MATH2901 Final Revision

Part II: Statistical Inference

Presenter: Henry Lam
Based off Rui Tong's 2018 Slides

UNSW Society of Statistics/UNSW Mathematics Society

13 August 2020

Table of Contents

1. Study Design
   - Samples
2. Statistical Inference
   - Estimators and their Properties
     - Properties of any estimator
     - Desirable features of estimators
   - More work with Convergence
   - Confidence Intervals
   - Distributions arising from $\mathcal{N}(0, 1)$
     - Working with $S = S_X$
   - Common Estimators
   - Hypothesis tests
3. Coolio Doolio

Table of Contents

1. Study Design
   - Samples

2. Statistical Inference
   - Estimators and their Properties
     - Properties of any estimator
     - Desirable features of estimators
   - More work with Convergence
   - Confidence Intervals
   - Distributions arising from $\mathcal{N}(0, 1)$
     - Working with $S = S_X$
   - Common Estimators
   - Hypothesis tests

3. Coolio Doolio

Random Sample

Definition

A random sample (with size n) is a set of n independent, identically distributed random variables:

$$X_1, \ldots, X_n$$

Note:

- $x_1, \ldots, x_n$ is the 'observed data', whereas $X_1, \ldots, X_n$ is the 'theoretical data'.
- Whenever we are working with statistics, we work with the assumption that we have 'theoretical data' first.

Definition (Statistic)

For a random sample $X_1, \ldots, X_n$, a statistic is just a function of the sample.

Example (Common Statistics)

- Sample mean: $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$
- Sample variance: $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$
- Sample median: $X_{\left(\frac{n}{2}\right)}$

The Sample Mean

Theorem (Properties of the Sample Mean)

Let $X_1, \ldots, X_n$ be a random sample, with mean μ and variance σ^2. Then,

$$\mathbb{E}[\overline{X}] = \mu \qquad \text{Var}(\overline{X}) = \frac{\sigma^2}{n}$$

Example

Prove that $\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}$ as stated just now.

$$\begin{aligned}
\mathrm{Var}\left(\overline{X}\right) &= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) && \text{(indep.)} \\
&= \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}
\end{aligned}$$

Table of Contents

1. Study Design
   - Samples
2. Statistical Inference
   - Estimators and their Properties
     - Properties of any estimator
     - Desirable features of estimators
   - More work with Convergence
   - Confidence Intervals
   - Distributions arising from $\mathcal{N}(0, 1)$
     - Working with $S = S_X$
   - Common Estimators
   - Hypothesis tests
3. Coolio Doolio

Estimators

Let $X_1, \dots, X_n$ be a random sample with model $\{f_X(x;\theta) : \theta \in \Theta\}$.

Definition (Estimator)

An estimator for the parameter θ, denoted $\hat{\theta}$, is just a real valued function of the random sample.

$$\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$$

Meaning, fundamentally it's just a statistic (and so it is a random variable).

Estimators

Basically, we want to narrow our focus to **useful** estimators.

To do this, we'll be looking at a figure that measures how well the estimator performs on 'average', the MSE.

Bias

Remember that θ is a parameter, so it's constant. Whereas $\hat{\theta}$ is an estimator, which is a r.v.

Definition: Bias

Given an estimator $\hat{\theta}$ for θ, its bias is

$$\text{Bias}(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta.$$

The estimator is 'unbiased' if $\text{Bias}(\hat{\theta}) = 0$.

Significance

'Bias' is essentially a measure by how *far off on average* an estimator is from the *true value*.

Bias

Example

Let $X_1, \ldots, X_7$ be a random Poisson(λ) sample, and consider the estimator

$$\hat{\lambda} = \frac{1}{28}\sum_{i=1}^{7} i\, X_i = \frac{X_1 + 2X_2 + \cdots + 7X_7}{28}$$

for λ. Is this estimator unbiased?

We compute:

$$\mathbb{E}[\hat{\lambda}] = \mathbb{E}\left[\frac{X_1 + 2X_2 + \cdots + 7X_7}{28}\right]$$

Bias

Example

Let $X_1, \ldots, X_7$ be a random Poisson(λ) sample, and consider the estimator

$$\hat{\lambda} = \frac{1}{28}\sum_{i=1}^{7} i\,X_i = \frac{X_1 + 2X_2 + \cdots + 7X_7}{28}$$

for λ. Is this estimator unbiased?

We compute:

$$\begin{aligned}
\mathbb{E}[\hat{\lambda}] &= \mathbb{E}\left[\frac{X_1 + 2X_2 + \cdots + 7X_7}{28}\right] \\
&= \frac{1}{28}\mathbb{E}[X_1 + 2X_2 + \cdots + 7X_7] \\
&= \frac{1}{28}\left(\mathbb{E}[X_1] + 2\mathbb{E}[X_2] + \cdots + 7\mathbb{E}[X_7]\right)
\end{aligned}$$

Bias

We compute:

$$\begin{aligned}
\mathbb{E}[\hat{\lambda}] &= \mathbb{E}\left[\frac{X_1 + 2X_2 + \cdots + 7X_7}{28}\right] \\
&= \frac{1}{28}\mathbb{E}[X_1 + 2X_2 + \cdots + 7X_7] \\
&= \frac{1}{28}\left(\mathbb{E}[X_1] + 2\mathbb{E}[X_2] + \cdots + 7\mathbb{E}[X_7]\right) \\
&= \frac{1}{28}\left(\lambda + 2\lambda + \cdots + 7\lambda\right) \\
&= \frac{1}{28} \times 28\lambda = \lambda
\end{aligned}$$

Hence $\text{Bias}(\hat{\lambda}) = \lambda - \lambda = 0$ and thus it *is* unbiased.

Standard Error

Definition (Standard Error)

$$\mathsf{Se}(\hat{\theta}) = \sqrt{\mathsf{Var}(\hat{\theta})}$$

Significance

This measures how sensitive the estimator is when estimating the parameter, similar to the concept of *variance of random variables*. Recall that an estimator is just a random variable!

Example

For the earlier example $\hat{\lambda} = \frac{X_1+2X_2+\cdots+7X_7}{28}$, find $\mathrm{Se}(\hat{\lambda})$.

$$\begin{aligned}
\mathrm{Var}(\hat{\lambda}) &= \mathrm{Var}\left(\frac{X_1 + 2X_2 + \cdots + 7X_7}{28}\right) \\
&= \frac{1}{28^2}\left(\mathrm{Var}(X_1) + 4\,\mathrm{Var}(X_2) + \cdots + 49\,\mathrm{Var}(X_7)\right) \qquad (\text{indep.}) \\
&= \frac{1}{28^2} \times 140\lambda = \frac{5}{28}\lambda.
\end{aligned}$$

Therefore $\mathrm{Se}(\hat{\lambda}) = \sqrt{\frac{5\lambda}{28}}$.

Estimated Standard Error

As seen in the previous slide, this *value actually depends on the parameter itself*, which is unknown. Naively we'll just plug in our **estimator** in the parameter's place (a later theorem justifies this - at least asymptotically). **This is actually something we'll be doing a lot in this course!**

Definition: Estimated Standard Error

$$\widehat{\mathrm{Se}}(\hat{\theta}) = \mathrm{Se}(\hat{\theta}) \mid_{\theta=\hat{\theta}}$$

Estimated Standard Error

As seen in the previous slide, this *value actually depends on the parameter itself*, which is unknown. Naively we'll just plug in our **estimator** in the parameter's place (a later theorem justifies this - at least asymptotically). **This is actually something we'll be doing a lot in this course!**

Example

For the earlier example $\hat{\lambda} = \frac{X_1+2X_2+\cdots+7X_7}{28}$, find $\mathrm{Se}(\hat{\lambda})$ and $\widehat{\mathrm{Se}}(\hat{\lambda})$.

From earlier we had,

$$\mathrm{Se}(\hat{\lambda}) = \sqrt{\frac{5\lambda}{28}}$$

and so we have:

$$\widehat{\mathrm{Se}}(\hat{\lambda}) = \sqrt{\frac{5\hat{\lambda}}{28}}.$$

Mean Squared Error

The previous two concepts: **bias** and **standard error** come from the following measure:

Definition: Mean Squared Error

Given an estimator $\hat{\theta}$ for θ, its mean squared error is

$$\mathsf{MSE}(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2].$$

Theorem: MSE Formula

$$\mathsf{MSE}(\hat{\theta}) = [\mathsf{Bias}(\hat{\theta})]^2 + \mathsf{Var}(\hat{\theta}).$$

Proof: MSE formula

$$\begin{aligned}\text{MSE}(\hat{\theta}) &= \mathbb{E}[(\hat{\theta}-\theta)^2] \\ &= \mathbb{E}\left[\left((\hat{\theta}-\mathbb{E}[\hat{\theta}]) + (\mathbb{E}[\hat{\theta}]-\theta)\right)^2\right] \\ &= \mathbb{E}\left[(\hat{\theta}-\mathbb{E}[\hat{\theta}])^2\right] + \mathbb{E}\left[(\mathbb{E}[\hat{\theta}]-\theta)^2\right] + 2\mathbb{E}\left[(\hat{\theta}-\mathbb{E}[\hat{\theta}])(\mathbb{E}[\hat{\theta}]-\theta)\right]\end{aligned}$$

from expanding the perfect square. Note that $\mathbb{E}[(\hat{\theta}-\mathbb{E}[\hat{\theta}])] = \text{Var}(\hat{\theta})$ by definition, and

$$\mathbb{E}\left[(\mathbb{E}[\hat{\theta}]-\theta)\right] = \mathbb{E}[\text{Bias}(\hat{\theta})^2] = \text{Bias}(\hat{\theta})^2.$$

(Q: Why was I allowed to take off the expected value brackets?)

