MATH2801/2901 Final Revision

Part II: Statistical Inference

Rui Tong

UNSW Society of Statistics

30 May 2018

Table of Contents

Before we start: Some thoughts on the finals...

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
   - Common Estimators
   - Distributions arising from $\mathcal{N}(0, 1)$
     - Working with $S = S_X$
   - Hypothesis tests
3. Acknowledgements

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
   - Common Estimators
   - Distributions arising from $\mathcal{N}(0,1)$
     - Working with $S = S_X$
   - Hypothesis tests
3. Acknowledgements

Random Sample

Definition

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

$$X_1, \dots, X_n$$

Extra notation:

- $x_1, \dots, x_n$ is the random data, which is the 'observed value' of the random sample.
- X is 'representative' for this sample if $f_X(x) = f_{X_i}(x)$ for all i

Statistics

Definition (Statistic)

For a random sample $X_1, \dots, X_n$, 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: $X_{0.5}$

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,

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

The Sample Mean

Example (2901)

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

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

Efficiency of Statistics (2801)

Definition (Efficiency)

Let $g(X_1, \ldots, X_n)$ and $h(Y_1, \ldots, Y_m)$ be two distinct **unbiased** statistics.

$g(X_1, \ldots, X_n)$ is more **efficient** than $h(Y_1, \ldots, Y_m)$ if it has smaller variance, i.e.

$$\text{Var}[g(X_1, \ldots, X_n)] < \text{Var}[h(Y_1, \ldots, Y_m)]$$

Remark: This means we can use different statistics, *or* sample differently, to increase efficiency.

Sampling methods

- Simple random sample - Sampling in a so that all possible samples are equally likely. (Can be hard to do in practice)
- Stratified random sample - As above, but dividing into subclasses of samples beforehand (e.g. age)
- Cluster sampling - Sampling in small groups

Experimental Design (2801)

- Observational study - We don't manipulate any variables.
- Experiment - We manipulate some variables and observe what happens to a 'response' variable.

Experimental Design (2801)

Important features to include in experiments:

- Compare - showing a change in one variable influences a change in another (e.g. via placebo)
- Randomise - minimise the influences of other factors (e.g. gender)
- Repetition

Remark (2801)

I've never seen this be examined...?

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
- Common Estimators
- Distributions arising from $\mathcal{N}(0, 1)$
  - Working with $S = S_X$
- Hypothesis tests

3 Acknowledgements

Estimators

Let $X_1, \ldots, X_n$ 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.

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

Meaning, fundamentally it's just a statistic.

Estimators

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

Because estimators are functions of random variables, the estimator itself is a random variable.

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

An estimator is 'biased' when it has a tendency of estimating *a little bit off* what the actual value of the parameter is. The bias measures how much it tends to be off by.

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 (2801 ver)

Definition (Standard Error)

$$\text{se}(\hat{\theta}) = \sqrt{\text{Var}_{\hat{\theta}}(\hat{\theta})}$$

Significance

Basically adapted from the significance of the variance; it measures just how much error the estimator is susceptible to.

Steps:

1. Compute $\text{Var}(\hat{\theta})$ the usual way
2. Square root it
3. For the standard error, replace θ with $\hat{\theta}$.

Standard Error (2801 ver)

Example

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

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

Therefore $\text{se}(\hat{\lambda}) = \sqrt{\frac{5\hat{\lambda}}{28}}$.

