AUA CS 108, Statistics, Fall 2019

Lecture 35

Michael Poghosyan
YSU, AUA
michael@ysu.am, mpoghosyan@aua.am

13 Nov 2019

Contents

Last Lecture ReCap

- Give the definition of the Asymptotic CI.

Last Lecture ReCap

- Give the definition of the Asymptotic CI.
- Give the $(1-\alpha)$-level AsymptoCI for μ in general case.

Example

Example: There was an ancient civilization called Etruscan Civilization, living in the area corresponding to Italy, around 8th century BC.

Example

Example: There was an ancient civilization called Etruscan Civilization, living in the area corresponding to Italy, around 8th century BC. Scientists have a data about 84 skull sizes of adult Etruscans, giving in average 143.8mm, with a Sample Standard Deviation 5.97mm.

Example

Example: There was an ancient civilization called Etruscan Civilization, living in the area corresponding to Italy, around 8th century BC. Scientists have a data about 84 skull sizes of adult Etruscans, giving in average 143.8mm, with a Sample Standard Deviation 5.97mm.

Scientists believe Etruscans were Italians.

Example

Example: There was an ancient civilization called Etruscan Civilization, living in the area corresponding to Italy, around 8th century BC. Scientists have a data about 84 skull sizes of adult Etruscans, giving in average 143.8mm, with a Sample Standard Deviation 5.97mm.

Scientists believe Etruscans were Italians. They believe that the skull size is not changing much through time.

Example

Example: There was an ancient civilization called Etruscan Civilization, living in the area corresponding to Italy, around 8th century BC. Scientists have a data about 84 skull sizes of adult Etruscans, giving in average 143.8mm, with a Sample Standard Deviation 5.97mm.

Scientists believe Etruscans were Italians. They believe that the skull size is not changing much through time. Nowadays, adult Italians skull size is in average 132.4mm.

Example

Example: There was an ancient civilization called Etruscan Civilization, living in the area corresponding to Italy, around 8th century BC. Scientists have a data about 84 skull sizes of adult Etruscans, giving in average 143.8mm, with a Sample Standard Deviation 5.97mm.

Scientists believe Etruscans were Italians. They believe that the skull size is not changing much through time. Nowadays, adult Italians skull size is in average 132.4mm.

We want to construct a 95% AsymptoCI to see if it is supporting scientists hypothesis.

Example

Example: There was an ancient civilization called Etruscan Civilization, living in the area corresponding to Italy, around 8th century BC. Scientists have a data about 84 skull sizes of adult Etruscans, giving in average 143.8mm, with a Sample Standard Deviation 5.97mm.

Scientists believe Etruscans were Italians. They believe that the skull size is not changing much through time. Nowadays, adult Italians skull size is in average 132.4mm.

We want to construct a 95% AsymptoCI to see if it is supporting scientists hypothesis.

We model our problem like this: we assume the skull sizes of Italians are coming from some Distribution with some Mean μ and Variance σ^2, σ^2 is unknown.

Example, Cont'd

If we believe that Etruscans are Italians, then we have a Sample from that Distrib:

$$X_1, X_2, ..., X_{84}.$$

where X_k is the skull size of k-th Etruscan person.

Example, Cont'd

If we believe that Etruscans are Italians, then we have a Sample from that Distrib:

$$X_1, X_2, ..., X_{84}.$$

where X_k is the skull size of k-th Etruscan person. We have that $n = 84$, $\overline{X} = 143.8$ and $S = 5.97$.

Example, Cont'd

If we believe that Etruscans are Italians, then we have a Sample from that Distrib:

$$X_1, X_2, ..., X_{84}.$$

where X_k is the skull size of k-th Etruscan person. We have that $n = 84$, $\overline{X} = 143.8$ and $S = 5.97$.

Then, we know how to construct an Asympto 95% CI for μ:

$$\overline{X} \pm t_{n-1,1-\alpha/2} \cdot \frac{S}{\sqrt{n}}$$

Example, Cont'd

If we believe that Etruscans are Italians, then we have a Sample from that Distrib:

$$X_1, X_2, ..., X_{84}.$$

where X_k is the skull size of k-th Etruscan person. We have that $n = 84$, $\overline{X} = 143.8$ and $S = 5.97$.

Then, we know how to construct an Asympto 95% CI for μ:

$$\overline{X} \pm t_{n-1,1-\alpha/2} \cdot \frac{S}{\sqrt{n}}$$

giving us

```
x.bar <- 143.8; s <- 5.97; n <- 84
a <- 0.05; t <- qt(1-a/2, df = n-1)
me <- t*s/sqrt(n)
c(x.bar - me, x.bar +me)
```

```
## [1] 142.5044 145.0956
```

Example, Cont'd

If we believe that Etruscans are Italians, then we have a Sample from that Distrib:

$$X_1, X_2, ..., X_{84}.$$

where X_k is the skull size of k-th Etruscan person. We have that $n = 84$, $\overline{X} = 143.8$ and $S = 5.97$.

