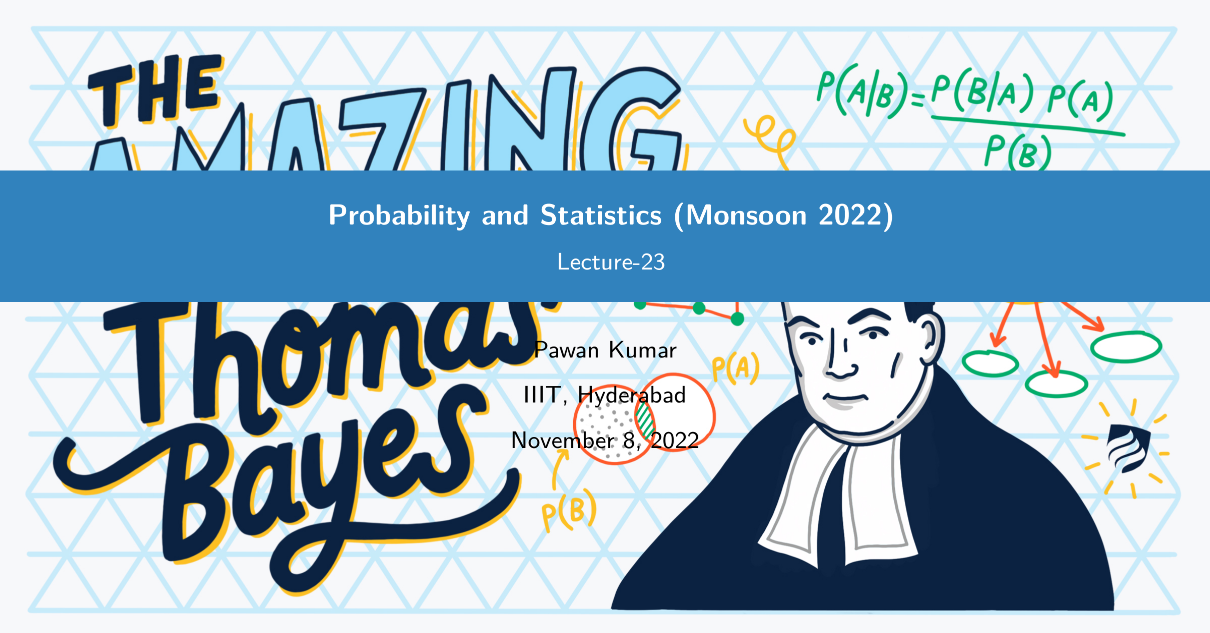


Probability and Statistics (Monsoon 2022)

Lecture-23

Pawan Kumar

IIIT, Hyderabad

November 8, 2022

1. Statistical Inference
   - Interval Estimation and Confidence Level
   - Pivotal Quantity

Table of Contents

Table of Contents

Example of Interval Estimation...

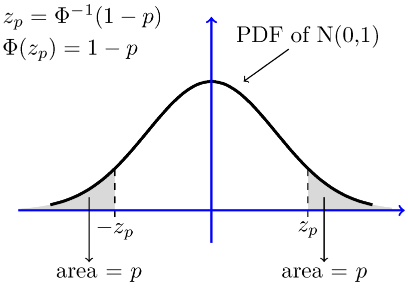
Example

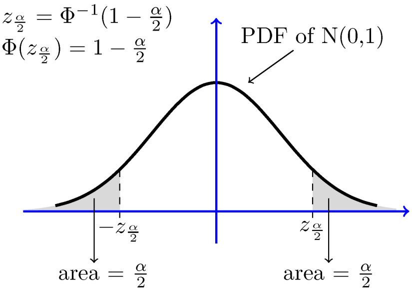
Let $Z \sim N(0,1)$, find x_l and x_h such that

$$P\left(x_l \leq Z \leq x_h\right) = 0.95.$$

https://www.probabilitycourse.com/chapter8/8_3_2_finding_interval_estimators.php

Plots of PDF and Confidence Interval...





How do we find interval estimators?

A general approach is to start with a point estimator $\hat{\Theta}$, such as the MLE, and create the interval $[\hat{\Theta}_l, \hat{\Theta}_h]$ around it such that

$$P\left(\theta \in [\hat{\Theta}_l, \hat{\Theta}_h]\right) \geq 1 - \alpha$$

How can we do this?

Find a Confidence Interval...

Example (Find a Confidence Interval)

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a normal distribution $N(\theta, 1)$. Find a 95% confidence interval for θ.

https://www.probabilitycourse.com/chapter8/8_3_2_finding_interval_estimators.php

Pivotal Quantity

Let $X_1, \ldots, X_n$ be a random sample from a distribution with a parameter that is to be estimated. The random variable Q is said to be a pivot or a pivotal quantity, if it has the following properties:

1. It is a function of the observed data $X_1, \ldots, X_n$ and the unknown parameter θ, but it does not depend on any other unknown parameters:

$$Q = Q(X_1, \ldots, X_n, \theta).$$

2. The probability distribution of Q does not depend on or any other unknown parameters.

Table of Contents

Pivotal Quantity...