Standard Error (2901 ver)

Definition (Standard Error)

$$\text{se}(\hat{\theta}) = \sqrt{\text{Var}(\hat{\theta})}$$

Definition (Estimated Standard Error)

$$\widehat{\text{se}}(\hat{\theta}) = \text{se}(\hat{\theta}), \text{ evaluated at } \theta = \hat{\theta}$$

Standard Error (2901 ver)

Example

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

Recycling earlier computations...

$$\text{se}(\hat{\lambda}) = \sqrt{\frac{5\lambda}{28}}$$

which implies that

$$\widehat{\text{se}}(\hat{\lambda}) = \sqrt{\frac{5\hat{\lambda}}{28}}.$$

Mean Squared Error

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)

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

Definition (Estimated Mean Square Error) (2801)

$$\widehat{\mathsf{MSE}}(\hat{\theta}) = [\mathsf{bias}(\hat{\theta})]^2 + [\mathsf{se}(\hat{\theta})]^2.$$

Mean Squared Error formula - Proof (2901)

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

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

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

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

Mean Squared Error formula - Proof (2901)

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 formula - "Proof" (2901)

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

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

"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 λ_2 better than λ_1?

We can compute:

$$\text{bias}(\hat{\lambda}_2) = 0$$
$$\text{Var}(\hat{\lambda}_2) = \frac{\lambda}{7}$$
$$\therefore \text{MSE}(\hat{\lambda}_2) = \frac{\lambda}{7}$$

"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 λ_2 better than λ_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 $X_1, \ldots, X_n$ converges in probability to X, i.e. $X_n \xrightarrow{P} X$, 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.

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

Verifying that an estimator is consistent

Theorem (Sufficient criteria for consistency)

If

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

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

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

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

Equivariance

Definition (Equivariance Estimator)

$\hat{\theta}_n$ is an equivariant estimator for θ if $g(\hat{\theta}_n)$ is an estimator for $g(\theta)$.

(Only really useful for the MLE.)

Asymptotic Normality

A sequence of random variables $X_1, \ldots, X_n$ 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

$$\frac{\hat{\theta}_n - \theta}{\text{se}(\hat{\theta})} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

2901 note: This is an abuse of notation. But we don't care.

Remark

You don't need to know how to prove these, just how to use it... (soon)

Convergence Theorems

Central Limit Theorem

For a random sample $X_1, \dots, X_n$ 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:

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

where c is a constant. Then,

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

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 $g'(\theta) \neq 0$,

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

The Delta Method

Example

Suppose $\hat{\beta}_1, \hat{\beta}_2, \ldots$ is a sequence of *i.i.d.* Exp(β) random variables. Find the 'asymptotic distribution' of $\ln \hat{\beta}_n$.

From the CLT and the formula sheet:

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

The Delta Method

Example

Suppose $\hat{\beta}_1, \hat{\beta}_2, \ldots$ is a sequence of *i.i.d.* $\text{Exp}(\beta)$ random variables. Find the 'asymptotic distribution' of $\ln \hat{\beta}_n$.

From the CLT and the formula sheet:

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

We know $\beta \in (0, \infty)$, so ln is differentiable at β. Also $(\ln \beta)'$, i.e. β^{-1}, never equals 0. So we can use the Delta method:

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

The Delta Method

Example

Suppose $\hat{\beta}_1, \hat{\beta}_2, \dots$ is a sequence of *i.i.d.* $\text{Exp}(\beta)$ random variables. Find the 'asymptotic distribution' of $\ln \hat{\beta}_n$.

Use the properties of the normal distribution!

$$\sqrt{n}\left(\ln \hat{\beta}_n - \ln \beta\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

The Delta Method

Example

Suppose $\hat{\beta}_1, \hat{\beta}_2, \dots$ is a sequence of *i.i.d.* $\text{Exp}(\beta)$ random variables. Find the 'asymptotic distribution' of $\ln \hat{\beta}_n$.

Use the properties of the normal distribution!

$$\sqrt{n}\left(\ln \hat{\beta}_n - \ln \beta\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

$$\ln \hat{\beta}_n - \ln \beta \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{n}\right)$$

The Delta Method

Example

Suppose $\hat{\beta}_1, \hat{\beta}_2, \dots$ is a sequence of *i.i.d.* $\text{Exp}(\beta)$ random variables. Find the 'asymptotic distribution' of $\ln \hat{\beta}_n$.

