3. Hypothesis Testing

3.1 Introduction

In this chapter we discuss how, by using data, we can prove statements about the parameters.

Consider for instance the following scenario. Train companies regularly collect passenger data on customer satisfaction. One may ask whether the frequent ticket price increases cause any drop in average customer satisfaction. The population of interest is the train passengers and the parameter of interest is the average customer satisfaction. We are interested in assessing whether the ticket price increase has an impact on the average customer satisfaction.

Definition 3.1 (Statistical hypothesis)**.**
A **statistical hypothesis** is a statement about the parameter value of the population under study.

Definition 3.2 (Test of significance)**.**
The **test of significance** is a rule, based on data, for deciding which hypothesis is true between two competing hypotheses: the *null hypothesis* (denoted by H_0), and the *alternative hypothesis* (denoted by H_1).

The **null hypothesis** corresponds to the common belief about the parameter in question. It is interpreted as *no change* in the value of the parameter.

The **alternative hypothesis** corresponds to a new claim which we wish to prove. It is interpreted as *a change* in the value of the parameter.

The outcome of a test of significance is the decision whether to reject or not the null hypothesis.

Example 3.1. A company is selling bathroom toiletries and cosmetics. Their daily sales is a normally distributed random variable $\mathrm{N}(\mu, \sigma^2)$ with mean $\mu = 2000$ products. They believe that their plan of giving a free sample of one type of their products when you buy another of their products will increase their average daily sales by 100.

- What are the null and alternative hypotheses?

  The claim is that with the offer the average daily sales will become 2100 (increase by 100). This statement determines the alternative hypothesis (a new claim). The null hypothesis corresponds to no change in the average daily sales, i.e. they will remain at 2000. Therefore, the two hypotheses are: H_0: $\mu = 2000$ and H_1: $\mu = 2100$.

- If the claim was that the new offer will increase the sales (without saying by how much), what would the two hypotheses be?

  In this case the claim is that the average daily sales will be some number $\mu > 2000$ *while the null hypothesis is as before. Then* H_0*:* $\mu = 2000$ *and* H_1*:* $\mu > 2000$*.*

- If the claim was that the new offer will have some impact on the sales, what would the two hypotheses be?

  In this case the claim is that the average daily sales will be some number different from 2000 *while the null hypothesis is as before. Then* H_0*:* $\mu = 2000$ *and* H_1*:* $\mu \neq 2000$*.*

- What can a statistician do to confirm the claim that the new offer improves sales?

  Here we want to compare H_0*:* $\mu = 2000$ *vs* H_1*:* $\mu > 2000$*. The company may consider the following experiment. Introduce the offer for a period of time, say* $n = 30$ *days, and record the number of sales on each day. Let* $x_1, \ldots, x_n$ *be the number of sales for each day, i.e. the sample, and let* $\bar{x}$ *be the sample average. If* $\bar{x}$ *turns out to be significantly larger than 2000, then there is evidence that the sales have improved.*

  There is no easy answer to what "significantly larger" means. That's a matter of personal opinion. For some people if $\bar{x} > 2010$ *is enough to indicate increase in average daily sales but others may require* $\bar{x} > 2100$*. If we set this "critical value" for* $\bar{x}$ *too low, say 2010, then there is the danger of a false positive conclusion, i.e. claiming that the average sales increased while in reality they stayed the same and the fact that* $\bar{x} > 2010$ *was just due to variability in the sample. On the other hand, if the critical value is set to a large number, say 2100, then there is the possibility of a false negative conclusion, i.e. claiming that the average daily sales haven't increased when in reality they did but not by that much as to bring* $\bar{x}$ *to exceed 2100.*

▶

In order to draw a conclusion in the example above, we had to summarise the data into one number, in that case $\bar{x}$, and compare this number against a critical boundary and depending on whether $\bar{x}$ exceeded or not this critical boundary then we decide which hypothesis to accept. This brings us to the concept of the *test statistic* and its *critical value*.

Definition 3.3 (Test statistic and critical value)**.**
The **test statistic** associated with a hypothesis test is the statistic, i.e. a number derived from the sample, which is used to make a decision in a hypothesis test. The value of the test statistic is compared against some predetermined number called the **critical value** of the statistic. If the value of the test statistic exceeds the critical value then our decision is to reject the H_0.

As with any decision we make, we may reach the wrong conclusion. These are commonly referred to as "false positive" and "false negative" conclusions but in the language of statistics they are called *Type I error* and *Type II error*.

