ST5201: Basic Statistical Theory
Chapter 9: Testing Hypotheses

CHOI Yunjin
stachoiy@nus.edu.sg

Department of Statistics and Applied Probability
National University of Singapore (NUS)

31st October, 2017

Announcement

- Assignment 4 due tomorrow (due on 1st November).

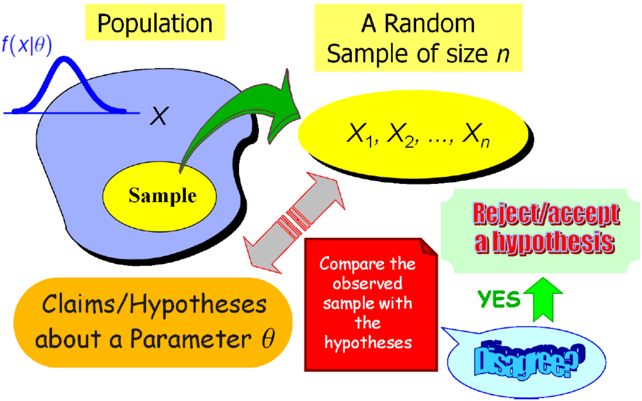
- We will review the midterm next week the slides will not be uploaded.

Outline

- Introduction
- The Null and the Alternative Hypothesis
- Significance level
- Likelihood Ratio Test
- p-value

Mandatory Reading

- Section 9.1 - Section 9.4



$f(x|\theta)$
Population
X
Sample
A Random Sample of size n
$X_1, X_2, \ldots, X_n$
Claims/Hypotheses about a Parameter θ
Compare the observed sample with the hypotheses
Disagree?
YES
Reject/accept a hypothesis

- A *statistical hypothesis test* is a method of statistical inference using data from a scientific study. In statistics, a result is called statistically significant if it has been predicted as unlikely to have occurred by chance alone, according to a pre-determined threshold probability, the significance level.
- These tests are used in determining what outcomes of a study would lead to a rejection of the null hypothesis for a pre-specified level of significance; this can help to decide whether results contain enough information to cast doubt on conventional wisdom, given that conventional wisdom has been used to establish the null hypothesis.

Drug effect

A neurologist is testing the effect of a new drug on response time by injecting 100 rats with a unit dose of the drug, injecting each to neurological stimulus, and recording its response time. The neurologist knows that the mean response time for rats not injected with the drug is 1.2 seconds. The mean of the 100 injected rats' response is 1.05 seconds with a sample standard deviation of 0.5 seconds. Do you think that the drug has an effect on response time?

- Drug has no effect ($H_0 : \mu = 1.2s$)
- Drug has an effect ($H_1 : \mu \neq 1.2s$)

Communication Time

- At the beginning of this semester, you were planning when to leave your apartment for an evening class starting at 7 PM at NUS on certain weekdays
- **A claim/hypothesis**: You may feel that the mean commuting time μ between your apartment and NUS in order to reach NUS at 7 PM is less than 30 minutes during that period of time
- **To verify this claim**: Collect some data/samples by recording the commuting times during the same period of time taken for, say, 20 days
- **To conclude**: Decide whether your hypothesis is correct based on your observed samples

Post Delivery Time

- A real example
  - I sent 60 international packages 05/2013-05/2014 from Geneva to Hangzhou, with one package being detoured to Brazil
  - The *La poste* claims that their mean time for any priority mail to arrive in Beijing Customs is no more than 14 days
- **A claim/hypothesis**:
  - $H_0 : \mu \leq 14$
  - $H_a : \mu > 14$
- **To verify this claim**: Collected receipts of the mails and all the recorded arrival time (the time of each package arriving in Beijing)
- **To conclude**: Decide whether the hypothesis is correct based on the samples

- To start with:
  - A random phenomenon of interest with population summarized by a parametric probability model $f(x|\theta)$ *with unknown parameter* $\theta \in \Theta$
  - 2 *complementary hypotheses/claims* describing the population in terms of θ
- **The goal**: is to decide, based on the sample from the population, which of the two complementary hypotheses is true
- **Elements in a hypothesis test**:
  - Null and alternative hypotheses, H_0 and H_a
  - Test statistic and testing criteria
  - Significance level
  - p value and interpretation

Definition

A hypothesis testing procedure or hypothesis test is a *rule* or *criterion* that specifies:

- For which sample values the decision is made not to reject H_0, and, in turn, accept H_0 as true
- For which sample values H_0 is rejected, and, in turn, H_a is accepted to be true

The set of the observation/sample values for which H_0 will be rejected is called the *rejection region (RR)* or *critical region*. The complement of the rejection region is called the acceptance region

- The rule is associated with a test statistic, which is a function of the sample and takes different forms for different $f(x|\theta)$'s or θ's

Type I and Type II Error

- We are expecting making some error, since no one knows the truth

$$\text{type I error} = \text{erroneous rejection of } H_0 \text{ while } H_0 \text{ is true.}$$
$$\text{type II error} = \text{erroneous retention of } H_0 \text{ while } H_1 \text{ is true.}$$