Use the properties of the normal distribution!

$$\sqrt{n}\left(\ln \hat{\beta}_n - \ln \beta\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$$

$$\ln \hat{\beta}_n - \ln \beta \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{n}\right)$$

$$\ln \hat{\beta}_n \xrightarrow{\mathcal{D}} \mathcal{N}\left(\ln \beta, \frac{1}{n}\right)$$

Confidence Intervals (Generic Definition)

In a confidence interval, we put the parameter in the middle, instead of the random variable.

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

$$(L, U)$$

Note: α is just a percentage!

Confidence Intervals (Generic Definition)

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

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

$$\mathbb{P}(L < \theta < U) = 0.95$$

for some estimators L and U, then a 95% confidence interval for θ is

$$(L, U)$$

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

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

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

Approximate CI's via Asymptotic Normality

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

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

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

is a 95% confidence interval for θ.

Approximate CI's via Asymptotic Normality

Example (Adapted from Tutorial)

Consider a random sample $X_1, \dots, X_n$ 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.

Method 1: Directly use the formula: The sample mean is always consistent and asymptotically normal. Recall 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

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

2901 note: This is actually $\widehat{\text{se}}(\hat{\lambda})$!

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.

(In case you forgot...) According to R,

$$z_{0.975} = \texttt{qnorm(0.975)} = 1.959964$$

so an approximate confidence interval is

$$\left(\overline{X} - 1.96\sqrt{\frac{\hat{\lambda}}{n}}, \overline{X} + 1.96\sqrt{\frac{\hat{\lambda}}{n}} \right)$$

Approximate CI's via Asymptotic Normality

Example (Adapted from Tutorial)

Consider a random sample $X_1, \dots, X_n$ 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.

Method 2: Derive it on the day: Again, because the sample mean is consistent and asymptotically normal, noting that $\text{Var}(X_i) = \lambda$:

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

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:

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

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

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

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

Be very careful going from line 2 to line 3!

Approximate CI's via Asymptotic Normality

Example (Adapted from Tutorial)

Consider a random sample $X_1, \dots, X_n$ 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:

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

so a 95% confidence interval is

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

(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?

Aside: Alternate forms of the CI (Mostly 2901)

The confidence interval

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

can be re-expressed as

$$\left(\hat{\theta}_n + z_{0.025}\,\mathrm{se}(\hat{\theta}), \hat{\theta}_n - z_{0.025}\,\mathrm{se}(\hat{\theta})\right)$$

or as

$$\left(\hat{\theta}_n + z_{0.025}\,\mathrm{se}(\hat{\theta}), \hat{\theta}_n + z_{0.975}\,\mathrm{se}(\hat{\theta})\right)$$

...what else? And why?

Aside: "Symmetry" of CI's (Mostly 2901)

They actually don't need to be symmetric.

In fact, confidence intervals are never unique!

Method of Moments

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

Definition (Method of Moments Estimator)

Consider the system of equations

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

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

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?

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.

Only 1 parameter, therefore we only need one equation:

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

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.

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

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:

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

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

Likelihood Function

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

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

Log-likelihood function

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

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

Maximum Likelihood Estimator (MLE)

Definition (Maximum Likelihood Estimator)

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

Theorem (Computation of the MLE)

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

Maximum Likelihood Estimator (MLE)

Example (2901 Assignment, 2017)

Let $\theta > -1$ and $X_1, \dots, X_n$ be i.i.d. r.v.s with PDF

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

Find θ_{MLE}

$$\ell(\theta) = \sum_{i=1}^{n} \ln[(\theta + 1)x_i^{\theta}]$$

Maximum Likelihood Estimator (MLE)

Example (2901 Assignment, 2017)

Let $\theta > -1$ and $X_1, \dots, X_n$ be i.i.d. r.v.s with PDF

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

Find θ_{MLE}

$$\begin{aligned}\ell(\theta) &= \sum_{i=1}^{n} \ln[(\theta+1)x_i^{\theta}] \\ &= \sum_{i=1}^{n} [\ln(\theta+1) + \theta \ln(x_i)] \\ &= n\ln(\theta+1) + \theta \sum_{i=1}^{n} \ln(x_i)\end{aligned}$$

