STAT0008 Lecture 9
Further Hypothesis Testing and Interval Estimation

Dr. Aidan O'Keeffe

Department of Statistical Science
University College London

3rd December 2018

Outline

- Generalised likelihood ratio tests involving more than one parameter
- Interval estimation (general)
- Confidence intervals
  - Construction
  - Interpretation
- Pivotal quantities
- Relationship between confidence intervals and hypothesis testing
- Likelihood-based confidence intervals
- Confidence regions (sets)

Likelihood Ratio Tests

Recall from Lecture 8, if we have a sample $X_1, \ldots, X_n$ where the distribution of X_i is parameterised by θ and the likelihood function is $\mathcal{L}(\theta \mid \mathbf{X})$ then...

The **Neyman Pearson lemma** tells us that the most powerful test (of size α) of the simple hypotheses

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

has critical (rejection) region given by

$$\mathcal{C} = \left\{ \mathbf{x}\colon \frac{\mathcal{L}(\theta_1 \mid \mathbf{x})}{\mathcal{L}(\theta_0 \mid \mathbf{x})} > k \right\}$$

with k such that

$$\mathbb{P}(\mathbf{X} \in \mathcal{C} \mid \mathsf{H}_0) = \alpha.$$

Generalised Likelihood Ratio Test

In addition, we saw that if the null hypothesis is simple and the alternative hypothesis general, for example

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

then the critical (rejection) region for a size α generalised likelihood ratio test is given by

$$\begin{aligned}\mathcal{C} &= \left\{\mathbf{x}\colon 2\log\left[\frac{\sup_{\theta\in\Theta}\mathcal{L}(\theta \mid \mathbf{x})}{\mathcal{L}(\theta_0 \mid \mathbf{x})}\right] > k\right\} \\ &= \left\{\mathbf{x}\colon 2\Big[\ell(\hat{\theta} \mid \mathbf{X}) - \ell(\theta_0 \mid \mathbf{X})\Big] > k\right\}\end{aligned}$$

with $\hat{\theta} =$ maximum likelihood estimate of θ under H_1 and k such that

$$\mathbb{P}(\mathbf{X} \in \mathcal{C} \mid \mathsf{H}_0) = \alpha.$$

Each of these tests rejects H_0 if the (maximised) likelihood function under the assumption of H_1 is significantly larger than the likelihood function under the assumption of H_0.

Likelihood Ratio Tests - Relationship to Sufficiency

We note that likelihood ratio tests will depend on the data $X_1, \ldots, X_n$ through the **sufficient statistic(s)** $T(\mathbf{X})$.

This is because, using the factorisation criterion, if $T(\mathbf{X})$ is sufficient for θ, then the likelihood ratio may be written in the form

$$\frac{\mathcal{L}(\theta_1 \mid \mathbf{x})}{\mathcal{L}(\theta_0 \mid \mathbf{x})} = \frac{g(T(\mathbf{x}), \theta_1)h(\mathbf{x})}{g(T(\mathbf{x}), \theta_0)h(\mathbf{x})} = \frac{g(T(\mathbf{x}), \theta_1)}{g(T(\mathbf{x}), \theta_0)}.$$

Note that if θ_1 is replaced by the mle $\hat{\theta}$ in the above, then the result still holds since we know that the mle is a function of the sufficient statistic $T(\mathbf{x})$ (see previous lectures).

But what about an even more general scenario where both the null and alternative hypotheses are composite hypotheses. . . ?

Generalised Likelihood Ratio Test

Suppose we have data $X_1, \ldots, X_n$ from a probability distribution parameterised by $\boldsymbol{\theta}$, where $\boldsymbol{\theta}$ is a $p \times 1$ vector $(p \in \mathbb{N}, p \geq 2)$. We shall consider a test of

$$\mathsf{H}_0 \colon \boldsymbol{\theta} \in \Theta_0 \quad \text{versus} \quad \mathsf{H}_1 \colon \boldsymbol{\theta} \in \Theta.$$

with $\Theta_0 \subset \Theta$.

