Point and Interval Estimation, and Testing of Hypothesis

by

Dr. Suchandan Kayal

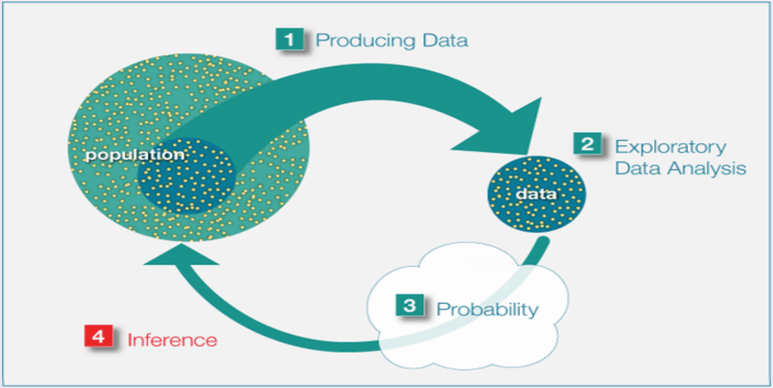
Department of Mathematics
National Institute of Technology Rourkela
Rourkela - 769008, Odisha, India

Autumn, 2020

Outline

- Motivated example
- Point estimation
  - Properties of a good estimator
  - Method of moment estimator
  - Method of likelihood estimator
- Interval estimation
  - Interval estimation of μ of a normal distribution $N(\mu, \sigma^2)$ when σ^2 is known/unknown
  - Interval estimation of σ^2 of a normal distribution $N(\mu, \sigma^2)$ when μ is known/unknown
  - Bootstrap confidence interval
- Testing of hypothesis
  - Neyman-Pearson fundamental lemma
  - Most powerful test
  - Uniformly most powerful test

Diagram of the Steps



(a)

Statistical inference

- Producing data- how data are obtained, and what considerations affect the data production process.
- Exploratory data analysis-tools that help us get a first feel for the data, by exposing their features using visual displays and numerical summaries which help us explore distributions, compare distributions, and investigate relationships.
- We use probability to quantify how much we expect random samples to vary.
- In statistical inference, we infer something about the population based on what is measured in the sample.

Goal!

In statistical inference, we draw conclusions about a population based on the data obtained from a sample chosen from it.

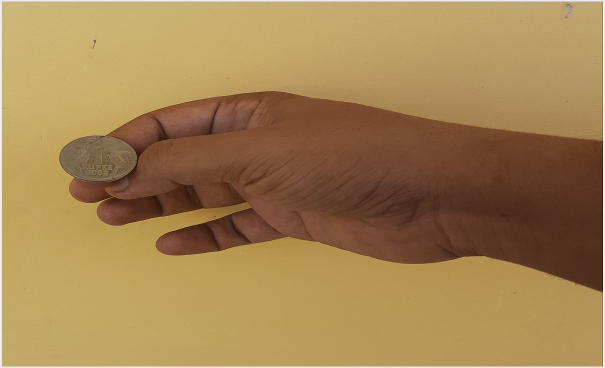
Models

Classification of models

$$\text{Models} \rightarrow \begin{cases} \text{Parametric Models} \\ \text{Non-parametric Models} \end{cases}$$

- Parametric model is associated with finite-dimensional parameter.
- In case of the non-parametric model, the number and nature of parameters is flexible and not fixed in advance.

Resolving disputed ownership of a painting by tossing a coin.



(b)

An example

Resolving disputed ownership of a painting by tossing a coin.

- Suppose two brothers Ramesh and Suresh agree to resolve their disputed ownership of a painting by tossing a coin.
- Ramesh produces a coin to Suresh. Suresh tosses the coin. Ramesh calls Head. The coin comes to rest with Head facing up. Thus, Ramesh takes possession of the property.
- At the same evening, Suresh got disappointed and decided to conduct an experiment. He tosses the same coin 100 times and observed 68 Head.
- Therefore, Suresh thought the coin he tossed in the morning may not be entirely fair. But, he is unwilling to accuse his brother for the coin.

How will Suresh proceed?

(Statistical Inference!)

Modeling the example

Suresh's experiment can be modeled as follows:

- Here, each toss is a Bernoulli trial and the experiment is a sequence of $n = 100$ trials. Let X_i denote the outcome of the toss i. Then,

$$X_i = \begin{cases} 1, & \text{if Head is observed} \\ 0, & \text{if Tail is observed.} \end{cases}$$

- Thus, we have $X_1, \ldots, X_{100} \sim Bernoulli(p)$, where p (probability of occurrence of head) is fixed but unknown to Suresh. Now,

$$Z = \sum_{i=1}^{100} X_i \sim Binomial(100, p).$$

Questions!

Now, Suresh is interested to draw inferences about this fixed but unknown quantity. We consider three questions he might ask.

- What is the true value of p? More precisely, what is the resonable guess as to the true value of p?
- What are the possible values of p? In particular, is there a subset of $[0, 1]$ that Suresh can confidently claim contains the true value of p?
- Is $p = 0.5$? Specifically, is there any evidence that $p \neq 0.5$? So that Suresh can completely accuse Ramesh for giving an unfair coin.

Point Estimation

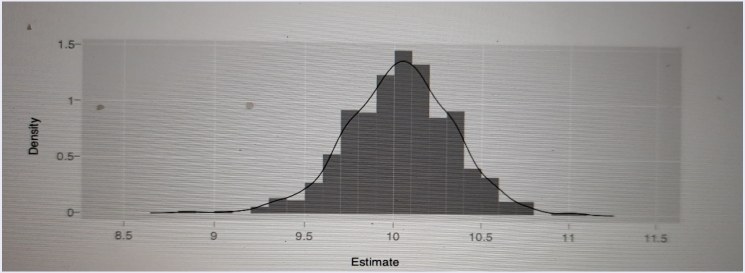
- Let $X_1, \ldots, X_n$ be a random sample drawn from a model (parametric).
- Based on this random sample, we wish to come up with a function that will estimate the unknown model parameters.
- Let $\delta_n = g(X_1, \ldots, X_n)$ be an estimator. It is only useful for inference if we know something about how it behaves.
- Note that δ_n is a function of the random sample. It is a random variable and its behavior depends on the sample size n.

Illustration

- Consider estimating the population mean of a normally distributed population, say $N(\mu, 9)$.
- The most obvious estimate is to simply draw a sample and calculate the sample mean.
- If we repeat this process with a new sample, we would expect to get a different estimate.
- The distribution that results from repeated sampling is called the sampling distribution of the estimate.

