ASDS Statistics, YSU, Fall 2020

Lecture 26

Michael Poghosyan

02 Dec 2020

Contents

Some Notes about MLE

Note: The MLE does not always exist;

Some Notes about MLE

Note: The MLE does not always exist;

Example: Consider the Model $Unif(a, b)$, i.e., Uniform over the open inteval (a, b).

Some Notes about MLE

Note: The MLE does not always exist;

Example: Consider the Model $Unif(a, b)$, i.e., Uniform over the open inteval (a, b).

Example: Consider the Model with PDF

$$f(x|\theta) = \begin{cases} e^{\theta - x}, & x > \theta \\ 0, & x \leq \theta. \end{cases}$$

Some Notes about MLE

Note: The MLE does not always exist;

Example: Consider the Model $Unif(a, b)$, i.e., Uniform over the open inteval (a, b).

Example: Consider the Model with PDF

$$f(x|\theta) = \begin{cases} e^{\theta - x}, & x > \theta \\ 0, & x \leq \theta. \end{cases}$$

Note: The MLE is not always unique;

Some Notes about MLE

Note: The MLE does not always exist;

Example: Consider the Model $Unif(a, b)$, i.e., Uniform over the open inteval (a, b).

Example: Consider the Model with PDF

$$f(x|\theta) = \begin{cases} e^{\theta - x}, & x > \theta \\ 0, & x \leq \theta. \end{cases}$$

Note: The MLE is not always unique;

Example: Consider the Model $Unif[\theta, \theta + 1]$.

Some Notes about MLE

Note: The MLE does not always exist;

Example: Consider the Model $Unif(a, b)$, i.e., Uniform over the open inteval (a, b).

Example: Consider the Model with PDF

$$f(x|\theta) = \begin{cases} e^{\theta - x}, & x > \theta \\ 0, & x \leq \theta. \end{cases}$$

Note: The MLE is not always unique;

Example: Consider the Model $Unif[\theta, \theta + 1]$.

Note: Sometimes it is necessary to find it numerically, say, using the NR Method

Examples

Example: It is not possible to find the MLE Estimator in a closed form for θ in the one-Parametric Cauchy Distribution $Cauchy(\theta)$ Model. Here, the PDF of $X \sim Cauchy(\theta)$ is given by

$$f(x|\theta) = \frac{1}{\pi(1 + (x - \theta)^2)}, \qquad x \in \mathbb{R},$$

and $\theta \in \mathbb{R}$ is called the *location parameter*.

Properties of the MLE

It can be proven that, under some regularity conditions on the Parametric Family $\mathcal{F}_\theta$,

- The MLE $\hat{\theta}_n^{MLE}$ is Consistent, i.e.,

Properties of the MLE

It can be proven that, under some regularity conditions on the Parametric Family $\mathcal{F}_\theta$,

- The MLE $\hat{\theta}_n^{MLE}$ is Consistent, i.e.,

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

Properties of the MLE

It can be proven that, under some regularity conditions on the Parametric Family $\mathcal{F}_\theta$,

- The MLE $\hat{\theta}_n^{MLE}$ is Consistent, i.e.,

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

- The MLE $\hat{\theta}_n^{MLE}$ is Asymptotically Normal and Efficient:

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

Properties of the MLE

It can be proven that, under some regularity conditions on the Parametric Family $\mathcal{F}_\theta$,

- The MLE $\hat{\theta}_n^{MLE}$ is Consistent, i.e.,

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

- The MLE $\hat{\theta}_n^{MLE}$ is Asymptotically Normal and Efficient:

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

  or, as people usually use:

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

Properties of the MLE

It can be proven that, under some regularity conditions on the Parametric Family $\mathcal{F}_\theta$,

- The MLE $\hat{\theta}_n^{MLE}$ is Consistent, i.e.,

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

- The MLE $\hat{\theta}_n^{MLE}$ is Asymptotically Normal and Efficient:

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

  or, as people usually use:

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

  or, put in other way,

$$\hat{\theta}_n^{MLE} \overset{D}{\approx} \mathcal{N}\left(\theta, \frac{1}{n \cdot \mathcal{I}(\theta)}\right)$$

Properties of the MLE, Cont'd

Let us give an interpretation to this last property:

$$\hat{\theta}_n^{MLE} \overset{D}{\approx} \mathcal{N}\left(\theta, \frac{1}{n \cdot \mathcal{I}(\theta)}\right)$$

Properties of the MLE, Cont'd

Let us give an interpretation to this last property:

$$\hat{\theta}_n^{MLE} \overset{D}{\approx} \mathcal{N}\left(\theta, \frac{1}{n \cdot \mathcal{I}(\theta)}\right)$$

We can interpret this as:

- $\hat{\theta}_n^{MLE}$ is approximately (Asymptotically) Normal;