Here, we have a null hypothesis where some restriction is imposed on the parameters $\boldsymbol{\theta}$, but $\boldsymbol{\theta}$ is unrestricted under the alternative hypothesis.

Suppose that the null hypothesis imposes r independent restrictions on $\boldsymbol{\theta}$ (for $r < p$). Then we might write

$$\boldsymbol{\theta} = (\boldsymbol{\theta}_r, \boldsymbol{\theta}_{p-r})^\top$$

where $\boldsymbol{\theta}_r$ is $r \times 1$ and $\boldsymbol{\theta}_{p-r}$ is $(p-r) \times 1$.

Here, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_r, \theta_{r+1}, \ldots, \theta_p)^\top$.

Generalised Likelihood Ratio Test

Our null and alternative hypotheses may then be written

$$\mathsf{H}_0\colon \theta_1 = \gamma_1, \ldots, \theta_r = \gamma_r;\ \boldsymbol{\theta}_{p-r} \text{ unrestricted}$$

$$\mathsf{H}_1\colon \theta_1 \neq \gamma_1, \ldots, \theta_r \neq \gamma_r;\ \boldsymbol{\theta}_{p-r} \text{ unrestricted}$$

We see that there are $p - r$ free parameters under the null hypothesis ($|\Theta_0| = p - r$) and there are p free parameters under the alternative hypothesis ($|\Theta_1| = p$).

Under regularity conditions (and with $X_1, \ldots, X_n$ iid), for a test of the above hypotheses the generalised likelihood ratio test has test statistic

$$2\log\left[\frac{\sup_{\theta\in\Theta}\mathcal{L}(\boldsymbol{\theta},\mathbf{x})}{\sup_{\theta\in\Theta_0}\mathcal{L}(\boldsymbol{\theta},\mathbf{x})}\right] = 2\left[\ell(\hat{\boldsymbol{\theta}} \mid \mathbf{x}) - \ell(\boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}_{p-r} \mid \mathbf{x})\right]. \tag{1}$$

Here

$\hat{\boldsymbol{\theta}}$ is the $p \times 1$ mle of $\boldsymbol{\theta}$ under H_1.

$\boldsymbol{\theta}_0 = (\gamma_1, \ldots, \gamma_r)^\top$ (pre-specified values of $\theta_1, \ldots, \theta_r$ under H_0)

$\hat{\boldsymbol{\theta}}_{p-r}$ is the $(p - r) \times 1$ mle of $\boldsymbol{\theta}_{p-r}$ under H_0.

Generalised Likelihood Ratio Test

As $n \to \infty$, the above test statistic (1) has a χ^2_r distribution.

Note: as we shall see, since the test rejects where the ratio of the likelihoods is large, we do not always need to use a chi-squared distribution *approximation*.

If the likelihood ratio can be used to form a test where the distribution of a test statistic is known exactly, then we would use this instead (e.g. t-tests and F-tests are two examples of likelihood ratio tests where an asymptotic chi-squared approximation is not used).

This type of test is very common, especially when fitting linear (or generalised linear) models and performing hypothesis tests of the null hypothesis that certain linear predictor coefficients are equal to zero versus an unrestricted alternative.

We shall show that the one-sample t-test is an example of a generalised likelihood ratio test. For this example, we don't test using a χ^2 distribution.

One-sample t-test: Derivation

Suppose that $X_1, \ldots, X_n$ are independent and identically distributed $\mathcal{N}(\mu, \sigma^2)$ random variables with σ^2 unknown and that we wish to test the hypotheses

$$\mathsf{H}_0 \colon \mu = \mu_0 \quad \text{versus} \quad \mathsf{H}_1 \colon \mu \neq \mu_0$$

using a generalised likelihood ratio test of size α. Construct the test and derive the distribution of the test statistic under H_0.

One-sample t-test: Derivation

One-sample t-test: Derivation

One-sample t-test: Derivation

Generalised Likelihood Ratio Test

In summary, to construct a generalised likelihood ratio test we use the following steps:

1. State hypotheses in terms of the parameter(s) of interest. H_0 should contain some restriction on the parameter(s) with H_1 not containing this restriction.

