6 Hypothesis testing

6.1 Motivation

This chapter will introduce you to another major activity in Statistical Inference, namely the testing of statistical hypotheses for the distribution's parameters and confidence region construction for these parameters.

These activities are naturally arising in inference because of the randomness of our sample. Given the sample $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ of n i.i.d. observations from $f(x, \theta), \theta \in \Theta$, it is natural to first try to evaluate, as closely as possible, the unknown parameter θ by suggesting its estimate $\hat{\theta}$.

Previously we have focused our efforts on finding good estimators $\hat{\theta}$ and discussing the criteria that describe such good estimators.

However, the value $\hat{\theta}$ is only one point in the space Θ of possible θ values. Because of the randomness of the sample, if we repeat the sampling, the point estimator value will change. Any particular realisation $\hat{\theta}$ of the estimator is an estimated value of θ which will be, in general, different from the true θ.

In fact, if Θ was not a finite set but an interval (as it often happens) then the estimator and the true value will coincide with probability zero! This fact alone is convincing enough to claim that it is not enough just to give a single (point) estimated value of the parameter.

The problem of constructing confidence intervals (if the parameter was one-dimensional), or confidence regions (if the parameter was multi-dimensional), and the problem of testing hypotheses about θ naturally arise. The confidence regions are (random) subsets of Θ that contain the "true" unknown θ values with a high pre-specified probability.

Hypothesis tests are used to test if any other θ values (possibly different from $\hat{\theta}$) are also likely to be the one that has generated the sample. The two activities (confidence region construction and hypothesis testing) are, in fact, interrelated. We will explore this relationship at the end of the chapter.

6.2 General terminology in hypothesis testing

Let us start with the case of testing a *simple* hypothesis against a *simple* alternative. This is the easiest case to discuss. Besides, the technique that is being used in this simple case (the Neyman–Pearson lemma below) is indeed fundamental and serves as a basis to deal with the more difficult cases.

Assume that the unknown parameter θ can be one of the two values $\{\theta_0, \theta_1\}$ only. In other words, we are testing a simple hypothesis $H_0 : \theta = \theta_0$ versus a simple alternative $H_1 : \theta = \theta_1$. The name simple is telling us that the parameter is completely specified.

A hypothesis or alternative that is not simple is called composite. For example: $H_0 : \theta = 5$ is a simple hypothesis but $H_0 : \theta < 5$ is a composite hypothesis.

In a hypothesis problem, after observing the sample the experimenter must decide either to accept H_0 as true, or to reject H_0 as false and decide H_1 is true.

A test $\varphi(\mathbf{x})$ is defined as

$$\varphi(\mathbf{x}) = P(\text{reject } H_0 \mid \mathbf{X} = \mathbf{x})$$

In other words, the test is giving us the probability to reject the null hypothesis given the observed sample. Generally, we would prefer deterministic decisions regarding the rejection, i.e. we would like $\varphi(\mathbf{x})$ to be equal to either zero or one. For example:

$$\varphi(\mathbf{x}) = \begin{cases} 1 & \text{if} \quad \bar{X} > c \quad \text{(reject)} \\ 0 & \text{if} \quad \bar{X} \leq c \quad \text{(accept)} \end{cases}$$

Based on the observations, we would then calculate $\varphi(\mathbf{x})$ and according to its value, we reject H_0 (if it happens that $\varphi(\mathbf{x}) = 1$) or do not reject it (if $\varphi(\mathbf{x}) = 0$). This means that we need to decompose the sample space $\mathcal{X}$ into two non-intersecting regions A and S:

$$\mathcal{X} = A \cup S \qquad \text{and} \qquad A \cap S = \emptyset$$

so that depending on $\mathbf{x} \in A$ or $\mathbf{x} \in S$ we decide for either H_0 or H_1.

We will be looking for a test (i.e., finding a function $\varphi(\mathbf{x})$ or equivalently, a decomposition of $\mathcal{X}$ into A and S) in some sort of optimal way.