- Significance level $0 < \alpha < 1$: the probability of committing a type I error.
  Probability of committing a type II error is usually denoted by $0 < \beta < 1$

Decision Based on Sample	Truth About the Population: H_0 True	Truth About the Population: H_A True
Reject H_0	Type I Error	Correct Decision
Fail to Reject H_0	Correct Decision	Type II Error

- The rejection region (RR) is derived according to both the sampling distribution of the test statistics under the assumption of H_0 and the significance level α

	H_0 is valid: Innocent	**H_0 is invalid: Guilty**
Reject H_0 **I think he is guilty!**	Type I error False positive Convicted!	Correct outcome Positive Convicted!
Don't reject H_0 **I think he is innocent!**	Correct outcome Negative Freed!	Type II error False negative Freed!

- Suppose H_0 means "the defendant is innocent". A positive correct outcome would be "letting the defendant who is innocent go free", and a negative correct outcome would be "convicting the guilty defendant"
- type I error can be thought of as "convicting an innocent person" (a more serious mistake) and type II error "letting a guilty person go free"

- The general format of H_0 and H_a is

$$H_0 : \theta \in \Theta_0 \text{ and } H_a : \theta \in \Theta_1$$

where $\Theta_0, \Theta_1 \in \Theta$ and $\Theta_0 \cap \Theta_1 = \emptyset$

- is equivalent to investigating whether the samples are generated from a probability model in

$$\{f(x|\theta)|\theta \in \Theta_0\} \text{ or } \{f(x|\theta)|\theta \in \Theta_1\}$$

assuming there is no overlap f between the above two collections

- Difference between estimation (Ch.8) and hypothesis testing (Ch.9):
    - **Recall**: In estimation (Ch.8), we aim at choosing the best value/values in the whole parameter space Θ to serve as a guess of the true but unknown value of a parameter θ, with confidence interval
    - There are 2 meaningful and contrasting collections of possible values of θ in Θ set as H_0 and H_a in the first place and our goal in a hypothesis test (Ch.9) is to choose one over the other, i.e., in which collection is the true value of the parameter θ?

Two hypotheses according to the nature of Θ_i, $i = 0, 1$

A hypothesis Θ_i consisting of only one value in Θ is a *simple hypothesis*; otherwise it is a composite hypothesis

- There exists an asymmetry between H_0 and H_a, illustrated from determination/choice of the two hypotheses:
  - H_0 is usually set as the status quo, i.e., a claim/norm believed to be true by most people
  - A hypothesis test is derived based on the assumption that H_0 is true
  - We have a nontrivial claim/assertion about θ_0 which we wish it to be correct and set it as H_a
  - H_a must be complementary to H_0 and is nontrivial compared to H_0 as it overturns the general belief if H_a was true

- If θ denotes the average change in a patient's blood pressure after taking a new drug, we are interested in testing $H_0 : \theta = 0$ versus $H_a : \theta \neq 0$ (i.e., no effect versus some effect)
- H_0 states that there is, on the average, the drug has no effect on blood pressure. Without any proof, a new drug is assumed to be ineffective by the general audience (which is the status quo)
- H_a states that there is some effect; in case this was true, it would bring in substantial changes and impact on usual practice in controlling blood pressure

One must have sufficient evidence to overturn a general belief!

- In the "Post Delivery Time" example, we have $H_0 : \mu \leq 14$ versus $H_a : \mu > 14$
- **The logic**:calculate the probability of having observed this sample (e.g., 60 packages with one taking exceptionally long time), if the null is true; reject the null if this probability is too small. "Too small" implies that we should not have collected this sample, if the null is true.
- Not rejecting the null does not mean the null is "actually" true; it only means that we do not have enough evidence to support the alternative.

Note: The consequence of rejecting a H_0 in the previous example about "effectiveness of a new drug" $\Rightarrow$ Usually, rejecting a true H_0/committing a type I error is more consequential than failing to reject a false H_0/committing a type II error

Definition

Significance level $0 < \alpha < 1$ of a hypothesis test is a prespecified small probability taken into account when the rejection region/rule of a test is derived s.t. when the decision of a test is to reject H_0, the probability that H_0 is actually correct would be α

- α is chosen to be as small as .05 or .01, before analyzing the data

- The likelihood ratio method of hypothesis testing is one possible approach/method in relation to the mles in finding a test procedure/test statistic
- The likelihood ratio tests (LRT's) are as widely applicable as the method of maximum likelihood
- **The Setup**:
  - Given an i.i.d. random sample of size n, with observed values $x_1, \cdots, x_n$ from a population with density $f(x|\theta)$ and $\theta \in \Theta$
  - The likelihood function (Ch.8) is defined as

$$\text{lik}(\theta) = f(x_1, \cdots, x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