Point estimation

500 estimates of the population mean based on a sample size 100



(c)

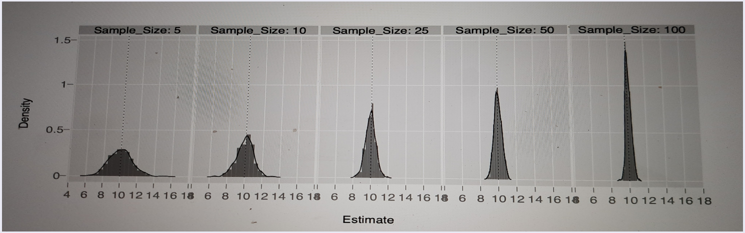
- Figure c illustrates 500 estimates of the population mean based on a sample of size 100.
- We can see that our estimates are generally centered around the true value of 10, but there is some variation: maybe a standard deviation of about 1.
- These observations translate to more formal ideas: the expectation of the estimator and the standard error of the estimator.

So, what do we look for in a good estimator?

Expected behaviours

- We want our estimate to be close to the true value.
- Also, we want δ_n to behave in a nice way as the sample size n increases.
- If we take a large sample, we would like the estimate to be more accurate than a small sample.

500 estimates of the population mean based on sample sizes of 5, 10, 25, 50 and 100



(d)

- Each histogram in Figure d represents 500 estimates of the population mean for sample sizes of 5, 10, 25, 50 and 100.
- We can see that the standard deviation of the estimate is smaller as the sample size increases. Formally, this is embodied in the principle of consistency.
- A consistent estimator will converge to the true parameter value as the sample size increases.
- Our estimator $\bar{X}$ for the population mean of a normal seems very well behaved.

Properties

- Consistency- An estimator δ_n of θ is said to be consistent if δ_n converges to θ in probability. Mathematically,

$$P(|\delta_n - \theta| > 0) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

- Unbiasedness- An estimator δ_n is said to be unbiased of θ if

$$E(\delta_n) = \theta.$$

- Efficiency-An estimator is efficient if it has the lowest possible variance among all unbiased estimators.

Method of moments (MOM) estimators

- Let X be a random variable following some distribution, say $f(x|\theta)$, where $\theta = (\theta_1, \ldots, \theta_k)$. Then the kth moment of the distribution is defined as

$$\mu_k = E(X^k).$$

- The sample moments based on the random sample $X_1, \ldots, X_n$ drawn from this distributon are given by

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

- The MOM estimator simply equates the moments of the distribution with the sample moments ($\mu_k = \hat{\mu}_k$) and solves for the unknown parameters. Note that this implies the distribution must have finite moments.

Example-1

Let $X_1, \ldots, X_n$ be a random sample drawn from a Poisson population with probability mass function

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!},\ x = 0, 1, 2\ldots,$$

where $\lambda > 0$. Obtain the MOM estimator of λ.

Solution

Here $E(X) = \lambda$. So that $\mu_1 = E(X) = \lambda = \bar{X} = \hat{\mu}_1$. Hence, the method of moments estimator of λ is the sample mean.

Example-2

Let $X_1, \ldots, X_n$ be a random sample drawn from a gamma population with probability density function

$$f_X(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x},\ x > 0,\ \alpha, \lambda > 0.$$

Obtain the MOM estimators of λ and α.

Solution

The first two moments of the gamma distribution are $\mu_1 = \frac{\alpha}{\lambda}$ and $\mu_2 = \frac{\alpha(\alpha+1)}{\lambda^2}$. After solving these two equations, we get the MOM estimators, which are given by

$$\hat{\lambda} = \frac{\hat{\mu}_1}{\hat{\mu}_2 - (\hat{\mu}_1)^2} = \frac{\bar{X}}{\bar{X^2} - (\bar{X})^2} \text{ and } \hat{\alpha} = \hat{\lambda}\hat{\mu}_1.$$

Example-3

Suppose X is a discrete random variable with the probability mass function

$$P(X=x)=\begin{cases}\frac{2\theta}{3}, & x=0\\ \frac{\theta}{3}, & x=1\\ \frac{2(1-\theta)}{3}, & x=2\\ \frac{1-\theta}{3}, & x=3,\end{cases}$$

where $0 \leq \theta \leq 1$. The following 10 independent observations were taken from this distribution: $(3,0,2,1,3,2,1,0,2,1)$. Use the MOM method to find the estimate of θ.

Solution

Here, the theoretical mean $\mu_1 = E(X) = \frac{7}{3} - 2\theta$. The sample mean is $\hat{\mu}_1 = \bar{X} = 1.5$. Equating these two means, we have $\hat{\theta} = \frac{5}{12}$.

MOM estimators

Example-4

Let $X_1, \ldots, X_n$ be a random sample from a population with probability density function

$$f_X(x|\sigma) = \begin{cases} \frac{1}{2\sigma} e^{-\frac{|x|}{\sigma}}, & -\infty < x < \infty \\ 0, & otherwise, \end{cases}$$

where $\sigma > 0$. Please use the method of moment to estimate σ.

Solution

If we calculate the first order theoretical moment, we would have

$$E(X) = \int_{-\infty}^{\infty} x \frac{1}{2\sigma} e^{-\frac{|x|}{\sigma}} dx = 0.$$

Thus, if we try to solve equation $E(X) = \bar{X}$, we will not get the estimator, because $E(X)$ does not contain the unknown parameter σ.

MOM estimators

Solution (cont...)

Now, let us calculate the second order theoretical moment, we have

$$\mu_2 = E(X^2) = \int_{-\infty}^{\infty} x^2 \frac{1}{2\sigma} e^{-\frac{|x|}{\sigma}} dx = 2\sigma^2.$$

The second order sample moment is

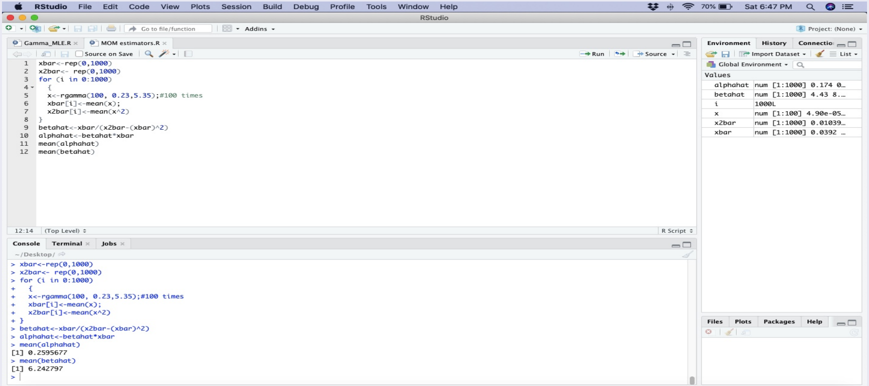
$$\hat{\mu}_2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2.$$

Solving the equation $\mu_2 = \hat{\mu}_2$, we get the estimate of σ as

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n} X_i^2}{2n}}.$$

R-progamming for Example-2

To investigate the MOM on simulated data using R software, we consider 1000 repetitions of 100 independent observations of a $Gamma(0.23, 5.35)$ random variable.



