Chapter 2

Methods of estimator construction

Objectives

After studying this chapter, you should:

- Understand the concepts and importance of point estimation in statistics.
- Differentiate between the method of moments and the maximum likelihood method for parameter estimation.
- Recognize key characteristics of an estimator, including bias, efficiency, and the Cramer-Rao bound.
- Learn how to construct and interpret confidence intervals for estimation.
- Apply interval estimation techniques to normal samples and populations.

Introduction

Estimation is a core aspect of statistical inference, providing the means to draw conclusions about population parameters based on sample data. This chapter delves into the various methods of estimator construction, which are crucial for accurately estimating these parameters. We will begin by exploring point estimation, where we seek a single value that best represents the unknown parameter. Two widely used techniques for point estimation—the Method of Moments and the Maximum Likelihood Method—will be examined in detail, highlighting their principles, advantages, and limitations.

Understanding the characteristics of estimators is essential for evaluating their performance. We will discuss key concepts such as bias, mean squared error, Fisher information, and the Cramer-Rao bound, which together provide a framework for assessing the quality of different estimators. Moreover, we will introduce the concept of efficiency, which indicates how well an estimator performs relative to others.

In addition to point estimation, this chapter will cover estimation through confidence intervals, offering a range of methods for determining intervals that capture the true parameter value with a specified level of confidence. We will also focus on constructing confidence intervals for normal samples and populations, enabling practitioners to make informed decisions based on the results of their analyses.

By the end of this chapter, readers will have a solid foundation in the methods of estimator construction, equipping them with the tools necessary for effective statistical analysis and interpretation. This knowledge is vital for researchers and practitioners who aim to make reliable inferences from sample data.

2.1 Point Estimation

2.1.1 Method of Moments

The method of moments is a commonly used approach for estimating the parameters of a distribution by matching sample moments to theoretical (population) moments. The idea is to calculate sample moments from the data and equate them to the corresponding population moments, which are functions of the unknown parameters of the distribution. Solving these equations provides estimates for the parameters.

Formula for the Method of Moments Estimator

Let θ represent the parameter of interest. The method of moments estimator $\hat{\theta}$ is obtained by solving the equation that sets the k-th sample moment equal to the k-th theoretical moment. For a sample $X_1, X_2, \ldots, X_n$, the sample moments are calculated as the averages of powers of the sample values.

For example, the first sample moment (the sample mean) is given by:

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

where X_i are the observed data points. The estimator $\hat{\theta}$ is the method of moments estimate of the parameter θ.

Steps for the Method of Moments:

1. **Compute the sample moments**: Based on the data, calculate the sample moments.
2. **Determine the population moments**: Express the population moments as functions of the unknown parameter(s) θ.
3. **Set sample moments equal to population moments**: Form equations by equating the sample moments to the corresponding population moments.
4. **Solve for θ**: Solve the system of equations to estimate the parameter(s) θ.

Example 1: Estimating the Parameter of an Exponential Distribution

Let's consider the case of an exponential distribution with unknown rate parameter λ. We have a random sample $X_1, X_2, \ldots, X_n$ from an exponential distribution, and we want to estimate λ using the method of moments.

Step 1: Population Moment of the Exponential Distribution For the exponential distribution, the first population moment (the expected value of the random variable X) is:

$$\mu_1 = \frac{1}{\lambda}$$

Step 2: Sample Moment The first sample moment (the sample mean) is:

$$\hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Step 3: Equate Sample Moment to Population Moment Equating the sample moment to the population moment gives:

$$\frac{1}{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Step 4: Solve for λ Solving for λ gives the method of moments estimator:

$$\hat{\lambda} = \frac{1}{\bar{X}} = \frac{n}{\sum_{i=1}^{n} X_i}$$

Thus, the method of moments estimator for λ is the inverse of the sample mean.