2. Find the maximum likelihood estimates of all unknown parameters under H_0 and then calculate the maximised likelihood under H_0 by substituting these mles into the likelihood function (together with parameter values assumed known under H_0). Call this maximised likelihood function $\hat{\mathcal{L}}_0$.

3. Find the maximum likelihood estimates of all unknown parameters under H_1 and then calculate the maximised likelihood under H_1 by substituting these mles into the likelihood function (together with parameter values assumed known under H_1 - if any). Call this maximised likelihood function $\hat{\mathcal{L}}_1$.

Generalised Likelihood Ratio Test

4. Determine the likelihood ratio

$$\Lambda(\mathsf{H}_0, \mathsf{H}_1) = \frac{\hat{\mathcal{L}}_1}{\hat{\mathcal{L}}_0}.$$

5. Reject H_0 if $\Lambda(\mathsf{H}_0, \mathsf{H}_1)$ is large, either by using an exact test statistic (best case) or (if an exact test is difficult to obtain) the asymptotic test statistic

$$2 \log \Lambda(\mathsf{H}_0, \mathsf{H}_1) \sim \chi^2_r$$

where

$r =$ no. of free parameters under $\mathsf{H}_1 -$ no. of free parameters under H_0.

Interval Estimation

In previous lectures, we considered point estimation of parameters and methods to obtain desirable (i.e. minimum variance) estimators.

However, a single point estimate of a parameter alone is not necessarily useful without some idea of the precision of this estimate.

For example, in class of 150 UCL students, I sampled 30 students at random and the sample mean height of these students is 5ft 8in.

So we'd interpret this as meaning that, 'on average, we'd expect someone in the class to be 5ft 8in tall'. However, not everyone will have the same height, and this point estimate gives us little idea of the possible range of values of heights (i.e. the variability!).

Of course, we might be tempted to quote the standard error estimate of the sample mean but, in this example, we might prefer a range of plausible values for the mean (an **interval estimate**) since this could be more interpretable than a standard error estimate alone.

Interval Estimation

Suppose we have $X_1, \ldots, X_n$ with $X_i \sim \mathcal{D}(\theta)$ and, as usual, we want to make inference about the unknown parameter θ.

Let $\theta_L(\mathbf{X})$ and $\theta_U(\mathbf{X})$ be two statistics (i.e. **random variables** - functions of $X_1, \ldots, X_n$) that satisfy

$$\theta_L(\mathbf{X}) < \theta_U(\mathbf{X})$$

for all $\mathbf{X}$.

Suppose that, on observing the data $\mathbf{X} = \mathbf{x}$, we make the inference

$$\theta_L(\mathbf{x}) < \theta < \theta_U(\mathbf{x}).$$

Then the interval

$$[\theta_L(\mathbf{x}), \theta_U(\mathbf{x})]$$

is called an **interval estimate** for θ and the random interval

$$[\theta_L(\mathbf{X}), \theta_U(\mathbf{X})]$$

is known as an **interval estimator** for θ.

Interval Estimation: Confidence Interval

A very common type of interval estimate (which you've seen in other courses) is a **confidence interval**.

Definition: Confidence Interval

A random interval $[\theta_L(\mathbf{X}), \theta_U(\mathbf{X})]$ is a $100(1-\alpha)\%$ **confidence interval** for a parameter θ if

$$\mathbb{P}\left(\theta \in [\theta_L(\mathbf{X}), \theta_U(\mathbf{X})] \mid \theta\right) = 1 - \alpha$$

for all θ. Here $\alpha \in [0,1]$

We note that $1-\alpha$ is known as the **confidence coefficient**. In addition, $\theta_L(\mathbf{X})$ and $\theta_U(\mathbf{X})$ are known as the **lower** and **upper confidence limits**, respectively.

It is crucial to note here that the *interval*, $[\theta_L(\mathbf{X}), \theta_U(\mathbf{X})]$, is the *random quantity* **NOT** θ.