LRT for two simple hypotheses

Suppose that $H_0 : \theta = \theta_0$ and $H_a : \theta = \theta_1$ are simple hypotheses with $\theta_0, \theta_1 \in \Theta$. The likelihood-ratio test statistic is defined by

$$R = \frac{\text{lik}(\theta_0)}{\text{lik}(\theta_1)}$$

which is a function of the sample $x_1, \cdots, x_n$. A LRT with significance level $0 < \alpha < 1$ is a test that has a rejection region of the form

$$R \leq c \text{ (reject if the ratio is too small)}$$

where $c \geq 0$ is chosen so that $P(R \leq c | H_0) = \alpha$

The Z Test for two Simple Hypotheses

Let $X_1, \cdots, X_n$ be a random sample from $N(\mu, \sigma^2)$ having known variance $\sigma^2 > 0$. Consider two simple hypotheses:

$$H_0 : \mu = \mu_0 \text{ and } H_a : \mu = \mu_1$$

where $-\infty < \mu_0 < \mu_1 < \infty$ are given constants. To derive the LRT with significance level $0 < \alpha < 1$:

- Calculate the likelihood ratio, or, LR statistic:

$$R = \frac{\exp\left\{\frac{-1}{2\sigma^2}\sum_{i=1}^n (X_i - \mu_0)^2\right\}}{\exp\left\{\frac{-1}{2\sigma^2}\sum_{i=1}^n (X_i - \mu_1)^2\right\}} = e^{\frac{-1}{2\sigma^2}\left\{\sum_{i=1}^n (X_i - \mu_0)^2 - \sum_{i=1}^n (X_i - \mu_1)^2\right\}}$$

  After some manipulation, it yields

$$R = \exp\left[-\frac{(\mu_1 - \mu_0)}{\sigma^2/n}\left(\bar{X} - \frac{\mu_0 + \mu_1}{2}\right)\right]$$

- The rejection region is given by

$$\{R \leq c\} = \left\{ \exp\left[-\frac{(\mu_1 - \mu_0)}{\sigma^2/n} \left(\bar{X} - \frac{\mu_0 + \mu_1}{2} \right) \right] \leq c \right\}$$
$$= \left\{ \bar{X} \geq \frac{\mu_0 + \mu_1}{2} - \frac{\sigma^2/n}{(\mu_1 - \mu_0)} \log(c) \right\} = \{\bar{X} \geq c^*\}$$

  where c^* is just another constant (μ_0, μ_1, n, σ are all fixed constant). That is, it is equivalently to reject H_0 when $\bar{X}$ is large

- To solve for the c^*, consider $P(R \leq c|H_0) = P(\bar{X} \geq c^*|H_0) = \alpha$. Note that

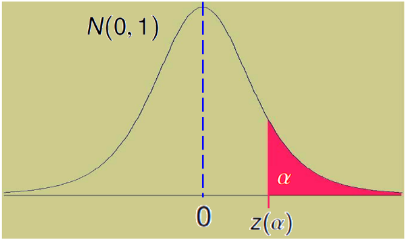
$$P(R \leq c|H_0) = P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \geq \frac{c^* - \mu_0}{\sigma/\sqrt{n}} |H_0 \right) = P\left(Z \geq \frac{c^* - \mu_0}{\sigma/\sqrt{n}} \right)$$

where the equality follows from the fact that $\bar{X} \sim N(\mu_0, \frac{\sigma^2}{n})$ under the null H_0

- Setting $P\left(Z \geq \frac{c^* - \mu_0}{\sigma/\sqrt{n}}\right) = \alpha$, we have

$$\frac{c^* - \mu_0}{\sigma/\sqrt{n}} = Z(\alpha) \Leftrightarrow c^* = \mu_0 + Z(\alpha)\frac{\sigma}{\sqrt{n}}$$

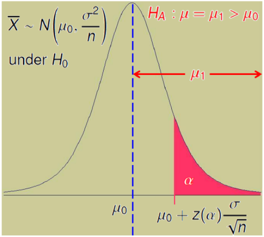
where $Z(\alpha)$ is the $100(1-\alpha)$-th percentile of the standard normal distribution Z.

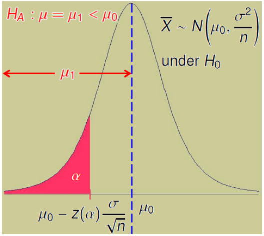


The Z Test for two Simple Hypotheses

The LRT with significance level α for testing $H_0 : \mu = \mu_0$ against $H_a : \mu = \mu_1 > \mu_0$(resp. $< \mu_0$) for $N(\mu, \sigma^2)$ with σ known is a test that has a rejection region (red in the below graph) given by

$$\left\{\bar{X} \geq \mu_0 + Z(\alpha)\frac{\sigma}{\sqrt{n}}\right\} \text{resp.} \left(\left\{\bar{X} \leq \mu_0 - Z(\alpha)\frac{\sigma}{\sqrt{n}}\right\}\right)$$





An application of the Z test for two simple hypotheses:

- A random sample of size 10 from a $N(\mu, 1)$ model with a sample mean $\bar{x} = 9$ is given
- To test $H_0 : \mu = 8$ against $H_a : \mu = 10$ at 5% significance level
- Calculate the RR as

$$\left\{\bar{X} \geq 8 + Z(.05)\frac{1}{\sqrt{10}} = 8 + 1.645\frac{1}{\sqrt{10}} = 8.52\right\}$$

- As the observed value $\bar{x} = 9$ is greater than 8.52, we reject H_0 at 5% level
- In rejecting H_0, we should keep in mind that the probability of rejecting a true H_0 is equal to 5%. That is, there is 5% chance that the truth is $\mu = 8$

Neyman-Pearson Lemma

Suppose that H_0 and H_1 are simple hypotheses and that the test that rejects H_0 whenever the likelihood is less than c and significance level α. Then *any other test* for which the significance level is less than or equal to α has power less than or equal to that of the likelihood ratio test.