Example 2: Estimating the Parameters of a Uniform Distribution

Now, consider a random sample $X_1, X_2, \ldots, X_n$ from a continuous uniform distribution on the interval $[a, b]$, where a and b are unknown parameters. We want to estimate a and b using the method of moments.

Step 1: Population Moments of the Uniform Distribution For a uniform distribution on $[a, b]$, the first population moment (mean) and second population moment (variance) are:

$$\mu_1 = \frac{a+b}{2}, \quad \mu_2 = \frac{(b-a)^2}{12}$$

Step 2: Sample Moments The first and second sample moments (mean and variance) are:

$$\hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu}_1)^2$$

Step 3: Equate Sample Moments to Population Moments Equating the sample moments to the population moments gives the following system of equations:

$$\frac{a+b}{2} = \hat{\mu}_1, \quad \frac{(b-a)^2}{12} = \hat{\mu}_2$$

Step 4: Solve for a and b From the first equation:

$$a + b = 2\hat{\mu}_1$$

From the second equation:

$$(b-a)^2 = 12\hat{\mu}_2$$

Solving this system of equations yields the method of moments estimators for a and b:

$$\hat{a} = \hat{\mu}_1 - \sqrt{3\hat{\mu}_2}, \quad \hat{b} = \hat{\mu}_1 + \sqrt{3\hat{\mu}_2}$$

Verification: Unbiasedness of the Estimators

To prove that the method of moments estimators for a and b are unbiased, we need to calculate the expected values of the estimators and verify that they equal the true values of a and b. For example, consider the estimator for a:

$$\mathbb{E}(\hat{a}) = \mathbb{E}\left(\hat{\mu}_1 - \sqrt{3\hat{\mu}_2}\right)$$

Since $\hat{\mu}_1$ and $\hat{\mu}_2$ are consistent estimators of the true mean and variance, we have:

$$\mathbb{E}(\hat{\mu}_1) = \mu_1, \quad \mathbb{E}(\hat{\mu}_2) = \mu_2$$

Therefore, $\hat{a}$ is an unbiased estimator of a, and similarly for b.

2.1.2 Maximum Likelihood Method

The Maximum Likelihood Estimation (MLE) method is a widely used approach in statistical inference for estimating the parameters of a statistical model. It aims to find the parameter θ that maximizes the likelihood function, which represents the probability of observing the given sample data under the model parameterized by θ.

In simpler terms, the MLE method finds the parameter value that makes the observed data most probable.

Likelihood Function

The likelihood function $L(\theta)$ is a key concept in MLE. For a given sample of size n, consisting of independent observations $X_1, X_2, \ldots, X_n$, the likelihood function is defined as the joint probability (or probability density) of observing the sample, assuming a particular value of the parameter θ.

Mathematically, if the observations $X_1, X_2, \ldots, X_n$ are drawn independently from a probability distribution with a probability density function (PDF) $f(X_i; \theta)$, the likelihood function is given by the product of the individual probabilities (or densities):

$$L(\theta) = f(X_1; \theta) \cdot f(X_2; \theta) \cdot \ldots \cdot f(X_n; \theta)$$

Or, equivalently:

$$L(\theta) = \prod_{i=1}^{n} f(X_i; \theta)$$

In this expression, $f(X_i; \theta)$ is the probability density (or mass) function of the observation X_i under the parameter θ.

Interpretation: The likelihood function can be interpreted as a measure of how likely it is to observe the given sample for different values of θ. The larger $L(\theta)$, the more likely the observed data is under the model with parameter θ. Thus, the goal of MLE is to find the value of θ that maximizes this likelihood function, leading to the "best fit" of the model to the observed data.

Log-Likelihood Function

Because the likelihood function involves a product of probabilities, it can sometimes lead to computational difficulties, especially with large sample sizes. Therefore, it is common to work with the *log-likelihood* function, which is the natural logarithm of the likelihood function.

