Week 8*

November 8, 2017

1 Hypothesis Testing

1.1 Principles of Hypothesis Testing

We consider a statistical problem involving a parameter θ whose value is unknown but must lie in a certain space Ω. We consider the testing problem

$$H_0 : \theta \in \Omega_0 \qquad \text{versus} \qquad H_1 : \theta \in \Omega_1, \tag{1}$$

where $\Omega_0 \cap \Omega_1 = \emptyset$ and $\Omega_0 \cup \Omega_1 = \Omega$.

Here the hypothesis H_0 is called the *null hypothesis* and H_1 is called the *alternative hypothesis*.

We are given data (say $X_1, \ldots, X_n$ i.i.d P_θ) from a model that is parametrized by θ.

Question: Is there enough evidence in the data against the null hypothesis (in which case we reject it) or should we continue to stick to it?

Such questions arise very naturally in many different fields of application.

Definition 1 (One-sided and two-sided hypotheses)**.** *Let θ be a one-dimensional parameter.*

- *one-sided hypotheses*
  - $H_0 : \theta \le \theta_0$, *and* $H_1 : \theta > \theta_0$, *or*

*Notes for Chapter 9 of DeGroot and Schervish adapted from Giovanni Motta's and Martin Lindquists notes for STAT W4109/W4105.

– *$H_0 : \theta \geq \theta_0$, and $H_1 : \theta < \theta_0$*

- *two-sided hypotheses $H_0 : \theta = \theta_0$, and $H_1 : \theta \neq \theta_0$.*

H_0 is *simple* if Ω_0 is a set with only one point; otherwise, H_0 is *composite*.

Testing for a normal mean: Suppose that $X_1, X_2, \ldots, X_n$ is a sample from a $N(\mu, \sigma^2)$ distribution and let, initially, σ^2 be known.

We want to test the *null hypothesis* $H_0 : \mu = \mu_0$ against the alternative $H_1 : \mu \neq \mu_0$.

Example: For concreteness, $X_1, X_2, \ldots, X_n$ could be the heights of n individuals in some tribal population. The distribution of heights in a (homogeneous) population is usually normal, so that a $N(\mu, \sigma^2)$ model is appropriate. If we have some a-priori reason to believe that the average height in this population is around 60 inches, we could postulate a null hypothesis of the form $H_0 : \mu = \mu_0 \equiv 60$; the alternative hypothesis is $H_1 : \mu \neq 60$.

1.2 Critical regions and test statistics

Consider a problem in which we wish to test the following hypotheses:

$$H_0 : \theta \in \Omega_0, \qquad \text{and} \qquad H_1 : \theta \in \Omega_1. \tag{2}$$

Question: How do we do the test?

The statistician must decide, after observing data, which of the hypothesis H_0 or H_1 appears to be true.

A procedure for deciding which hypothesis to choose is called a *test procedure* of simply a *test*. We will denote a test by δ.

Suppose we can observe a random sample $\mathbf{X} = (X_1, \ldots, X_n)$ drawn from a distribution that involves the unknown parameter θ, e.g., suppose that $X_1, \ldots, X_n$ are i.i.d P_θ, $\theta \in \Omega$.

Let S denote the set of all possible values of the n-dimensional random vector $\mathbf{X}$.

We specify a test procedure by partitioning S into two subsets: $S = S_0 \cup S_1$

– The *rejection region* (sometimes also called the *critical region*) S_1 contains the values of $\mathbf{X}$ for which we will reject H_0, and

- the other subset S_0 (usually called the *acceptance* region) contains the values of $\mathbf{X}$ for which we will not reject H_0.

A test procedure is determined by specifying the critical region S_1 of the test.

In most hypothesis-testing problems, the critical region is defined in terms of a statistic, $T = \varphi(\mathbf{X})$.

Definition 2 (Test statistic/rejection region)**.** *Let* $\mathbf{X}$ *be a random sample from a distribution that depends on a parameter* θ*. Let* $T = \varphi(\mathbf{X})$ *be a statistic, and let* R *be a subset of the real line. Suppose that a test procedure is of the form:*

$$\textit{reject } H_0 \quad \textit{if} \quad T \in R.$$

Then we call T *a test statistic, and we call* R *the rejection region of the test:*

$$S_1 = \{\mathbf{x} : \varphi(\mathbf{x}) \in R\}.$$

Typically, the rejection region for a test based on a test statistic T will be some fixed interval or the complement of some fixed interval.

If the test rejects H_0 when $T \geq c$, the rejection region is the interval $[c, \infty)$. Indeed, most of the tests can be written in the form:

$$\text{reject } H_0 \quad \text{if} \quad T \geq c.$$

Example: Suppose that $X_1, \ldots, X_n$ are i.i.d $N(\mu, \sigma^2)$ where $\mu \in \mathbb{R}$ is unknown, and $\sigma > 0$ is assumed *known*.

Suppose that we want to test $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$.

Some of these procedures can be justified using formal paradigms. Under the null hypothesis the X_i's are i.i.d $N(\mu_0, \sigma^2)$ and the sample mean $\overline{X}$ follows $N(\mu_0, \sigma^2/n)$.

Thus, it is reasonable to take $T = \varphi(\mathbf{X}) = |\overline{X} - \mu_0|$.

Large deviations of the observed value of $\overline{X}$ from μ_0 would lead us to suspect that the null hypothesis might not be true.

Thus, a reasonable test can be to reject H_0 if $T = |\overline{X} - \mu_0| > c$, for some "large" constant c.

But how large is large? We will discuss this soon...

Decision / State	Fail to reject H_0	Reject H_0
H_0 True	Correct	Type 1 error
H_1 True	Type 2 error	Correct

Table 1: Hypothesis Test

Associated with the test procedure δ are two different kinds of error that we can commit. These are called *Type 1 error* and *Type 2 error* (Draw the 2×2 table!).

Type 1 error occurs if we reject the null hypothesis when actually H_0 is true.

Type 2 error occurs if we do not reject the null hypothesis when actually H_0 is false.

