Estimation and Inference

Brian Vegetabile

2016 Statistics Bootcamp
Department of Statistics
University of California, Irvine

September 8th, 2016

General Idea of Estimation and Inference I

- Statistical inference revolves around observing some data which has been generated from an unknown probability distribution and some type of statement about the distribution will be made
  - How precisely can we estimate the expected values of the distribution?
  - What is an estimate of the spread of the distribution and can we reasonable provide bounds on these estimates?
- These insights about the distribution may inform us about how to expect similar experiments to behave in the future.

General Idea of Estimation and Inference II

- We will consider a few topics related to estimation and inference
  - Point Estimation of a Parameter - How we come up with estimators of parameters and can we evaluate these estimators?
  - Hypothesis testing -
    - Is our estimate different from some pre-believed value?
    - Are the parameters of the distributions which generated these two groups of data differ?
  - Confidence Intervals
    - Instead of a point estimate, can we obtain an estimate of an interval of likely values?

Idea of Point Estimation I

- Consider that we observe random variables $X_1, X_2, \ldots, X_n$,
- The distribution function of $X_1, \ldots, X_n$ depends on unknown parameters θ, which is known to be in some given set Θ (i.e. the parameter space).
- The problem of estimating θ is a problem of point estimation
- We will in turn estimate θ using some appropriate function of $X_1, \ldots, X_n$.

Reviewing a Statistic

- Recall the definition of a statistic...

Definition (Statistic (Casella))

Let $X_1, X_2, \dots, X_n$ be a random sample of size n from a population and let $T(x_1, \dots, x_n)$ be a real-valued or vector-valued function whose domain includes the sample space of $(X_1, X_2, \dots, X_n)$. Then the random variable or random vector $Y = T(X_1, \dots, X_n)$ is called a statistic. The probability distribution of a statistic Y is called the sampling distribution of Y.

- We emphasize again that there is a major difference between the distribution of a single data point X_i and the distribution of the sample

Relationship between Statistics, Estimators and Estimates

- It is not hard to imagine that all statistics can be used (in some fashion) as a point estimator.

Definition (Point Estimator (Casella))

A point estimator is any function $T_n(X_1, X_2, \ldots, X_n)$ of a sample; any statistic is a point estimator.

- An important distinction is that an *estimator* is a random variable (it is a function of the random sample), an **estimate** is a realization of an estimator, that is $T_n(x_1, \ldots, x_n)$.

Properties of Estimators

- We introduce some definitions to talk about properties of estimators

Definition (Unbiased Estimator)

We say $T_n(X_1, \ldots, X_n)$ is an unbiased estimator of θ if

$$E_\theta T_n(X_1, X_2, \ldots, X_n) = \theta \text{ for all } \theta \in \Theta$$

Definition (Bias of an Estimator)

The bias of the estimator T is given as

$$b_n(\theta) = E_\theta T_n(X_1, X_2, \ldots, X_n) - \theta$$

Example of Unbiased Estimators I

- Recall the sample mean and and consider a different estimator of the variance, from a population where $E(X) = \mu$ and $Var(X) = \sigma^2$.

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

- We already know $E(\bar{X}) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}n\mu = \mu$.
- Let us now consider the sample variance

Example of Unbiased Estimators II

$$
\begin{aligned}
E(S^2) &= \frac{1}{n} E\left(\sum_{i=1}^{n} (X_i - \bar{X})^2\right) \\
&= \frac{1}{n} E\left(\sum_{i=1}^{n} (X_i - \bar{X})^2\right) \\
&= \frac{1}{n} E\left(\sum_{i=1}^{n} (X_i^2 - 2X_i\bar{X} + \bar{X}^2)\right) \\
&= \frac{1}{n} E\left(\sum_{i=1}^{n} (X_i^2) - 2\bar{X}\sum_{i=1}^{n} (X_i) + n\bar{X}^2)\right)
\end{aligned}
$$

Example of Unbiased Estimators III

$$
\begin{aligned}
&= \frac{1}{n} E\left(\sum_{i=1}^{n}(X_i^2) - 2n\bar{X}\bar{X} + n\bar{X}^2)\right) \\
&= \frac{1}{n} E\left(\sum_{i=1}^{n}(X_i^2) - n\bar{X}^2\right) \\
&= \frac{1}{n} \left(\sum_{i=1}^{n} E(X_i^2) - nE(\bar{X}^2)\right) \\
&= \frac{1}{n} \left(n(\sigma^2 + \mu^2) - n(\frac{\sigma^2}{n} + \mu^2)\right) \\
&= \frac{1}{n} \left(n\sigma^2 - \sigma^2\right) = \frac{1}{n}(n-1)\sigma^2 \\
&= \frac{n-1}{n}\sigma^2
\end{aligned}
$$