- is the most powerful test at significance level α for a threshold c.
- If the test is most powerful for all $\theta_1 \in \Theta_1$, it is said to be uniformly most powerful (UMP) for alternatives in the set Θ_1

The LRT for two Simple Hypotheses

Let $X_1, \cdots, X_n$ be a random sample from $N(\mu, \sigma^2)$ having known mean μ. Consider two simple hypotheses:

$$H_0 : \sigma^2 = \sigma_0^2 \text{ and } H_a : \sigma^2 = \sigma_1^2$$

where $\sigma_0^2 < \sigma_1^2$. To derive the LRT with significance level $0 < \alpha < 1$:

- Calculate the likelihood ratio, or, LR statistic:

$$R = \frac{\sigma_0^2}{\sigma_1^2} \exp\left\{ \frac{-1}{2} (\sigma_0^{-2} - \sigma_1^{-2}) \sum_{i=1}^{n} (X_i - \mu)^2 \right\}$$

- This ratio only depends on the data through $\sum_{i=1}^{n} (x_i - \mu)^2$. Therefore, by the NeymanPearson lemma, the most powerful test of this type of hypothesis for this data will depend only on $\sum_{i=1}^{n} (x_i - \mu)^2$.

The LRT Test for two Simple Hypotheses

- If $\sigma_1^2 > \sigma_0^2$, then R is a non-decreasing function of $\sum_{i=1}^{n}(x_i-\mu)^2$. So we should reject H_0 if $\sum_{i=1}^{n}(x_i-\mu)^2$ is sufficiently large.

$$\sum_{i=1}^{n}(x_i-\mu)^2 > C^*$$

- The rejection threshold depends on the size of the test.
- In this example, the test statistic can be shown to be a scaled Chi-square distributed random variable and an exact critical value can be obtained.
    - under H_0, what is the distribution of the test statistic? e.g. find the value of C^* by controlling α

$$P(\sum_{i=1}^{n}(x_i-\mu)^2 > C^*|H_0) = \alpha$$

More commonly, $\Theta_0 \cup \Theta_1 = \Theta$, Θ_1 is composite and Θ_0 is either simple or composite (i.e., more than 3 possible values of θ in Θ are considered)

Generalized Likelihood Ratio Test

The LR test statistic for testing $H_0 : \theta \in \Theta_0$ versus $H_a : \theta \in \Theta_1 \equiv \Theta_0^c$ (i.e., $\Theta_0 \cup \Theta_1 = \Theta$) is defined

$$\Lambda = \frac{\max_{\theta \in \Theta_0} \text{lik}(\theta)}{\max_{\theta \in \Theta} \text{lik}(\theta)} = \frac{\max_{\theta \in \Theta_0} \text{lik}(\theta)}{\text{lik}(\hat{\theta})} \leq 1$$

which is a function of the sample $x_1, \cdots, x_n$, where $\hat{\theta}$ is the mle of θ. A generalized LRT with significance level $0 < \alpha < 1$ is a test that has a rejection region of the form

$$\{\Lambda \leq \lambda_0\}$$

where $0 \leq \lambda_0 \leq 1$ is chosen so that $P(\Lambda \leq \lambda_0 | H_0) = \alpha$

LRT-exponential population

Let $X_1, \ldots, X_n$ be a random sample from an exponential population with pdf

$$f(x|\theta) = \begin{cases} e^{-(x-\theta)} & x \geq \theta \\ 0 & x < \theta, \end{cases}$$

where $-\infty < \theta < \infty$. Consider testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$, where θ_0 is a value specified by the experimenter.

The likelihood function is

$$L(\theta|\mathbf{x}) = \begin{cases} e^{-\Sigma x_i + n\theta} & \theta \le x_{(1)} \\ 0 & \theta > x_{(1)}. \end{cases} \qquad (x_{(1)} = \min x_i)$$