1.3 Power function and types of error

Let δ be a test procedure. If S_1 denotes the critical region of δ, then the **power function** of the test δ, $\pi(\theta|\delta)$, is defined by the relation

$$\pi(\theta|\delta) = \mathbb{P}_\theta(\mathbf{X} \in S_1) \quad \text{for} \quad \theta \in \Omega.$$

Thus, the power function $\pi(\theta|\delta)$ specifies for each possible value of θ, the *probability that δ will reject H_0*. If δ is described in terms of a test statistic T and rejection region R, the power function is

$$\pi(\theta|\delta) = \mathbb{P}_\theta(T \in R) \quad \text{for} \quad \theta \in \Omega.$$

Example: Suppose that $X_1, \ldots, X_n$ are i.i.d Unif$(0, \theta)$, where $\theta > 0$ is unknown.

Suppose that we are interested in the following hypotheses:

$$H_0 : 3 \leq \theta \leq 4, \qquad \text{versus} \qquad H_1 : \theta < 3, \text{ or } \theta > 4.$$

We know that the MLE of θ is $X_{(n)} = \max\{X_1, \ldots, X_n\}$.

Note that $X_{(n)} < \theta$.

Suppose that we use a test δ given by the critical region

$$S_1 = \{\boldsymbol{x} : x_{(n)} \leq 2.9 \text{ or } x_{(n)} \geq 4\}.$$

Question: Find the power function $\pi(\theta|\delta)$?

Solution: The power function of δ is

$$\pi(\theta|\delta) = \mathbb{P}_\theta(X_{(n)} \leq 2.9 \text{ or } X_{(n)} > 4) = \mathbb{P}_\theta(X_{(n)} \leq 2.9) + \mathbb{P}_\theta(X_{(n)} \geq 4).$$

Case (i): Suppose that $\theta \leq 2.9$. Then

$$\pi(\theta|\delta) = \mathbb{P}_\theta(X_{(n)} \leq 2.9) = 1.$$

Case (ii): Suppose that $2.9 < \theta < 4$. Then

$$\pi(\theta|\delta) = \mathbb{P}_\theta(X_{(n)} \leq 2.9) = \left(\frac{2.9}{\theta}\right)^n.$$

Case (iii): Suppose that $\theta > 4$. Then

$$\pi(\theta|\delta) = \left(\frac{2.9}{\theta}\right)^n + \left[1 - \left(\frac{4}{\theta}\right)^n\right].$$

The ideal power function would be one for which

- $\pi(\theta|\delta) = 0$ for every value of $\theta \in \Omega_0$, and
- $\pi(\theta|\delta) = 1$ for every value of $\theta \in \Omega_1$.

If the power function of a test δ actually had these values, then regardless of the actual value of θ, δ would lead to the correct decision with probability 1.

In a practical problem, however, there would seldom exist any test procedure having this ideal power function.

- Type-I error: rejecting H_0 given that $\theta \in \Omega_0$. It occurs with probability $\pi(\theta|\delta)$.
- Type-II error: not rejecting H_0 given that $\theta \in \Omega_1$. It occurs with probability $1 - \pi(\theta|\delta)$.

Ideal goals: we would like the power function $\pi(\theta|\delta)$ to be **low** for values of $\theta \in \Omega_0$, and **high** for $\theta \in \Omega_1$.

Generally, these two goals work against each other. That is, if we choose δ to make $\pi(\theta|\delta)$ small for $\theta \in \Omega_0$, we will usually find that $\pi(\theta|\delta)$ is small for $\theta \in \Omega_1$ as well.

Examples:

- The test procedure δ_0 that never rejects H_0, regardless of what data are observed, will have $\pi(\theta|\delta_0) = 0$ for all $\theta \in \Omega_0$. However, for this procedure $\pi(\theta|\delta_0) = 0$ for all $\theta \in \Omega_1$ as well.

- Similarly, the test δ_1 that always rejects H_0 will have $\pi(\theta|\delta_1) = 1$ for all $\theta \in \Omega_1$, but it will also have $\pi(\theta|\delta_1) = 1$ for all $\theta \in \Omega_0$.

Hence, there is a need to strike an appropriate balance between the two goals of

low power in Ω_0 and *high power in* Ω_1.

1. The most popular method for striking a balance between the two goals is to choose a number $\alpha_0 \in (0, 1)$ and require that

$$\pi(\theta|\delta) \leq \alpha_0, \quad \text{for all} \quad \theta \in \Omega_0. \tag{3}$$

   This α_0 will usually be a small positive fraction (historically .05 or .01) and will be called the **level of significance** or simply *level*.

   Then, among all tests that satisfy (3), the statistician seeks a test whose power function is as high as can be obtained for $\theta \in \Omega_1$.

2. Another method of balancing the probabilities of type I and type II errors is to minimize a linear combination of the different probabilities of error.

1.4 Significance level

Definition 3 (level/size)**.** *(of the test)*

- *A test that satisfies* (3) *is called a level* α_0 *test, and we say that the test has level of significance* α_0*.*

- *The size* $\alpha(\delta)$ *of a test* δ *is defined as follows:*

$$\alpha(\delta) = \sup_{\theta \in \Omega_0} \pi(\theta|\delta).$$

It follows from Definition 3 that:

- A test δ is a level α_0 test iff $\alpha(\delta) \leq \alpha_0$.

- If the null hypothesis is simple (that is, $H_0 : \theta = \theta_0$), then $\alpha(\delta) = \pi(\theta_0|\delta)$.

Making a test have a specific significance level

Suppose that we wish to test the hypotheses

$$H_0 : \theta \in \Omega_0, \qquad \text{versus} \qquad H_1 : \theta \in \Omega_1.$$

Let T be a test statistic, and suppose that our test will reject the null hypothesis if $T \geq c$, for some constant c. Suppose also that we desire our test to have the level of significance α_0. The power function of our test is $\pi(\theta|\delta) = \mathbb{P}_\theta(T \geq c)$, and we want that

$$\sup_{\theta \in \Omega_0} \mathbb{P}_\theta(T \geq c) \leq \alpha_0. \tag{4}$$

Remarks:

1. It is clear that the power function, and hence the left side of (4), are non-increasing functions of c.

   Hence, (4) will be satisfied for large values of c, but not for small values.