The log-likelihood function is defined as:

$$\ell(\theta) = \ln L(\theta) = \sum_{i=1}^{n} \ln f(X_i; \theta)$$

Since the natural logarithm is a monotonic function, maximizing the log-likelihood function $\ell(\theta)$ leads to the same parameter estimate as maximizing the likelihood function $L(\theta)$. However, the log-likelihood function is often easier to work with because the product in the likelihood function becomes a sum in the log-likelihood.

Maximum Likelihood Estimator (MLE)

The Maximum Likelihood Estimator (MLE) of the parameter θ is the value of θ that maximizes the likelihood (or equivalently, the log-likelihood) function.

Formally, the MLE $\hat{\theta}_{MLE}$ is defined as:

$$\hat{\theta}_{MLE} = \arg\max_{\theta} L(\theta)$$

or, equivalently:

$$\hat{\theta}_{MLE} = \arg\max_{\theta} \ell(\theta)$$

This means that the MLE is the value of θ that maximizes the likelihood (or log-likelihood) function, making the observed data as probable as possible.

Steps to Obtain the MLE:

1. **Write the likelihood function**: Start by expressing the likelihood function $L(\theta)$ based on the PDF (or PMF) of the observed data.

2. **Take the logarithm**: Compute the log-likelihood function $\ell(\theta)$ to simplify the product into a sum, making the calculation more manageable.

3. **Differentiate**: Differentiate the log-likelihood function with respect to the parameter θ to find the critical points where the likelihood is maximized.

4. **Solve**: Set the derivative equal to zero and solve for θ. The solution $\hat{\theta}_{MLE}$ is the MLE, the parameter value that maximizes the likelihood.

5. **Verify**: Ensure that the critical point found corresponds to a maximum (by checking the second derivative or by other means, such as verifying that the likelihood is concave).

Example 1: Maximum Likelihood for Exponential Distribution

Let $X_1, X_2, \ldots, X_n$ be a sample of independent random variables following an exponential distribution with parameter λ. The probability density function of the exponential distribution is given by:

$$f(x; \lambda) = \lambda e^{-\lambda x}, \quad x \geq 0$$

The likelihood function $L(\lambda)$ for the sample is:

$$L(\lambda) = \prod_{i=1}^{n} f(X_i; \lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda X_i}$$

This simplifies to:

$$L(\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^{n} X_i}$$

The log-likelihood function $\ell(\lambda)$ is the logarithm of the likelihood function:

$$\ell(\lambda) = \ln L(\lambda) = n \ln(\lambda) - \lambda \sum_{i=1}^{n} X_i$$

To maximize the log-likelihood, we compute the derivative of $\ell(\lambda)$ with respect to λ:

$$\frac{\partial \ell(\lambda)}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} X_i$$

By solving $\frac{\partial \ell(\lambda)}{\partial \lambda} = 0$, we obtain the maximum likelihood estimator $\hat{\lambda}$:

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} X_i}$$

Thus, the estimator for the parameter λ is the inverse of the sample mean.

Example 2: Maximum Likelihood for Normal Distribution

Let $X_1, X_2, \ldots, X_n$ be a sample from a normal distribution with unknown mean μ and variance σ^2. The probability density function is:

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

The likelihood function $L(\mu, \sigma^2)$ is:

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(X_i - \mu)^2}{2\sigma^2}\right)$$

This simplifies to:

$$L(\mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2\right)$$

The log-likelihood function $\ell(\mu, \sigma^2)$ is:

$$\ell(\mu, \sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2$$

To find the MLE for μ and σ^2, we take the partial derivatives of the log-likelihood and set them to zero.

1. Maximizing with respect to μ:

$$\frac{\partial \ell(\mu, \sigma^2)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu) = 0$$

This gives:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

2. Maximizing with respect to σ^2:

$$\frac{\partial \ell(\mu, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(X_i - \mu)^2 = 0$$

This gives:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu})^2$$

Thus, the MLE for μ is the sample mean, and the MLE for σ^2 is the sample variance.