Properties of the MLE, Cont'd

Let us give an interpretation to this last property:

$$\hat{\theta}_n^{MLE} \overset{D}{\approx} \mathcal{N}\left(\theta, \frac{1}{n \cdot \mathcal{I}(\theta)}\right)$$

We can interpret this as:

- $\hat{\theta}_n^{MLE}$ is approximately (Asymptotically) Normal;
- The Mean of $\hat{\theta}_n^{MLE}$ is approximately θ, so it is Asymptotically Unbiased;

Properties of the MLE, Cont'd

Let us give an interpretation to this last property:

$$\hat{\theta}_n^{MLE} \overset{D}{\approx} \mathcal{N}\left(\theta, \frac{1}{n \cdot \mathcal{I}(\theta)}\right)$$

We can interpret this as:

- $\hat{\theta}_n^{MLE}$ is approximately (Asymptotically) Normal;
- The Mean of $\hat{\theta}_n^{MLE}$ is approximately θ, so it is Asymptotically Unbiased;
- The Variance of $\hat{\theta}_n^{MLE}$ is approximately $\dfrac{1}{n \cdot \mathcal{I}(\theta)}$. And we know that, by CR LB, this is the possible smallest Variance we can have with Unbiased Estimators.

Properties of the MLE, Cont'd

Let us give an interpretation to this last property:

$$\hat{\theta}_n^{MLE} \overset{D}{\approx} \mathcal{N}\left(\theta, \frac{1}{n \cdot \mathcal{I}(\theta)}\right)$$

We can interpret this as:

- $\hat{\theta}_n^{MLE}$ is approximately (Asymptotically) Normal;
- The Mean of $\hat{\theta}_n^{MLE}$ is approximately θ, so it is Asymptotically Unbiased;
- The Variance of $\hat{\theta}_n^{MLE}$ is approximately $\dfrac{1}{n \cdot \mathcal{I}(\theta)}$. And we know that, by CR LB, this is the possible smallest Variance we can have with Unbiased Estimators. So we interpret this as **MLE is Asymptotically Efficient**.

Properties of the MLE, Cont'd

Let us give an interpretation to this last property:

$$\hat{\theta}_n^{MLE} \overset{D}{\approx} \mathcal{N}\left(\theta, \frac{1}{n \cdot \mathcal{I}(\theta)}\right)$$

We can interpret this as:

- $\hat{\theta}_n^{MLE}$ is approximately (Asymptotically) Normal;
- The Mean of $\hat{\theta}_n^{MLE}$ is approximately θ, so it is Asymptotically Unbiased;
- The Variance of $\hat{\theta}_n^{MLE}$ is approximately $\dfrac{1}{n \cdot \mathcal{I}(\theta)}$. And we know that, by CR LB, this is the possible smallest Variance we can have with Unbiased Estimators. So we interpret this as **MLE is Asymptotically Efficient**.

So, MLE is **Consistent** and **Asymptotocally Efficient**.

Properties of the MLE, Cont'd

Let us give an interpretation to this last property:

$$\hat{\theta}_n^{MLE} \overset{D}{\approx} \mathcal{N}\left(\theta, \frac{1}{n \cdot \mathcal{I}(\theta)}\right)$$

We can interpret this as:

- $\hat{\theta}_n^{MLE}$ is approximately (Asymptotically) Normal;
- The Mean of $\hat{\theta}_n^{MLE}$ is approximately θ, so it is Asymptotically Unbiased;
- The Variance of $\hat{\theta}_n^{MLE}$ is approximately $\dfrac{1}{n \cdot \mathcal{I}(\theta)}$. And we know that, by CR LB, this is the possible smallest Variance we can have with Unbiased Estimators. So we interpret this as **MLE is Asymptotically Efficient**.

So, MLE is **Consistent** and **Asymptotocally Efficient**. And this is why, for large Sample Size n, MLE is the Top 1 Choice, is (almost) unbeatable.

Properties of the MLE, Cont'd

- Also,

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

Properties of the MLE, Cont'd

- Also,

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

or, in more non-rigorous terms,

$$\hat{\theta}_n^{MLE} \overset{D}{\approx} \mathcal{N}\left(\theta, \frac{1}{n \cdot \mathcal{I}\left(\hat{\theta}_n^{MLE}\right)}\right)$$

Properties of the MLE, Cont'd

- Also,

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

or, in more non-rigorous terms,

$$\hat{\theta}_n^{MLE} \overset{D}{\approx} \mathcal{N}\left(\theta, \frac{1}{n \cdot \mathcal{I}\left(\hat{\theta}_n^{MLE}\right)}\right)$$

Note: This is almost the above Property,

$$\hat{\theta}_n^{MLE} \overset{D}{\approx} \mathcal{N}\left(\theta, \frac{1}{n \cdot \mathcal{I}(\theta)}\right),$$

but, instead of $\mathcal{I}(\theta)$ we have $\mathcal{I}\left(\hat{\theta}_n^{MLE}\right)$.