   If T has a continuous distribution, then it is usually simple to find an appropriate c.

2. Whenever we choose a test procedure, we should also examine the power function. If one has made a good choice, then the power function should generally be larger for $\theta \in \Omega_1$ than for $\theta \in \Omega_0$.

Example: Suppose that $X_1, \ldots, X_n$ are i.i.d $N(\mu, \sigma^2)$ where $\mu \in \mathbb{R}$ is unknown, and $\sigma > 0$ is assumed *known*. We want to test $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$.

Suppose that the null hypothesis H_0 is true.

If the variance of the sample mean is, say, 100, a deviation of $\overline{X}$ from μ_0 by 15 is not really unusual.

On the other hand if the variance is 10, then a deviation of the sample mean from μ_0 by 15 is really sensational.

Thus the quantity $|\overline{X} - \mu_0|$ in itself is not sufficient to formulate a decision regarding rejection of the null hypothesis.

We need to adjust for the underlying variance. This is done by computing the so-called z-statistic,

$$Z := \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \equiv \frac{\sqrt{n}(\overline{X} - \mu_0)}{\sigma}$$

and rejecting the null hypothesis for large absolute values of this statistic.

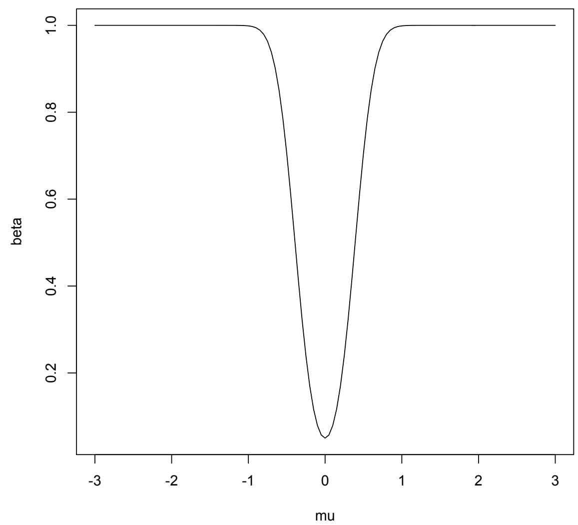

Figure 1: The power function $\pi(\mu|\delta)$ for $\mu_0 = 0$, $\sigma = 1$ and $n = 25$.

Under the null hypothesis Z follows $N(0, 1)$; thus an absolute Z-value of 3.5 is quite unlikely. Therefore if we observe an absolute Z-value of 3.5 we might rule in favor of the alternative hypothesis.

You can see now that we need a threshold value, or in other words a critical point such that if the Z-value exceeds that point we reject. Our test procedure δ then looks like,

$$\text{reject } H_0 \quad \text{if} \quad \left|\frac{\sqrt{n}(\overline{X} - \mu_0)}{\sigma}\right| > c_{n,\alpha_0}$$

where c_{n,α_0} is the *critical value* and will depend on α_0 which is the tolerance for the Type 1 error, i.e., the level that we set beforehand.

The quantity c_{n,α_0} is determined using the relation

$$\mathbb{P}_{\mu_0}\left(\left|\frac{\sqrt{n}(\overline{X} - \mu_0)}{\sigma}\right| > c_{n,\alpha_0}\right) = \alpha_0.$$

Straightforward algebra then yields that

$$P_{\mu_0}\left(-c_{n,\alpha_0}\frac{\sigma}{\sqrt{n}} \leq \overline{X} - \mu_0 \leq c_{n,\alpha_0}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha_0,$$

whence we can choose $c_{n,\alpha_0} = z_{\alpha_0/2}$, the $\frac{\alpha_0}{2}$-th quantile of the $N(0, 1)$ distribution.

The acceptance region $\mathcal{A}$ (or S_0) for the null hypothesis is therefore

$$\mathcal{A} = \left\{\boldsymbol{x} = (x_1, x_2, \ldots, x_n) : \mu_0 - \frac{\sigma}{\sqrt{n}}z_{\alpha_0/2} \leq \overline{x} \leq \mu_0 + \frac{\sigma}{\sqrt{n}}z_{\alpha_0/2}\right\}.$$

So we accept whenever $\overline{X}$ lies in a certain window of μ_0, the postulated value under the null, and reject otherwise which is in accordance with intuition.

The length of the window is determined by the tolerance level α_0, the underlying variance σ^2 and of course the sample size n.

1.5 P-value

The p-**value** is the smallest level α_0 such that we would reject H_0 at level α_0 with the observed data.

For this reason, the p-value is also called the *observed level of significance*.

Example: If the observed value of Z was 2.78, and that the corresponding p-value = 0.0054. It is then said that the observed value of Z is just significant at the level of significance 0.0054.

Advantages:

1. No need to select beforehand an arbitrary level of significance α_0 at which to carry out the test.
2. When we learn that the observed value of Z was just significant at the level of significance 0.0054, we immediately know that H_0 would be rejected for every larger value of α_0 and would not be rejected for any smaller value.

1.6 Testing simple hypotheses: optimal tests

Let the random vector $\mathbf{X} = (X_1, \ldots, X_n)$ come from a distribution for which the joint p.m.f/p.d.f is either $f_0(\mathbf{x})$ or $f_1(\mathbf{x})$. Let $\Omega = \{\theta_0, \theta_1\}$. Then,

- $\theta = \theta_0$ stands for the case in which the data have p.m.f/p.d.f $f_0(\mathbf{x})$,
- $\theta = \theta_1$ stands for the case in which the data have p.m.f/p.d.f $f_1(\mathbf{x})$.

We are then interested in testing the following simple hypotheses:

$$H_0 : \theta = \theta_0 \qquad \text{versus} \qquad H_1 : \theta = \theta_1.$$

In this case, we have special notation for the probabilities of type I and type II errors:

$$\begin{aligned} \alpha(\delta) &= \mathbb{P}_{\theta_0}(\text{Rejecting } H_0), \\ \beta(\delta) &= \mathbb{P}_{\theta_1}(\text{Not rejecting } H_0). \end{aligned}$$

1.6.1 Minimizing the $\mathbb{P}$(Type-II error)

Suppose that the probability $\alpha(\delta)$ of an error of type I is not permitted to be greater than a specified level of significance, and it is desired to find a procedure δ for which $\beta(\delta)$ will be a minimum.