2.2 Characteristics of an estimator

2.2.1 Bias, Mean squared error, Convergence

- **Bias**: The bias of an estimator $\hat{\theta}$ is defined as the difference between the expected value of $\hat{\theta}$ and the true parameter value θ:

$$\text{Bias}(\hat{\theta}) = \mathbb{E}(\hat{\theta}) - \theta$$

- **Mean Squared Error (MSE)**: The mean squared error (MSE) measures the average squared error of the estimator with respect to the true parameter value:
$$\text{MSE}(\hat{\theta}) = \mathbb{E}((\hat{\theta} - \theta)^2)$$
- **Convergence**: The convergence of an estimator refers to its behavior as the sample size increases. An estimator $\hat{\theta}_n$ is said to converge to θ if it approaches θ as n approaches infinity.

2.2.2 Fisher Information

The Fisher information quantity, denoted as $I(\theta)$, measures the information contained in the sample about the parameter θ. It is defined as:

$$I(\theta) = -\mathbb{E}\left(\frac{\partial^2 \ln f(X;\theta)}{\partial\theta^2}\right)$$

where $f(X;\theta)$ is the probability density (or mass) function of the probability distribution.

2.2.3 Cramer-Rao Bound

The Cramer-Rao bound establishes a lower limit on the variance of any unbiased estimator. For an unbiased estimator $\hat{\theta}$ of θ, the Cramer-Rao bound is given by:
$$\text{Var}(\hat{\theta}) \geq \frac{1}{I(\theta)}$$

2.2.4 Efficiency

An estimator $\hat{\theta}_1$ is said to be more efficient than an estimator $\hat{\theta}_2$ if it has a smaller or equal variance for all possible values of θ. That is, $\text{Var}(\hat{\theta}_1) \leq \text{Var}(\hat{\theta}_2)$ for all θ.

2.2.5 Completeness

An estimator $\hat{\theta}$ is said to be complete if it allows unbiased estimation of all functions of θ. This is an important property in the context of Bayesian estimation.

These concepts are fundamental for understanding the construction and evaluation of estimators in statistics. They play a crucial role in the selection and interpretation of estimation methods.

2.3 Estimation by confidence interval

Consider a scenario where we have a random sample, $X_1, X_2, X_3, \ldots, X_n$, originating from a distribution with an unknown parameter θ that requires estimation. We've already explored the concept of point estimation for θ. However, relying solely on the point estimate $\hat{\theta}$ doesn't provide a comprehensive understanding of θ. In essence, without additional context, we lack information about the proximity of $\hat{\theta}$ to the true θ. This leads us to introduce the concept of interval estimation.

In this approach, instead of presenting a single value $\hat{\theta}$ as the estimate for θ, we create an interval that is likely to encompass the actual value of θ. Rather than stating:

$$\hat{\theta} = 34.25,$$

we might present an interval like:

$$[\hat{\theta}_l, \hat{\theta}_h] = [30.69, 37.81],$$

with the expectation that it encompasses the true θ. In essence, we provide two estimations for θ: a higher estimate, $\hat{\theta}_h$, and a lower estimate, $\hat{\theta}_l$. Interval estimation introduces two crucial concepts. First, there's the length of the reported interval, $\hat{\theta}_h - \hat{\theta}_l$. The length of the interval reflects the precision of our θ estimation. A smaller interval signifies a more precise estimate of θ. The second critical factor is the confidence level, which indicates our confidence in the constructed interval. The confidence level represents the probability that our interval includes the true θ value. As a result, higher confidence levels are preferable. These concepts will be elaborated upon in this section.