Proof: MSE formula

As for the leftover bit:

$$2\mathbb{E}\left[(\hat{\theta} - \mathbb{E}[\hat{\theta}])(\mathbb{E}[\hat{\theta}] - \theta)\right] = 2\left(\mathbb{E}[\hat{\theta}] - \theta\right)\mathbb{E}\left[\hat{\theta} - \mathbb{E}[\hat{\theta}]\right]$$

...but

$$\mathbb{E}\left[\hat{\theta} - \mathbb{E}[\hat{\theta}]\right] = \mathbb{E}[\hat{\theta}] - \mathbb{E}[\hat{\theta}] = 0.$$

Make sure to remember all your properties of the expected value!

Mean Squared Error

Example

For the earlier example $\hat{\lambda} = \frac{X_1+2X_2+\cdots+7X_7}{28}$, find $\text{MSE}(\hat{\lambda})$.

$$\text{MSE}(\hat{\lambda}) = \text{Var}(\hat{\lambda}) + \text{Bias}(\hat{\lambda})^2 = \frac{5\lambda}{28} + 0^2 = \frac{5\lambda}{28}.$$

"Better" Estimators

Significance of MSE

Demonstrates a *trade-off* between the variance and the bias.

Better estimators in the MSE sense

Between two estimators $\hat{\theta}_1$ and $\hat{\theta}_2$, $\hat{\theta}_1$ is better (w.r.t. MSE), at some specific value of θ, if

$$\mathsf{MSE}(\hat{\theta}_1) < \mathsf{MSE}(\hat{\theta}_2)$$

"Better" Estimators

Example

Let $\hat{\lambda}_1$ be the estimator that we found earlier, with $\text{MSE}(\hat{\lambda}_1) = \frac{5\lambda}{28}$. Now let $\hat{\lambda}_2 = \overline{X}$. For what values of λ is $\hat{\lambda}_2$ better than $\hat{\lambda}_1$?

We can compute:

$$\begin{aligned} \text{Bias}(\hat{\lambda}_2) &= 0 \\ \text{Var}(\hat{\lambda}_2) &= \frac{\lambda}{7} \\ \therefore \text{MSE}(\hat{\lambda}_2) &= \frac{\lambda}{7} \end{aligned}$$

Example

Let $\hat{\lambda}_1$ be the estimator that we found earlier, with $\text{MSE}(\hat{\lambda}_1) = \frac{5\lambda}{28}$. Now let $\hat{\lambda}_2 = \overline{X}$. For what values of λ is $\hat{\lambda}_2$ better than $\hat{\lambda}_1$?

$$\text{MSE}(\hat{\lambda}_2) = \frac{\lambda}{7}$$

Solving $\text{MSE}(\hat{\lambda}_2) < \text{MSE}(\hat{\lambda}_1)$ gives

$$\frac{\lambda}{7} < \frac{5\lambda}{28} \implies \lambda > 0.$$

Application - Sample Proportion

Theorem (Properties of the Sample Mean)

Let $X_1, \ldots, X_n$ be a random sample from the $\text{Ber}(p)$ distribution. Then the sample proportion $\hat{p} = \frac{\text{No. of successes}}{\text{No. of trials}}$ satisfies:

$$\mathbb{E}[\hat{p}] = p$$

$$\text{Var}(\hat{p}) = \frac{p(1-p)}{n}$$

$$\text{Se}(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Consistency

A sequence of random variables $X_1, \dots, X_n$ **converges in probability** to X, i.e. $X_n \xrightarrow{\mathbb{P}} X$, if $\forall \varepsilon > 0$,

$$\lim_{n \to \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0.$$

Definition: Consistent Estimator

$\hat{\theta}_n$ is a consistent estimator for θ if it converges in probability to θ. i.e.

$$\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta$$

Consistency

A sequence of random variables $X_1, \ldots, X_n$ **converges in probability** to X, i.e. $X_n \xrightarrow{\mathbb{P}} X$, if $\forall \varepsilon > 0$,

$$\lim_{n \to \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0.$$

Definition: Consistent Estimator

$\hat{\theta}_n$ is a consistent estimator for θ if it converges in probability to θ. i.e.

$$\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta$$

Slight issue

Generally, this can be quite difficult to show. Luckily for us, we actually have another approach to show convergence in probability!

Verifying that an estimator is consistent

Theorem: Sufficient criteria for consistency

If

$$\lim_{n\to\infty} \mathrm{MSE}(\hat{\theta}_n) = 0$$

then $\hat{\theta}_n$ is a consistent estimator for θ.

Quick example: Consider the mean proportion $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ for μ. Then

$$\mathrm{MSE}(\hat{\theta}_n) = \mathrm{Var}(\hat{\theta}_n) + \mathrm{Bias}(\hat{\theta}_n)^2 = \frac{\sigma^2}{n} + 0^2 = \frac{\sigma^2}{n}.$$

Clearly $\lim_{n\to\infty} \mathrm{MSE}(\hat{\theta}_n) = 0$ so the sample mean is a consistent estimator for μ.

Asymptotic Normality

A sequence of random variables $X_1, \dots, X_n$ converges in distribution to X, i.e. $X_n \xrightarrow{\mathcal{D}} X$, if

$$\lim_{n\to\infty} F_{X_n}(x) \to F_X(x).$$

Definition: Asymptotically Normal Estimator

$\hat{\theta}_n$ is an asymptotically normal estimator for θ if

$$\frac{\hat{\theta}_n - \theta}{\mathrm{Se}(\hat{\theta})} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

Convergence Theorems

Central Limit Theorem

For a random sample $X_1, \ldots, X_n$ with mean μ and finite variance σ,

$$\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$$

Slutsky's Theorem

Suppose we have two sequences of random variables (or random samples) with:

$$X_n \xrightarrow{\mathcal{D}} X \qquad\qquad Y_n \xrightarrow{\mathbb{P}} c$$

where c is a constant. Then,

$$X_n + Y_n \xrightarrow{\mathcal{D}} X + c \qquad\qquad X_n Y_n \xrightarrow{\mathcal{D}} cX$$

The Delta Method

Theorem (Provided on formula sheet!!)

Let $\hat{\theta}_1, \hat{\theta}_2, \dots$ be a sequence of estimators (or a sequence of random variables) of θ such that

$$\frac{\hat{\theta}_n - \theta}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Then, for any function g that is differentiable at θ, with $g'(\theta) \neq 0$,

$$\frac{g(\hat{\theta}_n) - g(\theta)}{g'(\theta)\frac{\sigma}{\sqrt{n}}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

Delta Method Example

2015 FE Q2c)

Suppose $\{X_i\}_{i=1}^n$ are i.i.d. Ber(p) random variables and define $Y_n = \overline{X}_n$. What is the asymptotic distribution of $Y_n(1 - Y_n)$?

Delta Method Example

2015 FE Q2c)

Suppose $\{X_i\}_{i=1}^n$ are i.i.d. $\text{Ber}(p)$ random variables and define $Y_n = \overline{X}_n$. What is the asymptotic distribution of $Y_n(1 - Y_n)$?

Finding the distribution of $Y_n(1 - Y_n)$ may be difficult to do directly, but we note that $Y_n = \bar{X}_n$. Thus by the CLT we have:

$$\frac{Y_n - p}{\sqrt{\frac{p(1-p)}{n}}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Delta Method Example

2015 FE Q2c)

Suppose $\{X_i\}_{i=1}^n$ are i.i.d. Ber(p) random variables and define $Y_n = \overline{X}_n$. What is the asymptotic distribution of $Y_n(1 - Y_n)$?

Now using the delta method with transform of $g(x) = x(1 - x)$ on Y_n with $\theta = p$. $g'(x) = 1 - 2x$.

$$\frac{Y_n(1 - Y_n) - p(1 - p)}{(1 - 2p)\sqrt{\frac{p(1-p)}{n}}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

Delta Method Example

2015 FE Q2c)

Suppose $\{X_i\}_{i=1}^n$ are i.i.d. Ber(p) random variables and define $Y_n = \overline{X}_n$. What is the asymptotic distribution of $Y_n(1 - Y_n)$?

