YSU ASDS, Statistics, Fall 2019

Lecture 21

Michael Poghosyan

06 Nov 2019

Contents

- Maximum Likelihood Estimation
- Some Topics we'll cover soon
- Confidence Intervals (CI)

Last Lecture ReCap

- What is the problem that MLE is solving?

Last Lecture ReCap

- What is the problem that MLE is solving?
- What are the remarkable properties of the 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.,
$$\widehat{g(\theta)}^{MLE} = g\left(\hat{\theta}^{MLE}\right).$$

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 we will talk about soon

- 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 Parameteric Point Estimation Problem: we were assuming that our data comes 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 Parameteric Point Estimation Problem: we were assuming that our data comes 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 Parameteric Point Estimation Problem: we were assuming that our data comes 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 Parameteric Point Estimation Problem: we were assuming that our data comes 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 Parameteric Point Estimation Problem: we were assuming that our data comes 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

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

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 you were asking nice 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 you were asking nice 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 $(X, +\infty)$,... , **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

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

is a Random Interval.

CI Problem Setting, Again

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, Again

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, Again

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, Again

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

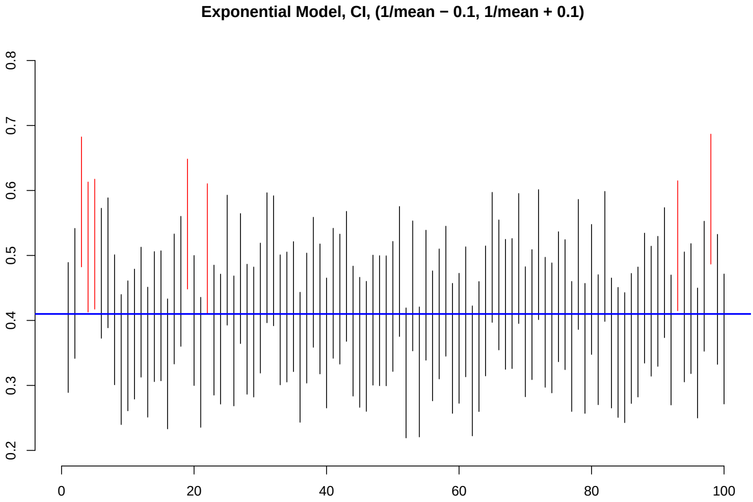
$$\hat{\lambda} = \frac{1}{\overline{X}}.$$

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

Methods to obtain Confidence Intervals

We will consider several methods to construct CIs:

- Chebyshev Inequality Based;
- Pivotal Quantity Based

And then we will talk about Asymptotic CIs.

Prob Refresher, Chebyshev Inequality

Recall the Cheby Inequality: If X is a r.v. with finite Mean $\mathbb{E}(X)$ and Variance $Var(X)$, then for any $\varepsilon > 0$,

$$\mathbb{P}\Big(|X - \mathbb{E}(X)| \geq \varepsilon\Big) \leq \frac{Var(X)}{\varepsilon^2},$$

Prob Refresher, Chebyshev Inequality

Recall the Cheby Inequality: If X is a r.v. with finite Mean $\mathbb{E}(X)$ and Variance $Var(X)$, then for any $\varepsilon > 0$,

$$\mathbb{P}\Big(|X - \mathbb{E}(X)| \geq \varepsilon\Big) \leq \frac{Var(X)}{\varepsilon^2},$$

or, which is the same,

$$\mathbb{P}\Big(|X - \mathbb{E}(X)| < \varepsilon\Big) \geq 1 - \frac{Var(X)}{\varepsilon^2}.$$

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 $Var(X_k) = \sigma^2$.

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. We will construct a CI by the Chebyshev inequality, 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. We will construct a CI by the Chebyshev inequality, of given Confidence Level $1 - \alpha$.

Clearly,

$$\mathbb{E}(\overline{X}) =$$

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. We will construct a CI by the Chebyshev inequality, of given Confidence Level $1 - \alpha$.