2.3.1 The General Concept of Interval Estimation

Consider a scenario where we have a set of observations: $X_1, X_2, X_3, \ldots, X_n$, drawn from a distribution with an unknown parameter θ that we wish to estimate. In this context, our objective is twofold:

1. We seek to establish two estimators for θ:
   (a) The lower estimator, $\Theta^l = \Theta^l(X_1, X_2, \ldots, X_n)$, and
   (b) The upper estimator, $\Theta^h = \Theta^h(X_1, X_2, \ldots, X_n)$.
2. The outcome is an interval estimator, which is denoted as the range $[\Theta^l, \Theta^h]$. These estimators, Θ^l and Θ^h, are carefully chosen to ensure that the probability of the interval $[\Theta^l, \Theta^h]$ containing θ surpasses $1 - \alpha$. Here, $1 - \alpha$ represents the confidence level, and it is preferable to have a small α. Common choices for α include 0.1, 0.05, and 0.01, corresponding to confidence levels of 90%, 95%, and 99%, respectively.

Hence, when tasked with determining a 95% confidence interval for a parameter θ, our goal is to identify Θ^l and Θ^h in such a way that:

$$P(\Theta^l < \theta \text{ and } \Theta^h > \theta) \geq 0.95.$$

This discussion will gain further clarity as we delve into practical examples. Before doing so, let's formally define the concept of interval estimation.

Definition 2.3.1. *(Interval Estimation): Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample originating from a distribution with an unknown parameter θ that requires estimation. An interval estimator, with a confidence level of $1-\alpha$, comprises two estimators, $\Theta^l(X_1, X_2, \ldots, X_n)$ and $\Theta^h(X_1, X_2, \ldots, X_n)$, satisfying the condition:*

$$P(\Theta^l \leq \theta \textit{ and } \Theta^h \geq \theta) \geq 1 - \alpha,$$

for all possible values of θ. Alternatively, we state that $[\Theta^l, \Theta^h]$ is a $(1-\alpha)100\%$ confidence interval for θ.

It's worth noting that the condition:

$$P(\Theta^l \leq \theta \text{ and } \Theta^h \geq \theta) \geq 1 - \alpha$$

can also be expressed as:

$$P(\Theta^l \leq \theta \leq \Theta^h) \geq 1 - \alpha, \text{ or } P(\theta \in [\Theta^l, \Theta^h]) \geq 1 - \alpha.$$

The variability in these expressions is due to Θ^l and Θ^h, not θ. In this context, θ is the unknown quantity, assumed to be non-random (following frequentist inference). On the other hand, Θ^l and Θ^h are random variables since they depend on the observed random variables $X_1, X_2, X_3, \ldots, X_n$.

2.3.2 Determining Interval Estimators

In this section, we explore the process of deriving interval estimators. But before we delve into that, let's revisit a fundamental concept related to random variables and their distributions. Imagine having a continuous random variable, denoted as X, with a cumulative distribution function (CDF) $F_X(x)$, representing the probability that X is less than or equal to x. Our objective is to identify two values, x_h and x_l, such that the probability of X falling within this interval $[x_l, x_h]$ is equal to $1-\alpha$.

One way to achieve this is by selecting x_l and x_h such that $P(X \leq x_l)$ equals $\alpha/2$ and $P(X \geq x_h)$ equals $\alpha/2$. In other words, $F_X(x_l)$ should be $\alpha/2$, and $F_X(x_h)$ should be $1-\alpha/2$. Expressing this using the inverse function, F_X^{-1}, we get

$$x_l = F_X^{-1}(\alpha/2) \text{and} x_h = F_X^{-1}(1 - \alpha/2)$$

. This interval, $[x_l, x_h]$, is known as a $(1-\alpha)$ interval for X.

Let's illustrate this concept visually, with Figure 8.2 depicting x_l and x_h using both the CDF and the PDF of X.

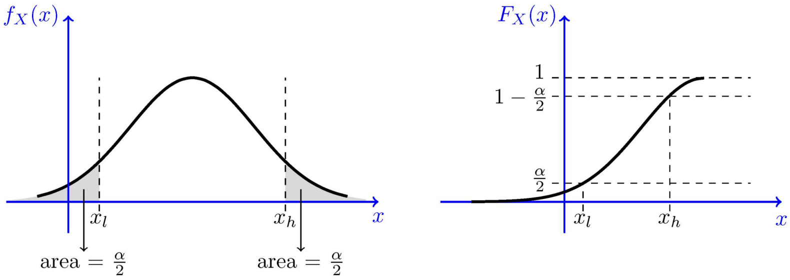

Figure 2.1: $[x_l, x_h]$ is a $(1-\alpha)$ interval for X, that is, $P(x_l \le X \le x_h) = 1-\alpha$.