Theorem 1.1 (Neyman-Pearson lemma)**.** *Suppose that δ' is a test procedure that has the following form for some constant $k > 0$:*

- *H_0 is not rejected if $f_1(\mathbf{x}) < kf_0(\mathbf{x})$,*
- *H_0 is rejected if $f_1(\mathbf{x}) > kf_0(\mathbf{x})$, and*
- *H_0 can be either rejected or not if $f_1(\mathbf{x}) = kf_0(\mathbf{x})$.*

Let δ be another test procedure. Then,

$$\begin{array}{llll} \textit{if} & \alpha(\delta) \leq \alpha(\delta'), & \textit{then it follows that} & \beta(\delta) \geq \beta(\delta') \\ \textit{if} & \alpha(\delta) < \alpha(\delta'), & \textit{then it follows that} & \beta(\delta) > \beta(\delta'). \end{array}$$

Example: Suppose that $\mathbf{X} = (X_1, \ldots, X_n)$ is a random sample from the normal distribution with unknown mean θ and known variance 1. We are interested in testing:

$$H_0 : \theta = 0 \qquad \text{versus} \qquad H_1 : \theta = 1.$$

We want to find a test procedure for which $\beta(\delta)$ will be a minimum among all test procedures for which $\alpha(\delta) \leq 0.05$.

We have,

$$f_0(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} x_i^2\right) \qquad \text{and} \qquad f_1(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i - 1)^2\right].$$

After some algebra, the likelihood ratio $f_1(\mathbf{x})/f_0(\mathbf{x})$ can be written in the form

$$\frac{f_1(\mathbf{x})}{f_0(\mathbf{x})} = \exp\left[n\left(\bar{x} - \frac{1}{2}\right)\right].$$

Thus, rejecting H_0 when the likelihood ratio is greater than a specified positive constant k is equivalent to rejecting H_0 when the sample mean $\bar{X}$ is greater than $k' := 1/2 + \log k/n$, another constant. Thus, we want to find, k' such that

$$\mathbb{P}_0(\bar{X} > k') = 0.05.$$

Now,

$$\begin{aligned} & \mathbb{P}_0(\bar{X} > k') = \mathbb{P}_0(\sqrt{n}\bar{X} > \sqrt{n}k') = \mathbb{P}_0(Z > \sqrt{n}k') = 0.05 \\ \Rightarrow \quad & \sqrt{n}k' = 1.645. \end{aligned}$$

1.7 Uniformly most powerful (UMP) tests

Let the null and/or alternative hypothesis be composite

- $H_0 : \theta \leq \theta_0$ and $H_1 : \theta > \theta_0$, or
- $H_0 : \theta \geq \theta_0$ and $H_1 : \theta < \theta_0$

We suppose that Ω_0 and Ω_1 are disjoint subsets of Ω, and the hypotheses to be tested are

$$H_0 : \theta \in \Omega_0 \qquad \text{versus} \qquad H_1 : \theta \in \Omega_1. \tag{5}$$

- The subset Ω_1 contains at least two distinct values of θ, in which case the alternative hypothesis H_1 is composite.
- The null hypothesis H_0 may be either simple or composite.

We consider *only* procedures in which

$$\mathbb{P}_\theta(\text{Rejecting } H_0) \leq \alpha_0 \qquad \forall\, \theta \in \Omega_0.$$

that is

$$\pi(\theta|\delta) \leq \alpha_0 \qquad \forall\, \theta \in \Omega_0$$

or

$$\alpha(\delta) \leq \alpha_0. \tag{6}$$

Finally, among all test procedures that satisfy the requirement (6), we want to find one such that

- the probability of type II error is as small as possible for every $\theta \in \Omega_1$, or
- the value of $\pi(\theta|\delta)$ is as large as possible for every value of $\theta \in \Omega_1$.

There might be no single test procedure δ that maximizes the power function $\pi(\theta|\delta)$ simultaneously for every value of $\theta \in \Omega_1$.

In some problems, however, there will exist a test procedure that satisfies this criterion. Such a procedure, when it exists, is called a UMP test.

Definition 4 (Uniformly most powerful (UMP) test)**.** *A test procedure δ^* is a uniformly most powerful (UMP) test of the hypotheses (5) at the level of significance α_0 if*

$$\alpha(\delta^*) \leq \alpha_0$$

and, for every other test procedure δ such that $\alpha(\delta) \leq \alpha_0$, it is true that

$$\pi(\theta|\delta) \leq \pi(\theta|\delta^*)$$

for every value of $\theta \in \Omega_1$.

Usually no test will uniformly most powerful against ALL alternatives, except in the special case of "monotone likelihood ratio" (MLR).

Example: Suppose that $X_1, \ldots, X_n$ form a random sample from a normal distribution for which the mean μ (unknown) and the variance σ^2 (known). Consider testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$. Even in this simple example, there is no UMP test.

1.8 The t-test

1.8.1 Testing hypotheses about the mean with unknown variance

- Problem: testing hypotheses about the **mean** of a normal distribution when both the mean and the variance are unknown.
- The random variables $X_1, \ldots, X_n$ form a random sample from a normal distribution for which the mean μ and the variance σ^2 are unknown.
- The parameter space Ω in this problem comprises every two-dimensional vector (μ, σ^2), where $-\infty < \mu < \infty$ and $\sigma^2 > 0$.
- $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$
- Define

  $$U_n = \frac{\bar{X}_n - \mu_0}{s_n/\sqrt{n}}, \tag{7}$$

  where $s_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^n \left(X_i - \bar{X}_n\right)^2}$.
- We reject H_0 if

  $$|U_n| \geq T_{n-1}^{-1}\left(1 - \frac{\alpha_0}{2}\right),$$

  the $(1 - \alpha_0/2)$-quantile of the t-distribution with $n-1$ degrees of freedom and U_n is defined in (7).
- p-values for t-tests: The p-value from the observed data and a specific test is the smallest α_0 such that we would reject the null hypothesis at level of significance α_0.