Interval Estimation: Confidence Interval

The probability statement

$$\mathbb{P}\left(\theta \in [\theta_L(\mathbf{X}), \theta_U(\mathbf{X})] \mid \theta\right) = 1 - \alpha$$

refers to the sampling distribution of $\mathbf{X}$.

It may help us to see this more easily if we write the above probability as (the equivalent)

$$\mathbb{P}\left(\{\theta_L(\mathbf{X}) < \theta\} \cap \{\theta_U(\mathbf{X}) > \theta\} \mid \theta\right) = 1 - \alpha.$$

Frequentist Interval Estimation

Recall from Lecture 2 that the frequentist approach to estimation regards θ as fixed but unknown and focuses somewhat on the concept of an unbiased estimator of θ.

As such, the frequentist approach to interval estimation aims to use the distribution of an unbiased estimator of θ to determine a confidence interval for θ.

Note that this approach does not rely on the formation of a likelihood function or an maximum likelihood estimator of θ.

Frequentist Interval Estimation: Example

Suppose that $X_1, \ldots, X_n$ are iid $\mathcal{N}(\theta, \sigma^2)$ random variables where σ^2 is known.

We know that $\bar{X}$ is unbiased for θ and that the distribution of $\bar{X}$ is

$$\bar{X} \sim \mathcal{N}\left(\theta, \frac{\sigma^2}{n}\right).$$

We see that $\bar{X}$ is consistent for θ so, as $n \to \infty$, $\bar{X}$ is increasingly likely to be close to θ. Furthermore

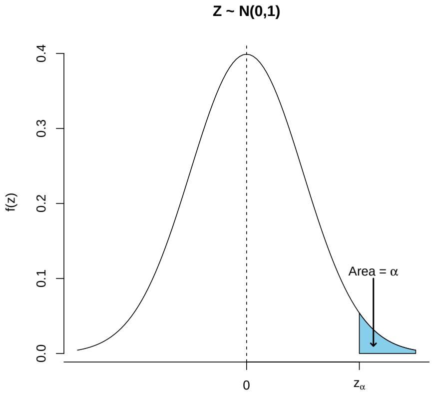
$$\frac{\bar{X} - \theta}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

Frequentist Interval Estimation: Example

Hence

$$\mathbb{P}\left(\frac{\bar{X}-\theta}{\sigma/\sqrt{n}} > z_\alpha\right) = \alpha \qquad (2)$$

with z_α the upper $100(1-\alpha)\%$ quantile of a standard normal random variable.



Frequentist Interval Estimation: Example

From (2), we see that

$$
\begin{aligned}
&\mathbb{P}\left(\left|\frac{\bar{X}-\theta}{\sigma/\sqrt{n}}\right| < z_{\frac{\alpha}{2}}\right) = 1-\alpha \\
&\implies \mathbb{P}\left(-z_{\frac{\alpha}{2}} < \frac{\bar{X}-\theta}{\sigma/\sqrt{n}} < z_{\frac{\alpha}{2}}\right) = 1-\alpha \\
&\implies \mathbb{P}\left(-\frac{z_{\frac{\alpha}{2}}\sigma}{\sqrt{n}} < \bar{X}-\theta < \frac{z_{\frac{\alpha}{2}}\sigma}{\sqrt{n}}\right) = 1-\alpha \\
&\implies \mathbb{P}\left(\frac{z_{\frac{\alpha}{2}}\sigma}{\sqrt{n}} > -\bar{X}+\theta > -\frac{z_{\frac{\alpha}{2}}\sigma}{\sqrt{n}}\right) = 1-\alpha \\
&\implies \mathbb{P}\left(\left\{\bar{X}+\frac{z_{\frac{\alpha}{2}}\sigma}{\sqrt{n}} > \theta\right\} \cap \left\{\bar{X}-\frac{z_{\frac{\alpha}{2}}\sigma}{\sqrt{n}} < \theta\right\}\right) = 1-\alpha
\end{aligned}
$$

Frequentist Interval Estimation: Example

It follows that a $100(1-\alpha)\%$ confidence interval for θ is

$$[\theta_L(\mathbf{X}), \theta_U(\mathbf{X})] = \left[\bar{X} - \frac{z_{\frac{\alpha}{2}}\sigma}{\sqrt{n}}, \bar{X} + \frac{z_{\frac{\alpha}{2}}\sigma}{\sqrt{n}}\right].$$

This is a random interval that covers the true parameter value, θ, with probability $1-\alpha$.