Now using the delta method with transform of $g(x) = x(1-x)$ on Y_n with $\theta = p$. $g'(x) = 1 - 2x$.

$$\frac{Y_n(1 - Y_n) - p(1-p)}{(1-2p)\sqrt{\frac{p(1-p)}{n}}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

Rearranging everything yields:

$$Y_n(1 - Y_n) - p(1-p) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, (1-2p)^2 \frac{p(1-p)}{n}\right)$$

Delta Method Example

2015 FE Q2c)

Suppose $\{X_i\}_{i=1}^n$ are i.i.d. Ber(p) random variables and define $Y_n = \overline{X}_n$. What is the asymptotic distribution of $Y_n(1-Y_n)$?

Now using the delta method with transform of $g(x) = x(1-x)$ on Y_n with $\theta = p$. $g'(x) = 1-2x$.

$$\frac{Y_n(1-Y_n) - p(1-p)}{(1-2p)\sqrt{\frac{p(1-p)}{n}}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Rearranging everything yields:

$$Y_n(1-Y_n) - p(1-p) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, (1-2p)^2 \frac{p(1-p)}{n}\right)$$

$$Y_n(1-Y_n) \xrightarrow{\mathcal{D}} \mathcal{N}\left(p(1-p), (1-2p)^2 \frac{p(1-p)}{n}\right).$$

Delta Method Example

2015 FE Q2c)

Suppose $\{X_i\}_{i=1}^n$ are i.i.d. Ber(p) random variables and define $Y_n = \overline{X}_n$. What is the asymptotic distribution of $Y_n(1 - Y_n)$?

We have found the asymptotic distribution! But we must also check the condition that: $g'(p) \neq 0 \Leftrightarrow p \neq 0.5$. Hence, we have:

$$Y_n(1 - Y_n) \xrightarrow{\mathcal{D}} \mathcal{N}\left(p(1-p), (1-2p)^2 \frac{p(1-p)}{n}\right)$$

for $p \in (0, 1), p \neq 0.5$.

Confidence Intervals (Generic Definition)

In a confidence interval, we put the parameter in the middle, instead of the random variable. We want to predict where the **true parameter lies in**, not the estimator!

Definition: Confidence Interval

For a random sample $X_1, \dots, X_n$ with parameter θ, if

$$\mathbb{P}(L < \theta < U) = 1 - \alpha$$

for some statistics (estimators) L and U, then a $100(1-\alpha)\%$ confidence interval for θ is

$$(L, U)$$

The method to find these CI's mostly depends on the **distribution of your random variables**.

Example (2012 FE Q4)

Consider a sequence of i.i.d. $\{X_i\}_{i=1}^n \sim \text{Unif}(0, \theta)$, from which we utilise the following estimator $X_{(n)} = \max\{X_1, X_2, \ldots, X_n\}$.

a) Find the density of $X_{(n)}$.

b) Find an exact 95% confidence interval for θ.

Example (2012 FE Q4)

Consider a sequence of i.i.d. $\{X_i\}_{i=1}^n \sim \mathrm{Unif}(0,\theta)$, from which we utilise the following estimator $X_{(n)} = \max\{X_1, X_2, \ldots, X_n\}$.

a) Find the density of $X_{(n)}$.

$$\begin{aligned}
\mathbb{P}(X_{(n)} \le x) &= \mathbb{P}(\cap_{i=1}^n X_i \le x) \\
&= \prod_{i=1}^n \mathbb{P}(X_i \le x) && \text{(indep.)} \\
&= (\mathbb{P}(X_1 \le x))^n && \text{(identically dist.)} \\
&= \left(\frac{x}{\theta}\right)^n
\end{aligned}$$

for $x \in (0,\theta)$.

Example (2012 FE Q4)

Consider a sequence of i.i.d. $\{X_i\}_{i=1}^n \sim \text{Unif}(0, \theta)$, from which we utilise the following estimator $X_{(n)} = \max\{X_1, X_2, \dots, X_n\}$.

a) Find the density of $X_{(n)}$.

Hence the PDF of $X_{(n)}$ is:

$$\begin{aligned} f_{X_{(n)}}(x) &= \frac{\mathrm{d}}{\mathrm{d}x}\mathbb{P}(X_{(n)} \leq x) \\ &= \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{x}{\theta}\right)^n \\ &= \frac{n}{\theta^n}x^{n-1} \qquad (x \in (0, \theta)) \end{aligned}$$

Example (2012 FE Q4)

Consider a sequence of i.i.d. $\{X_i\}_{i=1}^n \sim \text{Unif}(0, \theta)$, from which we utilise the following estimator $X_{(n)} = \max\{X_1, X_2, \ldots, X_n\}$.

b) Find an exact 95% confidence interval for θ.

We desire a 95% exact confidence interval for θ. In this case, the interpretation of θ is simply the maximum.

Example (2012 FE Q4)

Consider a sequence of i.i.d. $\{X_i\}_{i=1}^n \sim \text{Unif}(0,\theta)$, from which we utilise the following estimator $X_{(n)} = \max\{X_1, X_2, \ldots, X_n\}$.

b) Find an exact 95% confidence interval for θ.

We desire a 95% exact confidence interval for θ. In this case, the interpretation of θ is simply the maximum.

$$\begin{aligned}\mathbb{P}(X_{(n)} > a_{0.05}) &= 0.95\\ 1-\left(\frac{a_{0.05}}{\theta}\right)^n &= 0.95\\ a_{0.05} &= 0.05^{1/n}\theta.\end{aligned}$$

Exact Confidence Intervals

Example (2012 FE Q4)

Consider a sequence of i.i.d. $\{X_i\}_{i=1}^n \sim \text{Unif}(0, \theta)$, from which we utilise the following estimator $X_{(n)} = \max\{X_1, X_2, \ldots, X_n\}$.

b) Find an exact 95% confidence interval for θ.

Subbing that back in and re-arranging yields:

$$\begin{aligned}
\mathbb{P}(X_{(n)} > 0.05^{1/n}\theta) &= 0.95 \\
\mathbb{P}(\theta < 20^{1/n}X_{(n)}) &= 0.95 \\
\mathbb{P}(0 < \theta < 20^{1/n}X_{(n)}) &= 0.95
\end{aligned}$$

Hence an exact 95% confidence interval for θ would be:

$$\left(0, 20^{\frac{1}{n}}X_{(n)}\right).$$

Exact Confidence Intervals from Normality

Just like in pretty much every scenario, when dealing with confidence intervals of normally distributed random variables, we have some very nice properties. These allow us to build fairly easy confidence intervals.

Chi-Squared distribution

Definition (Chi-Squared distribution)

A random variable X follows a χ^2_ν distribution if for $x \geq 0$,

$$f_X(x) = \frac{e^{-\frac{x}{2}} x^{\frac{\nu}{2}-1}}{2^{\frac{\nu}{2}} \Gamma\left(\frac{\nu}{2}\right)}$$

Lemma (Chi-Squared as a 'special' distribution)

$$X \sim \chi^2_\nu \quad \iff \quad X \sim \text{Gamma}\left(\frac{\nu}{2}, 2\right)$$

Significance of ν: It is the number of degrees of freedom you have. (MATH2831/2931)

Chi-Squared distribution

Theorem (Origin of Chi-Squared)

If $Z \sim \mathcal{N}(0,1)$, then $Z^2 \sim \chi^2_1$.

Lemma (Sum of Chi-Squared is Chi-Squared)

Let $X_1 \sim \chi^2_{\nu_1}$, ..., $X_n \sim \chi^2_{\nu_n}$ be independently distributed. Then their sum satisfies:

$$\sum_{i=1}^{n} X_i = X_1 + \cdots + X_n \sim \chi^2_{\nu_1 + \cdots + \nu_n}$$

Example (Course pack)

If we have independent standard normal random variables Z_i, find the probability that $\sum_{i=1}^{6} Z_i^2 > 16.81$.

$\boxed{Z_i^2 \sim \chi_1^2}$ for all i, so

$$\boxed{\sum_{i=1}^{n} Z_i^2 \sim \chi_6^2}.$$

$$\mathbb{P}\left(\sum_{i=1}^{6} Z_i^2 > 16.81\right) = \texttt{pchisq(16.81,df=6,lower.tail=FALSE)} \approx 0.01$$

Student's t distribution

Definition (t-distribution)

A random variable T follows a t_ν distribution if for $t \in \mathbb{R}$,

$$f_T(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

Significance of ν: It is the number of degrees of freedom you have. (MATH2831/2931)

Student's t distribution

Theorem (Origin of t)

If $Z \sim \mathcal{N}(0,1)$ and $Q \sim \chi^2_\nu$, where Z and Q are independent, then $\frac{Z}{\sqrt{Q/\nu}} \sim t_\nu$

Theorem (Convergence of t)

As $\nu \to \infty$, $t_\nu \to \mathcal{N}(0,1)$

Example (Density of t_ν is an even function)

Just like the density of normal distributions, $f_T(-t) = f_T(t)$.

Sample Variance

Definition (Sample Variance)

For a random sample $X_1, \ldots, X_n$, the sample variance is

$$S_X^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

where $\overline{X}$ is the sample mean.

Key property

S^2 is an **unbiased** estimator for σ^2.

Sample Variance and Distributions

Theorem (Distribution of Sample Variance)

Suppose that $X_1, \ldots, X_n$ are i.i.d. random samples from the $\mathcal{N}(\mu, \sigma^2)$ distribution. Then,

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Dealing with two unknown variables simultaneously can be quite challenging, and so the following is a way to only deal with a single unknown.