Then, we know how to construct an Asympto 95% CI for μ:

$$\overline{X} \pm t_{n-1,1-\alpha/2} \cdot \frac{S}{\sqrt{n}}$$

giving us

```
x.bar <- 143.8; s <- 5.97; n <- 84
a <- 0.05; t <- qt(1-a/2, df = n-1)
me <- t*s/sqrt(n)
c(x.bar - me, x.bar +me)
```

```
## [1] 142.5044 145.0956
```

We know the true value of μ, it is $\mu = 132.4$.

Example, Cont'd

If we believe that Etruscans are Italians, then we have a Sample from that Distrib:

$$X_1, X_2, ..., X_{84}.$$

where X_k is the skull size of k-th Etruscan person. We have that $n = 84$, $\overline{X} = 143.8$ and $S = 5.97$.

Then, we know how to construct an Asympto 95% CI for μ:

$$\overline{X} \pm t_{n-1,1-\alpha/2} \cdot \frac{S}{\sqrt{n}}$$

giving us

```
x.bar <- 143.8; s <- 5.97; n <- 84
a <- 0.05; t <- qt(1-a/2, df = n-1)
me <- t*s/sqrt(n)
c(x.bar - me, x.bar +me)
```

```
## [1] 142.5044 145.0956
```

We know the true value of μ, it is $\mu = 132.4$. So we can deduce that ...

Example, Cont'd

If we believe that Etruscans are Italians, then we have a Sample from that Distrib:

$$X_1, X_2, ..., X_{84}.$$

where X_k is the skull size of k-th Etruscan person. We have that $n = 84$, $\overline{X} = 143.8$ and $S = 5.97$.

Then, we know how to construct an Asympto 95% CI for μ:

$$\overline{X} \pm t_{n-1,1-\alpha/2} \cdot \frac{S}{\sqrt{n}}$$

giving us

```
x.bar <- 143.8; s <- 5.97; n <- 84
a <- 0.05; t <- qt(1-a/2, df = n-1)
me <- t*s/sqrt(n)
c(x.bar - me, x.bar +me)
```

```
## [1] 142.5044 145.0956
```

We know the true value of μ, it is $\mu = 132.4$. So we can deduce that ... they were looong skul Italians $\ddot\smile$

Asymptotic CI for the general case, using MLE

Using the Asymptotic Normality of the MLE, we can obtain a $(1-\alpha)$-level Asymptotic CI for the general case.

Asymptotic CI for the general case, using MLE

Using the Asymptotic Normality of the MLE, we can obtain a $(1-\alpha)$-level Asymptotic CI for the general case.

Assume we have a (regular) Parametric Model $\mathcal{F}_\theta$, and a Random Sample from it

$$X_1, X_2, ..., X_n, ... \sim \mathcal{F}_\theta.$$

Asymptotic CI for the general case, using MLE

Using the Asymptotic Normality of the MLE, we can obtain a $(1-\alpha)$-level Asymptotic CI for the general case.

Assume we have a (regular) Parametric Model $\mathcal{F}_\theta$, and a Random Sample from it

$$X_1, X_2, ..., X_n, ... \sim \mathcal{F}_\theta.$$

We want to construct a $(1-\alpha)$-level Asymptotic CI for θ.

Asymptotic CI for the general case, using MLE

Using the Asymptotic Normality of the MLE, we can obtain a $(1-\alpha)$-level Asymptotic CI for the general case.

Assume we have a (regular) Parametric Model $\mathcal{F}_\theta$, and a Random Sample from it

$$X_1, X_2, ..., X_n, ... \sim \mathcal{F}_\theta.$$

We want to construct a $(1-\alpha)$-level Asymptotic CI for θ.

Let $\hat{\theta}_n^{MLE}$ be the ML Estimator for θ.

Asymptotic CI for the general case, using MLE

Using the Asymptotic Normality of the MLE, we can obtain a $(1-\alpha)$-level Asymptotic CI for the general case.

Assume we have a (regular) Parametric Model $\mathcal{F}_\theta$, and a Random Sample from it

$$X_1, X_2, ..., X_n, ... \sim \mathcal{F}_\theta.$$

We want to construct a $(1-\alpha)$-level Asymptotic CI for θ.

Let $\hat{\theta}_n^{MLE}$ be the ML Estimator for θ. Then we know that

$$\frac{\hat{\theta}_n^{MLE} - \theta}{\sqrt{\frac{1}{n \cdot \mathcal{I}\left(\hat{\theta}_n^{MLE}\right)}}} \xrightarrow{D} \mathcal{N}(0,1)$$

Asymptotic CI for the general case, using MLE

This means that

$$\mathbb{P}\left(-z_{1-\alpha/2} < \frac{\hat{\theta}_n^{MLE} - \theta}{\sqrt{\frac{1}{n \cdot \mathcal{I}\left(\hat{\theta}_n^{MLE}\right)}}} < z_{1-\alpha/2}\right) \to 1 - \alpha$$