Pivotal Quantity

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ that is to be estimated. The random variable Q is said to be a pivot or a pivotal quantity, if it has the following properties: It is a function of the observed data $X_1, X_2, X_3, \ldots, X_n$ and the unknown parameter θ, but it does not depend on any other unknown parameters:

$$Q = Q(X_1, X_2, \ldots, X_n, \theta).$$

The probability distribution of Q does not depend on θ or any other unknown parameters.

Find a Confidence Interval...

Example (Find a Confidence Interval)

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a normal distribution $N(\theta, 1)$. Find a 95% confidence interval for θ. Verify that the random variables

$$Q_1 = \bar{X} - \theta \quad \text{and} \quad Q_2 = \sqrt{n}(\bar{X} - \theta)$$

are both valid pivots.

https://www.probabilitycourse.com/chapter8/8_3_2_finding_interval_estimators.php

Steps for Pivotal Method...

Steps for Pivotal Method for Confidence Interval

1. First, find a pivotal quantity $Q(X_1, X_2, \ldots, X_n, \theta)$
2. Find an interval for Q such that
$$P(q_l \leq Q \leq q_h) = 1 - \alpha$$
3. Using algebraic manipulations, convert the above equation to an equation of the form
$$P(\hat{\Theta}_l \leq \theta \leq \hat{\Theta}_h) = 1 - \alpha$$

Some remarks on how exactly to perform these steps:

1. The most important is crucial one is the first step
2. For many important cases, statisticians have already found the pivotal quantities
3. Many of the interval estimation problems you encounter are of the forms for which general confidence intervals have been found previously
4. To solve many confidence interval problems, it suffices to write the problem in a format similar to a previously solved problem

Solved Example...

Example

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with known variance $\mathrm{Var}(X_i) = \sigma^2$, and unknown mean $E[X_i] = \theta$. Find a $(1-\alpha)$ confidence interval for θ. Assume that n is large.

https://www.probabilitycourse.com/chapter8/8_3_2_finding_interval_estimators.php

Summarize Interval Estimators...

Summary

1 Assumptions: A random sample $X_1, X_2, X_3, \ldots, X_n$ is given from a distribution with known variance $\text{Var}(X_i) = \sigma^2 < \infty$; n is large.

2 Parameter to be Estimated: $\theta = E[X_i]$.

3 Confidence Interval: $\left[\hat{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \quad \hat{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$ is approximately a $(1-\alpha)100\%$ confidence interval for θ

4 Above, we have used CLT, hence, we found an approximation to confidence interval.

Example

An engineer is measuring a quantity θ. It is assumed that there is a random error in each measurement, so the engineer will take n measurements and report the average of the measurements as the estimated value of θ. Here, n is assumed to be large enough so that the central limit theorem applies. If X_i is the value that is obtained in the ith measurement, we assume that

$$X_i = \theta + W_i,$$

where W_i is the error in the ith measurement. We assume that the W_i's are i.i.d. with $E[W_i] = 0$ and $\mathrm{Var}(W_i) = 4$ square units. The engineer reports the average of the measurements

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

How many measurements does the engineer need to make until he is 90% sure that the final error is less than 0.25 units? In other words, what should the value of n be such that

$$P(\theta - 0.25 \leq \bar{X} \leq \theta + 0.25) \geq 0.90?$$

https://www.probabilitycourse.com/chapter8/8_3_2_finding_interval_estimators.php

Confidence Interval with Unknown Variance...

Confidence interval with unknown variance

1 An upper bound for σ^2 : Suppose that we can somehow show that $\sigma \leq \sigma_{\max}$, where $\sigma_{\max} < \infty$. Then if we replace σ above by $\sigma_{\max}$, the interval gets bigger. The interval

$$\left[\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma_{\max}}{\sqrt{n}}, \quad \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma_{\max}}{\sqrt{n}}\right]$$

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

2 Estimate σ^2 : Note that here, since n is large, we should be able to find a relatively good estimate for σ^2. After estimating σ^2, we can use that estimate and

$$\left[\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \quad \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$$

to find an approximate $(1-\alpha)100\%$ confidence interval for θ.

Solved Problem

Example (Public Opinion Polling)

We would like to estimate the portion of people who plan to vote for Candidate A in an upcoming election. It is assumed that the number of voters is large, and θ is the portion of voters who plan to vote for Candidate A. We define the random variable X as follows. A voter is chosen uniformly at random among all voters and we ask her/him: "Do you plan to vote for Candidate A?" If she/he says "yes," then X $= 1$, otherwise $X = 0$. Then, $X \sim$ Bernoulli(). Let $X_1, X_2, X_3, ..., X_n$ be a random sample from this distribution, which means that the X_i's are i.i.d. and $X_i \sim$ Bernoulli(θ). In other words, we randomly select n voters (with replacement) and we ask each of them if they plan to vote for Candidate A. Find a $(1-\alpha)100\%$ confidence interval for θ based on $X_1, X_2, X_3, \ldots, X_n$.

https://www.probabilitycourse.com/chapter8/8_3_2_finding_interval_estimators.php