To define reasonably what optimality could mean in this setting, we need to examine the errors that we can encounter when deciding to reject or accept H_0. We could commit two types of errors:

- to reject H_0 given that H_0 is correct (first type of error)
- to accept H_0 given that H_1 is correct (second type of error)

The corresponding probabilities are denoted as follows:

$$P(\text{reject } H_0 \mid H_0 \text{ correct }) = \text{level of the test (significance)} = \alpha$$

$$P(\text{accept } H_0 \mid H_1 \text{ correct }) = 1 - (\text{power of the test}) = 1 - \beta$$

It is also well known and intuitively clear that minimising level and maximising power simultaneously is **not possible** since they are related, so increasing one of them typically leads to decreasing the other one.

Because of this fact, one possible way out is to formulate a **constrained optimisation** problem as follows:

We decide to fix a certain (small) value α (e.g., $\alpha = 0.005, 0.01, 0.05, 0.10$) (level of significance) that is not allowed to be exceeded for the first type error and in the set of all tests having first type error equal to α, we are looking for the one with a smallest possible second type error (or equivalently the highest possible power).

In the signal processing literature, the first type error is called, not unreasonably, the false alarm. This is because if the null hypothesis is about no (enemy) signal mixed with the recorded noise level then the rejection of the null implies that a false alarm has been raised.

You understand now that allowing the probability level for a false alarm (α) to be too high means that too many false alarms could be raised; on the other hand, choosing it too low implies that we are increasing the chance for a signal to be missed. Both events are undesirable.

The fire alarm example is why choosing a suitable level α represents a compromise. There is no unique recommendation for the choice of α and this choice is often related to the particular field of study.

Indeed, very often we have an idea of what highest first type error we could tolerate. Despite the above comments, the value $\alpha = 0.05$ is often considered as a "default".

Once having decided on the level of significance, we would like to perform optimally. Of course, after having constructed the optimal test, we would still like to examine its power to see if it is not too low for the purpose of our analysis. If this turns out to be the case, we might need to increase the sample size n in order to improve the power.

Unfortunately, the above described constrained optimisation problem sometimes turns out not to have a solution in cases of **discrete** observations. Sometimes, in the discrete case, it is not possible to decompose $\mathcal{X}$ in such a way that the first type error is exactly equal to the allowable α (we cannot exhaust the level) and then the power of such a test will not be optimal.

To be able to avoid this unpleasant situation, one needs to introduce randomised tests by allowing $\varphi(\mathbf{x})$ to take *any value* in $[0,1]$. With this extension of φ, the two types of errors discussed above have the following interpretation:

$$\begin{aligned} P(\text{reject } H_0 \mid H_0 \text{ correct}) &= \int \cdots \int P(\text{reject} H_0 \mid \mathbf{X} = \mathbf{x}) L(\mathbf{x}, \theta_0) \mathrm{d}\mathbf{x} \\ &= \int \cdots \int \varphi(\mathbf{x}) L(\mathbf{x}, \theta_0) \mathrm{d}\mathbf{x} \\ &= \mathbb{E}_{\theta_0}(\varphi) \end{aligned}$$

where $\mathrm{d}\mathbf{x}$ is a shorthand notation for $\mathrm{d}\mathbf{x} = \mathrm{d}x_1 \mathrm{d}x_2 \ldots \mathrm{d}x_n$. We also have:

$P(\text{accept } H_0 \mid H_1 \text{ true}) = \text{similarly to the argument above} = 1 - \mathbb{E}_{\theta_1}(\varphi)$.

These definitions of the two types of errors can be easily interpreted also in cases of composite hypotheses and will be used from now on. If

$$H_0 : \theta \in \Theta_0 \quad \text{and} \quad H_1 : \theta \in \Theta_1$$

then for a given test φ, $\mathbb{E}_\theta(\varphi)$ as a function of θ is its power function. It is clear that the power function belongs to [0,1] since φ is always in [0,1] and hence so is its expected value.