  Let u be the observed value of the statistic U_n. Thus the p-value of the test is

  $$\mathbb{P}(|U_n| > |u|),$$

  where $U_n \sim T_{n-1}$, under H_0.

- The p-value is $2[1 - T_{n-1}(|u|)]$, where u be the observed value of the statistic U_n.

The Complete power function

Before we study the case when $\sigma > 0$ is unknown, let us go back to the case when σ is known.

Our test δ is "reject H_0 if $\left|\frac{\sqrt{n}(\overline{X}-\mu_0)}{\sigma}\right| > z_{\alpha/2}$".

Thus we have,

$$\pi(\mu|\delta) = \mathbb{P}_\mu\left(\left|\frac{\sqrt{n}(\overline{X}-\mu_0)}{\sigma}\right| > z_{\alpha/2}\right),$$

which is just,

$$\mathbb{P}_\mu\left(\left|\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma} + \frac{\sqrt{n}(\mu-\mu_0)}{\sigma}\right| > z_{\alpha/2}\right).$$

But when μ is the population mean, $\sqrt{n}(\overline{X}-\mu)/\sigma$ is $N(0,1)$. If Z denotes a $N(0,1)$ variable then,

$$\begin{aligned}
\pi(\mu|\delta) &= \mathbb{P}_\mu\left(\left|Z + \frac{\sqrt{n}(\mu-\mu_0)}{\sigma}\right| > z_{\alpha/2}\right) \\
&= \mathbb{P}_\mu\left(Z + \frac{\sqrt{n}(\mu-\mu_0)}{\sigma} > z_{\alpha/2}\right) + \mathbb{P}\left(Z + \frac{\sqrt{n}(\mu-\mu_0)}{\sigma} < -z_{\alpha/2}\right) \\
&= 1 - \Phi\left(z_{\alpha/2} - \frac{\sqrt{n}(\mu-\mu_0)}{\sigma}\right) + \Phi\left(-z_{\alpha/2} - \frac{\sqrt{n}(\mu-\mu_0)}{\sigma}\right) \\
&= \Phi\left(-z_{\alpha/2} + \frac{\sqrt{n}(\mu-\mu_0)}{\sigma}\right) + \Phi\left(-z_{\alpha/2} - \frac{\sqrt{n}(\mu-\mu_0)}{\sigma}\right).
\end{aligned}$$

Check from the above calculations that $\pi(\mu_0|\delta) = \alpha$, the level of the test δ.

Notice that the test function δ depends on the value μ_0 under the null but it does not depend on any value in the alternative.

The power increases as the true value μ deviates further from μ_0.

It is easy to check that $\pi(\mu|\delta)$ diverges to 1 as μ diverges to ∞ or $-\infty$.

Moreover the power function is symmetric around μ_0. In other words, $\pi(\mu_0 + \Delta|\delta) = \pi(\mu_0 - \Delta|\delta)$ where $\Delta > 0$.

To see this, note that

$$\pi(\mu_0 + \Delta|\delta) = \Phi\left(-z_{\alpha/2} + \frac{\sqrt{n}\,\Delta}{\sigma}\right) + \Phi\left(-z_{\alpha/2} - \frac{\sqrt{n}\,\Delta}{\sigma}\right).$$

Check that you get the same expression for $\pi(\mu_0 - \Delta|\delta)$.

Exercise: What happens when $\sigma > 0$ is unknown?

We can rewrite U_n as

$$U_n = \frac{\sqrt{n}\big(\bar{X}_n - \mu_0\big)/\sigma}{s_n/\sigma},$$

- The numerator has the normal distribution with mean $\sqrt{n}(\mu - \mu_0)/\sigma$ and variance 1.
- The denominator is the square-root of a χ^2-random variable divided by its degrees of freedom, $n-1$.
- When the mean of the numerator is not 0, U_n has a *non-central* t-distribution.

Definition 5 (Noncentral t-distributions)**.** *Let W and Y_m be independent random variables $W \sim \mathcal{N}(\psi, 1)$ and $Y \sim \chi^2_m$. Then the distribution of*

$$X := \frac{W}{\sqrt{Y_m/m}}$$

is called the **non-central** *t-distribution with m degrees of freedom and non-centrality parameter ψ. We define*

$$T_m(t|\psi) = \mathbb{P}(X \le t)$$

as the c.d.f of this distribution.

- The non-central t-distribution with m degrees of freedom and non-centrality parameter $\psi = 0$ is also the t-distribution with m degrees of freedom.
- The distribution of the statistic U_n in (7) is the non-central t-distribution with $n-1$ degrees of freedom and non-centrality parameter

$$\psi := \sqrt{n}\frac{(\mu - \mu_0)}{\sigma}.$$

- The power function of δ (see Figure 9.14) is

$$\pi(\mu, \sigma^2|\delta) = T_{n-1}(-c|\psi) + 1 - T_{n-1}(c|\psi),$$

  where $c := T^{-1}_{n-1}(1 - \alpha_0/2)$.

Exercise: Prove this result.

1.8.2 One-sided alternatives

We consider testing the following hypotheses:

$$H_0 : \mu \leq \mu_0, \qquad \text{versus} \qquad H_1 : \mu > \mu_0. \tag{8}$$

- When $\mu = \mu_0$, $U_n \sim t_{n-1}$, regardless of the value of σ^2.
- The test rejects H_0 if
$$U_n \geq c,$$
where $c := T_{n-1}^{-1}(1 - \alpha_0)$ (the $(1 - \alpha_0)$-quantile) of the t-distribution with $n - 1$ degrees of freedom.
- $\pi(\mu, \sigma^2|\delta) = 1 - T_{n-1}(c|\psi)$.

Power function of the t-test

Let δ be the test that rejects H_0 in (8) if $U_n \geq c$.

The p-value for the hypotheses in (8) is $1 - T_{n-1}(u)$, where u is the observed value of the statistic U_n.