Properties of the MOM estimators

Nice properties:

- it is consistent
- sometimes easier to calculate than other methods say maximum likelihood estimates

Not so nice properties:

- sometimes not sufficient. (Sufficiency has a formal definition but intuitively it means that all the data that are relevant to estimating the parameter of interest are used.)
- sometimes gives estimates outside the parameter space

Maximum likelihood Estimator (MLE)

- Let $X_1, X_2, \ldots, X_n$ be a random vector of observations with joint density function $f_X(x_1, \ldots, x_n|\theta)$.
- Then, the likelihood of θ as a function of the observed values, $X_i = x_i$, is defined as,

$$L(\theta) = f_X(\theta|x_1, \ldots, x_n).$$

- The MLE of the parameter θ is the value of θ that maximizes the likelihood function.
- In general, it is easier to maximize the natural log of the likelihood. In the case that the X_i' are iid, the log likelihood is generally of the form,

$$l(\theta) = \log L(\theta).$$

MLE

Example-5

Let $X_1, \ldots, X_n$ be a random sample from a population with probability mass function

$$P(X = x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!}, & x = 0, 1, 2, \ldots \\ 0, & otherwise, \end{cases}$$

where $\lambda > 0$. Please use the method of maximum likelihood to estimate λ.

Solution

The log-likelihood function is

$$l(\lambda) = \log \lambda \sum_{i=1}^{n} x_i - n\lambda - \sum_{i=1}^{n} \log x_i!.$$

MLE

Solution (cont...)

To find the maximum we set the first derivative to zero,

$$l'(\lambda) = \frac{1}{\lambda}\sum_{i=1}^{n} x_i - n = 0.$$

Solving for λ we find that the MLE is

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

NOTE:- Note that this agrees with the MOM estimator (see Example-1).

Example-6

Let $X_1, \ldots, X_n$ be a random sample drawn from a gamma population with probability density function

$$f_X(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x},\ x > 0, \quad \alpha, \lambda > 0.$$

Obtain the MLEs of λ and α.

Solution

The log likelihood is

$$l(\lambda, \alpha) = n\alpha \log \lambda - n \log \Gamma(\alpha) + (\alpha - 1) \sum_{i=1}^{n} \log x_i - \lambda \sum_{i=1}^{n} x_i.$$

Solution (cont...)

In this case we have two parameters so we take the partial derivatives and set them both to zero.

$$\frac{\partial l}{\partial \alpha} = \sum_{i=1}^{n} \log x_i + n \log \lambda - n \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} = 0.$$

$$\frac{\partial l}{\partial \lambda} = \frac{n\alpha}{\lambda} - \sum_{i=1}^{n} x_i = 0.$$

MLE

Solution (cont...)

This second equality gives the MLE for λ as

$$\hat{\lambda} = \frac{\hat{\alpha}}{\bar{X}}.$$

Substituting this into the first equation we find that the MLE for α must satisfy,

$$n \log \hat{\alpha} - n \log \bar{X} + \sum_{i=1}^{n} \log x_i - n \frac{\Gamma'(\hat{\alpha})}{\Gamma(\hat{\alpha})} = 0.$$

This equation needs to be solved by numerical means.

NOTE:- Note however this is a different estimate to that given by the method of moments.

Example-7

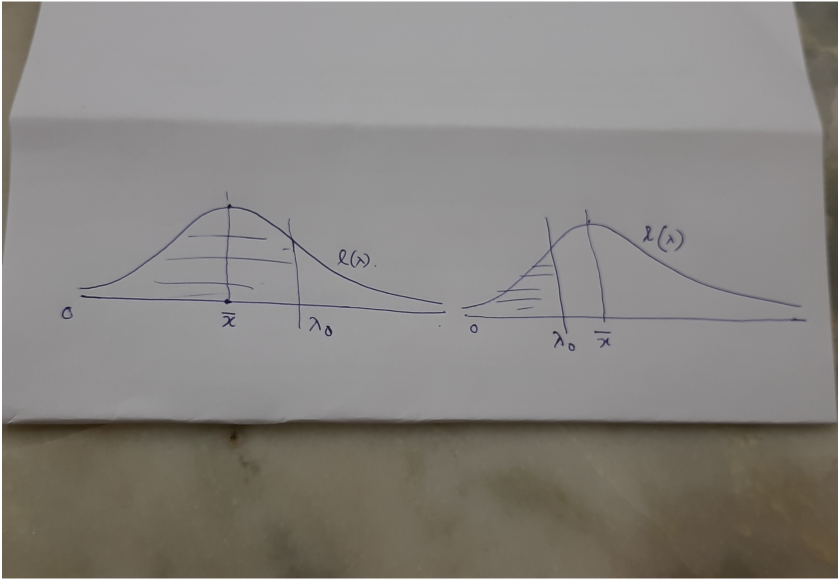
Consider Example-5 with the additional assumption that $\lambda \leq \lambda_0$. Obtain the MLE of λ.

Solution

In this case, the MLE of λ is

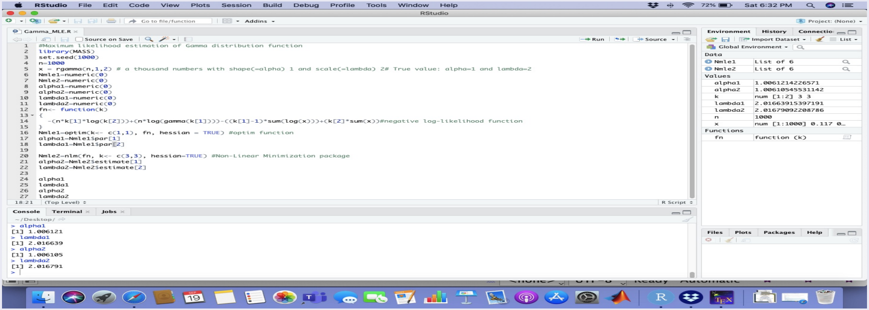
$$\hat{\lambda}_{RML} = \begin{cases} \bar{X}, & \bar{X} \leq \lambda_0 \\ \lambda_0, & \bar{X} > \lambda_0. \end{cases}$$

MLE



R-progamming for Example-6

To investigate the MLE on simulated data using R software, we consider independent observations of a $Gamma(1,2)$ random variable.