Asymptotic Unbiasedness

- Some estimators may be biased in small samples, but this bias may disappear when n grows large enough.
- This motivates us to define asymptotic unbiased estimators

Definition (asymptotically unbiased estimator)

We say $T_n(X_1, X_2, \ldots, X_n)$ is an asymptotically unbiased estimator of θ if

$$\lim_{n \to \infty} b_n(\theta) = 0 \text{ for all } \theta \in \Theta$$

- We see that the last estimator of the sample variance will be asymptotically unbiased!

Consistent Estimators

- Finally, we would like to talk about estimators of θ which converge in probability to θ (i.e. their probability "collects" close to θ)

Definition (consistent sequence of estimators)

If $X_1, \ldots, X_n$ are random variables to be observed and they have distribution function $F_{X_1,\ldots,X_n}(y_1, \ldots, y_n|\theta)$ where the unknown $\theta \in \Theta$ then a sequence $T_1, T_2, \ldots$ is called a consistent sequence of estimators of θ if and only if (as $n \to \infty$)

$$T_n(X_1, \ldots, X_n) \xrightarrow{P} \theta \text{ for all } \theta \in \Theta$$

- Under this definition $\bar{X}_n$ is a consistent sequence of estimators for the population mean μ.

Mean Square Consistency

- Finally, we define one more type of estimator. This is an estimator that is consistent in mean square (or consistent in quadratic mean).

Definition (Consistent in Mean Square)

We say that $T_n(\theta_1, \ldots, X_n)$ is consistent in mean square if

$$E_\theta(T_n - \theta)^2 \to 0 \text{ as } n \to \infty \text{ for all } \theta \in \Theta$$

- The quantity $E(T_n - \theta)^2$ is called the mean squared error of the estimator T_n.
- If an estimator is consistent in mean square both its variance and bias go to zero as n grows.

Method of Moments Estimators I

- One of the simplest ways of finding parameter estimates is through the Method of Moments Estimation.
- Dates back to Karl Pearson in the 1800s
- Relies on the Weak Law of Large Numbers.
- Recall the definition of the population moments

Definition (Moment of a Random Variable)

For each random variable X and every positive integer k, $E(X^k)$ is called the k^{th} moment of X, often denoted μ'_k.

- Note that μ'_k will be a function of the parameters.

Method of Moments Estimators II

- We now define the **sample** moments

Definition (Sample Moments)

Let $X_1, \ldots, X_n$ be n random variables (not necessarily independent or identically distributed). Their k^{th} (noncentral) sample moment is

$$m'_k = \frac{1}{n} \sum_{i=1}^{n} X_i^k$$

- Each one of these will be a consistent estimate of the population moment based upon the WLLN.

Method of Moments Estimators III

- To come up with estimators based on the sample moments, we equate the sample moment to the population moment and solve
- That is we solve the following system of equations

$$\begin{array}{rcl} m'_1 & = & \mu'_1(\theta_1, \dots, \theta_d) \\ m'_2 & = & \mu'_2(\theta_1, \dots, \theta_d) \\ \vdots & & \vdots \\ m'_d & = & \mu'_d(\theta_1, \dots, \theta_d) \end{array}$$

Example - Normal MOM Estimators I

- Suppose a sample $X_1, X_2, \ldots, X_n$, iid, from a $Normal(\theta, \sigma^2)$ distribution.
- We know we have two unknown parameters and that the population moments are

$$\mu_1' = E(X) = \theta \qquad \mu_2' = E(X^2) = \sigma^2 + \theta^2$$

- The corresponding sample moments are

$$m_1' = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad m_2' = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

Example - Normal MOM Estimators II

- Thus, we need to solve the following system of equations

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} X_i &= \theta \\
\frac{1}{n}\sum_{i=1}^{n} X_i^2 &= \theta^2 + \sigma^2
\end{aligned}
$$

Example - Normal MOM Estimators III

- Continuing,

$$\hat{\theta} = \bar{X}$$

and

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\end{aligned}$$

- Notice that this is an **biased** estimate of the population variance, but that it does indeed provide an estimator for the parameters of interest!