Asymptotic CI for the general case, using MLE

This means that

$$\mathbb{P}\left(-z_{1-\alpha/2} < \frac{\hat{\theta}_n^{MLE} - \theta}{\sqrt{\frac{1}{n \cdot \mathcal{I}\left(\hat{\theta}_n^{MLE}\right)}}} < z_{1-\alpha/2}\right) \to 1 - \alpha$$

giving us a $(1-\alpha)$-level Asymptotic CI for θ:

Asymptotic CI for the general case, using MLE

This means that

$$\mathbb{P}\left(-z_{1-\alpha/2} < \frac{\hat{\theta}_n^{MLE} - \theta}{\sqrt{\frac{1}{n \cdot \mathcal{I}\left(\hat{\theta}_n^{MLE}\right)}}} < z_{1-\alpha/2}\right) \to 1 - \alpha$$

giving us a $(1-\alpha)$-level Asymptotic CI for θ:

$$\left(\hat{\theta}_n^{MLE} - z_{1-\alpha/2} \cdot \sqrt{\frac{1}{n \cdot \mathcal{I}\left(\hat{\theta}_n^{MLE}\right)}};\;\; \hat{\theta}_n^{MLE} + z_{1-\alpha/2} \cdot \sqrt{\frac{1}{n \cdot \mathcal{I}\left(\hat{\theta}_n^{MLE}\right)}}\right).$$

Asymptotic CI for the Bernoulli Model

We have already constructed a CI for p in the *Bernoulli*(p) Model. Here we want to construct an Asymptotic CI for p, which usually gives smaller length interval.

Asymptotic CI for the Bernoulli Model

We have already constructed a CI for p in the $Bernoulli(p)$ Model. Here we want to construct an Asymptotic CI for p, which usually gives smaller length interval.

The Problem is the following: we have a Random Sample

$$X_1, X_2, ..., X_n, ... \sim Bernoulli(p),$$

and $\alpha \in (0, 1)$.

Asymptotic CI for the Bernoulli Model

We have already constructed a CI for p in the $Bernoulli(p)$ Model. Here we want to construct an Asymptotic CI for p, which usually gives smaller length interval.

The Problem is the following: we have a Random Sample

$$X_1, X_2, ..., X_n, ... \sim Bernoulli(p),$$

and $\alpha \in (0, 1)$. Our aim is to construct a $(1-\alpha)$-level Asymptotic CI for p.

Asymptotic CI for the Bernoulli Model

We have already constructed a CI for p in the *Bernoulli*(p) Model. Here we want to construct an Asymptotic CI for p, which usually gives smaller length interval.

The Problem is the following: we have a Random Sample

$$X_1, X_2, ..., X_n, ... \sim Bernoulli(p),$$

and $\alpha \in (0, 1)$. Our aim is to construct a $(1 - \alpha)$-level Asymptotic CI for p.

We know that $\hat{p}_n^{MLE} =$

Asymptotic CI for the Bernoulli Model

We have already constructed a CI for p in the $Bernoulli(p)$ Model. Here we want to construct an Asymptotic CI for p, which usually gives smaller length interval.

The Problem is the following: we have a Random Sample

$$X_1, X_2, ..., X_n, ... \sim Bernoulli(p),$$

and $\alpha \in (0, 1)$. Our aim is to construct a $(1-\alpha)$-level Asymptotic CI for p.

We know that $\hat{p}_n^{MLE} = \overline{X}_n$. Also, we have calculated the Fisher Information for the Bernoulli case: $\mathcal{I}(p) =$

Asymptotic CI for the Bernoulli Model

We have already constructed a CI for p in the *Bernoulli*(p) Model. Here we want to construct an Asymptotic CI for p, which usually gives smaller length interval.

The Problem is the following: we have a Random Sample

$$X_1, X_2, ..., X_n, ... \sim Bernoulli(p),$$

and $\alpha \in (0, 1)$. Our aim is to construct a $(1-\alpha)$-level Asymptotic CI for p.

We know that $\hat{p}_n^{MLE} = \overline{X}_n$. Also, we have calculated the Fisher Information for the Bernoulli case: $\mathcal{I}(p) = \dfrac{1}{p(1-p)}$.

Asymptotic CI for the Bernoulli Model

The above obtained CI is:

$$\left(\hat{p}_n^{MLE} - z_{1-\alpha/2} \cdot \sqrt{\frac{1}{n \cdot \mathcal{I}\left(\hat{p}_n^{MLE}\right)}};\ \ \hat{p}_n^{MLE} + z_{1-\alpha/2} \cdot \sqrt{\frac{1}{n \cdot \mathcal{I}\left(\hat{p}_n^{MLE}\right)}}\right).$$

Asymptotic CI for the Bernoulli Model

The above obtained CI is:

$$\left(\hat{p}_n^{MLE} - z_{1-\alpha/2} \cdot \sqrt{\frac{1}{n \cdot \mathcal{I}\left(\hat{p}_n^{MLE}\right)}}; \quad \hat{p}_n^{MLE} + z_{1-\alpha/2} \cdot \sqrt{\frac{1}{n \cdot \mathcal{I}\left(\hat{p}_n^{MLE}\right)}}\right).$$