Properties of the MLE, Cont'd

- Also,

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

  or, in more non-rigorous terms,

$$\hat{\theta}_n^{MLE} \overset{D}{\approx} \mathcal{N}\left(\theta, \frac{1}{n \cdot \mathcal{I}\left(\hat{\theta}_n^{MLE}\right)}\right)$$

Note: This is almost the above Property,

$$\hat{\theta}_n^{MLE} \overset{D}{\approx} \mathcal{N}\left(\theta, \frac{1}{n \cdot \mathcal{I}(\theta)}\right),$$

but, instead of $\mathcal{I}(\theta)$ we have $\mathcal{I}\left(\hat{\theta}_n^{MLE}\right)$.

Note: We will use this later, to construct an (approximate) Confidence Interval for θ and for testing Hypotheses about θ.

Properties of the MLE, Cont'd

- If $\hat{\theta}$ is the MLE for θ, then for any function g, the MLE of $g(\theta)$ is $g(\hat{\theta})$, i.e.,

$$\widehat{g(\theta)}^{MLE} = g\left(\hat{\theta}^{MLE}\right).$$

Properties of the MLE, Cont'd

- If $\hat{\theta}$ is the MLE for θ, then for any function g, the MLE of $g(\theta)$ is $g(\hat{\theta})$, i.e.,
$$\widehat{g(\theta)}^{MLE} = g\left(\hat{\theta}^{MLE}\right).$$

This is sometimes called the **invariance Principle of MLE**.

Properties of the MLE, Cont'd

- If $\hat{\theta}$ is the MLE for θ, then for any function g, the MLE of $g(\theta)$ is $g(\hat{\theta})$, i.e.,

$$\widehat{g(\theta)}^{MLE} = g\left(\hat{\theta}^{MLE}\right).$$

This is sometimes called the **invariance Principle of MLE**.

Example Find the MLE for σ in $\mathcal{N}(\mu, \sigma^2)$ Model.

Solution: OTB

Some topics to consider by yourself

- Multivariate Normal and MLE for MVNormal
- Kullback-Leibler Divergence and its relation to MLE
- MLE for the Mixture Model, EM Algorithm
- Bayesian Estimation: MAP and Bayes Estimator

Confidence Intervals

Prelude

So far we have studied the Parametric Point Estimation Problem: we were assuming that our data is coming from a Parametric Family of Distribution, and our aim was to find a best guess for that Parameter.

Prelude

So far we have studied the Parametric Point Estimation Problem: we were assuming that our data is coming from a Parametric Family of Distribution, and our aim was to find a best guess for that Parameter.

Assume θ is our Parameter to be Estimated, and $\hat{\theta}$ is a good Estimator for θ.

Prelude

So far we have studied the Parametric Point Estimation Problem: we were assuming that our data is coming from a Parametric Family of Distribution, and our aim was to find a best guess for that Parameter.

Assume θ is our Parameter to be Estimated, and $\hat{\theta}$ is a good Estimator for θ. Assume $\hat{\theta}$ is a Continuous r.v.

Prelude

So far we have studied the Parametric Point Estimation Problem: we were assuming that our data is coming from a Parametric Family of Distribution, and our aim was to find a best guess for that Parameter.

Assume θ is our Parameter to be Estimated, and $\hat{\theta}$ is a good Estimator for θ. Assume $\hat{\theta}$ is a Continuous r.v. If the True value of our Parameter is θ^*, then

$$\mathbb{P}(\hat{\theta} = \theta^*) = 0,$$

i.e., we will (almost) **never** be correct in our guess.

Prelude

So far we have studied the Parametric Point Estimation Problem: we were assuming that our data is coming from a Parametric Family of Distribution, and our aim was to find a best guess for that Parameter.

Assume θ is our Parameter to be Estimated, and $\hat{\theta}$ is a good Estimator for θ. Assume $\hat{\theta}$ is a Continuous r.v. If the True value of our Parameter is θ^*, then

$$\mathbb{P}(\hat{\theta} = \theta^*) = 0,$$

i.e., we will (almost) **never** be correct in our guess. Sad news!

Prelude No. 2

But the good news is that even when we cannot exactly find the True value of our Parameter using $\hat{\theta}$, if $\hat{\theta}$ possesses some good properties, we believe that the Estimate obtained is a good approximation/Estimate for θ^*.

[1] Recall the $\widehat{SE}$, the Estimated Standard Error reporting story.

Prelude No. 2

But the good news is that even when we cannot exactly find the True value of our Parameter using $\hat{\theta}$, if $\hat{\theta}$ possesses some good properties, we believe that the Estimate obtained is a good approximation/Estimate for θ^*.

In fact, we think as

$$\theta^* = \hat{\theta} \pm \textit{error},$$

where *error* is small[1],

[1] Recall the $\widehat{SE}$, the Estimated Standard Error reporting story.