Consider testing $H_0: \theta \le \theta_0$ versus $H_1: \theta > \theta_0$, where θ_0 is a value specified by the experimenter. Clearly $L(\theta|\mathbf{x})$ is an increasing function of θ on $-\infty < \theta \le x_{(1)}$. Thus, the denominator of $\lambda(\mathbf{x})$, the unrestricted maximum of $L(\theta|\mathbf{x})$, is

$$L(x_{(1)}|\mathbf{x}) = e^{-\Sigma x_i + n x_{(1)}}.$$

If $x_{(1)} \le \theta_0$, the numerator of $\lambda(\mathbf{x})$ is also $L(x_{(1)}|\mathbf{x})$. But since we are maximizing $L(\theta|\mathbf{x})$ over $\theta \le \theta_0$, the numerator of $\lambda(\mathbf{x})$ is $L(\theta_0|\mathbf{x})$ if $x_{(1)} > \theta_0$. Therefore, the likelihood ratio test statistic is

$$\lambda(\mathbf{x}) = \begin{cases} 1 & x_{(1)} \le \theta_0 \\ e^{-n(x_{(1)} - \theta_0)} & x_{(1)} > \theta_0. \end{cases}$$

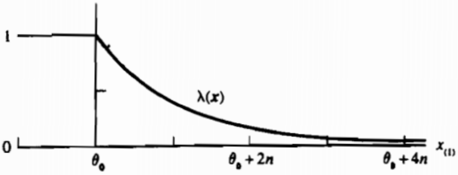


Figure 8.2.1. $\lambda(\mathbf{x})$, *a function only of* $x_{(1)}$.

A graph of $\lambda(\mathbf{x})$ is shown in Figure 8.2.1. An LRT, a test that rejects H_0 if $\lambda(\mathbf{X}) \leq c$, is a test with rejection region $\{\mathbf{x} : x_{(1)} \geq \theta_0 - \frac{\log c}{n}\}$. Note that the rejection region depends on the sample only through the sufficient statistic $X_{(1)}$.

The two-sided Z Test

Let $X_1, \cdots, X_n$ be a random sample from $N(\mu, \sigma^2)$ having known variance $\sigma^2 > 0$. Consider a two-sided test:

$$H_0 : \mu = \mu_0 \text{ and } H_a : \mu \neq \mu_0$$

where $-\infty < \mu_0 < \infty$ is a given constant. To derive the LRT with significance level $0 < \alpha < 1$:

- Calculate the numerator and the denominator of the LR statistic Λ:

$$\max_{\mu \in \Theta_0} \text{lik}(\theta) = \text{lik}(\mu_0) = \frac{1}{(\sigma\sqrt{2\pi})^n} \exp\left[\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu_0)^2\right]$$

and

$$\text{lik}(\hat{\mu}) = \frac{1}{(\sigma\sqrt{2\pi})^n} \exp\left[\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2\right]$$

The two-sided Z Test - cont'd

where the mle of μ is $\hat{\mu} = \bar{X}$. Then the LR test statistic is given by

$$\Lambda = \exp\left[\frac{-1}{2\sigma^2}\left\{\sum_{i=1}^{n}(X_i - \mu_0)^2 - \sum_{i=1}^{n}(X_i - \bar{X})^2\right\}\right] = e^{\frac{-1}{2\sigma^2/n}(\bar{X}-\mu_0)^2}$$

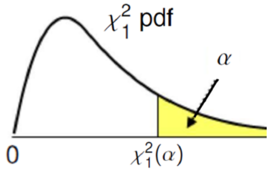
using the fact that $\sum_{i=1}^{n}(X_i - \mu_0)^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2 + n(\bar{X} - \mu_0)^2$

- The RR is given by

$$\{\Lambda \leq \lambda_0\} = \left\{e^{\frac{-1}{2\sigma^2/n}(\bar{X}-\mu_0)^2} \leq \lambda_0\right\} = \left\{\frac{n(\bar{X} - \mu_0)^2}{\sigma^2} \geq -2\log(\lambda_0)\right\}$$
$$= \left\{\frac{n(\bar{X} - \mu_0)^2}{\sigma^2} \geq k\right\}$$

where $k \geq 0$ is another constant. This is, it is equivalent to rejecting H_0 when $-2\log\Lambda = n(\bar{X} - \mu_0)^2/\sigma^2$ is large

- To solve for k, consider $P(\Lambda \leq \lambda_0 | H_0) = P\left(\frac{n(\bar{X}-\mu_0)^2}{\sigma^2} \geq k | H_0\right) = \alpha$
  Note that this probability is equal to
  $$P(Z^2 \geq k | H_0) = P(\chi_1^2 \geq k)$$
  as $Z = \sqrt{n}(\bar{X} - \mu_0)/\sigma \sim N(0,1)$ under the H_0
- Setting $P(\chi_1^2 \geq k) = \alpha$, we have $k = \chi_1^2(\alpha) = [z(\alpha/2)]^2$, where $\chi_1^2(\alpha)$ is the $100(1-\alpha)$-th percentile of a χ_1^2 r.v. and $z(\alpha/2)$ is the $100(1-\alpha/2)$-th percentile of the standard normal r.v. Z