The power function $\pi(\mu, \sigma^2|\delta)$ has the following properties:

1. $\pi(\mu, \sigma^2|\delta) = \alpha_0$ when $\mu = \mu_0$,
2. $\pi(\mu, \sigma^2|\delta) < \alpha_0$ when $\mu < \mu_0$,
3. $\pi(\mu, \sigma^2|\delta) > \alpha_0$ when $\mu > \mu_0$,
4. $\pi(\mu, \sigma^2|\delta) \to 0$ as $\mu \to -\infty$,
5. $\pi(\mu, \sigma^2|\delta) \to 1$ as $\mu \to \infty$,
6. $\sup_{\theta \in \Omega_0} \pi(\theta|\delta) = \alpha_0$.

When we want to test

$$H_0 : \mu \geq \mu_0 \qquad \text{versus} \qquad H_1 : \mu < \mu_0. \tag{9}$$

the test rejects H_0 if $U_n \leq c$, where $c = T_{n-1}^{-1}(\alpha_0)$ (the α_0-quantile) of the t-distribution with $n - 1$ degrees of freedom.

Power function of the t test

Let δ be the test that rejects H_0 in (9) if $U_n \leq c$.

The p-value for the hypotheses in (9) is $T_{n-1}(u)$. Observe that $\pi(\mu, \sigma^2|\delta) = T_{n-1}(c|\psi)$.

The power function $\pi(\mu, \sigma^2|\delta)$ has the following properties:

1. $\pi(\mu, \sigma^2|\delta) = \alpha_0$ when $\mu = \mu_0$,
2. $\pi(\mu, \sigma^2|\delta) > \alpha_0$ when $\mu < \mu_0$,
3. $\pi(\mu, \sigma^2|\delta) < \alpha_0$ when $\mu > \mu_0$,
4. $\pi(\mu, \sigma^2|\delta) \to 1$ as $\mu \to -\infty$,
5. $\pi(\mu, \sigma^2|\delta) \to 0$ as $\mu \to \infty$,
6. $\sup_{\theta \in \Omega_0} \pi(\theta|\delta) = \alpha_0$.

1.9 Comparing the means of two normal distributions (two-sample t test)

1.9.1 One-sided alternatives

Random samples are available from **two** normal distributions with common unknown variance σ^2, and it is desired to determine which distribution has the larger mean. Specifically,

- $\mathbf{X} = (X_1, \ldots, X_m)$ random sample of m observations from a normal distribution for which both the mean μ_1 and the variance σ^2 are unknown, and
- $\mathbf{Y} = (Y_1, \ldots, Y_n)$ form an independent random sample of n observations from another normal distribution for which both the mean μ_2 and the variance σ^2 are unknown.
- We shall assume that the variance σ^2 is the same for both distributions, even though the value of σ^2 is unknown.

If we are interested in testing hypotheses such as

$$H_0 : \mu_1 \leq \mu_2 \qquad \text{versus} \qquad H_1 : \mu_1 > \mu_2, \tag{10}$$

We reject H_0 in (10) if the difference between the sample means is large. For all values of $\theta = (\mu_1, \mu_2, \sigma^2)$ such that $\mu_1 = \mu_2$, the test statistics

$$U_{m,n} = \frac{\sqrt{m+n-2}\big(\bar{X}_m - \bar{Y}_n\big)}{\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right)(S_X^2 + S_Y^2)}}$$

follows the t-distribution with $m+n-2$ degrees of freedom, where

$$S_X^2 = \sum_{i=1}^{m} \left(X_i - \bar{X}_m\right)^2, \quad \text{and} \quad S_Y^2 = \sum_{j=1}^{n} \left(Y_j - \bar{Y}_n\right)^2.$$

We reject H_0 if

$$U_{m,n} \geq T_{m+n-2}^{-1}(1-\alpha_0).$$

The p-value for the hypotheses in (10) is $1 - T_{m+n-2}(u)$, where u is the observed value of the statistic $U_{m,n}$.

If we are interested in testing hypotheses such as

$$H_0 : \mu_1 \geq \mu_2 \qquad \text{versus} \qquad H_1 : \mu_1 < \mu_2, \tag{11}$$

we reject H_0 if

$$U_{m,n} \leq -T_{m+n-2}^{-1}(1-\alpha_0) = T_{m+n-2}^{-1}(\alpha_0).$$

The p-value for the hypotheses in (11) is $T_{m+n-2}(u)$, where u is the observed value of the statistic $U_{m,n}$.

1.9.2 Two-sided alternatives

If we are interested in testing hypotheses such as

$$H_0 : \mu_1 = \mu_2 \qquad \text{versus} \qquad H_1 : \mu_1 \neq \mu_2, \tag{12}$$

we reject H_0 if

$$|U_{m,n}| \geq T_{m+n-2}^{-1}(1 - \tfrac{\alpha_0}{2}).$$

The p-value for the hypotheses in (12) is $2[1 - T_{m+n-2}(|u|)]$, where u is the observed value of the statistic $U_{m,n}$.

The power function of the two-sided two-sample t test is based on the non-central t-distribution in the same way as was the power function of the one-sample two-sided t-test. The test δ that rejects H_0 when $|U_{m,n}| \geq c$ has power function

$$\pi(\mu_1, \mu_2, \sigma^2 | \delta) = T_{m+n-2}(-c|\psi) + 1 - T_{m+n-2}(c|\psi),$$

where $T_{m+n-2}(\cdot|\psi)$ is the c.d.f of the non-central t-distribution with $m+n-2$ degrees of freedom and non-centrality parameter ψ given by

$$\psi = \frac{\mu_1 - \mu_2}{\sqrt{\sigma^2\left(\frac{1}{m} + \frac{1}{n}\right)}}.$$

1.10 Comparing the variances of two normal distributions (F-test)

- $\mathbf{X} = (X_1, \ldots, X_m)$ random sample of m observations from a normal distribution for which both the mean μ_1 and the variance σ_1^2 are unknown, and
- $\mathbf{Y} = (Y_1, \ldots, Y_n)$ form an independent random sample of n observations from another normal distribution for which both the mean μ_2 and the variance σ_2^2 are unknown.