Clearly,

$$\mathbb{E}(\overline{X}) = \mu \qquad \textit{and} \qquad Var(\overline{X}) =$$

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. We will construct a CI by the Chebyshev inequality, of given Confidence Level $1 - \alpha$.

Clearly,

$$\mathbb{E}(\overline{X}) = \mu \qquad and \qquad Var(\overline{X}) = \frac{\sigma^2}{n},$$

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. We will construct a CI by the Chebyshev inequality, of given Confidence Level $1 - \alpha$.

Clearly,

$$\mathbb{E}(\overline{X}) = \mu \qquad \textit{and} \qquad Var(\overline{X}) = \frac{\sigma^2}{n},$$

so if we will plug $\overline{X}$ in the Cheby Inequality, we will obtain

$$\mathbb{P}\Big(|\overline{X} - \mu| \geq \varepsilon\Big) \leq \frac{\sigma^2}{n \cdot \varepsilon^2}$$

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. We will construct a CI by the Chebyshev inequality, of given Confidence Level $1 - \alpha$.

Clearly,

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

so if we will plug $\overline{X}$ in the Cheby Inequality, we will obtain

$$\mathbb{P}\Big(|\overline{X} - \mu| \geq \varepsilon\Big) \leq \frac{\sigma^2}{n \cdot \varepsilon^2} \qquad \text{or} \qquad \mathbb{P}\Big(|\overline{X} - \mu| < \varepsilon\Big) \geq 1 - \frac{\sigma^2}{n \cdot \varepsilon^2}$$

for any $\varepsilon > 0$.

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. We will construct a CI by the Chebyshev inequality, of given Confidence Level $1 - \alpha$.

Clearly,

$$\mathbb{E}(\overline{X}) = \mu \qquad \textit{and} \qquad Var(\overline{X}) = \frac{\sigma^2}{n},$$

so if we will plug $\overline{X}$ in the Cheby Inequality, we will obtain

$$\mathbb{P}\Big(|\overline{X} - \mu| \geq \varepsilon\Big) \leq \frac{\sigma^2}{n \cdot \varepsilon^2} \qquad \textit{or} \qquad \mathbb{P}\Big(|\overline{X} - \mu| < \varepsilon\Big) \geq 1 - \frac{\sigma^2}{n \cdot \varepsilon^2}$$

for any $\varepsilon > 0$. Now, take $\dfrac{\sigma^2}{n \cdot \varepsilon^2} = \alpha$.

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. We will construct a CI by the Chebyshev inequality, of given Confidence Level $1 - \alpha$.

Clearly,

$$\mathbb{E}(\overline{X}) = \mu \qquad and \qquad Var(\overline{X}) = \frac{\sigma^2}{n},$$

so if we will plug $\overline{X}$ in the Cheby Inequality, we will obtain

$$\mathbb{P}\Big(|\overline{X} - \mu| \geq \varepsilon\Big) \leq \frac{\sigma^2}{n \cdot \varepsilon^2} \qquad or \qquad \mathbb{P}\Big(|\overline{X} - \mu| < \varepsilon\Big) \geq 1 - \frac{\sigma^2}{n \cdot \varepsilon^2}$$

for any $\varepsilon > 0$. Now, take $\dfrac{\sigma^2}{n \cdot \varepsilon^2} = \alpha$. Here, σ, n and α are known, so this equality will give us the value for ε:

$$\varepsilon = \frac{\sigma}{\sqrt{n \cdot \alpha}}.$$

CI for the Mean, Variance is given, Cheby Method

Plugging this value into the above Cheby Inequality, we will get

$$\mathbb{P}\left(|\overline{X} - \mu| < \frac{\sigma}{\sqrt{n \cdot \alpha}}\right) \geq 1 - \alpha,$$