Maximum Likelihood Estimator (MLE)

Example (2901 Assignment, 2017)

Let $\theta > -1$ and $X_1, \dots, X_n$ be i.i.d. r.v.s with PDF

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

Find θ_{MLE}

Remember, ℓ is a function of θ.

$$\ell'(\theta) = \frac{n}{\theta + 1} + \sum_{i=1}^{n} \ln x_i$$

Maximum Likelihood Estimator (MLE)

Example (2901 Assignment, 2017)

Let $\theta > -1$ and $X_1, \dots, X_n$ be i.i.d. r.v.s with PDF

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

Find θ_{MLE}

Set $\ell'(\theta) = 0$.

$$\frac{n}{\theta+1} + \sum_{i=1}^{n} \ln x_i = 0$$

$$\frac{1}{\theta+1} = -\frac{1}{n}\sum_{i=1}^{n} \ln x_i$$

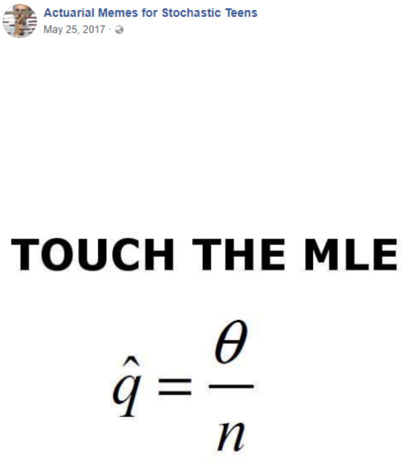
$$\theta = -1 - \left(\frac{1}{n}\sum_{i=1}^{n} \ln x_i\right)^{-1}$$

Where everybody loses marks

Use the second derivative!

Things are problematic if you found the minimum instead...

...


Actuarial Memes for Stochastic Teens
May 25, 2017 ·
TOUCH THE MLE
$\hat{q} = \frac{\theta}{n}$

memes...

TOUCH THE MLE

$$\hat{q} = \frac{\theta}{n}$$

OMG WHAT ARE YOU DOING!? DID YOU EVEN CHECK THE SECOND ORDER CONDITIONS??

$$\frac{\partial^2}{\partial q^2} \log L(q) = -\frac{\theta}{q^2} - \frac{n-\theta}{(1-q)^2} < 0 .$$

[2]

Maximum Likelihood Estimator (MLE)

Example

Let $\theta > -1$ and $X_1, \dots, X_n$ be i.i.d. r.v.s with PDF

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

Find θ_{MLE}

$$\ell'(\theta) = \frac{n}{\theta + 1} + \underbrace{\sum_{i=1}^{n} \ln(x_i)}_{\text{constant w.r.t. } \theta}$$

$$\ell''(\theta) = -\frac{n}{(\theta + 1)^2} < 0$$

Hence $\theta_{MLE} = -1 - \left(\frac{1}{n}\sum_{i=1}^{n} \ln X_i\right)^{-1}$ (note the capital X)

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 (2901)

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

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

Definition

The Fisher information is defined as

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

where we swap out x_i for X_i.

Theorem (Alternate definition of Fisher Information)

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

The Fisher Information

Example

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

$$-\mathbb{E}[\ell''(\theta)] = -\mathbb{E}\left[-\frac{n}{(\theta+1)^2}\right] = \frac{n}{(\theta+1)^2}.$$

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

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

Variance and Standard Error of the MLE

Theorem (Estimation for the Standard Error)

Given θ_{MLE},

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

Therefore

$$\boxed{\operatorname{se}(\theta_{MLE}) \approx \frac{1}{\sqrt{I_n(\theta)}}}$$

Variance and Standard Error of the MLE

Example

For the earlier example, with $I_n(\theta) = \frac{n}{(\theta+1)^2}$, estimate $\text{se}(\theta_{MLE})$

$$\text{se}(\theta_{MLE}) \approx \frac{\theta + 1}{\sqrt{n}}$$

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

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 i.i.d. Then their sum satisfies:

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

Revision: Probability

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 $X_1, \dots, X_n$, the sample variance is

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

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

2901: You know this as S_X^2.