(g)

Properties of the MLE

Nice properties:

- consistent
- unaffected by monotonic transformations of the data
- MLE of a function of the parameters, is that function of the MLE
- theory provides large sample properties
- asymptotically efficient estimators

Not so nice properties:

- may be slightly biased
- can be computationally demanding

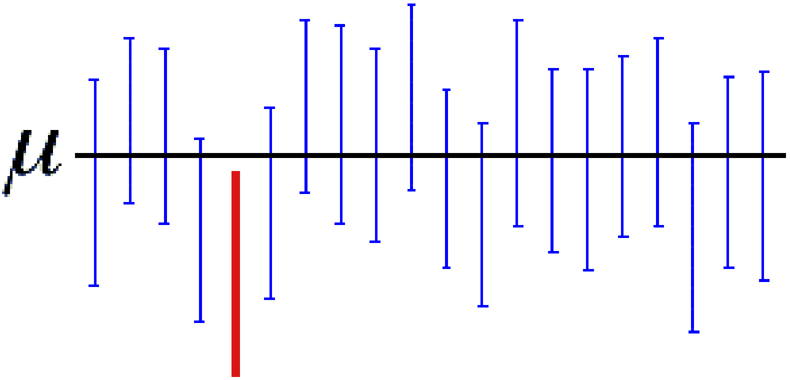
Interval estimation

Confidence interval

Let X be a random variable with distribution $P_{\boldsymbol{\theta}}, \ \boldsymbol{\theta} \in \Theta$. Consider a random sample $X_1 \ldots, X_n$ drawn from this distribution. Let $\delta_1(\boldsymbol{X})$ and $\delta_2(\boldsymbol{X})$ be two statistics such that

$$P(\delta_1(\boldsymbol{X}) \leq g(\boldsymbol{\theta}) \leq \delta_2(\boldsymbol{X})) = 1 - \alpha, \ \forall \ \theta \in \Theta.$$

Then, if the random sample $\boldsymbol{X} = \boldsymbol{x}$ is observed, we say that $[\delta_1(\boldsymbol{x}), \delta_2(\boldsymbol{x})]$ is a $100 \times (1 - \alpha)\%$ confidence interval for $g(\boldsymbol{\theta})$.

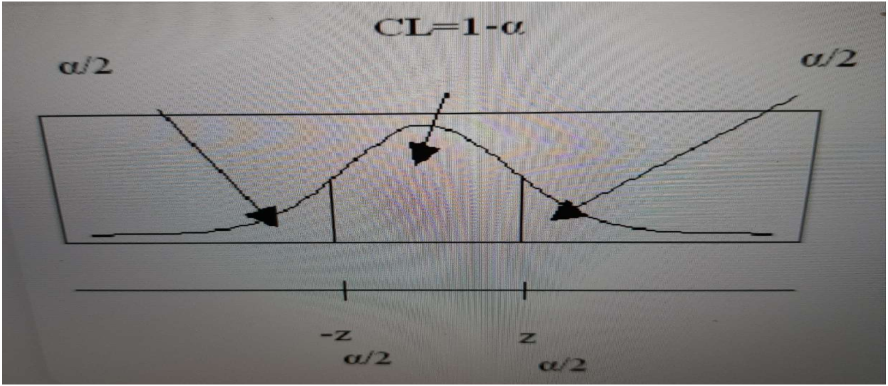


A 95% confidence interval indicates that 19 out of 20 samples (95%) from the same population will produce confidence intervals that contain the population parameter.

Interval estimation

I. Confidence interval of μ of $N(\mu, \sigma^2)$, when σ^2 is known

- Let $X_1, \ldots, X_n$ be a random sample drawn from normal population with unknown mean μ and known variance σ^2.
- Then, $\bar{X} \sim N(\mu, \sigma^2/n)$ and $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$.
- $P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha$.
- Then, $(\bar{x} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2}, \bar{x} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2})$ is a $100 \times (1-\alpha)\%$ confidence interval for μ.
- For example, if $\bar{x} = 2$, $\sigma = 1$, $n = 4$ and $\alpha = 0.05$, then $z_{\alpha/2} = 1.96$. Thus, the confidence interval is $(1.02, 2.98)$.