For a good α-size test we want $\mathbb{E}_\theta\varphi$ to be small when $\theta \in \Theta_0$ (and to be definitely $\leq \alpha$ since the test would not be an α test otherwise). On the other hand, when $\theta \in \Theta_1$ we want $E_\theta\varphi$ to be high (to achieve the goal of high power).

Example 6.46

Let X be a single observation from the density

$$f(x;\theta) = \theta x^{\theta-1}, \quad 0 \leq x \leq 1, \quad \theta > 0$$

For testing $H_0 : \theta \leq 1$ versus $H_1 : \theta > 1$, find the power function and size of the test with rejection region $x \geq \frac{1}{2}$. That is, the test is

$$\varphi(x) = \begin{cases} 1 & \text{if} \quad x \geq 1/2 \quad \text{(reject)} \\ 0 & \text{if} \quad x < 1/2 \quad \text{(accept)} \end{cases}$$

Solution:

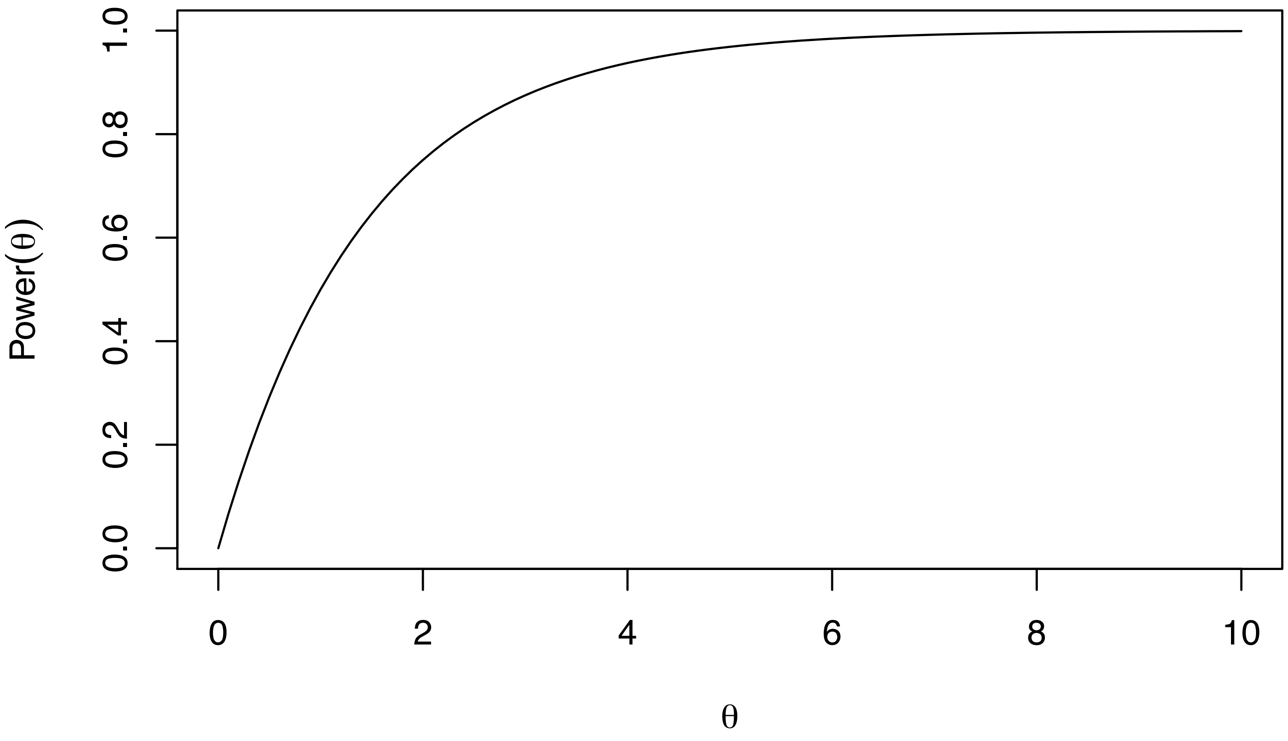
We have

$$\mathbb{E}_\theta(\varphi) = \int_{1/2}^{1} \theta x^{\theta-1} dx = 1 - \left(\frac{1}{2}\right)^\theta, \ \theta > 0$$