Theorem (S^2 to replace σ^2)

Suppose that $X_1, \ldots, X_n$ are i.i.d. random samples from the $\mathcal{N}(\mu, \sigma^2)$ distribution. Then,

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}.$$

These are exact! Not approximations!

Trap

Only ever use these if you know your original sample came from a normal distribution (or something that resembles it really well)!

Notation

Recall that z_α represents the α-th quantile of $Z \sim \mathcal{N}(0,1)$.

Notation (t-value)

$t_{n-1,\alpha}$ represents the α-th quantile of $T \sim t_{n-1}$, i.e. it satisfies

$$\mathbb{P}(T < t_{n-1,\alpha}) = \alpha$$

Normal Samples: Exact CI

Corollary (Exact CI for normal samples)

Suppose $X_1, \ldots, X_n$ are from a $\mathcal{N}(\mu, \sigma^2)$ sample. If we know what σ^2 is, a $100(1-\alpha)$ % confidence interval for μ is

$$\left(\overline{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$$

If we don't know what σ^2 is, then using the estimator S^2,

$$\left(\overline{X} - t_{n-1,1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \overline{X} + t_{n-1,1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right)$$

These can derived on the spot as well.

Example

The following data is taken from a normal random sample:

```
1.1633974 0.2623631 -2.0633406
```

By considering $\overline{X}$, find a 95% confidence interval for its mean μ.

The sample mean is $\overline{X} = -0.2125267$, and the sample variance is

$$\begin{aligned} S^2 &= \frac{1}{3-1}\Big((1.1633974 + 0.2125267)^2 \\ &\quad + (0.2523621 + 0.2125267)^2 \\ &\quad + (-2.0633406 + 0.2125267)^2\Big) \\ &= 2.7721 \end{aligned}$$

Confidence Intervals

Example

The following data is taken from a normal random sample:

```
1.1633974 0.2623631 -2.0633406
```

By considering $\overline{X}$, find a 95% confidence interval for its mean μ.

For the mean, $t_{2,0.975} = \texttt{qt(0.975,df=2)} = 4.302653$.

Therefore a 95% confidence interval is

$$\left(-0.2125267 - 4.302653\frac{\sqrt{2.7721}}{\sqrt{3}}, -0.2125267 + 4.302653\frac{\sqrt{2.7721}}{\sqrt{3}}\right)$$

i.e. (-4.34, 3.92)

Approximate CI's via Asymptotic Normality

Notation (z-value)

z_α represents the α-th quantile of $Z \sim \mathcal{N}(0,1)$, i.e it satisfies

$$\mathbb{P}(Z < z_\alpha) = \alpha$$

Corollary (Approximate CI)

For a random sample $X_1, \dots, X_n$ with parameter θ, if $\hat{\theta}_n$ is a consistent and asymptotically normal estimator of θ, then

$$\left(\hat{\theta}_n - z_{1-\frac{\alpha}{2}} \operatorname{Se}(\hat{\theta}_n), \hat{\theta}_n + z_{1-\frac{\alpha}{2}} \operatorname{Se}(\hat{\theta}_n)\right)$$

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

Approximate CI's via Asymptotic Normality

"Example" (Setting $\alpha = 0.05$)

For a random sample $X_1, \ldots, X_n$ with parameter θ, if $\hat{\theta}_n$ is a consistent and asymptotically normal estimator of θ, then

$$\left(\hat{\theta}_n - z_{0.975}\,\hat{\mathsf{Se}}(\hat{\theta}_n), \hat{\theta}_n + z_{0.975}\,\hat{\mathsf{Se}}(\hat{\theta}_n)\right)$$

is a 95% confidence interval for θ.

Example (Adapted from Tutorial)

Consider a random sample $X_1, \ldots, X_n$ from the Poisson(λ) distribution. Take $\hat{\lambda} = \overline{X}$, i.e. use the sample mean as an estimator. Find a 95% approximate confidence interval for λ.

As the estimator is the sample mean, we can utilise the CLT for very large n. Recalling that $\text{Var}(X_i) = \lambda$ and since our estimator is the sample mean,

$$\text{Var}(\hat{\lambda}) = \text{Var}(\overline{X}) = \frac{\lambda}{n}$$

so therefore

$$\text{Se}(\hat{\lambda}) = \sqrt{\frac{\hat{\lambda}}{n}}$$

This is actually $\widehat{\text{Se}}(\hat{\lambda})$! (which is fine as we are dealing with it asymptotically.)

Approximate CI's via Asymptotic Normality

Example (Adapted from Tutorial)

Consider a random sample $X_1, \ldots, X_n$ from the Poisson(λ) distribution. Take $\hat{\lambda} = \overline{X}$, i.e. use the sample mean as an estimator. Find a 95% approximate confidence interval for λ.

By CLT, we have:

$$\frac{\hat{\lambda} - \lambda}{\sqrt{\frac{\hat{\lambda}}{n}}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

Therefore

$$\mathbb{P}\left(z_{0.025} < \frac{\hat{\lambda} - \lambda}{\sqrt{\frac{\hat{\lambda}}{n}}} < z_{0.975} \right) = 0.95$$

Approximate CI's via Asymptotic Normality

Note that $\boxed{z_{0.025} = -z_{0.975}}$. Rearrange to make λ the subject:

$$-z_{0.975} < \frac{\hat{\lambda} - \lambda}{\sqrt{\frac{\hat{\lambda}}{n}}} < z_{0.975}$$

$$-z_{0.975}\sqrt{\frac{\hat{\lambda}}{n}} < \hat{\lambda} - \lambda < z_{0.975}\sqrt{\frac{\hat{\lambda}}{n}}$$

$$-z_{0.975}\sqrt{\frac{\hat{\lambda}}{n}} < \lambda - \hat{\lambda} < z_{0.975}\sqrt{\frac{\hat{\lambda}}{n}}$$

$$\hat{\lambda} - z_{0.975}\sqrt{\frac{\hat{\lambda}}{n}} < \lambda < \hat{\lambda} + z_{0.975}\sqrt{\frac{\hat{\lambda}}{n}}$$

Be very careful going from line 2 to line 3!

Approximate CI's via Asymptotic Normality

Example (Adapted from Tutorial)

Consider a random sample $X_1, \ldots, X_n$ from the Poisson(λ) distribution. Take $\hat{\lambda} = \overline{X}$, i.e. use the sample mean as an estimator. Find a 95% approximate confidence interval for λ.

Therefore we can rewrite:

$$\mathbb{P}\left(\hat{\lambda} - z_{0.975}\sqrt{\frac{\hat{\lambda}}{n}} < \lambda < \hat{\lambda} + z_{0.975}\sqrt{\frac{\hat{\lambda}}{n}}\right) = 0.95$$

so a 95% confidence interval is

$$\left(\hat{\lambda} - z_{0.975}\sqrt{\frac{\hat{\lambda}}{n}}, \hat{\lambda} + z_{0.975}\sqrt{\frac{\hat{\lambda}}{n}}\right)$$

(Then just sub everything in.)

Follow-up question

Example (contd. from Tutorial)

For the confidence interval above, suppose that for a sample size of 30 the observed values are:

8 2 5 5 8 6 7 2 4 8 4 2 8 4 5 3 3 6 8 3 6 5 5 4 6 3 7 5 1 5

Under these observed values, what is the relevant confidence interval?

From the calculator, the mean of this data is $\frac{148}{30}$, so subbing $\overline{X} = \frac{148}{30}$ and $n = 30$ gives

$$\left(148/30 - 1.96 \times \sqrt{\frac{148/30}{30}}, 148/30 + 1.96 \times \sqrt{\frac{148/30}{30}}\right)$$

which is approximately (4.1385, 5.7281)

Behaviour of the approximate CI

The confidence interval becomes smaller when we increase n, i.e. add more samples!

A 99% confidence interval is wider than a 95% confidence interval. Why?

Method of Moments

Compares theoretical value with estimated values to get parameter estimates.

Suppose we need to estimate k parameters: $\theta_1, \ldots, \theta_k$.

Definition (Method of Moments Estimator)

Consider the system of equations

$$\mathbb{E}[X] = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \mathbb{E}[X^2] = \frac{1}{n}\sum_{i=1}^{n} X_i^2, \quad \ldots \quad \mathbb{E}[X^k] = \frac{1}{n}\sum_{i=1}^{n} X_i^k.$$

The method of moments **estimator** is the solution to this system of equations.

The method of moments **estimate** is the observed value of the estimator. This is found by replacing X_i with x_i.

Method of Moments - Other way around

Compares theoretical value with estimated values to get parameter estimates.

Suppose we need to estimate k parameters: $\theta_1, \dots, \theta_k$.

Definition (Method of Moments Estimate)

Consider the system of equations

$$\mathbb{E}[X] = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \mathbb{E}[X^2] = \frac{1}{n}\sum_{i=1}^{n} x_i^2, \quad \dots \quad \mathbb{E}[X^k] = \frac{1}{n}\sum_{i=1}^{n} x_i^k.$$

The method of moments **estimate** is the solution to this system of equations.