Example 1. Consider a standard normal random variable $Z \sim N(0,1)$. We want to determine x_l and x_h such that $P(x_l \le Z \le x_h)$ equals 0.95.

Solution Here, α is 0.05, and we use the Φ function for the CDF of Z. Therefore, we can calculate $x_l = \Phi^{-1}(0.025) = -1.96$ and $x_h = \Phi^{-1}(1-0.025) = 1.96$. This means, for a standard normal random variable Z, $P(-1.96 \le Z \le 1.96)$ equals 0.95.

In general, we can find a $(1-\alpha)$ interval for a standard normal random variable by using the notation z_p. For any $p \in [0,1]$, z_p represents the real value for which $P(Z > z_p)$ equals p. It's also important to note that $z_{1-p} = -z_p$. Figure 8.3 visually illustrates this.

Interval Estimators: A General Approach

Now, let's discuss how we can create interval estimators. The typical approach involves starting with a point estimator $\hat{\Theta}$, such as the maximum likelihood estimator (MLE), and constructing an interval $[\hat{\Theta}_l, \hat{\Theta}_h]$ around it, ensuring that $P(\theta \in [\hat{\Theta}_l, \hat{\Theta}_h])$ is greater than or equal to $1-\alpha$. Let's consider an example to understand this process.

Example Suppose we have a random sample $X_1, X_2, \ldots, X_n$ from a normal distribution $N(\theta, 1)$. We need to find a 95% confidence interval for θ.

Solution First, we select $\hat{\Theta}$ as the point estimator for θ. Since θ represents the mean of the distribution, we can use the sample mean $\hat{\Theta} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Given that $X_i \sim N(\theta, 1)$ and the X_i's are independent, we conclude that $\bar{X} \sim N(\theta, 1/n)$.

By standardizing $\bar{X}$, we determine that the random variable $(\bar{X}-\theta)/(1/\sqrt{n})$ follows a $N(0,1)$ distribution. Therefore, by Example 8.12, we can establish $P(-1.96 \le (\bar{X}-\theta)/(1/\sqrt{n}) \le 1.96)$, which is equivalent to $P(\bar{X} - 1.96\sqrt{n} \le \theta \le \bar{X} + 1.96\sqrt{n}) = 0.95$. Hence, the 95% confidence interval for θ is $[\hat{\Theta}_l, \hat{\Theta}_h] = [\bar{X} - 1.96\sqrt{n}, \bar{X} + 1.96\sqrt{n}]$.

At first glance, this solution may seem unstructured, and you might wonder why we worked with the standardized variable $\bar{X}$. However, with more contemplation, we can develop a systematic method for solving confidence interval problems. The key insight is that the distribution of the random variable $\bar{X} - \theta$

does not depend on the unknown parameter θ but only on the observed data. Such a random variable is called a pivot or pivotal quantity.

Definition 2.3.2. *(Pivotal Quantity) A pivotal quantity Q is a function of the observed data $X_1, X_2, \ldots, X_n$ and the unknown parameter θ, but it does not depend on any other unknown parameters. Moreover, the probability distribution of Q does not rely on θ or any other unknown parameters.*

To summarize, in the pivotal method for finding confidence intervals:

1. *Identify 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. *Use algebraic manipulations to convert the above equation into one of the form $P(\hat{\Theta}_l \leq \theta \leq \hat{\Theta}_h) = 1 - \alpha$.*

In practice, for many common cases, statisticians have already determined pivotal quantities for which confidence intervals have been established. Therefore, you can often solve confidence interval problems by aligning them with previously solved problems.