CI for the Mean, Variance is given, Cheby Method

Plugging this value into the above Cheby Inequality, we will get

$$\mathbb{P}\left(|\overline{X} - \mu| < \frac{\sigma}{\sqrt{n \cdot \alpha}}\right) \geq 1 - \alpha,$$

or, in the expanded form,

$$\mathbb{P}\left(\overline{X} - \frac{\sigma}{\sqrt{n \cdot \alpha}} < \mu < \overline{X} + \frac{\sigma}{\sqrt{n \cdot \alpha}}\right) \geq 1 - \alpha$$

CI for the Mean, Variance is given, Cheby Method

Plugging this value into the above Cheby Inequality, we will get

$$\mathbb{P}\Big(|\overline{X} - \mu| < \frac{\sigma}{\sqrt{n \cdot \alpha}}\Big) \geq 1 - \alpha,$$

or, in the expanded form,

$$\mathbb{P}\Big(\overline{X} - \frac{\sigma}{\sqrt{n \cdot \alpha}} < \mu < \overline{X} + \frac{\sigma}{\sqrt{n \cdot \alpha}}\Big) \geq 1 - \alpha$$

This means, 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

Plugging this value into the above Cheby Inequality, we will get

$$\mathbb{P}\Big(|\overline{X} - \mu| < \frac{\sigma}{\sqrt{n \cdot \alpha}}\Big) \geq 1 - \alpha,$$

or, in the expanded form,

$$\mathbb{P}\Big(\overline{X} - \frac{\sigma}{\sqrt{n \cdot \alpha}} < \mu < \overline{X} + \frac{\sigma}{\sqrt{n \cdot \alpha}}\Big) \geq 1 - \alpha$$

This means, 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 $1 - \alpha$. URA!

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

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

Again we use the Chebyshev Inequality for $\overline{X}$: for any $\varepsilon > 0$,

$$\mathbb{P}(|\overline{X} - \mathbb{E}(\overline{X})| \geq \varepsilon) \leq \frac{Var(\overline{X})}{\varepsilon^2}$$

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

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

Again we use the Chebyshev Inequality for $\overline{X}$: for any $\varepsilon > 0$,

$$\mathbb{P}(|\overline{X}-\mathbb{E}(\overline{X})| \geq \varepsilon) \leq \frac{Var(\overline{X})}{\varepsilon^2} \quad \textit{or} \quad \mathbb{P}(|\overline{X}-\mathbb{E}(\overline{X})| < \varepsilon) \geq 1-\frac{Var(\overline{X})}{\varepsilon^2}.$$

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

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

Again we use the Chebyshev Inequality for $\overline{X}$: for any $\varepsilon > 0$,

$$\mathbb{P}(|\overline{X}-\mathbb{E}(\overline{X})| \geq \varepsilon) \leq \frac{Var(\overline{X})}{\varepsilon^2} \quad or \quad \mathbb{P}(|\overline{X}-\mathbb{E}(\overline{X})| < \varepsilon) \geq 1-\frac{Var(\overline{X})}{\varepsilon^2}.$$

For the *Bernoulli*(p) case, we have

$$\mathbb{E}(\overline{X}) =$$

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

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

Again we use the Chebyshev Inequality for $\overline{X}$: for any $\varepsilon > 0$,

$$\mathbb{P}(|\overline{X} - \mathbb{E}(\overline{X})| \geq \varepsilon) \leq \frac{Var(\overline{X})}{\varepsilon^2} \quad or \quad \mathbb{P}(|\overline{X} - \mathbb{E}(\overline{X})| < \varepsilon) \geq 1 - \frac{Var(\overline{X})}{\varepsilon^2}.$$

For the *Bernoulli*(p) case, we have

$$\mathbb{E}(\overline{X}) = p \qquad and \qquad Var(\overline{X}) =$$

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

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

Again we use the Chebyshev Inequality for $\overline{X}$: for any $\varepsilon > 0$,

$$\mathbb{P}(|\overline{X} - \mathbb{E}(\overline{X})| \geq \varepsilon) \leq \frac{Var(\overline{X})}{\varepsilon^2} \quad or \quad \mathbb{P}(|\overline{X} - \mathbb{E}(\overline{X})| < \varepsilon) \geq 1 - \frac{Var(\overline{X})}{\varepsilon^2}.$$