Type I error means that our decision was to reject H_0 when in reality the H_0 is true.

Type II error means that our decision was to accept H_0 when in reality the H_0 was false.

We would like to minimise the likelihood of reaching the wrong decision or maximise the likelihood of reaching the correct decision. Therefore for every hypothesis test we need to know the probabilities $\mathbb{P}$(Type I Error) and $\mathbb{P}$(Type II Error). We define

Definition 3.4 (Power)**.**
For every hypothesis test, we define the **power** to be

$$\text{power} = 1 - \mathbb{P}(\text{Type II Error}).$$

The power can be interpreted as the probability of correctly rejecting H_0. So we want to have a test with high power. When the alternative hypothesis is a range of values, the power can be defined for all those values, so, in this case, it is represented by a function on θ.

Example 3.2. Refer to Example 3.1. Suppose that the standard deviation is known to be $\sigma = 300$ and we want to test H_0: $\mu = 2000$ v.s. H_1: $\mu > 2000$. In a sample of $n = 30$ days we decide to reject the H_0 if $\bar{X} > 2070$.

- What is the probability of Type I error?

  By the definition of Type I error, the probability is $\mathbb{P}(\textit{Type I error}) = \mathbb{P}(\textit{Reject } H_0 | H_0 \textit{ is true})$.

  The statement "Reject H_0" is equivalent to $\bar{X} > 2070$ *and the statement "H_0 is true" is equivalent to* $\mu = 2000$. *Therefore,* $\mathbb{P}(\textit{Type I error}) = \mathbb{P}(\bar{X} > 2070 | \mu = 2000)$.

  In order to compute this probability, we need to know the distribution of the random variable inside the parentheses, namely $\bar{X}$. This distribution is the normal distribution with mean μ and variance σ^2/n. In this case $\mu = 2000$ *and* $\sigma^2/n = 300^2/30$.

  Then, $z = \frac{2070-2000}{300/\sqrt{30}} = 1.28$, *which corresponds to probability* $\Phi(1.28) = 0.8997$. *Therefore,* $\mathbb{P}(\textit{Type I error}) = 1 - 0.8997 = \mathbf{0.1003}$.

- What is the probability of Type II error if the true mean is 2100?

  By the definition of Type II error, the probability is $\mathbb{P}(\textit{Type II error}) = \mathbb{P}(\textit{Accept } H_0 | H_0 \textit{ is false})$.

  As before, the statement "Accept H_0" is equivalent to $\bar{X} \leq 2070$ *and the statement "H_0 is false" in this case is equivalent to* $\mu = 2100$. *Therefore,* $\mathbb{P}(\textit{Type I error}) = \mathbb{P}(\bar{X} \leq 2070 | \mu = 2100)$. *The distribution of $\bar{X}$ in this case is normal with mean* $\mu = 2100$ *and variance* $\sigma^2/n = 300^2/30$.

  Then, $z = \frac{2070-2100}{300/\sqrt{30}} = -0.55$, *which corresponds to probability* $\Phi(-0.55) = 0.2912$. *Therefore,*
  $\mathbb{P}(\textit{Type II error}) = \mathbf{0.2912}$ *and* $\text{power} = 1 - 0.2912 = \mathbf{0.7088}$.

- What is the power function?

  Let $\mu > 2000$. *Then, as before* $z = \frac{2070-\mu}{300/\sqrt{30}}$, *so, as in the previous section* $\text{power}(\mu) = 1 - \Phi(\frac{2070-\mu}{300/\sqrt{30}}) = \Phi(\frac{\mu-2070}{300/\sqrt{30}})$, *because of the symmetry of the function $\Phi(z)$ around 0, i.e.,* $1 - \Phi(z) = \Phi(-z)$.

  This function is plotted in Figure 3.1. We observe that the further μ is from 2000, the higher the power. This can be interpreted as being more likely to reject H_0 correctly when the true mean is further from 2000.