Using Estimators for σ^2

When dealing with an unknown variance σ^2, we can either find an upper bound for σ^2 or estimate it. Let's explore both approaches:

1. **Upper Bound for σ^2**: If you can demonstrate that $\sigma \leq \sigma_{\max}$, where $\sigma_{\max}$ is a known real number, then you can use $\sigma_{\max}$ instead of σ to determine a confidence interval. This conservative approach ensures that the interval is valid even if σ is larger than $\sigma_{\max}$.
2. **Estimate σ^2**: In many cases, you can estimate σ^2, particularly when the sample size is large. The sample variance, $S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$, serves as an estimate for σ^2. After estimating σ^2, you can use it to establish an approximate confidence interval.

To summarize, the steps for finding confidence intervals are as follows:

- **Assumptions**: Start with a random sample $X_1, X_2, \ldots, X_n$ from a distribution.
- **Parameter to be Estimated**: Determine the parameter θ you want to estimate.
- **Confidence Interval**: If you can find an upper bound for σ or estimate σ^2, construct an approximate $(1 - \alpha)100\%$ confidence interval for θ, typically of the form

$$[\hat{\Theta} - z_{\alpha/2}\sigma/\sqrt{n}, \hat{\Theta} + z_{\alpha/2}\sigma/\sqrt{n}]$$

or

$$[\hat{\Theta} - z_{\alpha/2}S/\sqrt{n}, \hat{\Theta} + z_{\alpha/2}S/\sqrt{n}]$$

depending on the situation.

This method often leads to approximate confidence intervals, especially when the Central Limit Theorem is applied. Nevertheless, it provides a practical way to estimate unknown parameters with a known level of confidence.

2.3.3 Confidence Intervals for Normal Samples

In the previous discussion, we made an assumption that the sample size, denoted as (n), is sufficiently large, allowing us to apply the Central Limit Theorem (CLT). One interesting feature of the confidence intervals we derived was that they often did not rely on the specifics of the distribution from which the random sample was drawn. However, what happens when (n) is not large? In such cases, we cannot invoke the CLT, and we need to rely on the probability distribution from which the random sample is drawn. This situation becomes particularly important when we are dealing with a sample $(X_1, X_2, X_3, \ldots, X_n)$ taken from a normal distribution. In this context, we will explore how to determine interval estimators for both the mean and the variance of a normal distribution. Before we do that, we will introduce two probability distributions that are closely related to the normal distribution.

Chi-Squared Distribution

First, let's recall the gamma distribution. A continuous random variable (X) is said to follow a gamma distribution with parameters $(\alpha > 0)$ and $(\lambda > 0)$, denoted as $(X \sim \text{Gamma}(\alpha, \lambda))$, if its probability density function (PDF) is given by:

$$f_X(x) = \begin{cases} \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Now, we'd like to introduce a closely related distribution known as the chi-squared distribution. Consider a set of independent standard normal random variables, denoted as $(Z_1, Z_2, \ldots, Z_n)$. If we sum these variables, i.e., $(X = Z_1 + Z_2 + \ldots + Z_n)$, the resulting random variable (X) is also normally distributed, specifically as $(X \sim N(0, n)$. Next, if we define a new random variable (Y) as the square of this sum:

$$Y = Z_1^2 + Z_2^2 + \ldots + Z_n^2,$$

then (Y) follows a chi-squared distribution with (n) degrees of freedom, represented as $(Y \sim \chi^2(n))$. It can be shown that the random variable (Y) actually has a gamma distribution with parameters $(\alpha = \frac{n}{2})$ and $(\lambda = \frac{1}{2})$:

$$Y \sim \text{Gamma}\left(\frac{n}{2}, \frac{1}{2}\right).$$

The probability density function for $(\chi^2(n))$ is depicted in Figure 8.5 for different values of (n). **The Chi-Squared Distribution**