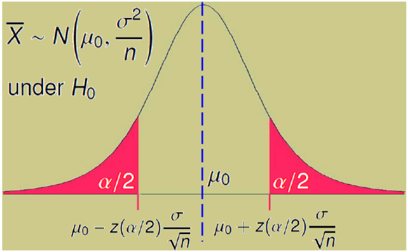


The two-sided Z Test

The LRT with significance level α for testing $H_0 : \mu = \mu_0$ against $H_a : \mu \neq \mu_0$ for $N(\mu, \sigma^2)$ with σ known is a test that has a rejection region (red in the below graph) given by

$$\left\{ |\bar{X} - \mu_0| \geq Z(\alpha/2)\frac{\sigma}{\sqrt{n}} \right\} =$$

$$\left\{ \bar{X} \leq \mu_0 - Z(\alpha/2)\frac{\sigma}{\sqrt{n}} \text{ or } \bar{X} \geq \mu_0 + Z(\alpha/2)\frac{\sigma}{\sqrt{n}} \right\}$$



An application of the two-sided Z test:

- A random sample of size 10 from a $N(\mu, 1)$ model with a sample mean $\bar{x} = 9$ is given
- To test $H_0 : \mu = 8$ against $H_a : \mu \neq 8$ at 5% significance level
- Calculate the RR as

$$\left\{\bar{X} \leq 8 - 1.96\frac{1}{\sqrt{10}} \text{ or } \bar{X} \geq 8 + 1.96\frac{1}{\sqrt{10}}\right\} = \{\bar{X} \leq 7.38 \text{ or } \bar{X} \geq 8.61\}$$

- As the observed value $\bar{x} = 9$ is greater than 8.61, we reject H_0 at 5% level
- In rejecting H_0, we should keep in mind that the probability of rejecting a true H_0 is equal to 5%

Let $X_1, \cdots, X_n$ be a random sample from $N(\mu, \sigma^2)$ having known variance $\sigma^2 > 0$. Consider a one-sided test:

$$H_0 : \mu \leq \mu_0 \text{ and } H_a : \mu > \mu_0$$

where $-\infty < \mu_0 < \infty$ are given constant. To derive the LRT with significance level $0 < \alpha < 1$:

- Calculate the numerator and the denominator of the LR statistic Λ:

$$\max_{\mu\in\Theta_0} \text{lik}(\theta) = \max_{\mu\in(-\infty,\mu_0)} \frac{1}{(\sigma\sqrt{2\pi})^n} \exp\left[\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2\right]$$

$$= \frac{1}{(\sigma\sqrt{2\pi})^n} e^{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2} \max_{\mu\in(-\infty,\mu_0)} \exp\left[\frac{-n}{2\sigma^2}(\bar{X}-\mu)^2\right]$$

and

$$\text{lik}(\hat{\mu}) = \frac{1}{(\sigma\sqrt{2\pi})^n} \exp\left[\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2\right]$$

The one-sided Z Test - cont'd

The LR test statistic is given by

$$\Lambda = \max_{\mu\in(-\infty,\mu_0)} \exp\left[\frac{-n}{2\sigma^2}(\bar{X}-\mu)^2\right] = \begin{cases} 1 & \text{if } \bar{X} \le \mu_0 \\ \exp\left[\frac{-n}{2\sigma^2}(\bar{X}-\mu_0)^2\right] & \text{if } \bar{X} > \mu_0 \end{cases}$$

- When $\bar{X} \le \mu_0$, we will never reject H_0,as $\Lambda = 1$ is always greater than any λ_0. The RR with $\bar{X} > \mu_0$ is given by

$$\begin{aligned}\{\Lambda \le \lambda_0\} &= \left\{\frac{n(\bar{X}-\mu_0)^2}{\sigma^2} \ge -2\log(\lambda_0)\right\} \\ &= \left\{\frac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma} \ge \sqrt{-2\log(\lambda_0)}\right\} = \{\bar{X} \ge k\}\end{aligned}$$

where $k \ge 0$ is another constant. This is, it is equivalent to reject H_0 when $\bar{X}$ is large

- To solve for k, consider $P(\Lambda \leq \lambda_0 | H_0) = P\left(\bar{X} \geq k | H_0\right) = \alpha$. Note that this probability is equal to

$$P(Z \geq \frac{k - \mu_0}{\sigma/\sqrt{n}})$$

as $Z = \sqrt{n}(\bar{X} - \mu_0)/\sigma \sim N(0,1)$ under the H_0

- Setting $P(Z \geq \frac{k-\mu_0}{\sigma/\sqrt{n}}) = \alpha$, we have

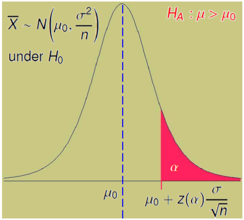
$$\frac{k - \mu_0}{\sigma/\sqrt{n}} = z(\alpha) \Leftrightarrow k = \mu_0 + z(\alpha)\frac{\sigma}{\sqrt{n}}$$

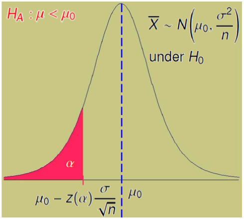
where $z(\alpha)$ is the $100(1-\alpha)$-th percentile of the standard normal r.v. Z