In this example, $\bar{X}$ is the test statistic and the number 2070 is the critical value c. The probabilities of Type I and Type II errors can be illustrated in Figure 3.2. The bell curve on the left is the distribution of $\bar{X}$ under H_0 and the dark grey area is the probability of Type I error. Similarly, the bell curve on the right is the distribution of $\bar{X}$ under H_1 when the mean is 2100 and the light grey area is the probability of Type II error. Both probabilities correspond to the critical value $c = 2070$. ▶

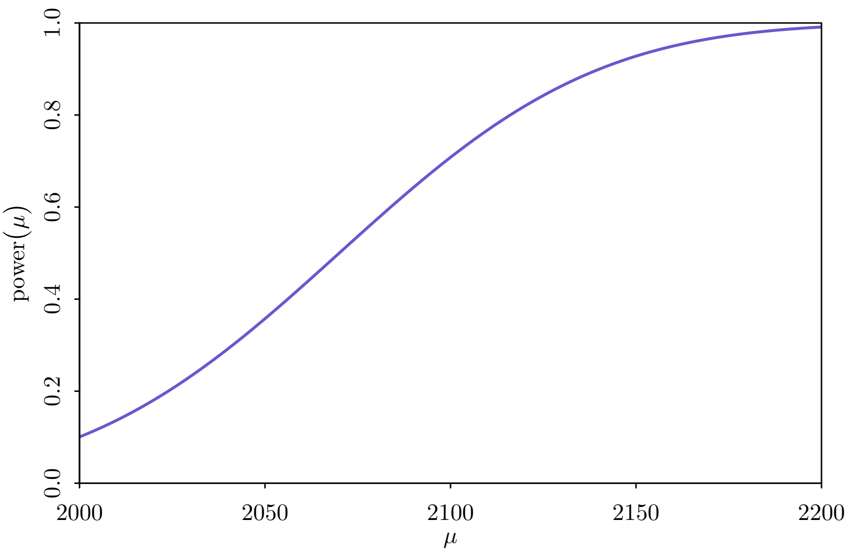

Figure 3.1: The power function for Example 3.2.

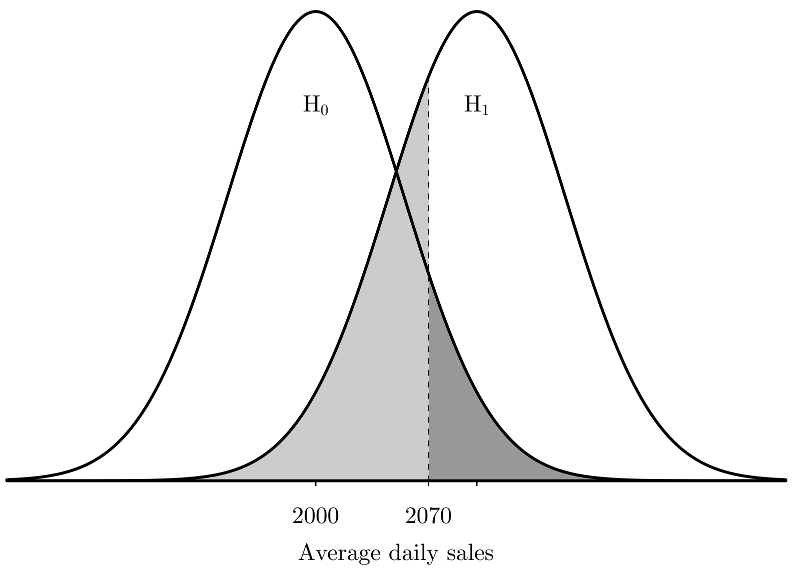

Figure 3.2: Illustration of the probabilities of Type I (dark gray area) and Type II (light gray area) errors for Example 3.2.

Now let's see what happens if the critical value changes. If we increase the critical value, the probability of Type I error (dark grey area) becomes smaller. This means that it becomes *less likely* to reject H_0 erroneously (a false positive). However, we also increase the probability of Type II error (light grey area) so it becomes *more likely* to accept H_0 when we shouldn't (a false negative). Unfortunately this is a usual impediment when performing hypothesis tests. In practice, we demand the probability of Type I error to be comfortably small typically around 5% which determines the critical value. This concept gives rise to the notion of *significance level* and *p-value*.

Definition 3.5 (Level of significance)**.**
It is desirable to have a rule for rejecting H_0 with small probability of Type I error. In practice this probability is set to a fixed, prescribed, value denoted by α (the greek letter "alpha") called the **level of significance** and a rule with the property $\mathbb{P}(\text{Type I error}) = \alpha$ is sought.

The smaller the value of α, the more evidence needed to reject the H_0. Typical values for α are 1%, 5% and 10%.