By plugging the values for our case, we'll obtain the following Asymptotic CI of level $(1-\alpha)$ for p:

$$\left(\overline{X}_n - z_{1-\alpha/2} \cdot \sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}}; \quad \overline{X}_n + z_{1-\alpha/2} \cdot \sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}}\right).$$

Example, Asymptotic CI for Poisson Model

Example: We consider the Poisson Model $Pois(\lambda)$, and a Random Sample from it:

$$X_1, X_2, ..., X_n, ... \sim Pois(\lambda).$$

Example, Asymptotic CI for Poisson Model

Example: We consider the Poisson Model $Pois(\lambda)$, and a Random Sample from it:

$$X_1, X_2, ..., X_n, ... \sim Pois(\lambda).$$

Taking an $\alpha \in (0, 1)$, we want to construct a $(1 - \alpha)$-level Asymptotic CI for λ.

Example, Asymptotic CI for Poisson Model

Example: We consider the Poisson Model $Pois(\lambda)$, and a Random Sample from it:

$$X_1, X_2, ..., X_n, ... \sim Pois(\lambda).$$

Taking an $\alpha \in (0,1)$, we want to construct a $(1-\alpha)$-level Asymptotic CI for λ.

We know that $\hat{\lambda}_n^{MLE} =$

Example, Asymptotic CI for Poisson Model

Example: We consider the Poisson Model $Pois(\lambda)$, and a Random Sample from it:

$$X_1, X_2, ..., X_n, ... \sim Pois(\lambda).$$

Taking an $\alpha \in (0,1)$, we want to construct a $(1-\alpha)$-level Asymptotic CI for λ.

We know that $\hat{\lambda}_n^{MLE} = \overline{X}_n$, and also, $\mathcal{I}(\lambda) =$

Example, Asymptotic CI for Poisson Model

Example: We consider the Poisson Model $Pois(\lambda)$, and a Random Sample from it:

$$X_1, X_2, ..., X_n, ... \sim Pois(\lambda).$$

Taking an $\alpha \in (0,1)$, we want to construct a $(1-\alpha)$-level Asymptotic CI for λ.

We know that $\hat{\lambda}_n^{MLE} = \overline{X}_n$, and also, $\mathcal{I}(\lambda) = \dfrac{1}{\lambda}$.

Example, Asymptotic CI for Poisson Model

Example: We consider the Poisson Model $Pois(\lambda)$, and a Random Sample from it:

$$X_1, X_2, ..., X_n, ... \sim Pois(\lambda).$$

Taking an $\alpha \in (0, 1)$, we want to construct a $(1-\alpha)$-level Asymptotic CI for λ.

We know that $\hat{\lambda}_n^{MLE} = \overline{X}_n$, and also, $\mathcal{I}(\lambda) = \dfrac{1}{\lambda}$. Then, using the formula above:

$$\left(\hat{\lambda}_n^{MLE} - z_{1-\alpha/2} \cdot \sqrt{\frac{1}{n \cdot \mathcal{I}\left(\hat{\lambda}_n^{MLE}\right)}};\ \ \hat{\lambda}_n^{MLE} + z_{1-\alpha/2} \cdot \sqrt{\frac{1}{n \cdot \mathcal{I}\left(\hat{\lambda}_n^{MLE}\right)}}\right),$$

Example, Asymptotic CI for Poisson Model

Example: We consider the Poisson Model $Pois(\lambda)$, and a Random Sample from it:

$$X_1, X_2, ..., X_n, ... \sim Pois(\lambda).$$

Taking an $\alpha \in (0,1)$, we want to construct a $(1-\alpha)$-level Asymptotic CI for λ.

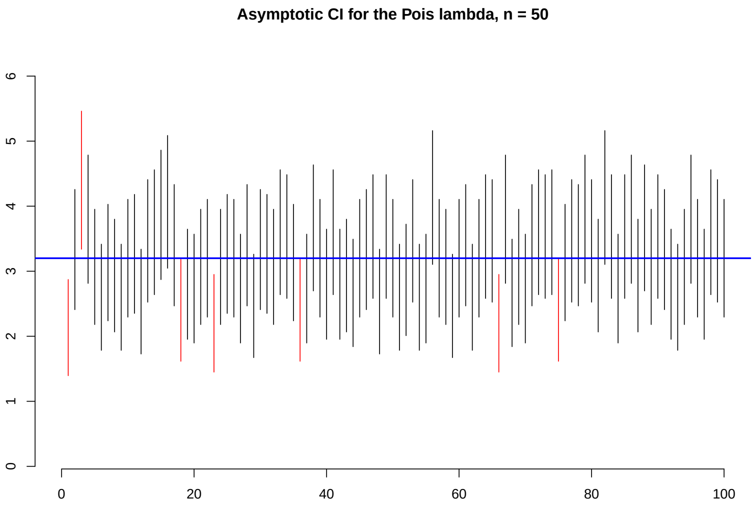
We know that $\hat{\lambda}_n^{MLE} = \overline{X}_n$, and also, $\mathcal{I}(\lambda) = \dfrac{1}{\lambda}$. Then, using the formula above:

$$\left(\hat{\lambda}_n^{MLE} - z_{1-\alpha/2} \cdot \sqrt{\frac{1}{n \cdot \mathcal{I}\left(\hat{\lambda}_n^{MLE}\right)}};\ \ \hat{\lambda}_n^{MLE} + z_{1-\alpha/2} \cdot \sqrt{\frac{1}{n \cdot \mathcal{I}\left(\hat{\lambda}_n^{MLE}\right)}}\right),$$

we'll obtain the following Asymptotic CI of level $(1-\alpha)$ for λ:

$$\left(\overline{X}_n - z_{1-\alpha/2} \cdot \sqrt{\frac{\overline{X}_n}{n}};\ \ \overline{X}_n + z_{1-\alpha/2} \cdot \sqrt{\frac{\overline{X}_n}{n}}\right).$$

Example, in **R**



Asymptotic CI for the Pois lambda, n = 50
0
1
2
3
4
5
6
0
20
40
60
80
100

Example, in **R**, Code

```
lambda <- 3.2
conf.level <- 0.95; a = 1 - conf.level
sample.size <- 15; no.of.intervals <- 100
z <- qnorm(1-a/2)

plot.new()
plot.window(xlim=c(0,no.of.intervals),ylim=c(lambda-3,lambda+3))
axis(1); axis(2)
title("Asymptotic CI for the Pois lambda, n = 50")
for(i in 1:no.of.intervals){
  x <- rpois(sample.size, lambda = lambda)
  ME <- z*sqrt(mean(x)/sample.size) #Margin of Error
  lo <- mean(x) - ME;  up <- mean(x) + ME
  if(lo > lambda || up < lambda){
    segments(c(i), c(lo), c(i), c(up), col = "red")
  }
  else{
    segments(c(i), c(lo), c(i), c(up))
  }
}
abline(h = lambda, lwd = 2, col = "blue")
```

Hypothesis Testing

RF

Sorry, no translation:

Экзамен, студентка валится безвозвратно. За дверью стоит толпа и думает, как ее выручить. Наконец в аудиторию врывается парень и кричит: — Иванова, у тебя сын родился! Препод ее, естественно, поздравляет, ставит оценку, расписывается.

Hypothesis Testing, Intro

Up to this point, we have considered the Point Estimation and CI construction Problems.

Hypothesis Testing, Intro

Up to this point, we have considered the Point Estimation and CI construction Problems. Here we will talk about another, very important Problem in Statistics, Hypothesis Testing.

Hypothesis Testing, Intro

Up to this point, we have considered the Point Estimation and CI construction Problems. Here we will talk about another, very important Problem in Statistics, Hypothesis Testing.

As always, we assume we have a Dataset coming as a realization of a Random Sample from some unknown Parametric Distribution $\mathcal{F}_\theta$:

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta.$$

In this case we want to Test a Hypothesis about θ: say, see whether $\theta = \theta_0$, a given number, or not.

Example

Example: We have a coin, and we want to see if it is fair or not.

Example

Example: We have a coin, and we want to see if it is fair or not.
We formulate a Hypothesis: coin is fair.

Example

Example: We have a coin, and we want to see if it is fair or not. We formulate a Hypothesis: coin is fair. To test the hypothesis, we toss that coin 100 times and obtain 56 Heads. Is this supporting our Hypothesis or not?

Example

Example: We have a coin, and we want to see if it is fair or not. We formulate a Hypothesis: coin is fair. To test the hypothesis, we toss that coin 100 times and obtain 56 Heads. Is this supporting our Hypothesis or not?

The Idea The idea is the following: even if our coin was fair, the Probability of Heads $p = 0.5$, it is possible to have some deviation from the expected number of Heads, 50 (in 100 tosses).

Example

Example: We have a coin, and we want to see if it is fair or not. We formulate a Hypothesis: coin is fair. To test the hypothesis, we toss that coin 100 times and obtain 56 Heads. Is this supporting our Hypothesis or not?

The Idea The idea is the following: even if our coin was fair, the Probability of Heads $p = 0.5$, it is possible to have some deviation from the expected number of Heads, 50 (in 100 tosses). The question is: how extreme, how unbelievable is that deviation.

Example

Example: We have a coin, and we want to see if it is fair or not. We formulate a Hypothesis: coin is fair. To test the hypothesis, we toss that coin 100 times and obtain 56 Heads. Is this supporting our Hypothesis or not?

The Idea The idea is the following: even if our coin was fair, the Probability of Heads $p = 0.5$, it is possible to have some deviation from the expected number of Heads, 50 (in 100 tosses). The question is: how extreme, how unbelievable is that deviation. If the deviation from the expected is large, we will not believe that the Hypothesis is True. Otherwise, we will keep believing that it is True.

Hypothesis Testing: Problem Setting and Formalization

So we have a Parametric Model $\mathcal{F}_\theta$, with $\theta \in \Theta$.