The one-sided Z Test

The LRT with significance level α for testing $H_0 : \mu \leq \mu_0$ (resp. $\mu \geq \mu_0$) against $H_a : \mu > \mu_0$ (resp. $\mu < \mu_0$) for $N(\mu, \sigma^2)$ with σ known is a test that has a rejection region (red in the below graph) given by

$$\left\{ \bar{X} \geq \mu_0 + Z(\alpha)\frac{\sigma}{\sqrt{n}} \right\} \text{resp.} \left\{ \bar{X} \leq \mu_0 - Z(\alpha)\frac{\sigma}{\sqrt{n}} \right\}$$





An application of the one-sided Z test:

- A random sample of size 10 from a $N(\mu, 1)$ model with a sample mean $\bar{x} = 9$ is given
- To test $H_0 : \mu \leq 8$ against $H_a : \mu > 8$ at 5% significance level
- Calculate the RR as

$$\left\{ \bar{X} \geq 8 + Z(.05)\frac{1}{\sqrt{10}} = 8 + 1.645\frac{1}{\sqrt{10}} = 8.52 \right\}$$

- As the observed value $\bar{x} = 9$ is greater than 8.52, we reject H_0 at 5%
- In rejecting H_0, we should keep in mind that the probability of rejecting a true H_0 is equal to 5%

Two sided t test

Let $X_1, \cdots, X_n$ be a random sample from $N(\mu, \sigma^2)$ having both μ and σ^2 unknown

- Consider a two-sided test:

$$H_0 : \mu = \mu_0 \text{ and } H_a : \mu \neq \mu_0$$

  where $-\infty < \mu_0 < \infty$ are given constant.

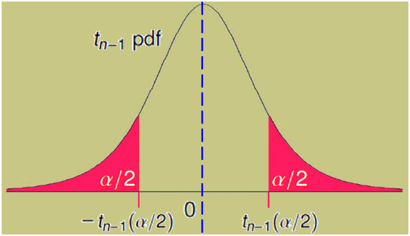
- The test statistic is constructed based on a t statistic; the derivation is omitted.

The two-sided t Test

The LRT with significance level α for testing $H_0 : \mu = \mu_0$ against $H_a : \mu \neq \mu_0$ for $N(\mu, \sigma^2)$ with unknown σ is a test that has a rejection region (red in the below graph) given by

$$\left\{ \bar{X} \leq \mu_0 - t_{n-1}(\alpha/2)\frac{S}{\sqrt{n}} \text{ or } \bar{X} \geq \mu_0 + t_{n-1}(\alpha/2)\frac{S}{\sqrt{n}} \right\}$$

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



An application of the two-sided t test:

- A random sample of size 10 from a $N(\mu, 1)$ model with a sample mean $\bar{x} = 9$ and a sample variance $S^2 = 1.5$ are given
- To test $H_0 : \mu = 8$ against $H_a : \mu \neq 8$ at 5% significance level
- Calculate the RR as

$$\left\{ \bar{X} \leq 8 - t_9(.025)\sqrt{\frac{1.5}{10}} \text{ or } \bar{X} \geq 8 + t_9(.025)\sqrt{\frac{1.5}{10}} \right\}$$

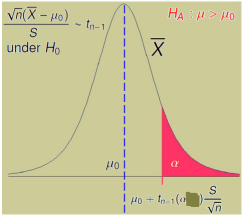
$$= \left\{ \bar{X} \leq 7.12 \text{ or } \bar{X} \geq 8.88 \right\}$$

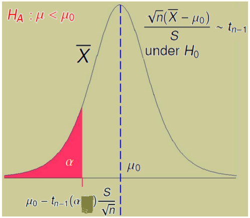
- As the observed value $\bar{x} = 9$ is greater than 8.88, we reject H_0 at 5%
- In rejecting H_0, we should keep in mind that the probability of rejecting a true H_0 is equal to 5%

The one-sided t Test

The LRT with significance level α for testing $H_0 : \mu \leq \mu_0$ (resp. $\mu \geq \mu_0$) against $H_a : \mu > \mu_0$ (resp. $\mu < \mu_0$) for $N(\mu, \sigma^2)$ with unknown σ is a test that has a rejection region (red in the below graph) given by

$$\left\{ \bar{X} \geq \mu_0 + t_{n-1}(\alpha)\frac{S}{\sqrt{n}} \right\} \text{ resp. } \left\{ \bar{X} \leq \mu_0 - t_{n-1}(\alpha)\frac{S}{\sqrt{n}} \right\}$$





An application of the one-sided t test:

- Assume a $N(\mu, \sigma^2)$ model for the random commutation time (in minutes) between office/home and NUS
- Commutation times on 10 days contribute to a random sample of size 10 with a sample mean $\bar{x} = 29.5$ and a sample variance $S^2 = 1.5$
- To test $H_0 : \mu \geq 30$ against $H_a : \mu < 30$ at 5% significance level
- Calculate the RR as

$$\left\{ \bar{X} \leq 30 - t_9(.05)\sqrt{\frac{1.5}{10}} \right\} = \left\{ \bar{X} \leq 30 - 1.833\sqrt{\frac{1.5}{10}} \right\} = \{\bar{X} \leq 29.29\}$$