The method of moments **estimator** is the original estimator in question. This is found by replacing x_i with X_i.

Method of Moments

Example (2901 Assignment, 2017)

Let θ be a parameter satisfying $\theta > -1$. Let $X_1, \ldots, X_n$ be i.i.d. random variables with PDF

$$f_{X_i}(\theta) = (\theta + 1)x^{\theta}, \quad 0 < x < 1$$

for $i = 1, \ldots, n$. Find the method of moments estimator for θ.

How many parameters to estimate?

Example (2901 Assignment, 2017)

Let θ be a parameter satisfying $\theta > -1$. Let $X_1, \ldots, X_n$ be i.i.d. random variables with PDF

$$f_{X_i}(\theta) = (\theta + 1)x^{\theta}, \quad 0 < x < 1$$

for $i = 1, \ldots, n$. Find the method of moments estimator for θ.

Only 1 parameter, therefore we only need one equation:

$$\mathbb{E}[X] = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Example (2901 Assignment, 2017)

Let θ be a parameter satisfying $\theta > -1$. Let $X_1, \ldots, X_n$ be i.i.d. random variables with PDF

$$f_{X_i}(\theta) = (\theta + 1)x^{\theta}, \quad 0 < x < 1$$

for $i = 1, \ldots, n$. Find the method of moments estimator for θ.

$$\begin{aligned}
\mathbb{E}[X] &= \int_0^1 x(\theta+1)x^{\theta}\,dx \\
&= \int_0^1 (\theta+1)x^{\theta+1}\,dx \\
&= \frac{\theta+1}{\theta+2}
\end{aligned}$$

Method of Moments

Example (2901 Assignment, 2017)

Let θ be a parameter satisfying $\theta > -1$. Let $X_1, \ldots, X_n$ be i.i.d. random variables with PDF

$$f_{X_i}(\theta) = (\theta + 1)x^{\theta}, \quad 0 < x < 1$$

for $i = 1, \ldots, n$. Find the method of moments estimator for θ.

So we solve:

$$\begin{aligned}
\frac{\hat{\theta}+1}{\hat{\theta}+2} &= \overline{x} \\
\hat{\theta}+1 &= \overline{x}\hat{\theta} + 2\overline{x} \\
\hat{\theta} - \overline{x}\hat{\theta} &= 2\overline{x} - 1 \\
\hat{\theta} &= \frac{2\overline{x}-1}{1-\overline{x}}
\end{aligned}$$

Method of Moments

Example (2901 Assignment, 2017)

Let θ be a parameter satisfying $\theta > -1$. Let $X_1, \dots, X_n$ be i.i.d. random variables with PDF

$$f_{X_i}(\theta) = (\theta + 1)x^{\theta}, \quad 0 < x < 1$$

for $i = 1, \dots, n$. Find the method of moments estimator for θ.

Therefore the method of moments estimator is

$$\hat{\theta} = \frac{2\overline{X} - 1}{1 - \overline{X}}$$

Properties of the Method of Moments Estimator

The estimator is

- Consistent
- Under 'nice' conditions, asymptotically normal

Likelihood function

The idea here is that we 'punish' models that have very unlikely observations heavily (low density values) and look for a general 'good' fit.

Likelihood Function

For observations $x_1, \dots, x_n$ in a random sample, the likelihood function is

$$\mathcal{L}(\theta; x) = \prod_{i=1}^{n} f(x_i; \theta)$$

Log-likelihood function

For observations $x_1, \dots, x_n$ in a random sample, the log-likelihood function is

$$\ell(\theta; x) = \ln \mathcal{L}(\theta; x) = \sum_{i=1}^{n} \ln[f(x_i; \theta)]$$

Maximum Likelihood Estimator (MLE)

Definition (Maximum Likelihood Estimator)

θ_{MLE} is the MLE of θ that maximises the likelihood function $\mathcal{L}(\theta; x)$.

Theorem (Computation of the MLE)

θ_{MLE} *also* maximises the log-likelihood function $\ell(\theta)$

Maximum Likelihood Estimator (MLE)

Example

Consider the following sequence of i.i.d. LogNormal$(0, \sigma^2)$ random variables, $\{X_i\}_{i=1}^n$ with PDF

$$f_{X_i}(x_i; \sigma) = \frac{1}{x_i \sigma \sqrt{2\pi}} e^{-\frac{(\ln x_i)^2}{2\sigma^2}}, \quad x > 0.$$

Find σ_{MLE}.

$$\ell(\sigma; x) = \sum_{i=1}^{n} \ln \left[\frac{1}{x_i \sigma \sqrt{2\pi}} e^{-\frac{(\ln x_i)^2}{2\sigma^2}} \right]$$

Maximum Likelihood Estimator (MLE)

Example

Consider the following sequence of i.i.d. LogNormal$(0, \sigma^2)$ random variables, $\{X_i\}_{i=1}^n$ with PDF

$$f_{X_i}(x_i; \sigma) = \frac{1}{x_i \sigma \sqrt{2\pi}} e^{-\frac{(\ln x_i)^2}{2\sigma^2}}, \quad x > 0.$$

Find σ_{MLE}.

$$\begin{aligned}
\ell(\sigma; x) &= \sum_{i=1}^n \ln \left[\frac{1}{x_i \sigma \sqrt{2\pi}} e^{-\frac{(\ln x_i)^2}{2\sigma^2}} \right] \\
&= -n \ln \sigma - \sum_{i=1}^n \frac{(\ln x_i)^2}{2\sigma^2} + C
\end{aligned}$$

Maximum Likelihood Estimator (MLE)

Example

Consider the following sequence of i.i.d. $\mathrm{LogNormal}(0, \sigma^2)$ random variables, $\{X_i\}_{i=1}^n$ with PDF

$$f_{X_i}(x_i; \sigma) = \frac{1}{x_i \sigma \sqrt{2\pi}} e^{-\frac{(\ln x_i)^2}{2\sigma^2}}, \quad x > 0.$$

Find σ_{MLE}.

We want to find the maximum, so we differentiate w.r.t. σ and find $\widehat{\sigma}$ such that the derative is 0:

$$\ell'(\sigma; x) = -\frac{n}{\sigma} + \sum_{i=1}^{n} \frac{(\ln x_i)^2}{\sigma^3}$$

Maximum Likelihood Estimator (MLE)

Example

Consider the following sequence of i.i.d. LogNormal$(0, \sigma^2)$ random variables, $\{X_i\}_{i=1}^n$ with PDF

$$f_{X_i}(x_i; \sigma) = \frac{1}{x_i \sigma \sqrt{2\pi}} e^{-\frac{(\ln x_i)^2}{2\sigma^2}}, \quad x > 0.$$

Find σ_{MLE}.

We want to find the maximum, so we differentiate w.r.t. σ and find $\widehat{\sigma}$ such that the derative is 0:

$$\ell'(\widehat{\sigma}; x) = -\frac{n}{\widehat{\sigma}} + \sum_{i=1}^{n} \frac{(\ln x_i)^2}{\widehat{\sigma}^3} = 0$$

$$\widehat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (\ln x_i)^2 \quad \text{Are we done?}$$

Maximum Likelihood Estimator (MLE)

Example

Consider the following sequence of i.i.d. LogNormal$(0, \sigma^2)$ random variables, $\{X_i\}_{i=1}^n$ with PDF

$$f_{X_i}(x_i; \sigma) = \frac{1}{x_i \sigma \sqrt{2\pi}} e^{-\frac{(\ln x_i)^2}{2\sigma^2}}, \quad x > 0.$$

Find σ_{MLE}.

We have to check that is indeed a maximum!

$$\ell''(\sigma; x) = \frac{n}{\sigma^2} - \sum_{i=1}^n \frac{3(\ln x_i)^2}{\sigma^4}$$

Example

Consider the following sequence of i.i.d. LogNormal$(0, \sigma^2)$ random variables, $\{X_i\}_{i=1}^n$ with PDF

$$f_{X_i}(x_i; \sigma) = \frac{1}{x_i \sigma \sqrt{2\pi}} e^{-\frac{(\ln x_i)^2}{2\sigma^2}}, \quad x > 0.$$

Find σ_{MLE}.

We have to check that is indeed a maximum!

$$\ell''(\sigma; x) = \frac{n}{\sigma^2} - \sum_{i=1}^{n} \frac{3(\ln x_i)^2}{\sigma^4}$$

$$\ell''(\widehat{\sigma}; x) = \frac{n}{\widehat{\sigma}^2} - \sum_{i=1}^{n} \frac{3(\ln x_i)^2}{\widehat{\sigma}^4}$$

Maximum Likelihood Estimator (MLE)

Example

Consider the following sequence of i.i.d. LogNormal$(0, \sigma^2)$ random variables, $\{X_i\}_{i=1}^n$ with PDF

$$f_{X_i}(x_i; \sigma) = \frac{1}{x_i \sigma \sqrt{2\pi}} e^{-\frac{(\ln x_i)^2}{2\sigma^2}}, \quad x > 0.$$

Find σ_{MLE}.