Definition 3.6 (p-value)**.**
The ***p*-value** is the smallest significance level which we can set and still be able to reject H_0 with the given data. By definition, the p-value is a probability which depends on the sample we are analysing. If p-value $< \alpha$ then the data provide enough evidence to reject H_0.

The two definitions suggest two paths one can follow to conduct a hypothesis test. In both cases, one the data and a significance level are provided. Then we could either

- Use the data to derive the test statistic and use the significance level to derive the critical value. If the value of the test statistic exceeds the critical value then we reject H_0, otherwise we accept it.
- Use the data to derive the test statistic and from there derive the corresponding p-value. If the p-value is smaller than the significance level then we reject H_0, otherwise we accept it.

These two paths are depicted below.

$$\begin{tikzcd}
\text{Data} \arrow[d] & \alpha \arrow[d] \\
\text{Test stat.} \arrow[r, leftrightarrow] & c
\end{tikzcd}
\qquad
\begin{tikzcd}
\text{Data} \arrow[d] & \alpha \arrow[dl, leftrightarrow] \\
\text{Test stat.} \arrow[d] & \\
p\text{-value} &
\end{tikzcd}$$

It shouldn't matter which of the two approaches we take to perform the hypothesis test as they are equivalent, meaning that the lead to the same conclusion regarding the rejection of the null hypothesis.

3.2 Simple vs simple tests and extensions

Let us consider a population $f(x|\theta)$, $\theta \in \Theta$ and two values $\theta_0, \theta_1 \in \Theta$ for which we want to perform a hypothesis test

$$\mathrm{H}_0\colon\ \theta = \theta_0 \text{ v.s. } \mathrm{H}_1\colon\ \theta = \theta_1.$$

These type of hypotheses, which consist of only equalities, are called **simple**. To that end, let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f(x|\theta)$ and consider the likelihood ratio

$$T(X_1, \ldots, X_n|\theta_0, \theta_1) = \frac{L(\theta_1|X_1, \ldots, X_n)}{L(\theta_0|X_1, \ldots, X_n)},$$

where $L(\theta|X_1, \ldots, X_n)$ denotes the likelihood function, $L(\theta|X_1, \ldots, X_n) = \prod_{i=1}^n f(X_i|\theta)$. If, for a particular sample $\mathbf{x} = (x_1, \ldots, x_n)$, $T(\mathbf{x}|\theta_0, \theta_1)$ is "small", then the data favour the null hypothesis H_0. If, on the other hand, $T(\mathbf{x}|\theta_0, \theta_1)$ is "large", then the data provide significant evidence to reject H_0. Therefore, the hypothesis test will have the form

$$\text{Reject } \mathrm{H}_0 \text{ if } T(\mathbf{x}|\theta_0, \theta_1) > c,$$

where c denotes the critical value which is to be determined for a given significance level α. In particular, considering the distribution of the random variable $T(\mathbf{X}|\theta_0, \theta_1)$, we choose c such that

$$\mathbb{P}(T(\mathbf{X}|\theta_0, \theta_1) > c \mid \theta = \theta_0) = \alpha.$$

Similarly, the p-value of the test corresponds to the probability

$$p\text{-value} = \mathbb{P}(T(\mathbf{X}|\theta_0, \theta_1) > t \mid \theta = \theta_0),$$

where $t = T(\mathbf{x}|\theta_0, \theta_1)$ denotes the observed value of the test statistic for the given sample.

Example 3.3. Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathrm{N}(\mu, \sigma^2)$ and we wish to test

$$\mathrm{H}_0\colon\ \mu = \mu_0 \text{ v.s. } \mathrm{H}_1\colon\ \mu = \mu_1,$$

where $\mu_1 > \mu_0$ are given possible values for μ. Suppose a sample $\mathbf{x} = (x_1, \ldots, x_n)$ is observed. The likelihood function is given by

$$\begin{aligned} L(\mu|\mathbf{x}) &= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu)^2\right\} \\ &= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^n x_i^2 - 2n\mu\bar{x} + \mu^2\right)\right\}, \end{aligned}$$

so the likelihood ratio is

$$\begin{aligned} \frac{L(\mu_1|\mathbf{x})}{L(\mu_0|\mathbf{x})} &= \frac{(2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^n x_i^2 - 2n\mu_1\bar{x} + \mu_1^2\right)\right\}}{(2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^n x_i^2 - 2n\mu_0\bar{x} + \mu_0^2\right)\right\}} \\ &= \exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^n x_i^2 - 2n\mu_1\bar{x} + \mu_1^2\right) + \frac{1}{2\sigma^2}\left(\sum_{i=1}^n x_i^2 - 2n\mu_0\bar{x} + \mu_0^2\right)\right\} \\ &= \exp\left\{\frac{n\bar{x}}{\sigma^2}(\mu_1 - \mu_0) - \frac{1}{2\sigma^2}(\mu_1^2 - \mu_0^2)\right\}. \end{aligned}$$