Prelude No. 2

But the good news is that even when we cannot exactly find the True value of our Parameter using $\hat{\theta}$, if $\hat{\theta}$ possesses some good properties, we believe that the Estimate obtained is a good approximation/Estimate for θ^*.

In fact, we think as

$$\theta^* = \hat{\theta} \pm \mathit{error},$$

where *error* is small[1], or, rather, as

$$\theta^* \in (\hat{\theta} - \mathit{error}, \hat{\theta} + \mathit{error}).$$

[1] Recall the $\widehat{SE}$, the Estimated Standard Error reporting story.

Prelude No. 2

But the good news is that even when we cannot exactly find the True value of our Parameter using $\hat{\theta}$, if $\hat{\theta}$ possesses some good properties, we believe that the Estimate obtained is a good approximation/Estimate for θ^*.

In fact, we think as

$$\theta^* = \hat{\theta} \pm \mathit{error},$$

where *error* is small[1], or, rather, as

$$\theta^* \in (\hat{\theta} - \mathit{error}, \hat{\theta} + \mathit{error}).$$

And we can ask questions about how small is the *error* or how much sure are we in our Estimate (for the Unknown Parameter), and how large n needs to be to have a good estimate.

[1] Recall the $\widehat{SE}$, the Estimated Standard Error reporting story.

Prelude No. 2

But the good news is that even when we cannot exactly find the True value of our Parameter using $\hat{\theta}$, if $\hat{\theta}$ possesses some good properties, we believe that the Estimate obtained is a good approximation/Estimate for θ^*.

In fact, we think as

$$\theta^* = \hat{\theta} \pm \mathit{error},$$

where *error* is small[1], or, rather, as

$$\theta^* \in (\hat{\theta} - \mathit{error}, \hat{\theta} + \mathit{error}).$$

And we can ask questions about how small is the *error* or how much sure are we in our Estimate (for the Unknown Parameter), and how large n needs to be to have a good estimate.

Here we want to develop the theory of Confidence Intervals, which will contain answers to these questions.

---

[1] Recall the $\widehat{SE}$, the Estimated Standard Error reporting story.

CI Problem Setting

Assume we have a Random Sample from a Parametric Model $\mathcal{F}_\theta$:

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

CI Problem Setting

Assume we have a Random Sample from a Parametric Model $\mathcal{F}_\theta$:

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

Our aim is to find an interval containing the Unknown Parameter value.

CI Problem Setting

Assume we have a Random Sample from a Parametric Model $\mathcal{F}_\theta$:

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

Our aim is to find an interval containing the Unknown Parameter value.

More rigorously, we want to have an interval

- which contains the True Parameter value with high Probability;

CI Problem Setting

Assume we have a Random Sample from a Parametric Model $\mathcal{F}_\theta$:

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

Our aim is to find an interval containing the Unknown Parameter value.

More rigorously, we want to have an interval

- which contains the True Parameter value with high Probability;
- which has the possible smallest length.

CI Problem Setting

Assume we have a Random Sample from a Parametric Model $\mathcal{F}_\theta$:

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

Our aim is to find an interval containing the Unknown Parameter value.

More rigorously, we want to have an interval

- which contains the True Parameter value with high Probability;
- which has the possible smallest length.

Let us state this in Mathematical terms. We will consider here only 1D case, i.e., we will assume $\theta \in \Theta \subset \mathbb{R}$.

Random Intervals and CI

We start by **Random Intervals:**

Random Intervals and CI

We start by **Random Intervals:**

Assume X and Y are two r.v. on the same probability space and $X \leq Y$ a.s., that is, $\mathbb{P}(X > Y) = 0$.

Random Intervals and CI

We start by **Random Intervals:**

Assume X and Y are two r.v. on the same probability space and $X \leq Y$ a.s., that is, $\mathbb{P}(X > Y) = 0$. We will call intervals of the form $[X, Y]$ or (X, Y) or $(X, Y]$ or $[X, Y)$ or $(X, +\infty)$,. . . , **random intervals**.

Random Intervals and CI

We start by **Random Intervals:**

Assume X and Y are two r.v. on the same probability space and $X \leq Y$ a.s., that is, $\mathbb{P}(X > Y) = 0$. We will call intervals of the form $[X, Y]$ or (X, Y) or $(X, Y]$ or $[X, Y)$ or $(X, +\infty)$,$\ldots$, **random intervals**.

Example: Assume $X \sim \mathit{Pois}(2.3)$. Then

$$[X, X + 2)$$

is a Random Interval.

Random Intervals and CI

We start by **Random Intervals:**

Assume X and Y are two r.v. on the same probability space and $X \leq Y$ a.s., that is, $\mathbb{P}(X > Y) = 0$. We will call intervals of the form $[X, Y]$ or (X, Y) or $(X, Y]$ or $[X, Y)$ or $(X, +\infty)$,... , **random intervals**.