Hypothesis Testing: Problem Setting and Formalization

So we have a Parametric Model $\mathcal{F}_\theta$, with $\theta \in \Theta$. We consider a Partition of Θ into Θ_0 and Θ_1:

$$\Theta = \Theta_0 \cup \Theta_1 \qquad \textit{and} \qquad \Theta_0 \cap \Theta_1 = \varnothing.$$

Hypothesis Testing: Problem Setting and Formalization

So we have a Parametric Model $\mathcal{F}_\theta$, with $\theta \in \Theta$. We consider a Partition of Θ into Θ_0 and Θ_1:

$$\Theta = \Theta_0 \cup \Theta_1 \qquad \textit{and} \qquad \Theta_0 \cap \Theta_1 = \varnothing.$$

Now we consider a Hypothesis:

$$\mathcal{H}_0 : \theta \in \Theta_0 \qquad \textit{vs} \qquad \mathcal{H}_1 : \theta \in \Theta_1.$$

Hypothesis Testing: Problem Setting and Formalization

So we have a Parametric Model $\mathcal{F}_\theta$, with $\theta \in \Theta$. We consider a Partition of Θ into Θ_0 and Θ_1:

$$\Theta = \Theta_0 \cup \Theta_1 \qquad \textit{and} \qquad \Theta_0 \cap \Theta_1 = \varnothing.$$

Now we consider a Hypothesis:

$$\mathcal{H}_0 : \theta \in \Theta_0 \qquad \textit{vs} \qquad \mathcal{H}_1 : \theta \in \Theta_1.$$

$\mathcal{H}_0$ is called the **Null Hypothesis**, and $\mathcal{H}_1$ (or, sometimes, $\mathcal{H}_a$) is the **Alternative Hypothesis**.

Hypothesis Testing: Problem Setting and Formalization

So we have a Parametric Model $\mathcal{F}_\theta$, with $\theta \in \Theta$. We consider a Partition of Θ into Θ_0 and Θ_1:

$$\Theta = \Theta_0 \cup \Theta_1 \qquad \textit{and} \qquad \Theta_0 \cap \Theta_1 = \varnothing.$$

Now we consider a Hypothesis:

$$\mathcal{H}_0 : \theta \in \Theta_0 \qquad \textit{vs} \qquad \mathcal{H}_1 : \theta \in \Theta_1.$$

$\mathcal{H}_0$ is called the **Null Hypothesis**, and $\mathcal{H}_1$ (or, sometimes, $\mathcal{H}_a$) is the **Alternative Hypothesis**.

Next, we have a Random Sample from $\mathcal{F}_\theta$:

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta, \qquad \theta \in \Theta,$$

and using this Sample, we want to Test if we can **Reject $\mathcal{H}_0$ in favor of $\mathcal{H}_1$ or not**, i.e., we want to see if **we have enough evidence in our Data to Reject $\mathcal{H}_0$**.

Example

Example: In the above example about the coin fairness, if p is the Probability of a Head, then our Hypotheses are:

$$\mathcal{H}_0: \ p = \frac{1}{2} \qquad vs \qquad \mathcal{H}_1: \ p \neq \frac{1}{2}.$$

Example

Example: In the above example about the coin fairness, if p is the Probability of a Head, then our Hypotheses are:

$$\mathcal{H}_0: \; p = \frac{1}{2} \qquad vs \qquad \mathcal{H}_1: \; p \neq \frac{1}{2}.$$

And here $\Theta = [0,1]$, $\Theta_0 = \{0.5\}$ and $\Theta_1 = \Theta \setminus \Theta_0$.

Example

Example: Assume that the Mean Stat grade for my courses is 71.

Example

Example: Assume that the Mean Stat grade for my courses is 71. This year I am using a new Methodology: I am lecturing using slides, working hard on homework, using **R** a lot.

Example

Example: Assume that the Mean Stat grade for my courses is 71. This year I am using a new Methodology: I am lecturing using slides, working hard on homework, using **R** a lot. And I want to see if this affects positively on the average grade or not.

Example

Example: Assume that the Mean Stat grade for my courses is 71. This year I am using a new Methodology: I am lecturing using slides, working hard on homework, using **R** a lot. And I want to see if this affects positively on the average grade or not. With high Probability, our average grade will not be exactly 71.

Example

Example: Assume that the Mean Stat grade for my courses is 71. This year I am using a new Methodology: I am lecturing using slides, working hard on homework, using **R** a lot. And I want to see if this affects positively on the average grade or not. With high Probability, our average grade will not be exactly 71. Say, it will be 73.5.

Example

Example: Assume that the Mean Stat grade for my courses is 71. This year I am using a new Methodology: I am lecturing using slides, working hard on homework, using **R** a lot. And I want to see if this affects positively on the average grade or not. With high Probability, our average grade will not be exactly 71. Say, it will be 73.5. Will this be a sign that my new approach is working, or that will be just because of randomness?