For the *Bernoulli*(p) case, we have

$$\mathbb{E}(\overline{X}) = p \qquad and \qquad Var(\overline{X}) = \frac{p(1-p)}{n}.$$

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

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

Again we use the Chebyshev Inequality for $\overline{X}$: for any $\varepsilon > 0$,

$$\mathbb{P}(|\overline{X}-\mathbb{E}(\overline{X})| \geq \varepsilon) \leq \frac{Var(\overline{X})}{\varepsilon^2} \quad or \quad \mathbb{P}(|\overline{X}-\mathbb{E}(\overline{X})| < \varepsilon) \geq 1-\frac{Var(\overline{X})}{\varepsilon^2}.$$

For the *Bernoulli*(p) case, we have

$$\mathbb{E}(\overline{X}) = p \qquad and \qquad Var(\overline{X}) = \frac{p(1-p)}{n}.$$

Hence,

$$\mathbb{P}(|\overline{X} - p| < \varepsilon) \geq 1 - \frac{p(1-p)}{n \cdot \varepsilon^2}.$$

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

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

Again we use the Chebyshev Inequality for $\overline{X}$: for any $\varepsilon > 0$,

$$\mathbb{P}(|\overline{X} - \mathbb{E}(\overline{X})| \geq \varepsilon) \leq \frac{Var(\overline{X})}{\varepsilon^2} \quad or \quad \mathbb{P}(|\overline{X} - \mathbb{E}(\overline{X})| < \varepsilon) \geq 1 - \frac{Var(\overline{X})}{\varepsilon^2}.$$

For the *Bernoulli*(p) case, we have

$$\mathbb{E}(\overline{X}) = p \qquad and \qquad Var(\overline{X}) = \frac{p(1-p)}{n}.$$

Hence,

$$\mathbb{P}(|\overline{X} - p| < \varepsilon) \geq 1 - \frac{p(1-p)}{n \cdot \varepsilon^2}.$$

Here, on the RHS, we have the unknown parameter value p, which is not desirable.

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

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

Again we use the Chebyshev Inequality for $\overline{X}$: for any $\varepsilon > 0$,

$$\mathbb{P}(|\overline{X}-\mathbb{E}(\overline{X})| \geq \varepsilon) \leq \frac{Var(\overline{X})}{\varepsilon^2} \quad or \quad \mathbb{P}(|\overline{X}-\mathbb{E}(\overline{X})| < \varepsilon) \geq 1-\frac{Var(\overline{X})}{\varepsilon^2}.$$

For the *Bernoulli*(p) case, we have

$$\mathbb{E}(\overline{X}) = p \qquad and \qquad Var(\overline{X}) = \frac{p(1-p)}{n}.$$

Hence,

$$\mathbb{P}(|\overline{X}-p| < \varepsilon) \geq 1 - \frac{p(1-p)}{n \cdot \varepsilon^2}.$$

Here, on the RHS, we have the unknown parameter value p, which is not desirable. To get rid of that, we use the estimate $p(1-p) \leq \dfrac{1}{4}$, so $\mathbb{P}(|\overline{X}-p| < \varepsilon) \geq 1 - \dfrac{1}{4n \cdot \varepsilon^2}$.