Example: Assume $X \sim Pois(2.3)$. Then

$$[X, X+2)$$

is a Random Interval.

Example: Let $X_1, X_2, ..., X_n$ are IID r.v.s. Then

$$(\overline{X} - 0.1,\ \overline{X} + 0.1)$$

is a Random Interval.

CI Problem Setting

Assume we have a Random Sample from a Parametric Model $\mathcal{F}_\theta$:

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

CI Problem Setting

Assume we have a Random Sample from a Parametric Model $\mathcal{F}_\theta$:

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

We take $\alpha \in (0, 1)$, and call the number $1 - \alpha$ the **confidence level**.

CI Problem Setting

Assume we have a Random Sample from a Parametric Model $\mathcal{F}_\theta$:

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

We take $\alpha \in (0,1)$, and call the number $1-\alpha$ the **confidence level**.

Problem: Using the Random Sample, construct an interval containing the Unknown Parameter value with Probability not less than $1-\alpha$.

CI Problem Setting

Assume we have a Random Sample from a Parametric Model $\mathcal{F}_\theta$:

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

We take $\alpha \in (0, 1)$, and call the number $1 - \alpha$ the **confidence level**.

Problem: Using the Random Sample, construct an interval containing the Unknown Parameter value with Probability not less than $1 - \alpha$.

The usual values of the confidence level are $90\%, 95\%, 99\%$, so the usual values of α are $0.1, 0.05$ and 0.01.

CI

Definition: Assume $0 < \alpha < 1$, and let $L = L(x_1, ..., x_n, \alpha)$, $U = U(x_1, ..., x_n, \alpha)$ be two functions with $L(x_1, ..., x_n, \alpha) \leq U(x_1, ..., x_n, \alpha)$ for all $(x_1, ..., x_n, \alpha)$.

CI

Definition: Assume $0 < \alpha < 1$, and let $L = L(x_1, ..., x_n, \alpha)$, $U = U(x_1, ..., x_n, \alpha)$ be two functions with $L(x_1, ..., x_n, \alpha) \leq U(x_1, ..., x_n, \alpha)$ for all $(x_1, ..., x_n, \alpha)$. The random interval

$$(L, U) = \Big(L(X_1, ..., X_n, \alpha)\ ,\quad U(X_1, ..., X_n, \alpha)\Big)$$

is called a **confidence interval (or confidence interval estimator) for θ of confidence level** $1 - \alpha$, if for any $\theta \in \Theta$,

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

CI

In the case we have a realization/observation of $X_1, ..., X_n$, say, $x_1, ..., x_n$, then the interval

$$\Big(L(x_1, ..., x_n, \alpha),\ \ U(x_1, ..., x_n, \alpha)\Big)$$

will be an **interval estimate for θ for the confidence level** $(1 - \alpha)$.

CI

In the case we have a realization/observation of $X_1, ..., X_n$, say, $x_1, ..., x_n$, then the interval

$$\Big(L(x_1, ..., x_n, \alpha),\ \ U(x_1, ..., x_n, \alpha)\Big)$$

will be an **interval estimate for θ for the confidence level $(1-\alpha)$.**

Going back to our CI, CI of the confidence level $1-\alpha$ is a Random Interval that contains θ in more than $(1-\alpha)\cdot 100\%$ of cases.

CI, Interpretation

Note: It is important to understand, that in the CI definition

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

θ is not our r.v., θ is our unknown constant Parameter, so we do not read this as "with high Probability, θ is in (L, U)".

CI, Interpretation

Note: It is important to understand, that in the CI definition

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

θ is not our r.v., θ is our unknown constant Parameter, so we do not read this as "with high Probability, θ is in (L, U)". Instead, the Interval (L, U) is Random, so we read this as **with high Probability, the interval (L, U) will contain θ.**

[2] But not different θ-s !!

CI, Interpretation

Note: It is important to understand, that in the CI definition

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

θ is not our r.v., θ is our unknown constant Parameter, so we do not read this as "with high Probability, θ is in (L, U)". Instead, the Interval (L, U) is Random, so we read this as **with high Probability, the interval (L, U) will contain θ.**

So, if we will have/generate different observations, we will have different Intervals[2] (L, U), and we want to have that most of the time that interval contains our unknown Parameter value.

[2] But not different θ-s !!

CI, **R** Simulation

Example: Consider an example: our Model is $Exp(\lambda)$, and we have an observation from it. Let us take a Random Sample for the general case: $X_1, X_2, ..., X_n$ from $Exp(\lambda)$.