Probability is considered with regard to the sampling distribution of $\bar{X}$.

The value of θ remains fixed and unknown.

But how do we interpret a $100(1-\alpha)\%$ confidence interval in light of sampled data $\mathbf{x}$?

Confidence Interval - Interpretation

Suppose that, as if often the case, we are interested in calculating and interpreting a 95% confidence interval for θ.

On **observing** a data sample of size n, the evaluated 95% confidence interval for θ is

$$\left[\bar{x} - \frac{z_{0.025}\sigma}{\sqrt{n}}, \bar{x} + \frac{z_{0.025}\sigma}{\sqrt{n}}\right].$$

In essence, this interval provides a range of plausible values of θ, based on our data. So what is the importance of the '95% confidence'...?

In answer, the 95% confidence is a property that relates to repeated data sampling.

Confidence Interval - Interpretation

Suppose we took M data samples of size n denoting these

$$\mathbf{x}_j = \{x_{j1}, \ldots, x_{jn}\} \text{ for } j = 1, \ldots, M.$$

Then, if $\bar{x}_j$ denotes the sample mean from sample j, we can calculate M 95% confidence intervals as follows

$$\begin{array}{llllll}
\text{Sample } 1\colon x_{11}, \ldots, x_{1n} \rightarrow \left[\bar{x}_1 - \frac{z_{0.025}\sigma}{\sqrt{n}}, \bar{x}_1 + \frac{z_{0.025}\sigma}{\sqrt{n}}\right] \\
\text{Sample } 2\colon x_{2n}, \ldots, x_{2n} \rightarrow \left[\bar{x}_2 - \frac{z_{0.025}\sigma}{\sqrt{n}}, \bar{x}_2 + \frac{z_{0.025}\sigma}{\sqrt{n}}\right] \\
\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\
\text{Sample } M\colon x_{Mn}, \ldots, x_{Mn} \rightarrow \left[\bar{x}_M - \frac{z_{0.025}\sigma}{\sqrt{n}}, \bar{x}_M + \frac{z_{0.025}\sigma}{\sqrt{n}}\right]
\end{array}$$

If M is large, then we would expect approximately 95% of these intervals to contain the true parameter value θ. Each interval will either contain θ or not.

Confidence Intervals

We note that any confidence interval is not unique.

In fact, in our example above, each of the following is a valid $100(1-\alpha)\%$ confidence interval for θ

$$(a) \quad \left[\bar{X} - \frac{z_{0.025}\sigma}{\sqrt{n}}, \bar{X} + \frac{z_{0.025}\sigma}{\sqrt{n}}\right]$$

$$(b) \quad \left[\bar{X} - \frac{z_{0.05}\sigma}{\sqrt{n}}, \infty\right]$$

$$(c) \quad \left[-\infty, \bar{X} + \frac{z_{0.05}\sigma}{\sqrt{n}}\right]$$

and there are infinitely many more possibilities!

So how do we choose which $100(1-\alpha)\%$ confidence interval to report?

Confidence Intervals

Obviously, if we just want a lower (or upper) bound for θ, then a one-sided confidence interval is fine.

For two-sided intervals, we might consider the following

1. That the tails have equal probability.
2. That the interval should be as narrow as possible.

These two criteria do not always give the same interval (e.g. if the sampling distribution of the statistic is skewed).

Pivotal Quantities

A certain class of random variable, known as a **pivotal quantity**, is useful for the construction of a confidence interval.

Definition: Pivotal Quantity

A random variable $T(\mathbf{X};\theta)$ is a **pivotal quantity** (or pivot) if the distribution of $T(\mathbf{X};\theta)$ is independent of θ.