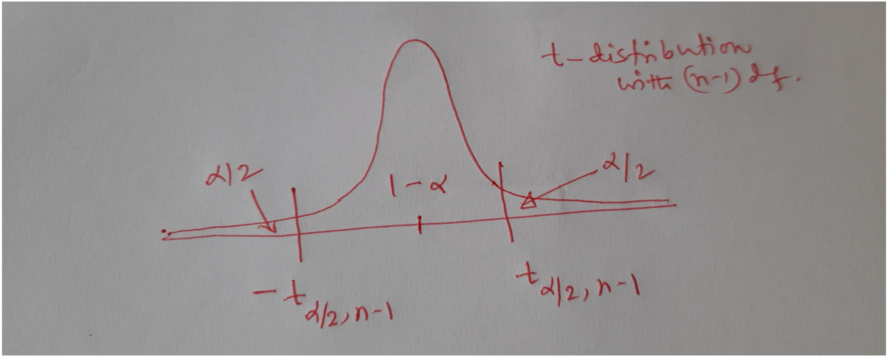


CL=1-α
α/2
α/2
-z
z
α/2
α/2

II. Confidence interval of μ of $N(\mu, \sigma^2)$, when σ^2 is unknown

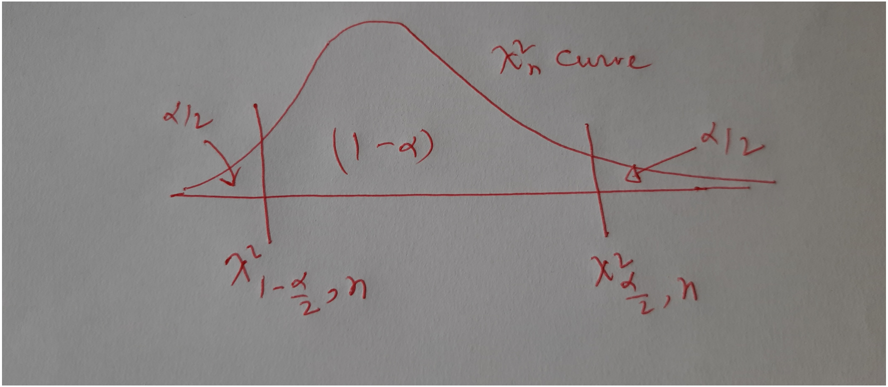
- Let $X_1, \ldots, X_n$ be a random sample drawn from normal population with unknown mean μ and unknown variance σ^2.
- $T = \frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t_{n-1}$, where s^2 is the sample variance, given by $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$.
- $P(-t_{\alpha/2,n-1} \leq T \leq t_{\alpha/2,n-1}) = 1 - \alpha$.
- Then, $(\bar{x} - \frac{s}{\sqrt{n}}t_{\alpha/2,n-1}, \bar{x} + \frac{s}{\sqrt{n}}t_{\alpha/2,n-1})$ is a $100 \times (1-\alpha)\%$ confidence interval for μ.
- For example, if $\bar{x} = 0.0506$, $s = 0.004$, $n = 10$ and $\alpha = 0.05$, then $t_{\alpha/2,9} = 2.262$. Thus, the confidence interval is $(0.0477, 0.0535)$.

Interval estimation



III. Confidence interval of σ^2 of $N(\mu, \sigma^2)$, when μ is known

- Let $X_1, \ldots, X_n$ be a random sample drawn from normal population with known mean μ and unknown variance σ^2.
- $W = \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} \sim \chi_n^2$.
- $P(\chi^2_{1-\alpha/2,n} \leq W \leq \chi^2_{\alpha/2,n}) = 1 - \alpha$.
- Then, $\left(\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\chi^2_{\alpha/2,n}}, \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\chi^2_{1-\alpha/2,n}}\right)$ is a $100 \times (1-\alpha)\%$ confidence interval for σ^2.
- Then, $\left(\sqrt{\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\chi^2_{\alpha/2,n}}}, \sqrt{\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\chi^2_{1-\alpha/2,n}}}\right)$ is a $100 \times (1-\alpha)\%$ confidence interval for σ.


χ^2_n curve
$\alpha/2$
$(1-\alpha)$
$\alpha/2$
$\chi^2_{1-\frac{\alpha}{2}, n}$
$\chi^2_{\frac{\alpha}{2}, n}$

IV. Confidence interval of σ^2 of $N(\mu, \sigma^2)$, when μ is unknown

- Let $X_1, \ldots, X_n$ be a random sample drawn from normal population with unknown mean μ and unknown variance σ^2.
- $W^* = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$.
- $P(\chi^2_{1-\alpha/2,n-1} \leq W^* \leq \chi^2_{\alpha/2,n-1}) = 1 - \alpha$.
- Then, $\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}\right)$ is a $100 \times (1-\alpha)\%$ confidence interval for σ^2.
- Then, $\left(\sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}}, \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}}\right)$ is a $100 \times (1-\alpha)\%$ confidence interval for σ.

V. Bootstrap confidence interval

- In 1979, Brad Efron invented a revolutionary new statistical procedure called the bootstrap.
- This is a computer-intensive procedure that substitutes fast computation for theoretical math. Surprisingly, the idea is quite simple.

Benefit

- The main benefit of the bootstrap is that it allows statisticians to set confidence intervals on parameters without having to make unreasonable assumptions.

Bootstrap confidence interval

- First, you gather the sample from $f(x|\theta)$ and use it to find $\hat{\theta}$, your estimate the unknown parameter θ.
- Then draw a new random sample of size n, with replacement, from $f(x|\hat{\theta})$. This is like drawing with replacement from a box in which each ticket is labeled with an observation in the initial random sample.
- This second sample is called a bootstrap sample. For that bootstrap sample, we can calculate an estimate of the parameter of interest for $f(x|\hat{\theta})$. Denote this new estimate by θ_1^*.
- We can draw as many bootstrap samples of size n as we want, obtaining M estimates $\theta_1^*, \ldots, \theta_M^*$. This will be used to get the bootstrap confidence interval.

$$
\begin{array}{ccccccccc}
f & \longrightarrow & \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} & \Longrightarrow & \hat{f} & \longrightarrow & \begin{bmatrix} X^*_{1(1)} \\ X^*_{2(1)} \\ \vdots \\ X^*_{n(1)} \end{bmatrix}, & \ldots, & \begin{bmatrix} X^*_{1(M)} \\ X^*_{2(M)} \\ \vdots \\ X^*_{n(M)} \end{bmatrix} \\
 & & \downarrow & & & & \downarrow & & \downarrow \\
 & & \hat{\theta} & & & & \hat{\theta}^*_1, & \ldots, & \hat{\theta}^*_M
\end{array}
$$

Bootstrap confidence interval

- Arrange $\theta_1^*, \ldots, \theta_M^*$ in acending order. For example,

$$\theta_{(1)}^*, \ldots, \theta_{(M)}^*.$$

- Thus, the $100(1-\alpha)\%$ approximate bootstrap (p) confidence interval for θ is given by

$$\left(\hat{\theta}^*_{\left(\frac{M\alpha}{2}\right)},\ \hat{\theta}^*_{\left(M\left(1-\frac{\alpha}{2}\right)\right)}\right).$$

  Then, the percentile bootstrap confidence interval of θ at 95% level of confidence is $(\hat{\theta}^*_{(25)}, \hat{\theta}^*_{(975)})$.

Testing of hypothesis

Testing of statistical hypothesis

Definition

A statistical hypothesis is an assertion or conjecture concerning one or more populations.

Examples (problems of testing of hypothesis)

- A brand of drug for a disease is applied to the patients and the rate of cure is denoted by p_0. It is known that $p_0 = 0.5$. A new drug is produced by another company. Somebody is interested to know the effectiveness of the new drug. Let p be the rate of cure using the new drug. Then, we are interested to check if $p > 0.5$.
- We have some data of weights of new born babies in a particular city. We are intested to test if the measurements follow a normal distribution, that is, $X \sim N(\mu, \sigma^2)$.