CI, **R** Simulation

Example: Consider an example: our Model is $Exp(\lambda)$, and we have an observation from it. Let us take a Random Sample for the general case: $X_1, X_2, ..., X_n$ from $Exp(\lambda)$. Then good **Estimator** for λ is

CI, **R** Simulation

Example: Consider an example: our Model is $Exp(\lambda)$, and we have an observation from it. Let us take a Random Sample for the general case: $X_1, X_2, ..., X_n$ from $Exp(\lambda)$. Then good **Estimator** for λ is

$$\hat{\lambda} = \frac{1}{\overline{X}}.$$

CI, **R** Simulation

Example: Consider an example: our Model is $Exp(\lambda)$, and we have an observation from it. Let us take a Random Sample for the general case: $X_1, X_2, ..., X_n$ from $Exp(\lambda)$. Then good **Estimator** for λ is

$$\hat{\lambda} = \frac{1}{\overline{X}}.$$

Now, let us take as CI

CI, **R** Simulation

Example: Consider an example: our Model is $Exp(\lambda)$, and we have an observation from it. Let us take a Random Sample for the general case: $X_1, X_2, ..., X_n$ from $Exp(\lambda)$. Then good **Estimator** for λ is

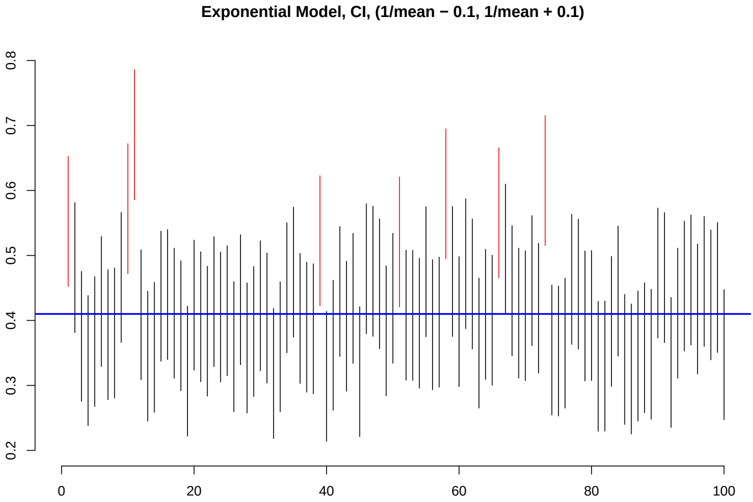
$$\hat{\lambda} = \frac{1}{\overline{X}}.$$

Now, let us take as CI

$$\left(\frac{1}{\overline{X}} - 0.1, \frac{1}{\overline{X}} + 0.1\right)$$

and do some simulations:

CI, **R** Simulation



CI, **R** Simulation, Code

```r
#CI Idea, Exponential Model
lambda <- 0.41
conf.level <- 0.95; a = 1 - conf.level
sample.size <- 50; no.of.intervals <- 100
epsilon <- 0.1

plot.new()
plot.window(xlim = c(0,no.of.intervals), ylim = c(0.2,0.8))
axis(1); axis(2)
title("Exponential Model, CI, (1/mean - 0.1, 1/mean + 0.1)")
for(i in 1:no.of.intervals){
  x <- rexp(sample.size, rate = lambda)
  lo <- 1/mean(x) - epsilon; up <- 1/mean(x) + epsilon
  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")
```

Methods to obtain Confidence Intervals

We will consider several methods to construct CIs:

- Chebyshev Inequality Based;
- Pivotal Quantity Based

CI for the Mean, Variance is given, Cheby Method

Example: Assume $X_1, X_2, ..., X_n$ are Independent r.v. with the same Mean $\mathbb{E}(X_k) = \mu$ and the same Variance $Var(X_k) = \sigma^2$.

CI for the Mean, Variance is given, Cheby Method

Example: Assume $X_1, X_2, ..., X_n$ are Independent r.v. with the same Mean $\mathbb{E}(X_k) = \mu$ and the same Variance $Var(X_k) = \sigma^2$. Assume σ^2 is known, and μ is unknown, and we want to estimate it using the Sample.

CI for the Mean, Variance is given, Cheby Method

Example: Assume $X_1, X_2, ..., X_n$ are Independent r.v. with the same Mean $\mathbb{E}(X_k) = \mu$ and the same Variance $Var(X_k) = \sigma^2$. Assume σ^2 is known, and μ is unknown, and we want to estimate it using the Sample. The problem is to construct a CI for μ of given Confidence Level $1 - \alpha$.

CI for the Mean, Variance is given, Cheby Method

Example: Assume $X_1, X_2, ..., X_n$ are Independent r.v. with the same Mean $\mathbb{E}(X_k) = \mu$ and the same Variance $Var(X_k) = \sigma^2$. Assume σ^2 is known, and μ is unknown, and we want to estimate it using the Sample. The problem is to construct a CI for μ of given Confidence Level $1 - \alpha$.

By the Chebyshev inequality method, we can obtain that the interval

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

is a CI for μ of Confidence Level $1 - \alpha$.