We have to check that is indeed a maximum!

$$\ell''(\sigma; x) = \frac{n}{\sigma^2} - \sum_{i=1}^n \frac{3(\ln x_i)^2}{\sigma^4}$$

$$\ell''(\widehat{\sigma}; x) = \frac{n}{\widehat{\sigma}^2} - \sum_{i=1}^n \frac{3(\ln x_i)^2}{\widehat{\sigma}^4} = \frac{1}{\widehat{\sigma}^2}(n - 3n) = -\frac{2n}{\widehat{\sigma}^2} < 0.$$

Maximum Likelihood Estimator (MLE)

Example

Consider the following sequence of i.i.d. LogNormal$(0, \sigma^2)$ random variables, $\{X_i\}_{i=1}^n$ with PDF

$$f_{X_i}(x_i; \sigma) = \frac{1}{x_i \sigma \sqrt{2\pi}} e^{-\frac{(\ln x_i)^2}{2\sigma^2}}, \quad x > 0.$$

Find σ_{MLE}.

Thus we have that this is indeed a maximum stationary point at $\widehat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\ln x_i)^2}$, i.e.

$$\sigma_{\text{MLE}} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\ln X_i)^2}.$$

Properties of the MLE

- Equivariant: $g(\theta_{MLE})$ is also the MLE of $g(\theta)$
- Asymptotically normal
- Consistent (in this course)
- *Asymptotically optimal

The Fisher Information

Definition

Let $\ell(\theta)$ be the log-likelihood function of a random sample. The Fisher score is just its defined as:

$$S_n(\theta) = \ell'(\theta; x).$$

Definition

The Fisher information is defined as

$$I_n(\theta) = -\mathbb{E}[\ell''(\theta; x)]$$

where we swap out x_i for X_i.

Theorem (Alternate definition of Fisher Information)

$$I_n(\theta) = \mathbb{E}[(\ell'(\theta; x))^2]$$

The Fisher Information

Example

For the earlier example, with $\ell'(\sigma; X) = -\frac{n}{\sigma} + \sum_{i=1}^{n} \frac{(\ln X_i)^2}{\sigma^3}$, what is its Fisher information?

$$-\mathbb{E}[\ell''(\sigma; X)] = -\mathbb{E}\left[\frac{n}{\sigma^2} - \sum_{i=1}^{n} \frac{3(\ln X_i)^2}{\sigma^4}\right]$$

$$= -\frac{n}{\sigma^2} + \frac{3n}{\sigma^4}\mathbb{E}[(\ln X_1)^2]$$

Recalling that $X_1 \sim \text{LogNormal}(0, \sigma^2)$ and so we have: $\ln X_1 \sim \mathcal{N}(0, \sigma^2)$.

The Fisher Information

Example

For the earlier example, with $\ell'(\sigma; X) = -\frac{n}{\sigma} + \sum_{i=1}^{n} \frac{(\ln X_i)^2}{\sigma^3}$, what is its Fisher information?

$$-\mathbb{E}[\ell''(\sigma; X)] = -\mathbb{E}\left[\frac{n}{\sigma^2} - \sum_{i=1}^{n} \frac{3(\ln X_i)^2}{\sigma^4}\right]$$

$$= -\frac{n}{\sigma^2} + \frac{3n}{\sigma^4}\mathbb{E}[(\ln X_1)^2]$$

Recalling that $X_1 \sim \text{LogNormal}(0, \sigma^2)$ and so we have: $\ln X_1 \sim \mathcal{N}(0, \sigma^2)$.

$$= -\frac{n}{\sigma^2} + \frac{3n}{\sigma^4}(\text{Var}(\ln X_1) + \mathbb{E}[\ln X_1]^2)$$

$$= -\frac{n}{\sigma^2} + \frac{3n}{\sigma^4}(\sigma^2 + 0^2) = \frac{2n}{\sigma^2}.$$

The Fisher Information

Example

Find $I_n(\theta)$ if you're told that $\mathbb{E}[X_i] = \theta$ and

$$\ell'(\theta) = e^{-\theta} + \theta \sum_{i=1}^{n} x_i.$$

The second derivative, with x_i replaced by X_i, is

$$\ell''(\theta) = -e^{-\theta} + \sum_{i=1}^{n} X_i$$

so its Fisher information is

$$I_n(\theta) = \mathbb{E}\left[e^{-\theta} - \sum_{i=1}^{n} X_i\right]$$

The Fisher Information

The second derivative, with x_i replaced by X_i, is

$$\ell''(\theta) = -e^{-\theta} + \sum_{i=1}^{n} X_i$$

so its Fisher information is

$$\begin{aligned} I_n(\theta) &= \mathbb{E}\left[e^{-\theta} - \sum_{i=1}^{n} X_i\right] \\ &= e^{-\theta} - \sum_{i=1}^{n} \mathbb{E}[X_i] \qquad \text{(Why?)} \end{aligned}$$

The Fisher Information

The second derivative, with x_i replaced by X_i, is

$$\ell''(\theta) = -e^{-\theta} + \sum_{i=1}^{n} X_i$$

so its Fisher information is

$$\begin{aligned} I_n(\theta) &= \mathbb{E}\left[e^{-\theta} - \sum_{i=1}^{n} X_i\right] \\ &= e^{-\theta} - \sum_{i=1}^{n} \mathbb{E}[X_i] && \text{(Why?)} \\ &= e^{-\theta} - \sum_{i=1}^{n} \theta \\ &= e^{-\theta} - n\theta \end{aligned}$$

Variance and Standard Error of the MLE

Theorem (Estimation for the Standard Error)

Given θ_{MLE},

$$I_n(\theta)\ \mathrm{Var}(\theta_{MLE}) \xrightarrow{\mathbb{P}} 1$$

Therefore

$$\boxed{\mathrm{Se}(\theta_{MLE}) \approx \frac{1}{\sqrt{I_n(\theta)}}}$$

Variance and Standard Error of the MLE

Example

For the earlier example, with $I_n(\sigma) = \frac{2n}{\sigma^2}$, estimate $\text{Se}(\sigma_{MLE})$

$$\text{Se}(\sigma_{MLE}) \approx \frac{\sigma}{\sqrt{2n}}$$

Approximate CI's: A remark

We can just replace $\text{Se}(\hat{\theta})$ with $\dfrac{1}{\sqrt{I_n(\theta)}}$ if $\hat{\theta}$ is the *MLE*.

Asymptotic Optimality: A remark

What it means in English:

If the *MLE* is asymptotically normal, then the variance of θ_{MLE} is less than the variance of **any other estimator** for θ

Hypothesis tests

The basic idea behind **Hypothesis testing** is that we are essentially performing a **proof by contradiction with some level of uncertainty involved.**

The initial assumption is known as the null hypothesis and we check it's probability based on the chosen alternate hypothesis.

The hypotheses

Definition (Null Hypothesis, Alternate Hypothesis)

In the null hypothesis H_0, we claim that our parameter θ takes a particular value, say θ_0. This is usually the **simpler assumption** to work under.

In the alternate hypothesis H_1, we claim some kind of different dependencies.

The 2901 alternate hypotheses:

- $H_1 : \theta \neq \theta_0$
- $H_1 : \theta > \theta_0$
- $H_1 : \theta < \theta_0$

p-values

Definition (p-value)

The p value tells you how much evidence there is against the null hypothesis (depends on alternative hypothesis choice).

The smaller the p-value, the more evidence against the null hypothesis there is (unlikely for initial assumption to be true).

We compare the p-value to a pre-determined acceptance level, α and reject the null if it is lower than this pre-determined α.

Set-up of a Hypothesis Test (Using p-value)

1. State the null and alternate hypotheses
2. State the test statistic, and its distribution if we assume H_0 is true
3. Find the observed value of the test statistic
4. Compute the corresponding p-value
5. Draw a conclusion

Exact Test Statistics (Normal samples, known variance)

Suppose we know what the variance σ^2 is. We test $H_0 : \mu = \mu_\theta$.

The null distribution is $\boxed{Z \sim \mathcal{N}(0,1)}$.

H_1 :	Test statistic	p-value	p-value
$\mu \neq \mu_0$	$\dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$	$\mathbb{P}(\lvert Z\rvert > \lvert\text{observed value}\rvert)$	$2\,\mathbb{P}(Z > \lvert\text{obs. value}\rvert)$
$\mu > \mu_0$	As above	$\mathbb{P}(Z > \text{observed value})$	
$\mu < \mu_0$	As above	$\mathbb{P}(Z < \text{observed value})$	

Exact Test Statistics (Normal samples, unknown variance)

Suppose we estimate the variance σ^2 via S^2. We test $H_0 : \mu = \mu_0$.

The null distribution is $\boxed{T \sim t_{n-1}}$.