Testing of statistical hypothesis

Null and alternative hypothesis

- Null hypothesis (H_0) is a first tentative specification (starting point) about the probability model (hypothesis actually to be tested).
- Alternative hypothesis (H_1) is a statement that directly contradicts a null hypothesis by stating that the actual value of a population parameter is less than, greater than, or not equal to the value stated in the null hypothesis.

Example

- Suppose in a coin tossing experiment, we want to test whether the coin is fair or not. Let p be the probability of occurrence of head, then we want to test whether $p = \frac{1}{2}$ or $p \neq \frac{1}{2}$.

$$H_0 : p = \frac{1}{2} \quad \text{vs} \quad H_1 : p \neq \frac{1}{2}$$

Testing of statistical hypothesis

Simple and composite hypothesis

- A hypothesis is called simple if it completely specifies the probability distribution and otherwise composite.

Example

- Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$ population.
- Simple hypothesis:

$$H_0 : \mu = 0,\ \sigma = 1 \text{ vs } H_1 : \mu = 1,\ \sigma = 1.$$

- Composite hypothesis:

$$H_0 : \mu = 0,\ \text{vs } H_1 : \mu = 1.$$

Test of statistical hypothesis

A test of hypothesis is a decision to accept or reject a hypothesis. The hypothesis we want to test is if H_1 is "likely" true. So, there are two possible outcomes:

- Reject H_0 and accept H_1 because of sufficient evidence in the sample in favor of H_1;
- Do not reject H_0 because of insufficient evidence to support H_1.

Very important!!

Note that failure to reject H_0 does not mean the null hypothesis is true. There is no formal outcome that says "accept H_0." It only means that we do not have sufficient evidence to support H_1.

Example

In a jury trial the hypotheses are:

$$H_0 : \text{defendant is innocent} \quad H_1 : \text{defendant is guilty.}$$

H_0 (innocent) is rejected if H_1 (guilty) is supported by evidence beyond "reasonable doubt." Failure to reject H_0 (prove guilty) does not imply innocence, only that the evidence is insufficient to reject it.

Critical or acceptance regions

Let S be the sample space of the random experiment.

- A critical region or the rejection region of the test is that part of the sample space that corresponds to the rejection of the null hypothesis. If $X \in S_R$, we reject the null hypothesis. S_R is called the rejection or critical region.
- $S_R^c = S_A$ is known as the aceptance region. If $X \in S_A$, we do not reject the null hypothesis.

Example

We have a coin. We toss the coin thrice to check whether the probability of head is 1/4 or 3/4. Here,

$$S = \{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT\}.$$

The hypotheses are

- $H_0 : p = 1/4$,
- $H_1 : p = 3/4$.

The acceptence and critical regions can be set as follows:
$S_A = \{HTT, THT, TTH, TTT\}$, $S_R = \{HHH, HHT, HTH, THH\}$.
That is, if 2 or 3 heads are observed, we go in favour of H_1 and if 0 or 1 head is observed we go for H_0.

Example

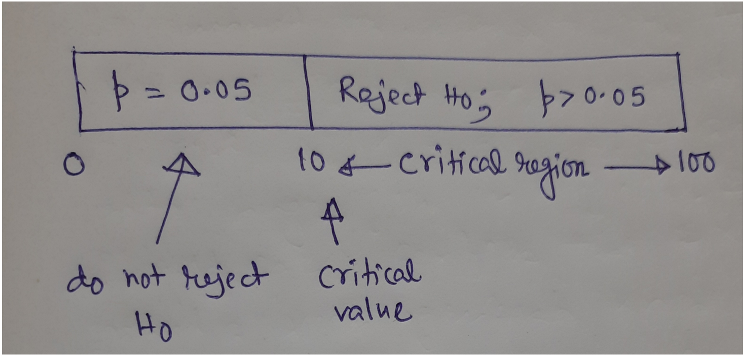
A company manufacturing RAM chips claims the defective rate of the population is 5%. Let p denote the true defective probability. We want to test if

- $H_0 : p = 0.05$
- $H_1 : p > 0.05$

We are going to use a sample of 100 chips from the production to test.

Let X denote the number of defective in the sample of 100. Reject H_0 if $X \geq 10$ (chosen "arbitrarily" in this case). X is called the test statistic.

Testing of hypothesis



Why did we choose a critical value of 10 for this example?

Because this is a Bernoulli process, the expected number of defectives in a sample is np. So, if $p = 0.05$ we should expect $100 \times 0.05 = 5$ defectives in a sample of 100 chips. Therefore, 10 defectives would be strong evidence that $p > 0.05$.

Errors

Because we are making a decision based on a finite sample, there is a possibility that we will make mistakes. The possible outcomes are:

	H_0 is true	H_1 is true
Do not reject H_0	Correct decision	Type II error
Reject H_0	Type I error	Correct decision

(n)

Type I error

The acceptance of H_1 when H_0 is true is called a Type I error. The probability of committing a Type I error is called the level of significance and is denoted by α.

Example

Convicting the defendant when he is innocent!

The lower significance level α, the less likely we are to commit a type I error. Generally, we would like small values of α; typically, 0.05 or smaller.

Type I error

$$
\begin{aligned}
\alpha &= P(\text{Type I error}) \\
&= P(\text{reject } H_0 \text{ when } H_0 \text{ is true}) \\
&= P(X \geq 10 | p = 0.05) \\
&= \sum_{x=10}^{100} b(x; n = 100, p = 0.05) \\
&= \sum_{x=10}^{100} \binom{100}{x} (0.05)^x (1 - 0.05)^{100-x} \\
&= 0.0282
\end{aligned}
$$

So, the level of significance is $\alpha = 0.0282$.

Type II error

Failure to reject H_0 when H_1 is true is called a Type II error. The probability of committing a type II error is denoted by β.

Note

It is impossible to compute β unless we have a specific alternate hypothesis.