This shows that the likelihood ratio depends on the data only through the sample mean, $\bar{x}$. In fact, because $\mu_1 > \mu_0$, the likelihood ratio is an increasing function of $\bar{x}$. Therefore, for every rule of the form

$$\text{Reject H}_0 \text{ if } \frac{L(\mu_1|\mathbf{x})}{L(\mu_0|\mathbf{x})} > c,$$

for a critical value c, there is an equivalent rule of the form

$$\text{Reject H}_0 \text{ if } \bar{x} > c^*,$$

for a different critical value c^*. It is easier to come up with the critical value c^* than c because the distribution of $\bar{X}$ is straightforward, it is $\bar{X} \sim \mathrm{N}(\mu, \sigma^2/n)$. Of course this distribution depends on σ^2. We will distinguish two cases, when σ^2 is known and when it is not. In either case we seek c^* such that $\mathbb{P}(\bar{X} > c^* \mid \mu = \mu_0) = \alpha$.

Let $Z = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$. Then, assuming H_0 holds, i.e., $\mu = \mu_0$, $Z \sim \mathrm{N}(0, 1)$ and $\mathbb{P}(\bar{X} > c^* \mid \mu = \mu_0) = \mathbb{P}\left(Z > \dfrac{c^* - \mu_0}{\sigma/\sqrt{n}}\right)$. Therefore, in order to $\mathbb{P}(\bar{X} > c^* \mid \mu = \mu_0) = \alpha$ we must choose $\dfrac{c^* - \mu_0}{\sigma/\sqrt{n}} = z_{1-\alpha}$. Rearranging terms gives us $c^* = \mu_0 + z_{1-\alpha} \times \sigma/\sqrt{n}$. Of course this is only possible if we know the true value of σ^2. In fact, we could even phrase our decision rule in terms of Z, which suggests

$$\text{Reject H}_0 \text{ if } z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > z_{1-\alpha}.$$

Then, the p-value of the test corresponding to the observed value z is

$$p\text{-value} = \mathbb{P}(Z > z), \text{ where } Z \sim \mathrm{N}(0, 1).$$

The power of the test is

$$\begin{aligned}
\text{power} &= \mathbb{P}\left(\bar{X} > \mu_0 + z_{1-\alpha} \times \sigma/\sqrt{n} \mid \mu = \mu_1\right) \\
&= \mathbb{P}\left(\frac{\bar{X} - \mu_1}{\sigma/\sqrt{n}} > \frac{\mu_0 - \mu_1 + z_{1-\alpha} \times \sigma/\sqrt{n}}{\sigma/\sqrt{n}} \mid \mu = \mu_1\right) \\
&= \mathbb{P}\left(\frac{\bar{X} - \mu_1}{\sigma/\sqrt{n}} > \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} + z_{1-\alpha} \mid \mu = \mu_1\right) \\
&= 1 - \Phi\left(\frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} + z_{1-\alpha}\right) = \Phi\left(\frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}} - z_{1-\alpha}\right).
\end{aligned}$$

If σ^2 is unknown then our hypothesis test is derived in terms of $T = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}}$, which, under H_0, $T \sim \mathrm{t}_{n-1}$. Our decision rule then becomes

$$\text{Reject H}_0 \text{ if } t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} > t_{n-1;1-\alpha},$$

and the p-value of the test corresponding to the observed value t is

$$p\text{-value} = \mathbb{P}(T > t), \text{ where } T \sim \mathrm{t}_{n-1}.$$

▶

Example 3.4. Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Exponential}(\mu)$, $\mu > 0$ and we wish to test

$$\text{H}_0\colon\ \mu = \mu_0 \text{ v.s. } \text{H}_1\colon\ \mu = \mu_1,$$

where $0 < \mu_1 < \mu_0$. Suppose a sample $\mathbf{x} = (x_1, \ldots, x_n)$ is observed. The likelihood function is given by