is the power function. The size is obtained at $\theta_0 = 1$, so $\mathbb{E}_{\theta_0}(\varphi) = \frac{1}{2}$.

We can plot the power function for a range of values of θ. We can see how the power of the test is increasing when the parameter θ is increasing, which illustrates that the probability of rejecting H_0 given that H_1 is true is high when $\theta \in \Theta_1$. We can also see that when $\theta = 1$ the power of the test is α.

```r
power.fn <- function(theta) 1 - 0.5^theta
curve(power.fn, 0, 10, xlab = expression(theta),
ylab = expression(Power(theta)))
```



Exercise 6.25 (at lecture)

Suppose X_1 and X_2 are independent random variables, each with density

$$f(x;\theta) = \theta x^{\theta-1}, \quad 0 \le x \le 1, \quad \theta > 0$$

For testing $H_0 : \theta \le 1$ versus $H_1 : \theta > 1$, find the size and the power function of the test with rejection region $3x_1 \le 4x_2$. Would you use this test by considering the size of the test?

6.3 Fundamental Lemma of Neyman-Pearson

We are interested in finding optimal tests. When dealing with testing a simple hypothesis against a simple alternative, such optimal tests exist and are given below by the Fundamental Lemma of Neyman and Pearson.

Admittedly, testing a simple hypothesis against a simple alternative is not the most interesting scenario in hypothesis testing. However, we will see that the ideas in the Neyman–Pearson lemma can be extended and can also lead to formulating optimality results in more interesting and practically relevant scenarios when the alternative is composite or when both the hypothesis and the alternative are composite.

This is why we start with the Neyman–Pearson lemma first. We will now formulate and prove this Lemma and consider examples of its application.

Recall that the unknown parameter θ can be one of the two values $\{\theta_0, \theta_1\}$ only. In other words, we are testing a simple hypothesis $H_0 : \theta = \theta_0$ versus a simple alternative $H_1 : \theta = \theta_1$.

Lemma (Fundamental Lemma of Neyman and Pearson)

i) For every $\alpha \in (0,1)$ there exists a constant C and a test

$$\varphi^* = \begin{cases} 1 & \text{if} \quad x \in S = \{x : L(x,\theta_1)/L(x,\theta_0) > C\}, \\ \gamma & \text{if} \quad x \in R = \{x : L(x,\theta_1)/L(x,\theta_0) = C\}, \\ 0 & \text{if} \quad x \in A = \{x : L(x,\theta_1)/L(x,\theta_0) < C\} \end{cases}$$

with $\mathbb{E}_{\theta_0}(\varphi^*) = \alpha$. The constant $\gamma \in (0,1)$ in the definition of the test is equal to $\gamma = \frac{\alpha - P_{\theta_0}(S)}{P_{\theta_0}(R)}$;

ii) φ^* is the best α-test, i.e. $\mathbb{E}_{\theta_1}(\varphi^*)$ is maximal among all tests $\varphi \in \Phi_\alpha = \{\varphi \mid \mathbb{E}_{\theta_0}\varphi \le \alpha\}$.

iii) φ^* is essentially unique, i.e. all other "best" α-tests in the sense of ii) must coincide with φ^* on S and A.

Proof.