That is, if $\mathbf{X} \sim \mathcal{D}(\theta)$, then $T(\mathbf{X};\theta)$ has the same distribution for all values of θ.

Pivotal quantities are useful because, if $T(\mathbf{X};\theta)$ is a pivotal quantity then, for any set $\mathcal{A}$, we know that

$$\mathbb{P}(T(\mathbf{X};\theta) \in \mathcal{A})$$

cannot depend on θ.

Pivotal Quantities

Hence, if $T(\mathbf{X};\theta)$ is a pivotal quantity and if we can find a set $\mathcal{A}$ such that

$$\mathbb{P}(T(\mathbf{X};\theta) \in \mathcal{A}) = 1-\alpha$$

then the set

$$\{\theta \colon T(\mathbf{x},\theta) \in \mathcal{A}\}$$

is a $100(1-\alpha)\%$ confidence interval (or set) for θ.

Look back at our normal distribution example and you'll see that we used the pivotal quantity

$$\frac{\bar{X}-\theta}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1) \;\leftarrow\; \text{'Distribution does not depend on } \theta\text{!'}$$

to find a $100(1-\alpha)\%$ confidence interval for θ.

Common Pivotal Quantities

If $X_1, \ldots, X_n$ are iid $\mathcal{N}(\mu, \sigma^2)$ random variables, then some common pivotal quantities that you've seen before are given in the table below.

Pivotal Quantity	**Distribution**
$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$	$\mathcal{N}(0,1)$
$\frac{\bar{X}-\mu}{S/\sqrt{n}}$	t_{n-1}
$\frac{(n-1)S^2}{\sigma^2}$	χ^2_{n-1}

Here

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Pivotal Quantity: Confidence Interval Example

Suppose that $X_1, \ldots, X_n$ are iid $\mathcal{N}(\mu, \sigma^2)$ random variables with μ and σ^2 unknown. Derive a $100(1-\alpha)\%$ confidence interval for σ^2.

Pivotal Quantity: Confidence Interval Example

Relationship to Hypothesis Testing

There is a direct relationship between a $100(1-\alpha)\%$ confidence interval and a size α hypothesis test.

First, we define the **acceptance region** of a hypothesis test as the complement of the test's critical region. That is, the acceptance region $\mathcal{A} = \bar{\mathcal{C}}$, where $\mathcal{C}$ is the critical region. If the test statistic belongs to $\mathcal{A}$, then we retain (accept) the null hypothesis.

Suppose that, for all θ_0, there exists a size α test of $\mathrm{H}_0\colon \theta = \theta_0$ versus some alternative. If $\mathcal{A}(\theta_0)$ denotes the acceptance region of this test then the set

$$\mathcal{B}(\mathbf{X}) = \{\theta_0 \colon \mathbf{X} \in \mathcal{A}(\theta_0)\}$$

is a $100(1-\alpha)\%$ confidence interval (or set) for θ_0.

Conversely, if $\mathcal{B}(\mathbf{X})$ is a $100(1-\alpha)\%$ confidence interval for θ, then an acceptance region for a size α test of $\mathrm{H}_0\colon \theta = \theta_0$ is given by

$$\mathcal{A}(\theta_0) = \{\mathbf{x} \colon \theta_0 \in \mathcal{B}(\mathbf{x})\}.$$

Relationship to Hypothesis Testing

Given a sample of data $\mathbf{x}$, we have the tautology

$$\mathbf{x} \in \mathcal{A}(\theta_0) \iff \theta_0 \in \mathcal{B}(\mathbf{x}).$$

In words, if the sample $\mathbf{x}$ would result in the null hypothesis $\mathrm{H}_0\colon \theta = \theta_0$ being retained (accepted) in a test of size α, then θ_0 lies within the corresponding $100(1-\alpha)\%$ confidence interval constructed using $\mathbf{x}$.