CI for the Proportion, Cheby Method, Cont'd

On the RHS, for the CI, we want to have $1 - \alpha$. So, as in the previous Example, we choose

$$\frac{1}{4n \cdot \varepsilon^2} = \alpha,$$

CI for the Proportion, Cheby Method, Cont'd

On the RHS, for the CI, we want to have $1 - \alpha$. So, as in the previous Example, we choose

$$\frac{1}{4n \cdot \varepsilon^2} = \alpha, \qquad \textit{i.e.} \qquad \varepsilon = \frac{1}{2\sqrt{n \cdot \alpha}}.$$

CI for the Proportion, Cheby Method, Cont'd

On the RHS, for the CI, we want to have $1-\alpha$. So, as in the previous Example, we choose

$$\frac{1}{4n\cdot\varepsilon^2}=\alpha, \qquad \textit{i.e.} \qquad \varepsilon=\frac{1}{2\sqrt{n\cdot\alpha}}.$$

Plugging into the inequality above, this will give

$$\mathbb{P}(|\overline{X}-p|<\frac{1}{2\sqrt{n\cdot\alpha}})\geq 1-\alpha,$$

CI for the Proportion, Cheby Method, Cont'd

On the RHS, for the CI, we want to have $1 - \alpha$. So, as in the previous Example, we choose

$$\frac{1}{4n \cdot \varepsilon^2} = \alpha, \qquad i.e. \qquad \varepsilon = \frac{1}{2\sqrt{n \cdot \alpha}}.$$

Plugging into the inequality above, this will give

$$\mathbb{P}(|\overline{X} - p| < \frac{1}{2\sqrt{n \cdot \alpha}}) \geq 1 - \alpha,$$

or,

$$\mathbb{P}\left(\overline{X} - \frac{1}{2\sqrt{n \cdot \alpha}} < p < \overline{X} + \frac{1}{2\sqrt{n \cdot \alpha}}\right) \geq 1 - \alpha.$$

CI for the Proportion, Cheby Method, Cont'd

On the RHS, for the CI, we want to have $1 - \alpha$. So, as in the previous Example, we choose

$$\frac{1}{4n \cdot \varepsilon^2} = \alpha, \qquad i.e. \qquad \varepsilon = \frac{1}{2\sqrt{n \cdot \alpha}}.$$

Plugging into the inequality above, this will give

$$\mathbb{P}(|\overline{X} - p| < \frac{1}{2\sqrt{n \cdot \alpha}}) \geq 1 - \alpha,$$

or,

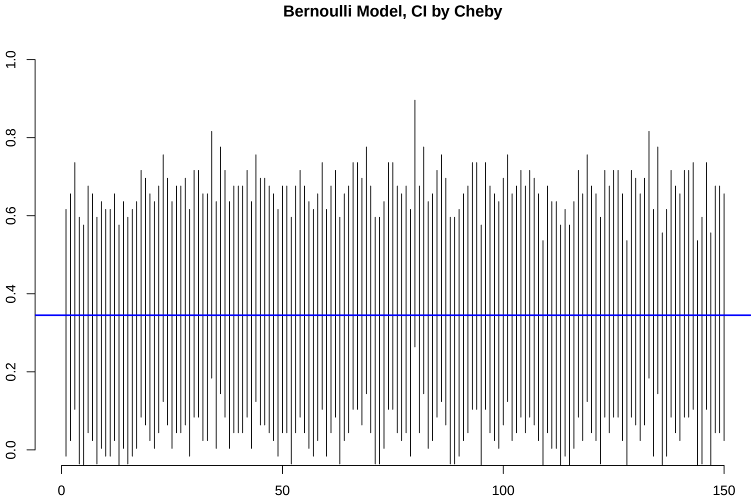
$$\mathbb{P}\left(\overline{X} - \frac{1}{2\sqrt{n \cdot \alpha}} < p < \overline{X} + \frac{1}{2\sqrt{n \cdot \alpha}}\right) \geq 1 - \alpha.$$