Maximum Likelihood Estimation

- We'll now talk about one of the most popular estimation methods since it was introduced by RA Fisher in 1912.
- Maximum likelihood estimation can and is applied to most problems, though it performs best when the sample size is large.
- We'll first define the Likelihood Function

The Likelihood Function

Definition (Likelihood Function)

Let $X_1, X_2, \ldots, X_n$ be n random variables with joint pdf or pmf $f(\mathbf{x}|\theta)$ where $\theta \in \Theta$ is unknown. The likelihood function is

$$L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta)$$

- Note: That while it appears that the likelihood function is just the joint density function there is a subtle difference.
- The difference is that the likelihood function is a function of θ, where the $x_1, x_2, \ldots, x_n$ are fixed and instead we let θ vary.

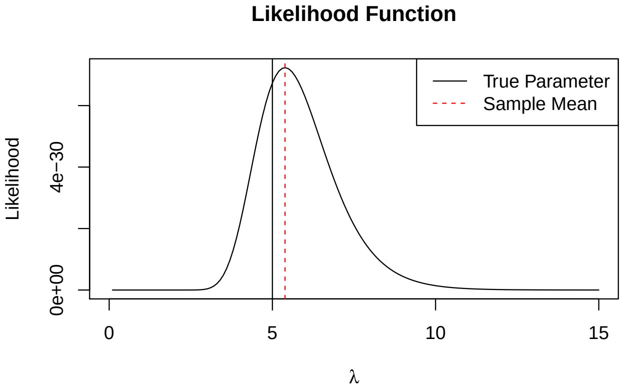
Example - Visualizing the Likelihood Function I

- Consider a random sample $X_1, \ldots, X_n$ from an $exponential(\lambda)$.
- We can find the likelihood function as follows,

$$
\begin{aligned}
L(\theta|\mathbf{x}) &= f(\mathbf{x}|\theta) \\
&= f(x_1|theta)f(x_2|theta)\ldots f(x_n|theta) \\
&= \prod_{i}^{n} f(x_i|\theta) \\
&= \left(\frac{1}{\lambda}\right)^n \exp\left\{-\frac{\sum_{i=1}^n x_i}{\lambda}\right\}
\end{aligned}
$$

Example - Visualizing the Likelihood Function II

- Let's visualize what this looks like...



- We now see that the likelihood is a function of the parameter, for fixed data

Understanding Likelihood

- Consider a discrete random variable and let $L(\theta|\mathbf{x}) = P(\mathbf{X} = \mathbf{x})$
- Comparing the likelihood functions at two different parameter values and find that

$$P_{\theta_1}(\mathbf{X} = \mathbf{x}) = L(\theta_1|\mathbf{x}) > L(\theta_2|\mathbf{x}) = P_{\theta_2}(\mathbf{X} = \mathbf{x})$$

  then the observed sample is 'more likely' to have occurred if $\theta = \theta_1$ than if $\theta = \theta_2$.
- There is a similar calculation that can be shown for continuous distributions
- This implies that we will want to find the value of θ such that the observed data is most likely to have come from that distribution

Maximum Likelihood Estimation

- We now define the maximum likelihood estimator

Definition (Maximum Likelihood Estimator)

For each sample realization $\mathbf{x}$, let $\hat{\theta}(\mathbf{x})$ be a parameter value at which $L(\theta|\mathbf{x})$ attains its maximum as a function of θ with $\mathbf{x}$ held fixed. A maximum likelihood estimator (MLE) of the parameter θ based on a sample $\mathbf{X}$ is $\hat{\theta}(\mathbf{X})$.

- Recall again that a maximum likelihood estimate is a realization of the function $\hat{\theta}(\mathbf{X})$.
- Mathematically, we want any $\hat{\theta} = \hat{\theta}_n(X_1, \ldots X_n) \in \Theta$ such that

$$L(\hat{\theta}) = \sup\{L(\theta) : \theta \in \Theta\}$$

Example - Maximizing The Likelihood I

- Recall our favorite example using the Exponential distribution
- We have shown that the likelihood is as follows,

$$L(\theta|\mathbf{x}) \;=\; \left(\frac{1}{\lambda}\right)^n \exp\left\{-\frac{\sum_{i=1}^n x_i}{\lambda}\right\}$$

- Let's maximize this function

Example - Maximizing The Likelihood II

$$\log L(\theta|\mathbf{x}) = -n \log \lambda - \frac{\sum_i^n x_i}{\lambda}$$

$$\frac{d}{d\lambda} \log L(\theta|\mathbf{x}) = -\frac{n}{\lambda} + \frac{\sum_{i=1}^n x_i}{\lambda^2}$$

- Setting to 0

$$\begin{aligned} 0 &= -\frac{n}{\lambda} + \frac{\sum_{i=1}^n x_i}{\lambda^2} \\ \frac{n}{\lambda} &= \frac{\sum_{i=1}^n x_i}{\lambda^2} \\ \hat{\lambda} &= \frac{1}{n} \sum_{i=1}^n x_i = \bar{x} \end{aligned}$$

Invariance of the MLE

- As with the creation of the delta method, we may not always want to estimate a parameter itself.
- What may be more interesting is some function of the MLE.