CI for the Mean, Variance is given, Cheby Method

Example: Assume $X_1, X_2, ..., X_n$ are Independent r.v. with the same Mean $\mathbb{E}(X_k) = \mu$ and the same Variance $Var(X_k) = \sigma^2$. Assume σ^2 is known, and μ is unknown, and we want to estimate it using the Sample. The problem is to construct a CI for μ of given Confidence Level $1 - \alpha$.

By the Chebyshev inequality method, we can obtain that the interval

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

is a CI for μ of Confidence Level $1 - \alpha$.

Note: Here

$$\frac{\sigma}{\sqrt{n \cdot \alpha}}$$

is called the **Margin of Error** (for the Interval Estimate of μ, given σ^2).

Some Notes

Two notes about the obtained CI - in fact, these notes will work also for other cases too:

Some Notes

Two notes about the obtained CI - in fact, these notes will work also for other cases too:

The CI length obtained above is

$$\frac{2\sigma}{\sqrt{n \cdot \alpha}}.$$

Note: If we increase n, the CI gets narrower. This is intuitive: if we collect more data, we can estimate the parameter more precisely, we can enclose it in a smaller length interval.

Some Notes

Two notes about the obtained CI - in fact, these notes will work also for other cases too:

The CI length obtained above is

$$\frac{2\sigma}{\sqrt{n \cdot \alpha}}.$$

Note: If we increase n, the CI gets narrower. This is intuitive: if we collect more data, we can estimate the parameter more precisely, we can enclose it in a smaller length interval.

Note: If we increase the Confidence Level, i.e., if we decrease α, then the length of CI increases. This is intuitive too: if we want to be more sure where our unknown Parameter is lying, we will get a larger interval.

CI for the Proportion, Cheby Method

Example: Now, let us construct a CI of CLevel $1 - \alpha$ for p in the *Bernoulli*(p) Model.

CI for the Proportion, Cheby Method

Example: Now, let us construct a CI of CLevel $1-\alpha$ for p in the *Bernoulli*(p) Model.We assume we have a Random Sample

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

CI for the Proportion, Cheby Method

Example: Now, let us construct a CI of CLevel $1-\alpha$ for p in the *Bernoulli*(p) Model.We assume we have a Random Sample

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

The CI for p by Chebyshev Inequality will be

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

is a CI for p of level $1-\alpha$.

CI for the Proportion, Cheby Method

Example: Now, let us construct a CI of CLevel $1-\alpha$ for p in the *Bernoulli*(p) Model.We assume we have a Random Sample

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

The CI for p by Chebyshev Inequality will be

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

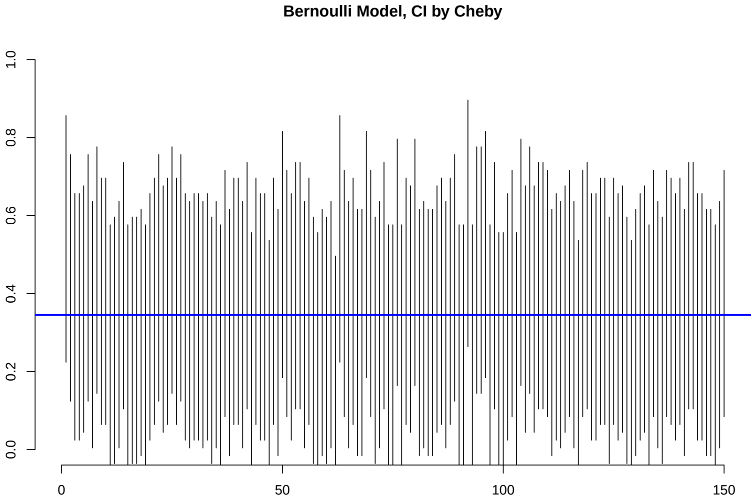
is a CI for p of level $1-\alpha$.

Note: Here

$$\frac{1}{2\sqrt{n \cdot \alpha}}$$

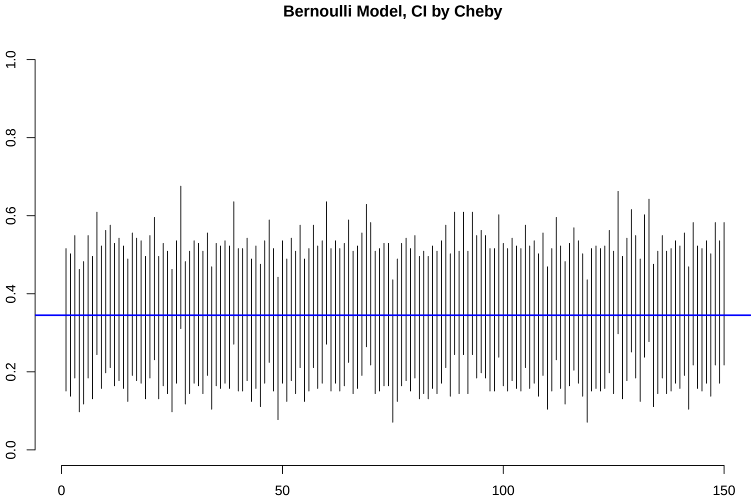
is called the **Margin of Error** (for the Interval Estimate of p).