Example

Example: Assume that the Mean Stat grade for my courses is 71. This year I am using a new Methodology: I am lecturing using slides, working hard on homework, using **R** a lot. And I want to see if this affects positively on the average grade or not. With high Probability, our average grade will not be exactly 71. Say, it will be 73.5. Will this be a sign that my new approach is working, or that will be just because of randomness? So we want to see if the **difference is significant or not**.

Example

Example: Assume that the Mean Stat grade for my courses is 71. This year I am using a new Methodology: I am lecturing using slides, working hard on homework, using **R** a lot. And I want to see if this affects positively on the average grade or not. With high Probability, our average grade will not be exactly 71. Say, it will be 73.5. Will this be a sign that my new approach is working, or that will be just because of randomness? So we want to see if the **difference is significant or not**.

To state the Hypotheses, let us denote the Average Stat Grade by μ.

Example

Example: Assume that the Mean Stat grade for my courses is 71. This year I am using a new Methodology: I am lecturing using slides, working hard on homework, using **R** a lot. And I want to see if this affects positively on the average grade or not. With high Probability, our average grade will not be exactly 71. Say, it will be 73.5. Will this be a sign that my new approach is working, or that will be just because of randomness? So we want to see if the **difference is significant or not**.

To state the Hypotheses, let us denote the Average Stat Grade by μ. Then our Hypotheses will be:

$$\mathcal{H}_0: \ \mu = 71 \qquad \textit{vs} \qquad \mathcal{H}_1: \ \mu > 71.$$

Example

Example: Assume that the Mean Stat grade for my courses is 71. This year I am using a new Methodology: I am lecturing using slides, working hard on homework, using **R** a lot. And I want to see if this affects positively on the average grade or not. With high Probability, our average grade will not be exactly 71. Say, it will be 73.5. Will this be a sign that my new approach is working, or that will be just because of randomness? So we want to see if the **difference is significant or not**.

To state the Hypotheses, let us denote the Average Stat Grade by μ. Then our Hypotheses will be:

$$\mathcal{H}_0: \ \mu = 71 \qquad \textit{vs} \qquad \mathcal{H}_1: \ \mu > 71.$$

So, when we will have our Data, we will see if we can **Reject** $\mathcal{H}_0$.

Example

Example: Assume that the Mean Stat grade for my courses is 71. This year I am using a new Methodology: I am lecturing using slides, working hard on homework, using **R** a lot. And I want to see if this affects positively on the average grade or not. With high Probability, our average grade will not be exactly 71. Say, it will be 73.5. Will this be a sign that my new approach is working, or that will be just because of randomness? So we want to see if the **difference is significant or not**.

To state the Hypotheses, let us denote the Average Stat Grade by μ. Then our Hypotheses will be:

$$\mathcal{H}_0: \ \mu = 71 \qquad \textit{vs} \qquad \mathcal{H}_1: \ \mu > 71.$$

So, when we will have our Data, we will see if we can **Reject** $\mathcal{H}_0$. Btw, hopefully, we will Reject, and not because of my new methodology, but **because of you**.

Notes

Some Notes:

Note: Usually, Θ_0 consists of just one point, so our Null Hypothesis is of the form

$$\mathcal{H}_0: \ \theta = \theta_0.$$

Notes

Some Notes:

Note: Usually, Θ_0 consists of just one point, so our Null Hypothesis is of the form

$$\mathcal{H}_0 : \theta = \theta_0.$$

In this case it is called a *Simple Hypothesis*.

Note: Continuing the previous Note, usually we consider three type of Alternative Hypotheses:

- $\mathcal{H}_1 : \theta \neq \theta_0$ (Two-Tailed, Two-Sided Test)

Notes

Some Notes:

Note: Usually, Θ_0 consists of just one point, so our Null Hypothesis is of the form

$$\mathcal{H}_0: \ \theta = \theta_0.$$

In this case it is called a *Simple Hypothesis*.

Note: Continuing the previous Note, usually we consider three type of Alternative Hypotheses:

- $\mathcal{H}_1: \ \theta \neq \theta_0$ (Two-Tailed, Two-Sided Test)
- $\mathcal{H}_1: \ \theta > \theta_0$ (Upper-Tailed Test)

Notes

Some Notes:

Note: Usually, Θ_0 consists of just one point, so our Null Hypothesis is of the form

$$\mathcal{H}_0 : \ \theta = \theta_0.$$

In this case it is called a *Simple Hypothesis*.

Note: Continuing the previous Note, usually we consider three type of Alternative Hypotheses:

- $\mathcal{H}_1 : \ \theta \neq \theta_0$ (Two-Tailed, Two-Sided Test)
- $\mathcal{H}_1 : \ \theta > \theta_0$ (Upper-Tailed Test)
- $\mathcal{H}_1 : \ \theta < \theta_0$ (Lower-Tailed Test)

Notes

Note: In Hypothesis Testing, our Strategy is, based on Data, to **try to Reject** $\mathcal{H}_0$.

Notes

Note: In Hypothesis Testing, our Strategy is, based on Data, to **try to Reject** $\mathcal{H}_0$. If our Data contains enough evidence against $\mathcal{H}_0$, we are Rejecting it.