Type II error

We cannot compute β for $H_1 : p > 0.05$ because the true p is unknown. However, we can compute it for testing

- $H_0 : p = 0.05$ against the alternative hypothesis that
- $H_1 : p = 0.1$, for instance.

$$\begin{aligned}\beta &= P(\text{type II error}) \\ &= P(\text{reject } H_1 \text{ when } H_1 \text{ is true}) \\ &= P(X < 10 | p = 0.1) \\ &= \sum_{x=0}^{9} b(x; n = 100, p = 0.1) \\ &= 0.4513.\end{aligned}$$

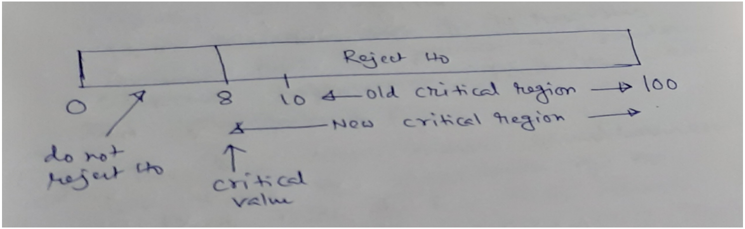
Type II error

What is the probability of a type II error if $p = 0.15$?

$$
\begin{aligned}
\beta &= P(\text{type II error}) \\
&= P(X < 10 | p = 0.15) \\
&= \sum_{x=0}^{9} b(x; n = 100, p = 0.15) \\
&= 0.0551.
\end{aligned}
$$

Moving the critical value provides a trade-off between α and β. A reduction in β is always possible by increasing the size of the critical region, but this increases α. Likewise, reducing α is possible by decreasing the critical region.

Let us see what happens when we change the critical value from 10 to 8. That is, we reject H_0 if $X \geq 8$.



(o)

The new significance level is

$$\begin{aligned}
\alpha &= P(\text{Type I error}) \\
&= P(\text{reject } H_0 \text{ when } H_0 \text{ is true}) \\
&= P(X \geq 8 | p = 0.05) \\
&= \sum_{x=8}^{100} b(x; n = 100, p = 0.05) \\
&= \sum_{x=8}^{100} \binom{100}{x} (0.05)^x (1 - 0.05)^{100-x} \\
&= 0.128
\end{aligned}$$

As expected, this is a large value than before (it was 0.0282).

Type II error

$$
\begin{aligned}
\beta &= P(\text{type II error}) \\
&= P(\text{reject } H_1 \text{ when } H_1 \text{ is true}) \\
&= P(X < 8 | p = 0.1) \\
&= \sum_{x=0}^{7} b(x; n = 100, p = 0.1) \\
&= 0.206,
\end{aligned}
$$

which is lower than before. Again, testing against the alternate hypothesis $H_1 : p = 0.15$, $\beta = 0.012$, again, lower than before.

Effect of the sample size

Both α and β can be reduced simultaneously by increasing the sample size.

Consider that now the sample size is $n = 150$ and the critical value is 12. Then, reject H_0 if $X \geq 12$, where X is now the number of defectives in the sample of 150 chips.

The new significance level is

$$\begin{aligned}\alpha &= P(X \geq 12 | p = 0.05) \\ &= \sum_{x=12}^{150} b(x; n = 150, p = 0.05) \\ &= 0.074.\end{aligned}$$

Note that this value is lower than 0.128 for $n = 100$ and critical value of 8.

Testing against the alternate hypothesis $H_1 : p = 0.1$

$$\begin{aligned}\beta &= P(X < 12 | p = 0.01)\\ &= \sum_{x=0}^{11} b(x; n = 150, p = 0.01)\\ &= 0.171,\end{aligned}$$

which is also lower than before (it was 0.206).

Power of a test

The power of a test is the probability of rejecting H_0 given that a specific alternate hypothesis is true. That is,

$$Power = 1 - \beta.$$

Properties of hypothesis testing

- α and β are related; decreasing one generally increases the other.
- α can be set to a desired value by adjusting the critical value. Typically, α is set at 0.05 or 0.01.
- Increasing n decreases both α and β.
- β decreases as the distance between the true value and hypothesized value (H_1) increases.

Neyman-Pearson fundamental lemma

- Randomized test procedure- For any value x, a randomized procedure chooses between two decisions rejection or acceptance with certain probabilities say $\phi(x)$ and $1-\phi(x)$.
- $\phi(x)$ is known as the probability of rejecting H_0 when $X=x$ is observed. This is also known as the critical function or test function.

Neyman-Pearson fundamental lemma

Let π_0 and π_1 be populations with distributions say f_0 and f_1, respectively. Then for testing

- $H_0 : f = f_0$
- $H_1 : f = f_1$,

where f_0 and f_1 are known (fixed), we can define a test ϕ with constant k such that

$$E_0\phi(X) = \alpha$$

and

$$\phi(x) = \begin{cases} 1, & f_1(x) > kf_0(x) \\ 0, & f_1(x) \leq kf_0(x). \end{cases}$$

Neyman-Pearson fundamental lemma

If ϕ satisfies above two conditions for some k, then it is most powerful test for H_0 against H_1 and level α. Conversely, if ϕ is the most powerful test of level α for testing H_0 against H_1, then for some k, ϕ satisfies second condition almost everywhere. It also satisfies the first condition unless there exists a test of size less than α and power 1.

Example-1

Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, 1)$ population. The hypotheses are

- $H_0 : \mu = \mu_0$
- $H_1 : \mu = \mu_1$.

We want to find out the most powerful test of size α for H_0 against H_1 when $\mu_0 < \mu_1$.

Solution

The joint density function of $X_1, \ldots, X_n$ is

$$f_\mu(\boldsymbol{x}) = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}\sum_{i=1}^n x_i^2 - \frac{n\mu^2}{2} + \mu \sum_{i=1}^n x_i}.$$

Solution (cont..)

By Neyman-Pearson fundamental lemma, the test is reject H_0 when

$$\begin{aligned}\frac{f_1(\boldsymbol{x})}{f_0(\boldsymbol{x})} &\geq k \\ \Rightarrow e^{\frac{n\mu_1^2}{2}-\frac{n\mu_0^2}{2}} e^{(\mu_1-\mu_0)n\bar{x}} &\geq k \\ \Rightarrow e^{(\mu_1-\mu_0)n\bar{x}} &\geq k_1 \\ \Rightarrow (\mu_1-\mu_0)n\bar{x} &\geq k_2 \\ \Rightarrow \bar{x} &\geq k_3\end{aligned}$$

It is known that $\bar{X} \sim N(\mu_0, \frac{1}{n})$, when H_0 is true.

Solution (cont..)

$$\begin{aligned}
\alpha &= P(\bar{X} \geq k_3 | H_0 \text{ is true}) \\
&= P(\sqrt{n}(\bar{X} - \mu_0) \geq \sqrt{n}(k_3 - \mu_0)) \\
&= P(Z \geq \sqrt{n}(k_3 - \mu_0)) \\
&= P(Z \geq z_\alpha).
\end{aligned}$$

So, the test is reducing to

- Reject H_0 if $\sqrt{n}(\bar{X} - \mu_0) \geq z_\alpha$
- Do not reject H_0 if $\sqrt{n}(\bar{X} - \mu_0) < z_\alpha$.