- As the observed value $\bar{x} = 29.5$ is not less than 29.29, we do not reject H_0 at 5%
- In not rejecting H_0, we should keep in mind that it is possible that we have accepted a false H_0

- Every test is based on a prespecified significance level α, but no guidance is there about how to choose it
- Our test result depends on $\alpha \Leftrightarrow$ larger α easier to reject H_0
- We must either reject or do not reject H_0 wrt α based on the data
- When we reject H_0, we have no idea how strong the evidence against H_0

p-value is a way to summarize the evidence against H_0

Definition

Based on any sample, a p-value is defined to be the smallest significance level at which the null hypothesis would be rejected

- The probability under H_0 of observing the current result as or more extreme result
- The smaller the p-value, the stronger the evidence against H_0

- The p-value is computed to be the probability

$$P(\Lambda \leq a | H_0)$$

where a is the observed value of the LR test statistic based on the sample

- For the "two-sided t test" example, the p-value is computed as

$$P(|\bar{X}| \geq 9 | H_0) = 2P(\bar{X} \geq 9 | H_0) = 2P(t_9 \geq \frac{9-8}{\sqrt{1.5/10}}) = 2P(t_9 \geq 2.58)$$

- Resorting to computer, it is given by $2 \times .01484831 = .02969662$. As long as $\alpha \geq .02969662$, we would reject H_0
- For the one-sided t test of $H_a : \mu > 8$, the p-value is computed as

$$P(\bar{X} \geq 9 | H_0) = P(t_9 \geq 2.58) = .01484831$$

Comparing these two p-values, there is stronger evidence that μ is greater than 8

- In order for the LRT to have the desirable significance level α, λ_0 must be chosen so that $P(\Lambda \leq \lambda_0 | H_0) = \alpha$. If the sampling distribution of Λ under H_0 is known or can be derived easily, we can determine λ_0
- In practice, the sampling distribution is usually not of a simple form

Resort to an approximation of the null distribution

Asymptotic Property of Generalized LRT

Under smoothness conditions on the density involved, the null distribution of $-2\log\Lambda$ tends to a χ^2 distribution with degree of freedom equal to $df = \dim\Theta - \dim\Theta_0$as the sample size tends to infinity, i.e.,

$$P(\Lambda \leq \lambda_0 | H_0) \approx P(\chi^2_{df} \geq -2\log\lambda_0),\ n \to \infty$$

- $\dim\Theta$ and $\dim\Theta_0$ are the number of free parameters under Θ and Θ_0, respectively

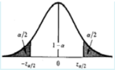
Motivational example

- Suppose that $X_1, \ldots, X_n$ are i.i.d. from $N(\mu, \sigma^2)$, where σ^2 is known.
- Consider the hypotheses
$$H_0 : \mu = \mu_0 \quad \text{v.s.} \quad H_A : \mu \neq \mu_0.$$
Then $\Omega_0 = \{\mu_0\}$, $\Omega_A = \{\mu : \mu \neq \mu_0\}$, $\Omega = \{-\infty < \mu < \infty\}$.
- The likelihood ratio statistic is
$$\begin{aligned}
\Lambda &= \frac{(\sqrt{2\pi}\sigma)^{-n} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_0)^2\right]}{\max_{-\infty<\mu<\infty} \left[(\sqrt{2\pi}\sigma)^{-n} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right]\right]} \\
&= \frac{(\sqrt{2\pi}\sigma)^{-n} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_0)^2\right]}{(\sqrt{2\pi}\sigma)^{-n} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \overline{x})^2\right]} \\
&= \exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^n (x_i - \mu_0)^2 - \sum_{i=1}^n (x_i - \overline{x})^2\right]\right) \\
&= \exp\left[-\frac{n}{2\sigma^2}(\overline{x} - \mu_0)^2\right]
\end{aligned}$$
- Thus the likelihood ratio test rejects H_0 for small values of Λ, i.e. large values of
$$-2\log\Lambda = \frac{(\overline{x} - \mu_0)^2}{\sigma^2/n}.$$

Motivational example -cont'd

- Under H_0, $\overline{X} \sim N(\mu_0, \frac{\sigma^2}{n})$ and $-2\log\Lambda \sim \chi_1^2$.
- Thus, the likelihood ratio test rejects when

$$\frac{(\overline{X}-\mu_0)^2}{\sigma^2/n} > \chi_1^2(\alpha) \quad \text{or equivalently} \quad \frac{\left|\overline{X}-\mu_0\right|}{\sigma/\sqrt{n}} > z(\alpha/2)\,.$$



- σ is a known constant
- The null hypothesis completely specifies (μ, σ^2) as $\mu = \mu_0$; There are no free parameter under Θ_0, so $\dim\Theta_0 = 0$
- Under Θ, σ^2 is fixed but μ is free, so $\dim\Theta = 1$
- One can apply the previous result to obtain an approximation of the sampling distribution of $-2\log\Lambda$ as a χ_1^2 r.v.
- In fact, we applied the exact sampling distribution of $-2\log\Lambda$ which is a χ_1^2 r.v. in the example to come up with the exact rejection region