Key property

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

...don't try using n instead of $n - 1$...

...you will live a life of regret.

Sample Variance and Distributions

Theorem (Distribution of Sample Variance) (2801 formula sheet)

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

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

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 $X_1, \dots, X_n$ 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,

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

Again, you can derive this on the spot.

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.

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

$$S^2 = \frac{1}{3-1}\bigg((1.1633974 + 0.2125267)^2 + (0.2523621 + 0.2125267)^2 + (-2.0633406 + 0.2125267)^2\bigg) = 2.7721$$

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)

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.

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

The 2801/2901 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.

The smaller the p-value, the more evidence against the null hypothesis there is.

If there's more evidence against the null hypothesis, we reject it.

Set-up of a Hypothesis Test (mostly 2801)

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

Test Statistics in Exact tests (Normal samples)

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
$\theta \neq \theta_0$	$\dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$	$\mathbb{P}(\lvert Z\rvert > \lvert\text{observed value}\rvert)$	$2\,\mathbb{P}(Z > \lvert\text{obs. value}\rvert)$
$\theta > \theta_0$	As above	$\mathbb{P}(Z > \text{observed value})$	
$\theta < \theta_0$	As above	$\mathbb{P}(Z < \text{observed value})$	

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	p-value	p-value
$\theta \neq \theta_0$	$\dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}$	$\mathbb{P}(\lvert T\rvert > \lvert\text{observed value}\rvert)$	$2\,\mathbb{P}(T > \text{obs. value})$
$\theta > \theta_0$	As above	$\mathbb{P}(T > \text{observed value})$	
$\theta < \theta_0$	As above	$\mathbb{P}(T < \text{observed value})$	

Exact Tests (Example)

Example (2901 Course Pack)

A popular brand of yoghurt claims to contain 120 calories per serving. A consumer watchdog group randomly sampled 14 servings of the yoghurt and obtained the following numbers of calories per serving:

160 200 220 230 120 180 140 130 170 190 80 120 100 170

Use this data to test the claim.

Step 1: State the hypotheses.

$$H_0 : \mu = 120 \text{ v.s. } H_1 : \mu \neq 120$$

Exact Tests (Example)

Example (2901 Course Pack)

160	200	220	230	120	180	140	130	170	190	80	120	100	170

Use this data to test the claim.

Step 2: State the test statistic, and its null distribution.

We will consider

$$T = \frac{\overline{X} - \mu}{S/\sqrt{14}}$$

and under H_0,

$$T = \frac{\overline{X} - 120}{S/\sqrt{14}} \sim t_{13}$$

Exact Tests (Example)

Example (2901 Course Pack)

160 200 220 230 120 180 140 130 170 190 80 120 100 170

Use this data to test the claim.

Step 3: Find the observed value of the statistic:

$$\overline{x} = 157.8571$$
$$s = 44.75206$$

so the observed value is

$$\frac{\overline{x} - 120}{s/\sqrt{14}} = \frac{157.8571 - 120}{44.75206/\sqrt{14}} = 3.165183$$

Exact Tests (Example)

Example (2901 Course Pack)

160 200 220 230 120 180 140 130 170 190 80 120 100 170

Use this data to test the claim.

Steps 4/5: Compute the p-value and arrive at a conclusion.

$$\begin{aligned} p\text{-value} &= \mathbb{P}\left(\left| \frac{\overline{X} - 120}{S/\sqrt{14}} \right| > 3.165183 \right) \\ &= 2\mathbb{P}(T > 3.165183) \\ &= 2 \texttt{ * pt(3.165183, df=13, lower.tail=FALSE)} \\ &= 0.00745 \end{aligned}$$

Strong evidence against H_0. The company lied to us...

