Hypothesis Tests

1 Basic Concepts of Hypothesis Tests

We will now examine hypothesis tests. To become familiar with them, we first look at some basic concepts. After that, we consider the simple case where there are only two hypotheses.

1.1 Definitions

Let's suppose we have a random vector $\underline{\mathbf{y}}$. Its PDF is $f_{\underline{\mathbf{y}}}(\mathbf{y}|\mathbf{x})$, where the vector $\mathbf{x}$ is not known. We can now assume a certain value for $\mathbf{x}$. Afterwards, we can use a measurement $\mathbf{y}$ to check whether our original assumption of $\mathbf{x}$ made sense. We now have a **statistical hypothesis**, denoted by $H : \underline{\mathbf{y}} \sim f_{\underline{\mathbf{y}}}(\mathbf{y}|\mathbf{x})$.

The **(process) state space** contains all the possible values for $\mathbf{x}$. Often an hypothesis H states that $\mathbf{x}$ has a certain value. It now completely specifies the distribution of $\underline{\mathbf{y}}$. It is therefore called a **simple hypothesis**. H can also state that $\mathbf{x}$ is among a certain group of values. In this case $\underline{\mathbf{y}}$ is not completely specified. H is then called a **composite hypothesis**. In this course we only deal with simple hypotheses.

1.2 The binary decision problem

Usually, when you examine hypotheses, you have two hypothesis. It is possible to have multiple hypothesis $H_1, H_2, \ldots, H_n$, but we will treat that later. For now we assume we have just two hypotheses. First there is the **null hypothesis** $H_0 : \underline{\mathbf{y}} \sim f_{\underline{\mathbf{y}}}(\mathbf{y}|\mathbf{x_0})$, representing the **nominal state**. Second, there is the **alternative hypothesis** $H_a : \underline{\mathbf{y}} \sim f_{\underline{\mathbf{y}}}(\mathbf{y}|\mathbf{x_a})$. Both distributions state that the random variable $\underline{\mathbf{y}}$ has a certain PDF $f_{\underline{\mathbf{y}}}$.

Let's examine the **binary decision problem**. We have a single observation $\mathbf{y}$. Based on this observation, we have to choose whether we accept H_0 (assume it to be correct) or reject it. The procedure used to decide whether to accept H_0 or not is called a **test**.

How do we decide whether to accept H_0? For that, we define the **critical region** K. If $\mathbf{y} \in K$, then we reject H_0. On the other hand, if $\mathbf{y} \notin K$ (or equivalently, $\mathbf{y} \in K^c$), then we accept H_0. We can also define the **test statistic** $T = h(\mathbf{y})$, where T is a scalar and $h(\mathbf{y})$ some function. Corresponding to the (multi-dimensional) region K is also a scalar region K. We now reject H_0 if $T \in K$ and accept H_0 is $T \notin K$.

1.3 Four situations

In the binary decision problem, we have two options: accept or reject. In this choice, we can be either right or wrong. There are now four possible situations:

- We reject H_0, when in reality H_0 is true. So we made an error. This is called the **type 1 error**. Its probability, called the **probability of false alarm** α, can be found using
$$\alpha = P(\underline{\mathbf{y}} \in K|H_0) = \int_K f_{\underline{\mathbf{y}}}(\mathbf{y}|H_0)\, dy = \int_K f_{\underline{\mathbf{y}}}(\mathbf{y}|x_0)\, dy, \tag{1.1}$$
where the latter part is just a different way of writing things. α is also called the **size** of the test, or the **level of significance**.
- We accept H_0, when in reality H_0 is false. This time we made a **type 2 error**. The so-called **probability of missed detection** β is given by
$$\beta = P(\underline{\mathbf{y}} \notin K|H_a) = \int_{K^c} f_{\underline{\mathbf{y}}}(\mathbf{y}|H_a)\, dy = 1 - \int_K f_{\underline{\mathbf{y}}}(\mathbf{y}|H_a). \tag{1.2}$$

- We accept H_0, and were right in doing so. The so-called **probability of correct dismissal** (which doesn't have its own sign) is now given by

$$P(\underline{\mathbf{y}} \notin K|H_0) = \int_{K^c} f_{\underline{\mathbf{y}}}(\mathbf{y}|H_0)\,dy = 1 - \int_K f_{\underline{\mathbf{y}}}(\mathbf{y}|H_0)\,dy = 1 - \alpha \tag{1.3}$$

- We reject H_0, and were right in doing so. The **probability of detection**, also called the **power** of the test γ, now is

$$\gamma = P(\underline{\mathbf{y}} \in K|H_a) = \int_K f_{\underline{\mathbf{y}}}(\mathbf{y}|H_a)\,dy = 1 - \beta. \tag{1.4}$$