Suppose that we want to test the hypothesis of equality of the population variances, i.e., $H_0 : \sigma_1^2 = \sigma_2^2$.

Definition 6 (F-distribution)**.** *Let Y and W be independent random variables such that $Y \sim \chi^2_m$ and $W \sim \chi^2_n$. Then the distribution of*

$$X = \frac{Y/m}{W/n}$$

is called the F-distribution with m and n degrees of freedom.

The test statistic

$$V^*_{m,n} = \frac{\frac{S_X^2}{\sigma_1^2}/(m-1)}{\frac{S_Y^2}{\sigma_2^2}/(n-1)} = \frac{\sigma_2^2 S_X^2/(m-1)}{\sigma_1^2 S_Y^2/(n-1)}$$

follows the F-distribution with $m-1$ and $n-1$ degrees of freedom. In particular, if $\sigma_1^2 = \sigma_2^2$, then the distribution of

$$V_{m,n} = \frac{S_X^2/(m-1)}{S_Y^2/(n-1)}$$

is the F-distribution with $m-1$ and $n-1$ degrees of freedom.

Let ν be the observed value of the statistic $V_{m,n}$ below, and let $G_{m-1,n-1}(\cdot)$ be the c.d.f of the F-distribution with $m-1$ and $n-1$ degrees of freedom.

1.10.1 One-sided alternatives

If we are interested in testing hypotheses such as

$$H_0: \quad \sigma_1^2 \leq \sigma_2^2 \qquad \text{versus} \qquad H_1: \quad \sigma_1^2 > \sigma_2^2, \tag{13}$$

we reject H_0 if

$$V_{m,n} \geq G^{-1}_{m-1,n-1}(1-\alpha_0).$$

The p-value for the hypotheses in (13) when $V_{m,n} = \nu$ is observed equals $1 - G_{m-1,n-1}(\nu)$.

1.10.2 Two-sided alternatives

If we are interested in testing hypotheses such as

$$H_0: \quad \sigma_1^2 = \sigma_2^2, \qquad \text{versus} \qquad H_1: \quad \sigma_1^2 \neq \sigma_2^2, \tag{14}$$

we reject H_0 if either $V_{m,n} \leq c_1$ or $V_{m,n} \geq c_2$, where c_1 and c_2 are constants such that

$$\mathbb{P}(V_{m,n} \leq c_1) + \mathbb{P}(V_{m,n} \geq c_2) = \alpha_0$$

when $\sigma_1^2 = \sigma_2^2$. The most convenient choice of c_1 and c_2 is the one that makes

$$\mathbb{P}(V_{m,n} \leq c_1) = \mathbb{P}(V_{m,n} \geq c_2) = \frac{\alpha_0}{2},$$

that is,

$$c_1 = G^{-1}_{m-1,n-1}(\alpha_0/2) \quad \text{and} \quad c_2 = G^{-1}_{m-1,n-1}(1-\alpha_0/2).$$

1.11 Likelihood ratio test

A very popular form of hypothesis test is the **likelihood ratio test**.

Suppose that we want to test

$$H_0: \theta \in \Omega_0, \quad \text{and} \quad H_1: \theta \in \Omega_1. \tag{15}$$

In order to compare these two hypotheses, we might wish to see whether the likelihood function is higher on Ω_0 or on Ω_1.

The *likelihood ratio statistic* is defined as

$$\Lambda(\mathbf{X}) = \frac{\sup_{\theta \in \Omega_0} L_n(\theta, \mathbf{X})}{\sup_{\theta \in \Omega} L_n(\theta, \mathbf{X})}, \tag{16}$$

where $\Omega = \Omega_0 \cup \Omega_1$.

A likelihood ratio test of the hypotheses (15) rejects H_0 when

$$\Lambda(\mathbf{x}) \leq k,$$

for some constant k.

Interpretation: we reject H_0 if the likelihood function on Ω_0 is sufficiently small compared to the likelihood function on all of Ω.

Generally, k is to be chosen so that the test has a desired level α_0.

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

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

at the level α_0. Show that the likelihood ratio test is equivalent to the z-test.

Exercise: Suppose that $X_1, \ldots, X_n$ from a normal distribution $N(\mu, \sigma^2)$ where both μ and σ^2 are unknown. We wish to test the hypotheses

$$H_0 : \sigma^2 = \sigma_0^2 \qquad \text{versus} \qquad H_a : \sigma^2 \neq \sigma_0^2$$

at the level α. Show that the likelihood ratio test is equivalent to the χ^2-test.

Exercise: Suppose that $X_1, \ldots, X_n$ from a normal distribution $N(\mu, \sigma^2)$ where both μ and σ^2 are unknown. We wish to test the hypotheses

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

at the level α. Show that the likelihood ratio test is equivalent to the t-test.

Theorem 1.2. *Let Ω be a open set of a p-dimensional space, and suppose that H_0 specifies that k coordinates of θ are equal to k specific values. Assume that H_0 is true and that the likelihood function satisfies the conditions needed to prove that the MLE is asymptotically normal and asymptotically efficient. Then, as $n \to \infty$,*

$$-2 \log \Lambda(\mathbf{X}) \xrightarrow{d} \chi_k^2.$$

Exercise: Let $X_1, \ldots, X_n$ be a random sample from the p.d.f

$$f_\theta(x) = e^{-(x-\theta)} \mathbf{1}_{[\theta,\infty)}(x).$$

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

Show that the likelihood ratio test statistic is

$$\Lambda(\boldsymbol{X}) = \begin{cases} 1 & X_{(1)} \leq \theta_0 \\ e^{-n(X_{(1)}-\theta_0)} & X_{(1)} > \theta_0. \end{cases}$$

1.12 Equivalence of tests and confidence sets

Suppose that $X_1, \ldots, X_n$ are i.i.d $N(\mu, \sigma^2)$ where μ is unknown and σ^2 is known.

We now illustrate how the testing procedure ties up naturally with the CI construction problem.

Consider testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$.