CI for Bernoulli, **R** Simulation



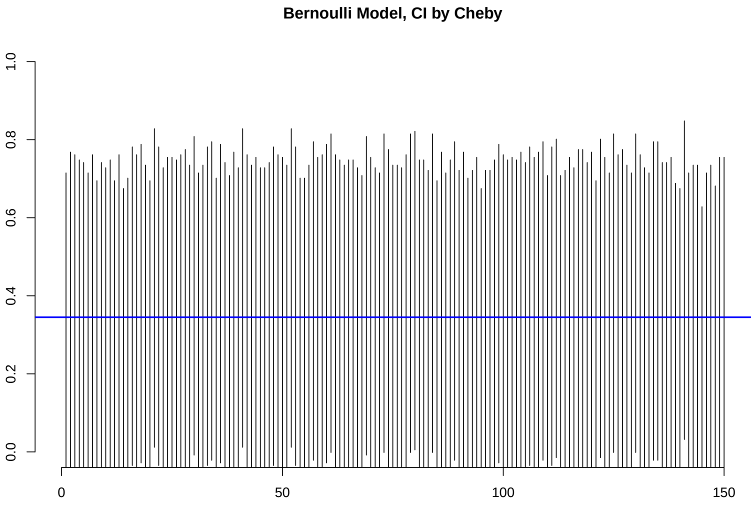
Sample Size $= 50$, $CL = 95\%$

CI for Bernoulli, **R** Simulation



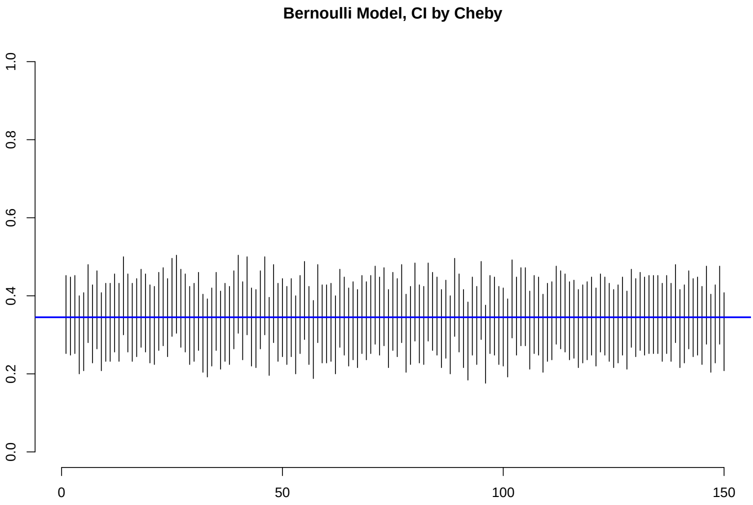
Sample Size $= 150$, $CL = 95\%$

CI for Bernoulli, **R** Simulation



Sample Size $= 150$, $CL = 99\%$

CI for Bernoulli, **R** Simulation



Sample Size $= 250$, $CL = 90\%$

CI, **R** Simulation, Code

```r
#CI Idea, Bernoulli Model
p <- 0.345
conf.level <- 0.9; a = 1 - conf.level
sample.size <- 250; no.of.intervals <- 150
ME <- 1/(2*sqrt(sample.size*a)) #Margin of Error
plot.new()
plot.window(xlim = c(0,no.of.intervals), ylim = c(0,1))
axis(1); axis(2)
title("Bernoulli Model, CI by Cheby")
for(i in 1:no.of.intervals){
  x <- rbinom(sample.size, size = 1, prob = p)
  lo <- mean(x) - ME
  up <- mean(x) + ME
  if(lo > p || up < p){
    segments(c(i), c(lo), c(i), c(up), col = "red")
  }
  else{
    segments(c(i), c(lo), c(i), c(up))
  }
}
abline(h = p, lwd = 2, col = "blue")
```

Examples

Example: Assume we are interested in the proportion of smokers in AUA. We ask 120 persons at AUA and learn that 55 of them are smokers. Construct a CI for the proportion of smokers in AUA of 95% confidence level.

Solution: OTB

Examples

Example: Assume we are interested in the proportion of smokers in AUA. We ask 120 persons at AUA and learn that 55 of them are smokers. Construct a CI for the proportion of smokers in AUA of 95% confidence level.

Solution: OTB

Example: Continuing the above Example: now assume we want to find that Proportion within the Error Margin 0.1, with the CL 95%. At least, how many persons at AUA we need to ask?

Solution: OTB