1.4 Defining the critical region

The size of α, β and γ depends on the critical region K. It is our task to define K. According to what criteria should we do that? Often, in real life, we want to minimize costs. A false alarm has a certain (positive) cost c_0, while a missed detection has a certain (also positive) cost c_a. The **average cost** in an experiment, also called the **Bayes risk**, is then $c_0\alpha + c_a\beta$. (We assume a correct choice has no costs, nor any special benefits.) We want to find the K for which the costs are at a minimum. (In fact, the **Bayes criterion** states that the Bayes risk should be minimized.) So we want to minimize

$$c_0\alpha + c_a\beta = c_0 \int_K f_{\underline{\mathbf{y}}}(\mathbf{y}|H_0)\,dy + c_a\left(1 - \int_K f_{\underline{\mathbf{y}}}(\mathbf{y}|H_a)\,dy\right) = c_a + \int_K \left(c_0 f_{\underline{\mathbf{y}}}(\mathbf{y}|H_0) - c_a f_{\underline{\mathbf{y}}}(\mathbf{y}|H_a)\right)\,dy. \tag{1.5}$$

We know that c_a is constant. So we should minimize the integral on the right. Note that an integral is something that adds up infinitely many numbers. By choosing K, we choose what numbers this integral adds up. We want to minimize the value of the integral. So we should make sure it only adds up negative numbers. (Any positive number would make its value only bigger.) So, we only have $\mathbf{y} \in K$ if

$$c_0 f_{\underline{\mathbf{y}}}(\mathbf{y}|H_0) - c_a f_{\underline{\mathbf{y}}}(\mathbf{y}|H_a) < 0. \tag{1.6}$$

The critical region K thus consists of all $\mathbf{y}$ for which

$$\frac{f_{\underline{\mathbf{y}}}(\mathbf{y}|H_0)}{f_{\underline{\mathbf{y}}}(\mathbf{y}|H_a)} < \frac{c_a}{c_0}. \tag{1.7}$$

In other words, if the above equation holds for the measurement $\mathbf{y}$, then we reject H_0.

1.5 A-priori probabilities

Let's complicate the situation a bit more. Previously we have made an assumption. We assumed that we didn't have a clue whether H_0 or H_a would be true in reality. Let's suppose we do have a clue now. The probability that H_0 is correct (and thus that $\mathbf{x} = \mathbf{x_0}$) is $P(\mathbf{x} = \mathbf{x_0})$. (Abbreviated this is $P(\mathbf{x_0})$.) Similarly, we know that the chance for H_a to be correct is $P(\mathbf{x_a})$. The probabilities $P(\mathbf{x_0})$ and $P(\mathbf{x_a})$ are called **a-priori probabilities** — probabilities we already know before the experiment. We know that either H_0 or H_a is true, so we have $P(\mathbf{x_0}) + P(\mathbf{x_a}) = 1$.

If H_0 is true, then we have a chance α to lose c_0. Similarly, if H_a is true, then we have a chance β to lose c_a. Therefore our average costs now become $P(\mathbf{x_0})c_0\alpha + P(\mathbf{x_a})c_a\beta$. From this we can find that $\mathbf{y} \in K$ (we should reject H_0) if

$$\frac{f_{\underline{\mathbf{y}}}(\mathbf{y}|H_0)}{f_{\underline{\mathbf{y}}}(\mathbf{y}|H_a)} < \frac{P(\mathbf{x_a})c_a}{P(\mathbf{x_0})c_0}. \tag{1.8}$$

If we don't have a clue which hypothesis will be correct, then $P(\mathbf{x_0}) = P(\mathbf{x_a}) = 1/2$. Note that, in this case, the above equation reduces to the result of the previous paragraph.

2 Multiple Hypothesis

Previously we have only considered two hypotheses, being H_0 and H_a. But what should we do if we have p hypotheses $H_1, H_2, \ldots, H_p$? How do we know which one to choose then? Let's take a look at that.

2.1 Deciding the hypothesis

First we will make a few definitions. Let's define the **discrete decision** δ as our choice of vector $\mathbf{x_i}$. Also, there is the **cost function** $C(\mathbf{x}, \delta)$ (not be confused with the covariance operator). Suppose we accept hypothesis H_j (and thus $\delta = \mathbf{x_j}$), but in reality we have $\mathbf{x} = \mathbf{x_i}$. In this case our costs are $C(\mathbf{x} = \mathbf{x_i}, \delta = \mathbf{x_j})$. This can also be written as $C(\mathbf{x_i}, \mathbf{x_j})$, or even as C_{ij}. We also assume that $C(\mathbf{x_i}, \mathbf{x_i}) = 0$. In words, this says that if we accept the right hypothesis, we don't have any costs.

Suppose we have a measurement $\mathbf{y}$. It is now rather difficult to decide which hypothesis we accept. We therefore make an assumption. We assume that the costs for all errors are equal. (So C_{ij} = constant for all i, j with $i \neq j$.) This is part of the so-called **Maximum A Posteriori probability criterion** (MAP). Now we can decide which hypothesis to accept. We should accept H_i if for all $j \neq i$ we have

$$f_{\underline{\mathbf{y}}}(\mathbf{y}|\mathbf{x_j})P(\mathbf{x_j}) < f_{\underline{\mathbf{y}}}(\mathbf{y}|\mathbf{x_i})P(\mathbf{x_i}). \tag{2.1}$$

So we should accept H_i if the number i gives the maximum value for $f_{\underline{\mathbf{y}}}(\mathbf{y}|\mathbf{x_i})P(\mathbf{x_i})$. This is, in fact, quite logical. If costs don't matter, we should simply choose the hypothesis which gives us the biggest chance that we're right. This also causes the chance that we're wrong to be the smallest.