H_1 :	Test statistic	p-value	p-value
$\mu \neq \mu_0$	$\dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}$	$\mathbb{P}(\lvert T\rvert > \lvert\text{observed value}\rvert)$	$2\,\mathbb{P}(T > \lvert\text{obs. value}\rvert)$
$\mu > \mu_0$	As above	$\mathbb{P}(T > \text{observed value})$	
$\mu < \mu_0$	As above	$\mathbb{P}(T < \text{observed value})$	

Exact Tests (Example)

Example 2011 FE Q1

Consider the following table which denotes the number of major earthquakes each year since 2016. We want to see whether there have been more earthquakes over these past few years.

Year	2016	2017	2018	2019
# of major earthquakes	17	12	17	21

Long-term records suggest that we should expect an average of 16.25 earthquakes per year. We assume that this data is normally distributed.

Exact Tests (Example)

Example 2011 FE Q1

Year	2016	2017	2018	2019
# of major earthquakes	17	12	17	21

We assume that this data is normally distributed.

Step 1: State the hypotheses.

$$H_0 : \mu = 16.25 \text{ v.s. } H_1 : \mu > 16.25$$

Exact Tests (Example)

Example 2011 FE Q1

Year	2016	2017	2018	2019
# of major earthquakes	17	12	17	21

We assume that this data is normally distributed.

Step 2: Test statistic under assumption of H_0.
Using the exact test, we have the following statistic:

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{4}}$$

which follows the following distribution under the null,

$$T = \frac{\overline{X} - 16.25}{S/\sqrt{4}} \sim t_3.$$

Example 2011 FE Q1

Year	2016	2017	2018	2019
# of major earthquakes	17	12	17	21

We assume that this data is normally distributed.

Step 3: Find the observed value of the statistic.

$$\overline{x} = 16.75$$
$$s = 3.68556.$$

Thus the observed statistic is:

$$T = \frac{16.75 - 16.25}{3.68556/\sqrt{4}} = 0.27133.$$

Example 2011 FE Q1

Year	2016	2017	2018	2019
# of major earthquakes	17	12	17	21

We assume that this data is normally distributed.

Step 4: Compute the corresponding p-value.

$$\begin{aligned}\text{p-value} &= \mathbb{P}\left(\frac{\overline{X}-\mu_0}{S/\sqrt{4}} > 0.27133\right)\\ &= \mathbb{P}(T > 0.27133)\\ &= \texttt{pt(0.27133, df = 3, lower.tail = FALSE)}\\ &= 0.40187 > 0.05.\end{aligned}$$

Exact Tests (Example)

Example 2011 FE Q1

Year	2016	2017	2018	2019
# of major earthquakes	17	12	17	21

We assume that this data is normally distributed.

Step 4: Compute the corresponding p-value.

$$\begin{aligned}
\text{p-value} &= \mathbb{P}\left(\frac{\overline{X} - \mu_0}{S/\sqrt{4}} > 0.27133\right) \\
&= \mathbb{P}(T > 0.27133) \\
&= \texttt{pt(0.27133, df = 3, lower.tail = FALSE)} \\
&= 0.40187 > 0.05.
\end{aligned}$$

Step 5: Draw a conclusion

Can't refute initial assumption, i.e. fail to reject the null hypothesis under a significance level of 0.05.

Rejection Region

Definition (α-level)

The α-level, sets a standard upon which we reject H_0.

Definition (Rejection region)

Under an α-level, we reject H_0 if our observed value lies in the relevant rejection region.

Test Statistics in Exact tests (Normal samples)

Suppose we know what the variance σ^2 is. We test $H_0 : \mu = \mu_0$.

The null distribution is $\boxed{Z \sim \mathcal{N}(0,1)}$.

H_1 :	Test statistic	Rejection region
$\mu \neq \mu_0$	$\dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$	$\left\{ \lvert\text{observed value}\rvert > z_{1-\frac{\alpha}{2}} \right\}$
$\mu > \mu_0$	As above	$\{\text{observed value} > z_{1-\alpha}\}$
$\mu < \mu_0$	As above	$\{\text{observed value} < -z_{1-\alpha}\}$

Test Statistics in Exact tests (Normal samples)

Suppose we estimate the variance σ^2 via S^2. We test $H_0 : \mu = \mu_0$.

The null distribution is $\boxed{T \sim t_{n-1}}$.

H_1 :	Test statistic	Rejection region
$\mu \neq \mu_0$	$\dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}$	$\left\{ \lvert\text{observed value}\rvert > t_{n-1,1-\frac{\alpha}{2}} \right\}$
$\mu > \mu_0$	As above	$\{\text{observed value} > t_{n-1,1-\alpha}\}$
$\mu < \mu_0$	As above	$\{\text{observed value} < -t_{n-1,1-\alpha}\}$

Set-up of a Hypothesis Test (Using Rejection regions)

1. State the null and alternate hypotheses
2. State the test statistic with its distribution if we assume H_0 is true, and the α-value
3. Determine the relevant rejection region
4. Find the observed value of the test statistic
5. Draw a conclusion

Earlier Example

We wish to test $H_0 : \mu = 16.25$ v.s $H_1 : \mu > 16.25$. Our null distribution is

$$T = \frac{\overline{X} - 16.25}{S/\sqrt{4}} \sim t_3.$$

Set the α level to 5%. Our rejection region is

$$R = \left\{ \frac{\overline{x} - 16.25}{s/\sqrt{4}} > t_{3,0.95} \right\}$$

Earlier Example

$t_{3,0.95} = \texttt{qt(0.95, df=3)} = 2.35336$ (can also find this in given table) so

$$R = \{\text{observed value} > 2.35336\}.$$

Our observed value was 0.27133, which doesn't lie in R. Therefore under a 5% level we fail to reject H_0.

Earlier Example

$t_{3,0.95} = \texttt{qt(0.95, df=3)} = 2.35336$ (can also find this in given table) so

$$R = \{\text{observed value} > 2.35336\}.$$

Our observed value was 0.27133, which doesn't lie in R. Therefore under a 5% level we fail to reject H_0.

Same conclusion reached when comparing p-values. As such the chosen method doesn't matter that much, but selecting the correct alternate hypothesis does.

Error

Definition (Type I Error)

Type I error is when H_0 is true, but was rejected.

Definition (Type II Error)

Type II error is when H_0 is false, but was accepted.

Lemma (The whole point of α)

The α-level is the significance level. The smaller α is, the more **type I** error is controlled.

Asymptotic Test

Assume that $\hat{\theta}$ is an asymptotically normal estimator of θ. We wish to test $H_0 : \theta = \theta_0$ v.s. $H_1 : \theta \neq \theta_0$.

Definition (Wald Test Statistic)

The Wald test statistic is

$$W = \frac{\hat{\theta} - \theta_0}{\mathrm{Se}(\hat{\theta})}$$

with null distribution $\mathcal{N}(0, 1)$.

The p-value is

$$\mathbb{P}(|Z| > |\text{observed value}|) = 2\,\mathbb{P}(Z > |\text{observed value}|)$$

Quick Example

Example

Suppose in the earlier example we computed $\sigma_{MLE} = 0.72706$ under a sample size $n = 20$. Assume that σ_{MLE} is asymptotically normal and that we can estimate $\mathrm{Se}(\sigma_{MLE}) \approx \frac{\sigma_{MLE}}{\sqrt{2n}}$. Test the hypotheses

$$H_0 : \sigma = 0.5 \text{ v.s. } H_1 : \sigma \neq 0.5$$

at a significance level of 5%.

Quick Example

Example

Suppose in the earlier example we computed $\sigma_{MLE} = 0.72706$ under a sample size $n = 20$. Assume that σ_{MLE} is asymptotically normal and that we can estimate $\mathrm{Se}(\sigma_{MLE}) \approx \frac{\sigma_{MLE}}{\sqrt{2n}}$. Test the hypotheses

$$H_0 : \sigma = 0.5 \text{ v.s. } H_1 : \sigma \neq 0.5$$

at a significance level of 5%.

We use the Wald statistic

$$W = \frac{\hat{\sigma} - 0.5}{\mathrm{Se}(\hat{\sigma})} = \frac{\sqrt{2 \times 20}(\sigma_{MLE} - 0.5)}{\sigma_{MLE}} \sim \mathcal{N}(0, 1)$$

under the null hypothesis.

The observed value is 1.97515.

Quick Example

Under the null, we know that this is asymptotically standard normal and so we look at the critical values for a standard normal r.v. $Z \sim \mathcal{N}(0, 1)$.

$$R = \left\{ W = \left| \frac{\hat{\sigma} - 0.5}{\text{Se}(\hat{\sigma})} \right| > |Z_{0.975}| = 1.96 \right\}$$

As the observed value is in the rejection region $1.97515 > 1.96$, we can reject the null hypothesis at a significance level of 5%. Thus we conclude the $\sigma \neq 0.5$.

Likelihood Ratio Test (Asymptotic Version)

Consider the following hypothesis test: $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$.

Asymptotic Likelihood Ratio Test

$$\Lambda(\hat{\theta}_n(X)) = -2\ln\left[\frac{L(\theta_0; X)}{L(\hat{\theta}_n; X)}\right] \xrightarrow{\mathcal{D}} \chi^2_1$$

under certain regularity conditions (this is pretty much assumed to be true within this course).

Note: $\hat{\theta}_n$ is the MLE of θ based on n observations.

