 \mid \mathbf{x})}{\mathcal{L}(\theta_0 \mid \mathbf{x})} > \frac{a\pi_0}{b(1 - \pi_0)}.$$

Bayesian Hypothesis Testing

We obtain a familiar likelihood ratio test of the form 'Reject H_0 if'

$$\frac{\mathcal{L}(\theta_1 \mid \mathbf{x})}{\mathcal{L}(\theta_0 \mid \mathbf{x})} > k$$

where

$$k = \frac{a\pi_0}{b(1-\pi_0)}$$

is specified using a combination of the loss function and the prior probabilities.

For this test, the critical region is given by

$$\mathcal{C} = \left\{ \mathbf{x} \colon \frac{\mathcal{L}(\theta_1 \mid \mathbf{x})}{\mathcal{L}(\theta_0 \mid \mathbf{x})} > \frac{a\pi_0}{b(1-\pi_0)} \right\}$$

Compare with the frequentist setting where we used the probability of a Type I error to determine k.

Bayesian Interval Estimation

In a non-Bayesian setting, suppose that $(\theta_L(\mathbf{x}), \theta_U(\mathbf{x}))$ is a $100(1-\alpha)\%$ confidence interval for the unknown parameter θ.

From Lecture 9 (and other courses) we know that **we do not** interpret this as implying that the probability that θ lies in the interval $(\theta_L(\mathbf{x}), \theta_U(\mathbf{x}))$ is equal to $1-\alpha$.

This is because, in a non-Bayesian setting, θ has a fixed value and does not have a probability distribution. as such. We cannot make probabilistic statements about θ in this way.

Conversely, in a Bayesian setting, θ has a probability distribution by design and, as such, we can make probabilistic statements using the posterior (or prior) distribution of θ.

One such probabilistic statement is a $100(1-\alpha)\%$ **Bayesian credible interval**.

Bayesian Credible Interval

Definition: Bayesian Credible Interval

A $100(1-\alpha)\%$ Bayesian credible interval for a parameter, θ, is a range of values (a, b), with $a < b$, such that the probability that θ lies within the interval (a, b) is equal to $1-\alpha$ (for some $\alpha \in (0, 1)$).

That is

$$\mathbb{P}(a < \theta < b) = 1 - \alpha.$$

Note here that we haven't conditioned on any data, $\mathbf{x}$, in the above definition. This is because the credible interval is a statement made about θ which can be done a priori if desired.

In practice, and for inferential purposes, we are usually interested in calculating a posterior credible interval. In other words, using the posterior distribution of θ, conditional on data $\mathbf{x}$, to determine a and b.

Example: Bayesian Credible Interval

Suppose that $X_1, \ldots, X_n$ are iid Exp(θ) random variables and we observe $T = \sum_{i=1}^{n} X_i = 22.2$ for a sample of size $n = 50$.

Our prior belief about θ is expressed using a Gamma distribution with mean 2 and variance 1.

(a) Determine the posterior distribution of θ.

(b) Determine a 95% posterior credible interval for θ.

Example: Bayesian Credible Interval

Example: Bayesian Credible Interval

Bayesian Credible Interval

There are many possible choices for a $100(1-\alpha)\%$ credible interval (often, in fact, an infinite number of choices). As such, we might consider how to choose a particular credible interval.

Often, we would be most interested in the shortest (most narrow) interval (a, b). Such an interval is known as a **central credible interval**.

Otherwise, we might want to consider one-sided credible intervals e.g. $(0, b)$, $(-\infty, b)$, (a, ∞) etc.

Once we have the full specification of the distribution of θ (prior or posterior) we may be as creative as we choose to suit our needs.

Another (final!) Example

Suppose that $X_1, \ldots, X_n$ are iid $\mathcal{N}(\mu, \sigma^2)$ with σ^2 known. Define $\bar{x}$ to be the observed sample mean from this sample.

(a) Construct a 95% *confidence* interval for μ using the distribution of the maximum likelihood estimator $\bar{X}$.

(b) Now, in a Bayesian context, suppose that the prior distribution for μ is $\mu \sim \mathcal{N}(\psi, \phi^2)$. Determine the posterior distribution of μ

(c) Using the distribution from (c), construct a 95% posterior central credible interval for μ.

(d) Compare your answers to (a) and (b).

Another (final!) Example

Another (final!) Example

Another (final!) Example

Another (final!) Example

Comparison between confidence and posterior credible intervals

$100(1-\alpha)\%$ confidence interval	$100(1-\alpha)\%$ posterior credible interval
Calculated using the **sampling distribution**: $\mathbf{X} \mid \theta$	Calculated using the **posterior distribution**: $\pi(\theta \mid \mathbf{x})$
θ has a fixed, unknown quantity	θ has a probability distribution (i.e. θ is a **random variable**)
$1-\alpha$ interpretation: in repeated sampling, we would expect the proportion of intervals that contain θ to be equal to $1-\alpha$. i.e. if we took M samples and calculated M confidence intervals, approximately $(1-\alpha)M$ intervals should contain θ.	$1-\alpha$ interpretation: having observed the sample $\mathbf{x}$, the posterior probability that θ lies in the interval is equal to $1-\alpha$.

Comparison between confidence intervals and posterior credible intervals

All in all, both interval types provide us with a range of plausible values for θ.

However, the statistical interpretations and methods used to construct the intervals are different!

Learning Outcomes

- Understand how to calculate a confidence interval using the probability integral transform method
- Understand how to derive Bayesian hypothesis tests
- Be able to carry out Bayesian hypothesis tests
- Understand how to derive Bayesian credible intervals
- Compare Bayesian and frequentist approaches to both hypothesis testing and interval estimation