2.2 Acceptance regions

Given a measurement $\mathbf{y}$, we now know which hypothesis H_i to choose. Let's look at all $\mathbf{y}$ for which we will accept H_i. These $\mathbf{y}$ form the **acceptance region** A_i. We can also look at this definition the other way around: If $\mathbf{y} \in A_i$, then we accept H_i.

Let's ask ourselves something. Suppose that in reality H_i is true. What is then the chance that we accept H_j? Let's call this chance β_{ij}. Its value depends on the acceptance region A_j and can be found using

$$\beta_{ij} = \int_{A_j} f_{\underline{\mathbf{y}}}(\mathbf{y}|\mathbf{x_i})\, dy. \tag{2.2}$$

The chance that we make any wrong decision (given that H_i is true) is denoted as β_i. It can be found by simply adding up all the β_{ij}s with $i \neq j$. So,

$$\beta_i = \sum_{j=1, j\neq i}^{p} \beta_{ij} \tag{2.3}$$

On the other hand, the chance that we make the right decision (given that H_i is true) is written as γ_i. Note that we have $\gamma_i = 1 - \beta_i$. (You might see that we also have $\gamma_i = \beta_{ii}$. This is correct. However, the sign β is normally used to indicate errors. So that's why we use the sign γ_i now, and not β_{ii}.)

You probably already expect the next question we will ask to ourselves. What would be the chance that we are wrong in general? This chance, called the **average probability of incorrect decision**, can be found using

$$\sum_{i=1}^{p} P(\mathbf{x_i})\beta_i = \sum_{i=1}^{p} \left(P(\mathbf{x_i}) \sum_{j=1, j\neq i}^{p} \beta_{ij} \right) = \sum_{i=1}^{p} \left(P(\mathbf{x_i}) \sum_{j=1, j\neq i}^{p} \int_{A_j} f_{\underline{\mathbf{y}}}(\mathbf{y}|\mathbf{x_i})\, dy \right). \tag{2.4}$$

If the acceptance regions are well defined, then this chance is minimal.

3 Other Methods of Testing

We just saw one way in which we can choose a hypothesis. Naturally, there are more ways. In this part we will examine another way to choose from two hypotheses H_0 and H_a.

3.1 The simple likelihood ratio test

You probably remember the maximum likelihood estimation (MLE) method, from the previous chapter. In that method, we looked for the $\mathbf{x}$ for which $f_{\underline{\mathbf{y}}}(\mathbf{y}|\mathbf{x})$ was maximal. We can do the same now. However, now we only have two possible values of $\mathbf{x}$, being $\mathbf{x_0}$ and $\mathbf{x_a}$. We thus accept H_0 if $f_{\underline{\mathbf{y}}}(\mathbf{y}|\mathbf{x_0}) > f_{\underline{\mathbf{y}}}(\mathbf{y}|\mathbf{x_a})$. We reject H_0 if

$$\frac{f_{\underline{\mathbf{y}}}(\mathbf{y}|\mathbf{x_0})}{f_{\underline{\mathbf{y}}}(\mathbf{y}|\mathbf{x_a})} < 1. \tag{3.1}$$

The critical region K can be derived from the above criterion. This testing method is called the **maximum likelihood test**.

Let's adjust the above method slightly. Instead of taking 1 as a boundary, we now take some (positive) constant c. We thus reject H_0 if

$$\frac{f_{\underline{\mathbf{y}}}(\mathbf{y}|\mathbf{x_0})}{f_{\underline{\mathbf{y}}}(\mathbf{y}|\mathbf{x_a})} < c. \tag{3.2}$$

We now have arrived at the **simple likelihood ratio** (SLR) test.

3.2 The most powerful test

We find another way of testing when we apply the **Neyman-Pearson** testing principle. To apply this principle, we should first give the probability of false alarm α (the size of the test) a certain value. We then examine all tests (or equivalently, all critical regions K) with size α. We select the one for which the probability of missed detection β is minimal. (Or equivalently, the one for which the power of the test γ is maximal.) The selected test is called the **most powerful** (MP) test of size α.

Let's take another look at the conditions. The value of α should be set, and the value of γ should be maximal. Now let's look at the simple likelihood ratio test. We can choose our ratio c such that the test has size α. This makes sure the first condition is satisfied. Now comes a surprising fact. The SLR test also always satisfies the second condition. In other words, the SLR test is always the test with maximal γ — it is always the most powerful test.

So, although we may have believed we had two new testing methods, we only have one. But we do always know which test is the most powerful one: the simple likelihood ratio test.