Likelihood Ratio Test (Asymptotic Version)

Consider the following hypothesis test: $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$.

Asymptotic Likelihood Ratio Test

$$\Lambda(\hat{\theta}_n(X)) = -2 \ln \left[\frac{L(\theta_0; X)}{L(\hat{\theta}_n; X)} \right] \xrightarrow{\mathcal{D}} \chi_1^2$$

Intuitively speaking, if the null hypothesis assumption is very unlikely to be true, then $L(\theta_0)$ would be much smaller than $L(\hat{\theta}_n)$, i.e. we would reject the null if $\Lambda(\hat{\theta}_n)$ is large enough.

One should note that $\Lambda(\hat{\theta}_n) \geq 0$, and so we end up having a one-sided test in this case.

P-value is

$$\mathbb{P}(\Lambda(\hat{\theta}_n(X) > \text{observed value}).$$

Asymptotic LRT Example

Example

Consider the same hypothesis test as before, but this time we'll be using the asymptotic likelihood ratio test instead. Hypotheses are:

$$H_0 : \sigma = 0.5 \text{ v.s. } H_1 : \sigma \neq 0.5$$

with $\sigma_{MLE} = 0.72706$.

Asymptotic LRT Example

Example

$$H_0 : \sigma = 0.5 \text{ v.s. } H_1 : \sigma \neq 0.5$$

with $\sigma_{MLE} = 0.72706$.

We need to sub in the observed values into the relevant distributions to find $L(0.5; x)$ and $L(0.72706; x)$.
Noting that $X_i \sim \text{LogNormal}(0, \sigma^2)$ for $i = 1, 2, \ldots, 20$ we have for $\sigma = \sigma_1$

$$L(\sigma_1; x) = \frac{1}{\sigma_1^{20} \times (2\pi)^{10}} \prod_{i=1}^{20} x_i^{-1} \times \exp\left[-\frac{1}{2\sigma_1^2} \sum_{i=1}^{20} (\ln x_i)^2\right]$$

Asymptotic LRT Example

Example

$$H_0 : \sigma = 0.5 \text{ v.s. } H_1 : \sigma \neq 0.5$$

with $\sigma_{MLE} = 0.72706$.

We need to sub in the observed values into the relevant distributions to find $L(0.5; x)$ and $L(0.72706; x)$.
Noting that $X_i \sim \text{LogNormal}(0, \sigma^2)$ for $i = 1, 2, \ldots, 20$ we have for $\sigma = \sigma_1$

$$\begin{aligned} L(\sigma_1; x) &= \frac{1}{\sigma_1^{20} \times (2\pi)^{10}} \prod_{i=1}^{20} x_i^{-1} \times \exp\left[-\frac{1}{2\sigma_1^2} \sum_{i=1}^{20} (\ln x_i)^2\right] \\ &\propto \sigma_1^{-20} \times \exp\left[-\frac{1}{2\sigma_1^2} \sum_{i=1}^{20} (\ln x_i)^2\right]. \end{aligned}$$

Asymptotic LRT Example

Subbing everything into the fraction yields:

$$-2\ln\left\{\frac{L(\sigma_0;x)}{L(\sigma_{MLE};x)}\right\} = -2\ln\left\{\frac{\sigma_0^{-20}\times\exp\left[-\frac{1}{2\sigma_0^2}\sum_{i=1}^{20}(\ln x_i)^2\right]}{\sigma_{MLE}^{-20}\times\exp\left[-\frac{1}{2\sigma_{MLE}^2}\sum_{i=1}^{20}(\ln x_i)^2\right]}\right\}$$

Asymptotic LRT Example

Subbing everything into the fraction yields:

$$-2\ln\left\{\frac{L(\sigma_0;x)}{L(\sigma_{MLE};x)}\right\} = -2\ln\left\{\frac{\sigma_0^{-20}\times\exp\left[-\frac{1}{2\sigma_0^2}\sum_{i=1}^{20}(\ln x_i)^2\right]}{\sigma_{MLE}^{-20}\times\exp\left[-\frac{1}{2\sigma_{MLE}^2}\sum_{i=1}^{20}(\ln x_i)^2\right]}\right\}$$

$$= -2\ln\left\{\left(\frac{\sigma_{MLE}}{\sigma_0}\right)^{20}\exp\left[\left(\frac{1}{2\sigma_{MLE}^2}-\frac{1}{2\sigma_0^2}\right)\sum_{i=1}^{20}(\ln x_i)^2\right.\right.$$

Subbing everything into the fraction yields:

$$
\begin{aligned}
-2\ln\left\{\frac{L(\sigma_0;x)}{L(\sigma_{MLE};x)}\right\} &= -2\ln\left\{\frac{\sigma_0^{-20}\times\exp\left[-\frac{1}{2\sigma_0^2}\sum_{i=1}^{20}(\ln x_i)^2\right]}{\sigma_{MLE}^{-20}\times\exp\left[-\frac{1}{2\sigma_{MLE}^2}\sum_{i=1}^{20}(\ln x_i)^2\right]}\right\}\\
&= -2\ln\left\{\left(\frac{\sigma_{MLE}}{\sigma_0}\right)^{20}\exp\left[\left(\frac{1}{2\sigma_{MLE}^2}-\frac{1}{2\sigma_0^2}\right)\sum_{i=1}^{20}(\ln x_i)^2\right]\right\}\\
&= -40\ln\left(\frac{\sigma_{MLE}}{\sigma_0}\right)-\left[\left(\frac{1}{\sigma_{MLE}^2}-\frac{1}{\sigma_0^2}\right)\sum_{i=1}^{20}(\ln x_i)^2\right]
\end{aligned}
$$

We sub things in now and deal with it rickety-split.

Asymptotic LRT Example

Example

$$H_0 : \sigma = 0.5 \text{ v.s. } H_1 : \sigma \neq 0.5$$

with $\sigma_{MLE} = 0.72706$.

Recall from way earlier:

$$\sigma_{MLE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\ln x_i)^2}$$

$$\sum_{i=1}^{n}(\ln x_i)^2 = n\sigma_{MLE}^2$$

and so we just replace n with 20 and we can get rid of the log-sum thing.

Asymptotic LRT Example

Example

$$H_0 : \sigma = 0.5 \text{ v.s. } H_1 : \sigma \neq 0.5$$

with $\sigma_{MLE} = 0.72706$.

Thus, our test statistic is:

$$\begin{aligned}\Lambda(\sigma_{MLE}(x)) &= -2\ln\left\{\frac{L(\sigma_0; x)}{L(\sigma_{MLE}; x)}\right\} \\ &= -40\ln\left(\frac{\sigma_{MLE}}{\sigma_0}\right) - \left[\left(\frac{1}{\sigma_{MLE}^2} - \frac{1}{\sigma_0^2}\right) 20\sigma_{MLE}^2\right]\end{aligned}$$

Asymptotic LRT Example

Example

$$H_0 : \sigma = 0.5 \text{ v.s. } H_1 : \sigma \neq 0.5$$

with $\sigma_{MLE} = 0.72706$.

Thus, our test statistic is:

$$\begin{aligned}
\Lambda(\sigma_{MLE}(x)) &= -2 \ln \left\{ \frac{L(\sigma_0; x)}{L(\sigma_{MLE}; x)} \right\} \\
&= -40 \ln \left(\frac{\sigma_{MLE}}{\sigma_0} \right) - \left[\left(\frac{1}{\sigma_{MLE}^2} - \frac{1}{\sigma_0^2} \right) 20\sigma_{MLE}^2 \right] \\
&= -40 \ln \left(\frac{0.72706}{0.5} \right) - 20 \left[\left(1 - \frac{0.72706^2}{0.5^2} \right) \right] \\
&= 7.31326.
\end{aligned}$$

Asymptotic LRT Example

Example

$$H_0 : \sigma = 0.5 \text{ v.s. } H_1 : \sigma \neq 0.5$$

with $\sigma_{MLE} = 0.72706$.

Reject or fail to reject the null?

Asymptotic LRT Example

Example

$$H_0 : \sigma = 0.5 \text{ v.s. } H_1 : \sigma \neq 0.5$$

with $\sigma_{MLE} = 0.72706$.

Reject or fail to reject the null?

p-value:

$$\begin{aligned}
\text{p-value} &= \mathbb{P}(\Lambda(\sigma_{MLE}(X)) > \Lambda(\sigma_{MLE}(x))) \\
&= \mathbb{P}(\chi_1^2 > 7.31326) \\
&= \texttt{pchisq(7.31326, df = 1, lower.tail = FALSE)} \\
&= 0.0068448 < 0.05.
\end{aligned}$$

Thus at a significance level of 5%, we can reject the null hypothesis.

Table of Contents

1 Study Design
- Samples

2 Statistical Inference
- Estimators and their Properties
  - Properties of any estimator
  - Desirable features of estimators
- More work with Convergence
- Confidence Intervals
- Distributions arising from $\mathcal{N}(0,1)$
  - Working with $S = S_X$
- Common Estimators
- Hypothesis tests

3 Coolio Doolio

Good Luck on your finals!