Theorem (Invariance Property of MLEs)

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

- This theorem (unproven) allows us to talk about MLE's of functions of parameters.

Bayesian Estimation I

- We won't develop all of Bayesian estimation here, but we will talk about some of the fundamentals.
- Ultimately Bayesian statistics is interested in the distribution of the parameter given the observed data

$$p(\theta|\mathbf{x})$$

- We also have the distribution of the data if we specify a model for that data, that is we have

$$f(\mathbf{x}|\theta)$$

Bayesian Estimation II

- Finally we may be able to use expert opinion to generate a distribution for likely values of the parameter

$$p(\theta)$$

- Under Bayes Theorem

$$\begin{aligned} p(\theta|\mathbf{x}) &= \frac{f(\mathbf{x}|\theta)p(\theta)}{p(\mathbf{x})} \\ &\propto f(\mathbf{x}|\theta)p(\theta) \end{aligned}$$

Bayesian Estimation III

- We define $p(\theta)$ as the prior distribution, $f(\mathbf{x}|\theta)$ is the likelihood,
- We define $p(\theta|\mathbf{x})$ as the posterior distribution
- Simplifying things

$$posterior \propto likelihood \times prior$$

- Bayesian estimation is done by choosing properties of the distribution which represent high posterior density
- Often the posterior median or mean are used, both offer nice theoretical properites

Bayesian Statistics

We often use probabilities to informally express our beliefs about unknown quantities. Bayesian statistics formalizes a way to bring *prior* information or personal belief into an analysis.

Bayesian vs. Frequentist Philosophy

Frequentist

- Data are a repeatable random sample - there is a *frequency*
- Underlying parameters remain constant during this repeatable process
- Parameters are fixed

Bayesian

- Data are observed from the realized sample.
- Parameters are unknown and described probabilistically
- Data are fixed

http://www.stat.ufl.edu/archived/casella/Talks/BayesRefresher.pdf

Bayesian Statistics Example I

- Consider a random sample $X_1, X_2, \ldots, X_n$ from an exponential distribution parameterized such that $E(X_i) = 1/\lambda$.

$$f(x_i|\theta) = \lambda \exp\left(-\lambda x_i\right)$$

- Notice the form of this distribution, as a function of λ it appears to have a gamma distribution
- Additionally the parameter λ is restricted from 0 to ∞.
- This implies that a gamma distribution may be a good prior distribution for exponential data!

$$p(\lambda) \propto \lambda^{\alpha-1} \exp\left(-\lambda\beta\right)$$

Bayesian Statistics Example II

- From our examples we know that the likelihood is

$$f(\mathbf{x}|\lambda) = \lambda^n \exp(-\lambda \sum_{i=1}^{n} x_i)$$

- Now from earlier we'll have that

$$\begin{aligned} p(\lambda|\mathbf{x}) &\propto f(\mathbf{x}|\lambda)p(\lambda) \\ &\propto \lambda^{(n+\alpha)-1} \exp\left(-\lambda\left(\beta + \sum_{i=1}^{n} x_i\right)\right) \end{aligned}$$

- We see that this is a $Gamma(n+\alpha, \beta + \sum x_i)$
- Thus a point estimate may be the posterior mean $E(\lambda|\mathbf{x}) = \frac{n+\alpha}{\beta+\sum x_i}$

The General Idea of Hypothesis Testing I

- We begin by talking about the general idea of hypothesis testing
- A hypothesis test mainly is concerned with deciding whether the parameter θ lies in one subset of the parameter space or another

The General Idea of Hypothesis Testing II

- A hypothesis testing problem usually results from asking specific questions...
  - Does smoking cause cancer?
  - Do atomic power plants increase radiation levels?
  - Are the heights of students the same at two different schools?
- There is some underlying parameter θ in each one of these examples
- We will wish to determine whether it changes in specified ways when an element of a system is changed.