- Suppose $\mu_0 = 0$ and $\mu_1 = 1$; $n = 25$, $\bar{x} = 0.2$ and $\alpha = 0.05$. Then, $z_{0.05} = 1.645$. Here $z = \sqrt{n}(\bar{x} - \mu_0) = 1$. Thus, we can not reject the null hypothesis $H_0 : \mu = 0$.
- Let $\bar{x} = 0.4$. Then, $z = 2$. In this case, the null hypothesis is rejected at 5% level of significance.
- Let us change $\alpha = 0.01$. Then $z_\alpha = 2.32$. Here, H_0 is accepted.

Example-2

Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, 1)$ population. The hypotheses are

- $H_0 : \mu = \mu_0$
- $H_1 : \mu = \mu_1$.

We want to find out the most powerful test of size α for H_0 against H_1 when $\mu_0 > \mu_1$.

Solution

The MP test is

- Reject H_0 if $\sqrt{n}(\bar{X} - \mu_0) \leq -z_\alpha$
- Do not reject H_0 if $\sqrt{n}(\bar{X} - \mu_0) > -z_\alpha$.

- Suppose $\mu_0 = 0$ and $\mu_1 = -1$; $n = 25$, $\bar{x} = -0.6$ and $\alpha = 0.05$. Then, $-z_{0.05} = -1.645$. Here $z = \sqrt{n}(\bar{x} - \mu_0) = -3$. Thus, we reject the null hypothesis $H_0 : \mu = 0$.

Example-3

Let $X_1, \ldots, X_n$ be a random sample from $N(0, \sigma^2)$ population. The hypotheses are

- $H_0 : \sigma^2 = \sigma_0^2$
- $H_1 : \sigma^2 = \sigma_1^2$.

We want to find out the most powerful test of size α for H_0 against H_1 when $\sigma_0^2 < \sigma_1^2$.

Solution

The joint density function of $X_1, \ldots, X_n$ is

$$f_\mu(\boldsymbol{x}) = \frac{1}{(\sigma_0\sqrt{2\pi})^n} e^{-\frac{1}{2\sigma_0^2}\sum_{i=1}^n x_i^2}.$$

Solution (cont..)

By Neyman-Pearson fundamental lemma, the test is reject H_0 when

$$\frac{f_1(\boldsymbol{x})}{f_0(\boldsymbol{x})} \geq k$$

$$\Rightarrow \sum_{i=1}^{n} x_i^2 \geq k_2$$

It is known that $Y_i = \frac{X_i}{\sigma_0} \sim N(0,1)$, when H_0 is true. Thus, $\frac{\sum_{i=1}^{n} X_i^2}{\sigma_0^2} \sim \chi_n^2$.

Solution (cont..)

So, the test is

- Reject H_0 if $\frac{\sum_{i=1}^{n} X_i^2}{\sigma_0^2} \geq \chi^2_{n,\alpha}$
- Do not reject H_0 if $\frac{\sum_{i=1}^{n} X_i^2}{\sigma_0^2} < \chi^2_{n,\alpha}$

For the case of $\sigma_0^2 > \sigma_1^2$, the test is

- Reject H_0 if $\frac{\sum_{i=1}^{n} X_i^2}{\sigma_0^2} \leq \chi^2_{n,1-\alpha}$
- Do not reject H_0 if $\frac{\sum_{i=1}^{n} X_i^2}{\sigma_0^2} > \chi^2_{n,1-\alpha}$

Uniformly most powerful test

- The UMP test occurs for some particular family of distributions, called monotone likelihood ratio property (MLR).

MLR

Let $f(x, \theta)$ be a probability density function of a random variable X. Define $r(x) = \frac{f(x,\theta_1)}{f(x,\theta_2)}$, $\theta_1 > \theta_2$. If $r(x)$ is an increasing function of $T(x)$, then we say that the family of densities $\{f(x, \theta) : \theta \in \Omega\}$ has MLR in $(\theta, T(x))$.

Example

Let $X \sim N(\theta, 1)$. Then, $r(x) = e^{\frac{1}{2}(\theta_2^2 - \theta_1^2) + (\theta_1 - \theta_2)x}$ is an increasing function of x (if $\theta_1 > \theta_2$). So, $\{N(\theta, 1) : \theta \in \mathbb{R}\}$ has MLR in (θ, x).

Theorem (Lehmann and Romano, 2005)

Let the random variable X have density function $f(x,\theta)$ with MLR in $(\theta, T(x))$, $\theta \in \Theta$. For testing

- $H_0 : \theta \leq \theta_0$
- $H_1 : \theta > \theta_0$

there exists a uniformly most powerful test, given by

$$\phi(x) = \begin{cases} 1, & T(x) > c \\ \gamma, & T(x) = c \\ 0, & T(x) < c, \end{cases}$$

where c and γ are determined by $E_{\theta_0}\phi(X) = \alpha$.

Example

Let $X_1, \ldots, X_n$ be a random sample from double exponential distribution with density function

$$f(x, \theta) = \frac{1}{2\theta} e^{-\frac{|x|}{\theta}},\ x \in \mathbb{R},\ \theta > 0.$$

The hypothesis are $H_0 : \theta \leq \theta_0$ and $H_1 : \theta > \theta_0$. Obtain the UMP test of size α.

Solution

The given density if MLR in $(\theta, \sum_{i=1}^{n} |X_i|)$. So, the UMP test for this problem is reject H_0 if $\sum_{i=1}^{n} |X_i| \geq c$, where c is to be determined from

$$E_{\theta_0} \phi(\boldsymbol{X}) = \alpha.$$

Solution

It is known that when $\theta = \theta_0$, $\frac{2\sum_{i=1}^n |X_i|}{\theta_0} \sim \chi^2_{2n}$. Thus,

$$P_{\theta=\theta_0}\left(\frac{2\sum_{i=1}^n |X_i|}{\theta_0} \geq \frac{2c}{\theta_0}\right) = \alpha \Rightarrow \frac{2c}{\theta_0} = \chi^2_{2n,\alpha}.$$

Thus, the UMP test of size α is

- reject H_0 if $\frac{2\sum_{i=1}^n |X_i|}{\theta_0} \geq \chi^2_{2n,\alpha}$
- accept H_0 if $\frac{2\sum_{i=1}^n |X_i|}{\theta_0} < \chi^2_{2n,\alpha}$.

Thank You