Notes

Note: In Hypothesis Testing, our Strategy is, based on Data, to **try to Reject** $\mathcal{H}_0$. If our Data contains enough evidence against $\mathcal{H}_0$, we are Rejecting it. **Otherwise, we are Failing to Reject** $\mathcal{H}_0$, and we are not talking about we are accepting $\mathcal{H}_1$.

Notes

Note: In Hypothesis Testing, our Strategy is, based on Data, to **try to Reject** $\mathcal{H}_0$. If our Data contains enough evidence against $\mathcal{H}_0$, we are Rejecting it. **Otherwise, we are Failing to Reject** $\mathcal{H}_0$, and we are not talking about we are accepting $\mathcal{H}_1$.

So the conclusion of the Hypothesis Testing need to be either:

Reject $\mathcal{H}_0$ or **Fail to Reject** $\mathcal{H}_0$.

Notes

At this point, Statisticians always compare this with the process in the court:

Notes

At this point, Statisticians always compare this with the process in the court: if the person is suspected, say, in murder, then, by default, judges assume that he is **not guilty** (anmeghutyan kankhavarkac $\ddot\smile$).

Notes

At this point, Statisticians always compare this with the process in the court: if the person is suspected, say, in murder, then, by default, judges assume that he is **not guilty** (anmeghutyan kankhavarkac $\ddot\smile$). This is our Null Hypothesis:

$$\mathcal{H}_0: \text{ He is not guilty} \quad \text{vs} \quad \mathcal{H}_1: \text{ He is guilty.}$$

Notes

At this point, Statisticians always compare this with the process in the court: if the person is suspected, say, in murder, then, by default, judges assume that he is **not guilty** (anmeghutyan kankhavarkac $\ddot\smile$). This is our Null Hypothesis:

$$\mathcal{H}_0: \text{ He is not guilty} \quad \text{vs} \quad \mathcal{H}_1: \text{ He is guilty.}$$

Now, based on the evidence, Data, the judges either will Reject $\mathcal{H}_0$, meaning that, upon the light of a lot of facts, they will decide that our person is guilty,

Notes

At this point, Statisticians always compare this with the process in the court: if the person is suspected, say, in murder, then, by default, judges assume that he is **not guilty** (anmeghutyan kankhavarkac $\ddot\smile$). This is our Null Hypothesis:

$$\mathcal{H}_0: \text{ He is not guilty} \quad \text{vs} \quad \mathcal{H}_1: \text{ He is guilty.}$$

Now, based on the evidence, Data, the judges either will Reject $\mathcal{H}_0$, meaning that, upon the light of a lot of facts, they will decide that our person is guilty, or, if the Data will not be enough, the judges will Not Reject $\mathcal{H}_0$, so they will free our person.

Notes

At this point, Statisticians always compare this with the process in the court: if the person is suspected, say, in murder, then, by default, judges assume that he is **not guilty** (anmeghutyan kankhavarkac $\ddot\smile$). This is our Null Hypothesis:

$$\mathcal{H}_0: \text{ He is not guilty} \quad \text{vs} \quad \mathcal{H}_1: \text{ He is guilty.}$$

Now, based on the evidence, Data, the judges either will Reject $\mathcal{H}_0$, meaning that, upon the light of a lot of facts, they will decide that our person is guilty, or, if the Data will not be enough, the judges will Not Reject $\mathcal{H}_0$, so they will free our person. **But this does not mean that the murder was not of our person's hands affair!**

Notes

At this point, Statisticians always compare this with the process in the court: if the person is suspected, say, in murder, then, by default, judges assume that he is **not guilty** (anmeghutyan kankhavarkac $\ddot{\smile}$). This is our Null Hypothesis:

$$\mathcal{H}_0 : \text{ He is not guilty} \quad \text{vs} \quad \mathcal{H}_1 : \text{ He is guilty.}$$

Now, based on the evidence, Data, the judges either will Reject $\mathcal{H}_0$, meaning that, upon the light of a lot of facts, they will decide that our person is guilty, or, if the Data will not be enough, the judges will Not Reject $\mathcal{H}_0$, so they will free our person. **But this does not mean that the murder was not of our person's hands affair!** This just means that the **Data was not showing strong Evidence against that he is not guilty** - maybe this is because we do not have enough Data.

Notes

At this point, Statisticians always compare this with the process in the court: if the person is suspected, say, in murder, then, by default, judges assume that he is **not guilty** (anmeghutyan kankhavarkac $\ddot{\smile}$). This is our Null Hypothesis:

$$\mathcal{H}_0: \text{ He is not guilty} \quad \text{vs} \quad \mathcal{H}_1: \text{ He is guilty.}$$

Now, based on the evidence, Data, the judges either will Reject $\mathcal{H}_0$, meaning that, upon the light of a lot of facts, they will decide that our person is guilty, or, if the Data will not be enough, the judges will Not Reject $\mathcal{H}_0$, so they will free our person. **But this does not mean that the murder was not of our person's hands affair!** This just means that the **Data was not showing strong Evidence against that he is not guilty** - maybe this is because we do not have enough Data.