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
$\theta \neq \theta_0$	$\dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$	$\left\{ \lvert\text{observed value}\rvert > z_{1-\frac{\alpha}{2}} \right\}$
$\theta > \theta_0$	As above	$\left\{ \text{observed value} > z_{1-\frac{\alpha}{2}} \right\}$
$\theta < \theta_0$	As above	$\left\{ \text{observed value} < z_{1-\frac{\alpha}{2}} \right\}$

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
$\theta \neq \theta_0$	$\frac{\overline{X} - \mu_0}{S/\sqrt{n}}$	$\left\{ \lvert\text{observed value}\rvert > t_{n-1,1-\frac{\alpha}{2}} \right\}$
$\theta > \theta_0$	As above	$\left\{ \text{observed value} > t_{n-1,1-\frac{\alpha}{2}} \right\}$
$\theta < \theta_0$	As above	$\left\{ \text{observed value} < t_{n-1,1-\frac{\alpha}{2}} \right\}$

Set-up of a Hypothesis Test (mostly 2901)

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 $H_0 : \mu = 120$ v.s $H_1 : \mu \neq 120$. Our null distribution was

$$T = \frac{\overline{X} - 120}{S/\sqrt{14}}.$$

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

$$R = \left\{ \left| \frac{\overline{x} - \mu}{s/\sqrt{14}} \right| > t_{13,0.975} \right\}$$

Earlier Example

$t_{13,0.975} = \texttt{qt(0.975, df=13)} = 2.160369$ so

$$R = \{|\text{observed value}| > 2.160369\}.$$

Our observed value was 3.165183, which lies in R. Therefore under a 5% level we reject H_0.

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

$$W = \frac{\hat{\theta} - \theta_0}{\text{se}(\hat{\theta})}$$

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 $\theta_{MLE} = 0.59366396$ under a sample size $n = 15$.. Assume that θ_{MLE} is asymptotically normal and that we can estimate $\text{se}(\theta_{MLE}) \approx \frac{\theta_{MLE}+1}{\sqrt{n}}$. Test the hypotheses

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

Quick Example

Example

Suppose in the earlier example we computed $\theta_{MLE} = 0.59366396$ under a sample size $n = 15$.. Assume that θ_{MLE} is asymptotically normal and that we can estimate $\text{se}(\theta_{MLE}) \approx \frac{\theta_{MLE}+1}{\sqrt{n}}$. Test the hypotheses

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

We use the Wald statistic $W = \frac{\hat{\theta}-0.5}{\text{se}(\hat{\theta})} = \frac{\sqrt{15}(\theta_{MLE}-0.5)}{\theta_{MLE}+1}$; null distribution $\mathcal{N}(0,1)$.

The observed value is 0.2276258.

Quick Example

Let $Z \sim \mathcal{N}(0, 1)$. Our p-value is then

$$\begin{aligned}\mathbb{P}(|Z| > 0.2276258) &= 2\,\mathbb{P}(Z > 0.2276258) \\ &= \texttt{2*(1 - pnorm(0.2276258))} \\ &= 0.8199372\end{aligned}$$

There is *huge* evidence in favour of H_0 here, so we accept it.

Remark (mostly 2901)

Similar analogies exist for one-sided tests.

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
   - Common Estimators
   - Distributions arising from $\mathcal{N}(0,1)$
     - Working with $S = S_X$
   - Hypothesis tests
3. Acknowledgements

Acknowledgements

Thank you to:

- Matthew Di Giorgio, Rachel Ha - Delivering 2801 materials and assessing content difficulty
- Samir Mustavi - Managing the event; delivering 2901 materials
- Rumi Salazar - Correctedness of the slides; delivering 2901 materials

Best of luck in your studies (and possible statistics career)!