This means that the interval

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

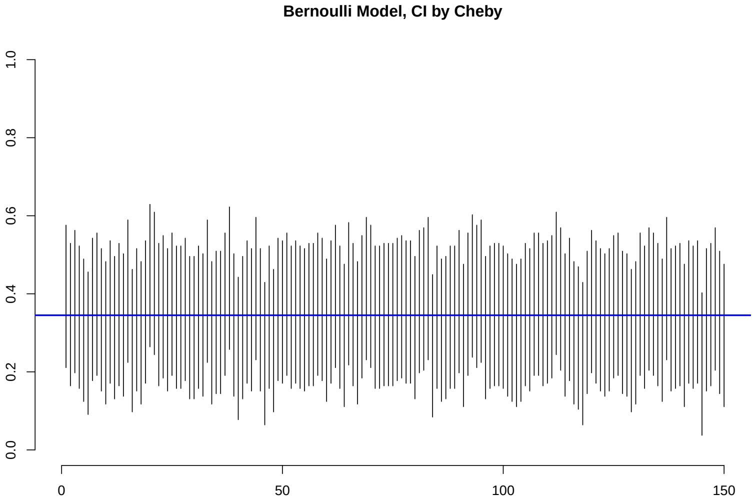
is a CI for p of level $1 - \alpha$.

CI for Bernoulli, **R** Simulation



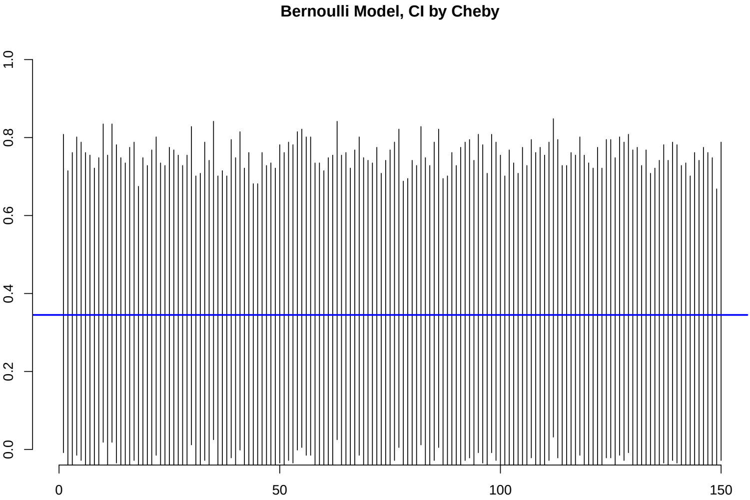
Sample Size $= 50$, $CL = 95\%$

CI for Bernoulli, **R** Simulation



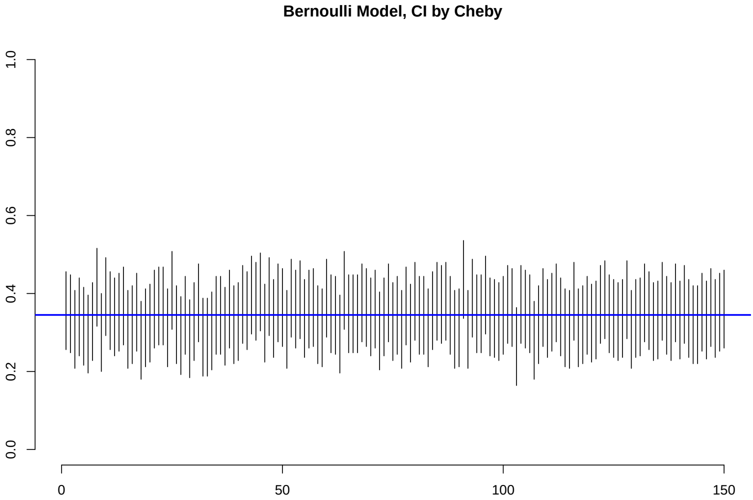
Sample Size $= 150$, $CL = 95\%$

CI for Bernoulli, **R** Simulation



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

CI for Bernoulli, **R** Simulation



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

CI, R Simulation, Code

```
#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")
```

CI for the Proportion, Cheby Method

Recall that if we have a Random Sample

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

then the interval

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

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

CI for the Proportion, Cheby Method

Recall that if we have a Random Sample

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

then the interval

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

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).

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