First note that the acceptance region of the derived test δ can be written as:

$$S_0 = \mathcal{A}_{\mu_0} = \left\{ \boldsymbol{x} = (x_1, x_2, \ldots, x_n) : \overline{x} - \frac{\sigma}{\sqrt{n}} z_{\alpha_0/2} \leq \mu_0 \leq \overline{x} + \frac{\sigma}{\sqrt{n}} z_{\alpha_0/2} \right\}.$$

Now, consider a fixed data set $(X_1, X_2, \ldots, X_n)$ and based on this consider testing a family of null hypotheses:

$$\{H_{0,\tilde{\mu}} : \mu = \tilde{\mu} : \tilde{\mu} \in \mathbb{R}\}.$$

We can now ask the following question: Based on the observed data and the above testing procedure, *what values of $\tilde{\mu}$ would fail to be rejected by the level α_0 test*? This means that $\tilde{\mu}$ would have to fall in the interval

$$\overline{X} - \frac{\sigma}{\sqrt{n}} z_{\alpha_0/2} \leq \tilde{\mu} \leq \overline{X} + \frac{\sigma}{\sqrt{n}} z_{\alpha_0/2}\,.$$

Thus, the set of $\tilde{\mu}$'s for which the null hypothesis would fail to be rejected by the level α_0 test is the set:

$$\left[\overline{X} - \frac{\sigma}{\sqrt{n}} z_{\alpha_0/2}, \overline{X} + \frac{\sigma}{\sqrt{n}} z_{\alpha_0/2} \right].$$

But this is precisely the level $1 - \alpha_0$ CI that we obtained before!

Thus, we obtain a level $1-\alpha_0$ CI for μ, the population mean, by *compiling all possible $\tilde{\mu}$'s for which the null hypothesis $H_{0,\tilde{\mu}} : \mu = \tilde{\mu}$ fails to be rejected by the level α_0 test.*

From hypothesis testing to CIs: Let $X_1, X_2, \ldots, X_n$ be i.i.d observations from some underlying distribution F_θ; here θ is a "parameter" indexing a family of distributions.

For each $\tilde{\theta}$ consider testing the null hypothesis $H_{0,\tilde{\theta}} : g(\theta) = g(\tilde{\theta})$. Suppose, there exists a level α_0 test $\delta_{\tilde{\theta}}$ for this problem with

$$\mathcal{A}_{\tilde{\theta}} = \{\boldsymbol{x} : T_{\tilde{\theta}}(\boldsymbol{x}) \leq c_{\alpha_0}\}$$

being the acceptance region of $\delta_{\tilde{\theta}}$ and

$$\mathbb{P}_{\tilde{\theta}}(\boldsymbol{X} \in \mathcal{A}_{\tilde{\theta}}) \geq 1 - \alpha_0.$$

Then a level $1-\alpha$ confidence set for $g(\theta)$ is:

$$\mathcal{S}(\boldsymbol{X}) = \{g(\tilde{\theta}) : \boldsymbol{X} \in \mathcal{A}_{\tilde{\theta}}\}.$$

We need to verify that for any θ,

$$\mathbb{P}_\theta[g(\theta) \in \mathcal{S}(\boldsymbol{X})] \geq 1-\alpha.$$

But

$$\mathbb{P}_\theta(g(\theta) \in \mathcal{S}(\boldsymbol{X})) = \mathbb{P}_\theta(\boldsymbol{X} \in \mathcal{A}_\theta) \geq 1-\alpha_0.$$

Theorem 1.3. *For each $\theta_0 \in \Omega$, let $\mathcal{A}(\theta_0)$ be the acceptance region of a level α test of $H_0 : \theta = \theta_0$. For each $\boldsymbol{x} \in \mathcal{X}$ ($\mathcal{X}$ is the space of all data values), define a set $\mathcal{S}(\boldsymbol{x})$ in the parameter space by*

$$\mathcal{S}(\boldsymbol{x}) = \{\theta_0 : \boldsymbol{x} \in \mathcal{A}(\theta_0)\}.$$

Then the random set $\mathcal{S}(\boldsymbol{X})$ is a $1-\alpha$ confidence set. Conversely, let $\mathcal{S}(\boldsymbol{X})$ be a $1-\alpha$ confidence set. For any $\theta_0 \in \Omega$, define

$$\mathcal{A}(\theta_0) = \{\boldsymbol{x} : \theta_0 \in \mathcal{S}(\boldsymbol{x})\}.$$

Then $\mathcal{A}(\theta_0)$ is the acceptance region of a level α test of $H_0 : \theta = \theta_0$.

Proof. The first part is essentially done above!

For the second part, the type I error probability for the test of $H_0 : \theta = \theta_0$ with acceptance region $\mathcal{A}(\theta_0)$ is

$$\mathbb{P}_{\theta_0}(\boldsymbol{X} \notin \mathcal{A}_{\theta_0}) = \mathbb{P}_{\theta_0}[\theta_0 \notin \mathcal{S}(\boldsymbol{X})] \leq \alpha.$$

□

Remark: The more useful part of the theorem is the first part, i.e., given a level α test (which is usually easy to construct) we can get a confidence set by inverting the family of tests.

Example: Suppose that $X_1, \ldots, X_n$ are i.i.d Exp(λ). We want to test $H_0 : \lambda = \lambda_0$ versus $H_1 : \lambda \neq \lambda_0$.

Find the LRT.

The acceptance region is given by

$$\mathcal{A}(\lambda_0) = \left\{\boldsymbol{x} : \left(\frac{\sum x_i}{\lambda_0}\right)^n e^{-\sum x_i/\lambda_0} \geq k^*\right\},$$

where k^* is a constant chosen to satisfy

$$\mathbb{P}_{\lambda_0}(\boldsymbol{X} \in \mathcal{A}(\lambda_0)) = 1 - \alpha.$$

Inverting this acceptance region gives the $1 - \alpha$ confidence set

$$\mathcal{S}(\boldsymbol{x}) = \left\{ \lambda : \left(\frac{\sum x_i}{\lambda} \right)^n e^{-\sum x_i / \lambda_0} \geq k^* \right\}.$$

This can be shown to be an interval in the parameter space.