General Outline of a Hypothesis Test

- This is the general outline of hypothesis testing which is provided in a basic introductory statistics book (Mind on Statistics - Utts)
    1. Determine the null and alternative hypotheses.
    2. Verify necessary data conditions, and if they are met, summarize the data into an appropriate test statistic.
    3. Assuming that the null hypothesis is true, find the p-value.
    4. Decide whether the result is statistically significant based on the p-value.
    5. Report the conclusion in the context of the situation.

Defining Hypotheses I

- The goal of a hypothesis test is to decide, based upon our sample, which of two complementary hypotheses is true.

Definition (Hypothesis)

A hypothesis is a statement about a population parameter

Definition (Null and Alternative Hypotheses)

Two complementary hypotheses in a hypothesis testing problem are called the null hypothesis and the alternative hypothesis. Denoted H_0 and H_1 respectively.

Defining Hypotheses II

- If θ denotes a population parameter, the general format of the null and alternative hypotheses is $H_0 : \theta \in \Theta_0$ and $H_1 : \theta \in \Theta_0^C$
- In the above Θ_O is a subset of the parameter space.
- As an example consider testing to see if a drug affects blood pressure, we may be interested if the parameter θ which denotes if a change in BP is equal to 0 or not.

$$H_0 : \theta = 0 \quad vs. \quad H_1 : \theta \neq 0$$

- In a hypothesis test we will either reject the null hypothesis or fail to reject the null hypothesis.

Deciding to Accept or Reject a Null Hypothesis

- We will ultimately be interested in values of that of the random sample for which we will reject the null hypothesis.
- Consider a random sample, $X_1, X_2, \ldots, X_n$ and let Ω be the sample space of $\mathbf{X} = (\mathbf{X_1}, \ldots, \mathbf{X_n})$.
  - There is a subset of Ω where we will fail to reject H_0 and a subset where will reject H_1.

Definition (critial region)

The subset of Ω for which H_0 will be rejected is called the critical region of the test

- In most hypothesis testing problems the critical region is defined in terms of a test statistic, $T(\mathbf{X})$, a function of the data. "Reject H_0 if $T \geq c$.

Types of Errors of Tests I

- Once we have a critical region we will be interested in characteristics of the test
- We must consider the types of errors that could arise from this testing procedure.

Definition (Type I Error)

In a hypothesis testing problem, rejecting H_0 when it is true is called an error of type I.

- The probability of rejecting H_0 depends on the test used and the true value of θ and is called the level of significance of the test.

Types of Errors of Tests II

Definition

Type II Error Failing to reject H_0 when it is false is called an error of type II.

	Accept H_0	Reject H_0
H_0	Correct Decision	Type I Error
H_1	Type II Error	Correct Decision

Power I

- In designing a test it will be necessary to have a test which exhibits good properties in relation to these errors.
- Consider a rejection region R then for $\theta \in \Theta_0$, the test will make a mistake if $\mathbf{x} \in R$, so the probability of a type I error is $P_\theta(\mathbf{X} \in R)$.
- For $\theta \in \Theta_0^C$, it can be shown that the probability of a Type II Error is $1 - P_\theta(\mathbf{X} \in R)$.
- There for $P_\theta(\mathbf{X} \in R)$ contains all the information about the test.

Power II

Definition (power function)

The power function of a hypothesis test with rejection region R is the function of θ defined by

$$\beta(\theta) = P_\theta(\mathbf{X} \in R)$$

- An ideal power function is 0 for all $\theta \in \Theta_0$ and 1 for $\theta \in \Theta_0^C$. This clearly unattainable in real life.
- See Hal Stern's notes from 210 for a good demonstration of a power calculation.
- There are many more ways of evaluating tests, but these are the basics.

Example of a Hypothesis Test I

- Suppose $\mathbf{X} = (X_1, \ldots, X_n)$ is a random sample from normal distribution with unknown mean μ and known variance σ^2.
- We would like to test the hypothesis

$$\begin{aligned} H_0 &: \mu = \mu_0 \\ H_1 &: \mu \neq \mu_0 \end{aligned}$$

- A reasonable test statistic that would provide an estimate of our parameter would be the sample mean
- Additionally it would be reasonable to reject H_0 if $\bar{X}$ is far from the null μ_0, or if $|\bar{X} - \mu_0| > c$.
- Therefore we could create a testing procedure that rejects H_0 is $|\bar{X} - \mu_0| > c$.