$$\begin{aligned} L(\mu|\mathbf{x}) &= \mu^{-n} \exp\left\{-\sum_{i=1}^{n} x_i/\mu\right\} \\ &= \mu^{-n} \exp\{-n\bar{x}/\mu\} \end{aligned}$$

so the likelihood ratio is

$$\begin{aligned} \frac{L(\mu_1|\mathbf{x})}{L(\mu_0|\mathbf{x})} &= \frac{\mu_1^{-n} \exp\{-n\bar{x}/\mu_1\}}{\mu_0^{-n} \exp\{-n\bar{x}/\mu_0\}} \\ &= \left(\frac{\mu_1}{\mu_0}\right)^{-n} \exp\left\{n\bar{x}\left(\frac{1}{\mu_0} - \frac{1}{\mu_1}\right)\right\}. \end{aligned}$$

This shows that the likelihood ratio depends on the data only through the sample mean, $\bar{x}$. In fact, because $0 < \mu_1 < \mu_0$, the likelihood ratio is a decreasing function of $\bar{x}$. Therefore, for every rule of the form

$$\text{Reject H}_0 \text{ if } \frac{L(\mu_1|\mathbf{x})}{L(\mu_0|\mathbf{x})} > c,$$

for a critical value c, there is an equivalent rule of the form

$$\text{Reject H}_0 \text{ if } \bar{x} < c^*,$$

for a different critical value c^*. If we let $W = \sum X_i$, the distribution of W/μ is $W/\mu \sim \text{Gamma}(n, 1)$ (see Example 2.12). Let γ_α denote the α left quantile of the $\text{Gamma}(n, 1)$ distribution. Then, under H_0,

$$\alpha = \mathbb{P}(W/\mu < \gamma_\alpha \mid \mu = \mu_0) = \mathbb{P}(W/\mu_0 < \gamma_\alpha) = \mathbb{P}(W < \gamma_\alpha \mu_0) = \mathbb{P}(\bar{X} < \gamma_\alpha \mu_0/n),$$

so choosing the critical value $c^* = \gamma_\alpha \mu_0/n$ will ensure we have the correct level.

Let $\bar{x}$ be the observed value out of a random sample of size n. The p-value is computed by

$$p\text{-value} = \mathbb{P}(\bar{X} < \bar{x}) = \mathbb{P}(W/\mu_0 < n\bar{x}/\mu_0) = \mathbb{P}(Y < n\bar{x}/\mu_0),$$

where $Y \sim \text{Gamma}(n, 1)$.

If instead $\mu_1 > \mu_0$, then the same calculation shows that the likelihood ratio is now an increasing function of $\bar{x}$ so the equivalent rejection rule in terms of $\bar{x}$ becomes

$$\text{Reject H}_0 \text{ if } \bar{x} > c^{**}.$$

To find c^{**}, we proceed as previously and define $\gamma_{1-\alpha}$ to be the $(1-\alpha)$ left quantile of the $\text{Gamma}(n, 1)$ distribution which allows probability α to its right. Then, under H_0,

$$\alpha = \mathbb{P}(W/\mu > \gamma_{1-\alpha} \mid \mu = \mu_0) = \mathbb{P}(W/\mu_0 > \gamma_{1-\alpha}) = \mathbb{P}(\bar{X} > \gamma_{1-\alpha}\mu_0/n),$$

so the critical value is now $c^{**} = \gamma_{1-\alpha}\mu_0/n$ and

$$p\text{-value} = \mathbb{P}(\bar{X} > \bar{x}) = \mathbb{P}(W/\mu_0 > n\bar{x}/\mu_0) = \mathbb{P}(Y > n\bar{x}/\mu_0),$$

where $Y \sim \text{Gamma}(n, 1)$. ▶

3.2.1 One-sided alternative hypothesis

Now, consider the hypothesis test for

$$\mathrm{H}_0\colon \theta = \theta_0 \text{ v.s. } \mathrm{H}_1\colon \theta > \theta_0,$$

or

$$\mathrm{H}_0\colon \theta = \theta_0 \text{ v.s. } \mathrm{H}_1\colon \theta < \theta_0.$$

This type of alternative hypothesis is called one-sided. Each of these alternatives is like equivalent to $\theta = \theta_1$ with $\theta_1 > \theta_0$ and with $\theta_1 < \theta_0$ respectively. Therefore, a one-sided alternative hypothesis is tested as the equivalent simple hypothesis.