Likewise, if a $100(1-\alpha)\%$ confidence interval constructed using $\mathbf{x}$ contains θ_0, then a size α test of the null hypothesis $\mathrm{H}_0\colon \theta = \theta_0$, that used $\mathbf{x}$ to form the test statistic, would not reject the null hypothesis.

Note that both the hypothesis test and the confidence interval must rely on the same distributional assumptions about $\mathbf{X}$.

Likelihood-based Confidence Intervals

Given the iid sample $X_1, \ldots, X_n$ with $X_i \sim \mathcal{D}(\theta)$, we have seen that, under regularity conditions, the maximum likelihood estimator of θ, written $\hat{\theta}$, has the asymptotic distribution

$$\hat{\theta} \sim \mathcal{N}\left(\theta, \frac{1}{\mathcal{I}(\theta)}\right).$$

It follows that

$$\sqrt{\mathcal{I}(\theta)}\left(\hat{\theta} - \theta\right) \sim \mathcal{N}(0, 1).$$

The above statistic can be used as a pivotal quantity to determine an approximate confidence inteval for θ based on the mle $\hat{\theta}$.

An approximate $100(1-\alpha)\%$ confidence interval for θ would be given by

$$\left[\hat{\theta}(\mathbf{X}) - \frac{z_{\frac{\alpha}{2}}}{\sqrt{\mathcal{I}(\theta)}}, \hat{\theta}(\mathbf{X}) + \frac{z_{\frac{\alpha}{2}}}{\sqrt{\mathcal{I}(\theta)}}\right].$$

Likelihood-based Confidence Intervals

However, since θ is unknown, we may not be able to evaluate $\mathcal{I}(\theta)$. Instead, we could use the approximation

$$\left[\hat{\theta}(\mathbf{X}) - \frac{z_{\frac{\alpha}{2}}}{\sqrt{\mathcal{I}(\hat{\theta})}}, \hat{\theta}(\mathbf{X}) + \frac{z_{\frac{\alpha}{2}}}{\sqrt{\mathcal{I}(\hat{\theta})}}\right]$$

as a $100(1-\alpha)\%$ likelihood-based confidence interval for θ.

Here $\mathcal{I}(\hat{\theta})$ is just the Fisher information but with the unknown θ replaced by the mle $\hat{\theta}$.

Likelihood-based confidence interval: Example

Suppose that $X \sim \text{Bin}(n, \theta)$. Construct a $100(1-\alpha)\%$ confidence interval for θ.

Likelihood-based confidence interval: Example

Confidence Regions/Sets

Confidence intervals can be extended to situations where θ has dimension greater than one.

We know that, if $\boldsymbol{\theta}$ is $k \times 1$, then, under regularity conditions, the following result holds as $n \to \infty$

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top \mathcal{I}(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \sim \chi^2_k.$$

Therefore, if $u_k(1-\alpha)$ is the upper $100(1-\alpha)\%$ point of a χ^2_k density function (i.e. $\mathbb{P}(U < u_k(1-\alpha)) = 1-\alpha$ for $U \sim \chi^2_k$) then

$$\mathbb{P}\left((\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top \mathcal{I}(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) < u_k(1-\alpha)\right) = 1-\alpha$$

and the set

$$\left\{\boldsymbol{\theta} \colon (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top \mathcal{I}(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) < u_k(1-\alpha)\right\}$$

is a $100(1-\alpha)\%$ confidence set for $\boldsymbol{\theta}$.

As before, we would usually approximate $\mathcal{I}(\boldsymbol{\theta})$ as $\mathcal{I}(\hat{\boldsymbol{\theta}})$, whereupon we would obtain a confidence set as a $k-$dimensional ellipsoid.

Learning Outcomes

- Understand how to derive and carry out a generalised likelihood ratio test (restricted null versus unrestricted alternative) with more than one parameter.
- Know what is meant by ‘interval estimation’
- Understand how to construct and interpret $100(1-\alpha)\%$ confidence intervals.
- Understand the relationship between hypothesis testing and confidence intervals.
- Know what is meant by a *pivotal quantity* and be able to construct confidence intervals using pivotal quantities.
- Understand how to construct likelihood-based confidence interval.