Example of a Hypothesis Test II

- We additionally know that $\bar{X} - \mu_0$ will be a normal distribution centered at 0 with distribution σ^2/n
- We can use this to find a value c such that we achieve a specified significance level, α
- Traditionally, we will use the statistic

  $$Z = \sqrt{n}\frac{\bar{X}_n - \mu_0}{\sigma}$$

  and reject H_0 if

  $$|Z| \geq \Phi^{-1}(1 - \alpha/2)$$

  for a symmetric test.
- Here $\Phi(\cdot)$ is the CDF of a standard normal distribution.

Confidence Intervals

- Confidence intervals provide a little more insight than what is received by a point estimate
- That is we may be more interested in say an ‘interval estimate’ of probable values for a parameter θ.
- Additionally, there is an intimate relationship between confidence intervals and interval estimation and hypothesis testing.

Defining a Internval Estimate

- From DeGroot and Schervish, suppose that $X_1, \ldots, X_n$ form a random sample from a distribution that involves a parameter θ whose value is unknown.
- Suppose two statistics $L(\mathbf{X})$ and $U(\mathbf{X})$ can be found such that

$$P(L(\mathbf{X}) < \theta < U(\mathbf{X})) = \gamma$$

where γ is a fixed probability $(0 < \gamma < 1)$.

Coverage Probability of an Interval Estimator

- We can also consider how often we should our intervals to capture the true parameter
- Clearly both $L(\mathbf{X})$ and $U(\mathbf{X})$ are random and thus we will not always capture the true parameter θ
- The coverage probability is defined

$$P(\theta \in [L(\mathbf{X}), U(\mathbf{X})] | \theta)$$

- We see that in the definition from DeGroot γ will be our coverage probability.

Inverting a Test Statistic I

- If we know the distribution of a test statistic, this can often aid in finding a confidence interval.
- Specifically, we will know the distribution of the test statistic **as a function of the parameter of interest**

$$P(a < T(\mathbf{X}) < b) = 1 - \alpha$$

- In the case of a symmetric distribution we could find the points a and b of the CDF such that we have equal tails

Inverting a Test Statistic II

- For example consider a random sample from a Normal distribution and we want a CI for the mean
- We know that

$$\sqrt{n}\frac{\bar{X} - \mu}{\sigma} \sim N(0,1)$$

- Therefore the points a and b will be $Z_{\alpha/2}$ and $Z_{1-\alpha/2}$.
- Thus

$$P(Z_{\alpha/2} < \sqrt{n}\frac{\bar{X} - \mu}{\sigma} < Z_{1-\alpha/2}) = 1 - \alpha$$

Inverting a Test Statistic III

$$
\begin{aligned}
P\left(Z_{\alpha/2} < \sqrt{n}\frac{\bar{X}-\mu}{\sigma} < Z_{1-\alpha/2}\right) &= 1-\alpha \\
P\left(Z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}} < \bar{X}-\mu < Z_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}}\right) &= 1-\alpha \\
P\left(-\bar{X}+Z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}} < -\mu < -\bar{X}+Z_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}}\right) &= 1-\alpha \\
P\left(\bar{X}-Z_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}} < \mu < \bar{X}-Z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}\right) &= 1-\alpha
\end{aligned}
$$

Inverting a Test Statistic IV

- Thus we have created our two test statistics which satisfy the desired definition of an interval estimate

$$\begin{aligned} L(\mathbf{X}) &= \bar{X} - Z_{1-\alpha/2}\sqrt{\frac{\sigma}{n}} \\ U(\mathbf{X}) &= \bar{X} - Z_{\alpha/2}\sqrt{\frac{\sigma}{n}} \end{aligned}$$

- While symmetric distributions are easiest to do this with, CIs can be found for most distributions

Interpretations of Confidence Intervals

- We first state what a CI is not: It is not correct to say that θ lies in the observed interval (a, b) with probability $1 - \alpha$.
- Prior to observing values we would expect with probability $1 - \alpha$ that the parameter will be contained in the interval.
- Thus a more appropriate interpretation is that in many many repeated experiments, where in each experiment we calculate a confidence interval, $1 - \alpha$ of these experiments will contain the true paramater.

One last note on Confidence Intervals

- Once last thing about confidence intervals is that they can intimately be related to hypothesis testing
- An interpretation of a confidence interval is that it is the set of all null hypotheses which are consistent with the observed data
- This implies that if the confidence interval captures the null hypothesis then it is also consistent with the observed data and we will fail to reject the null hypothesis.
- DeGroot and Schervish discuss this in detail.

References