Example 3.5. Let $X_1, \ldots, X_n$ be as in Example 3.3 and we wish to test

$$\mathrm{H}_0\colon \mu = \mu_0 \text{ v.s. } \mathrm{H}_1\colon \mu > \mu_0.$$

Then, the hypothesis test is as derived in Example 3.3. However, the power is now defined for all values $\mu_1 > \mu_0$ so, in the case where σ^2 is known, this becomes

$$\text{power}(\mu_1) = \Phi\left(\frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}} - z_{1-\alpha}\right).$$

►

3.2.2 Two-sided alternative hypothesis

Here, we consider the hypothesis test of the form

$$\mathrm{H}_0\colon \theta = \theta_0 \text{ v.s. } \mathrm{H}_1\colon \theta \neq \theta_0.$$

Let α be the desired significance level. Our hypothesis test can be split into two separate hypothesis tests

$$\mathrm{H}_0\colon \theta = \theta_0 \text{ v.s. } \mathrm{H}_2\colon \theta < \theta_0,$$

and

$$\mathrm{H}_0\colon \theta = \theta_0 \text{ v.s. } \mathrm{H}_3\colon \theta > \theta_0,$$

with different significance levels α_2 and α_3, such that the aggregate significance level is $\alpha_2 + \alpha_3 = \alpha$. It is common to choose $\alpha_2 = \alpha_3 = \alpha/2$ for each of the two separate tests. Then the null hypothesis, H_0, corresponding to the original test is rejected if H_0 is rejected for either of the two separate hypothesis tests.

Let p_2, p_3 denote the p-values of the tests corresponding to the alternatives H_2 and H_3 respectively. Then, for the original test,

$$p\text{-value} = 2\min\{p_2, p_3\}.$$

Example 3.6. Let $X_1, \dots, X_n$ be as in Example 3.4 and we wish to test

$$\mathrm{H}_0\colon\ \mu = \mu_0 \text{ v.s. } \mathrm{H}_1\colon\ \mu \neq \mu_0.$$

Consider first the hypothesis test

$$\mathrm{H}_0\colon\ \mu = \mu_0 \text{ v.s. } \mathrm{H}_2\colon\ \mu < \mu_0,$$

with desired level α_2. Then, following Example 3.4, the H_0 for this test is rejected if $\bar{x} < \mu_0 \gamma_{\alpha_2}/n$.

Next, consider the hypothesis test

$$\mathrm{H}_0\colon\ \mu = \mu_0 \text{ v.s. } \mathrm{H}_3\colon\ \mu > \mu_0,$$

with desired level α_3. Again, following Example 3.4, the H_0 for this test is rejected if $\bar{x} > \mu_0 \gamma_{1-\alpha_3}/n$.

Combining the two separate tests, and choosing $\alpha_2 = \alpha_3 = \alpha/2$, we arrive to the following rejection rule

$$\text{Reject } \mathrm{H}_0 \text{ if } \bar{x} < \mu_0 \gamma_{\alpha/2}/n \text{ or if } \bar{x} > \mu_0 \gamma_{1-\alpha/2}/n,$$

and the p-value for this test is

$$p\text{-value} = 2 \min \left\{ \mathbb{P}(Y < n\bar{x}/\mu_0), \mathbb{P}(Y > n\bar{x}/\mu_0) \right\},$$

where $Y \sim \mathrm{Gamma}(n, 1)$. ▶

3.3 Exercises

1. Consider the setup of Example 3.3 but with $\mu_1 < \mu_0$. Derive the hypothesis test in this case for σ^2 known and for σ^2 unknown.

2. Let $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathrm{Bernoulli}(\theta)$, $\theta \in (0, 1)$. Derive a hypothesis test for testing $\mathrm{H}_0\colon\ \theta = \theta_0$ v.s. $\mathrm{H}_1\colon\ \theta > \theta_0$.

3. Let $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathrm{Bernoulli}(\theta)$, $\theta \in (0, 1)$. Derive a hypothesis test for testing $\mathrm{H}_0\colon\ \theta = \theta_0$ v.s. $\mathrm{H}_1\colon\ \theta \neq \theta_0$.

4. Let $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathrm{N}(\mu, \sigma^2)$. Derive a hypothesis test for testing $\mathrm{H}_0\colon\ \mu = \mu_0$ v.s. $\mathrm{H}_1\colon\ \mu \neq \mu_0$ for the cases where σ^2 is known and when it is not. What is the p-value of the test?