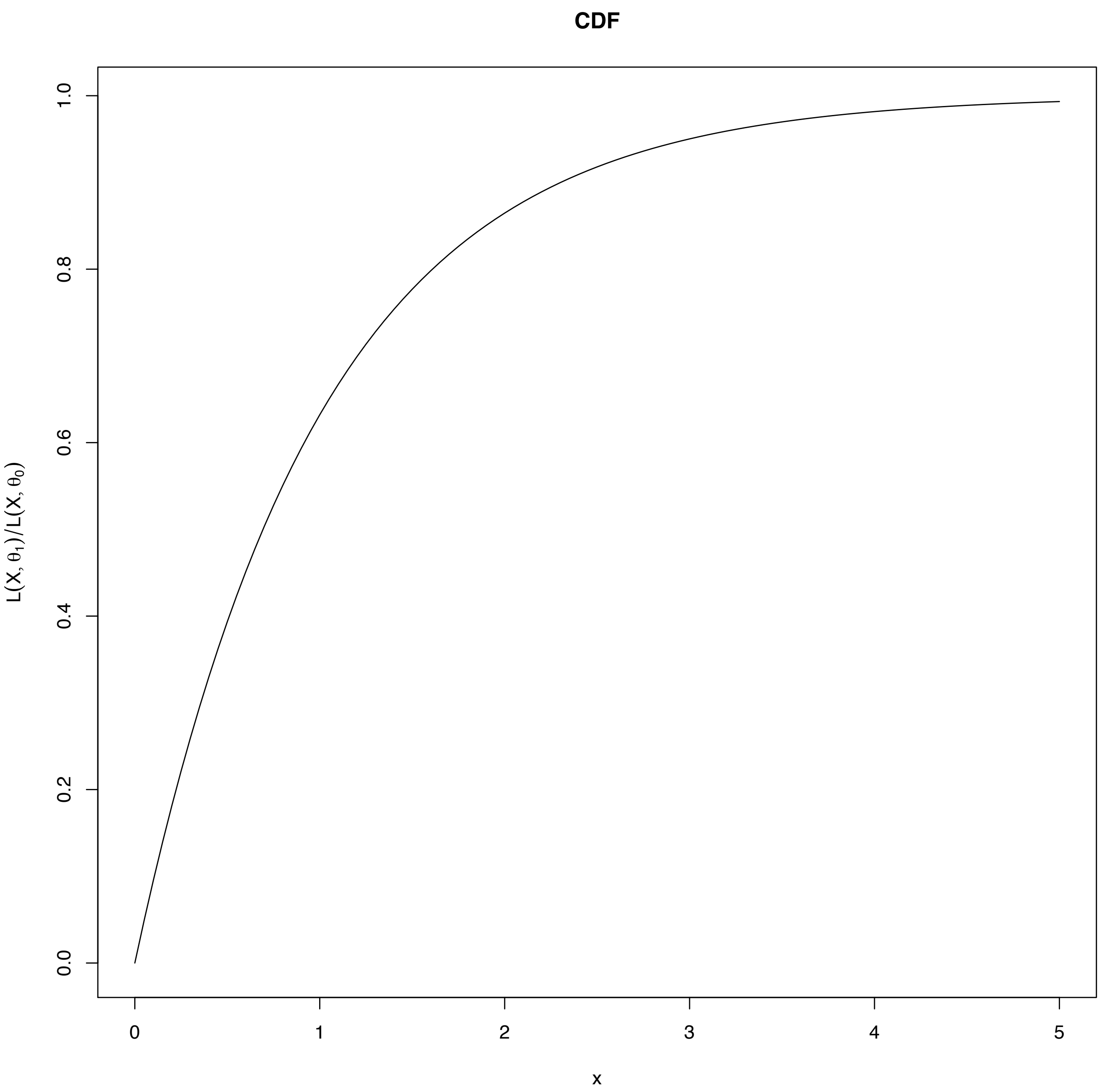
i) Given α, we define C to be the smallest value on the real line for which

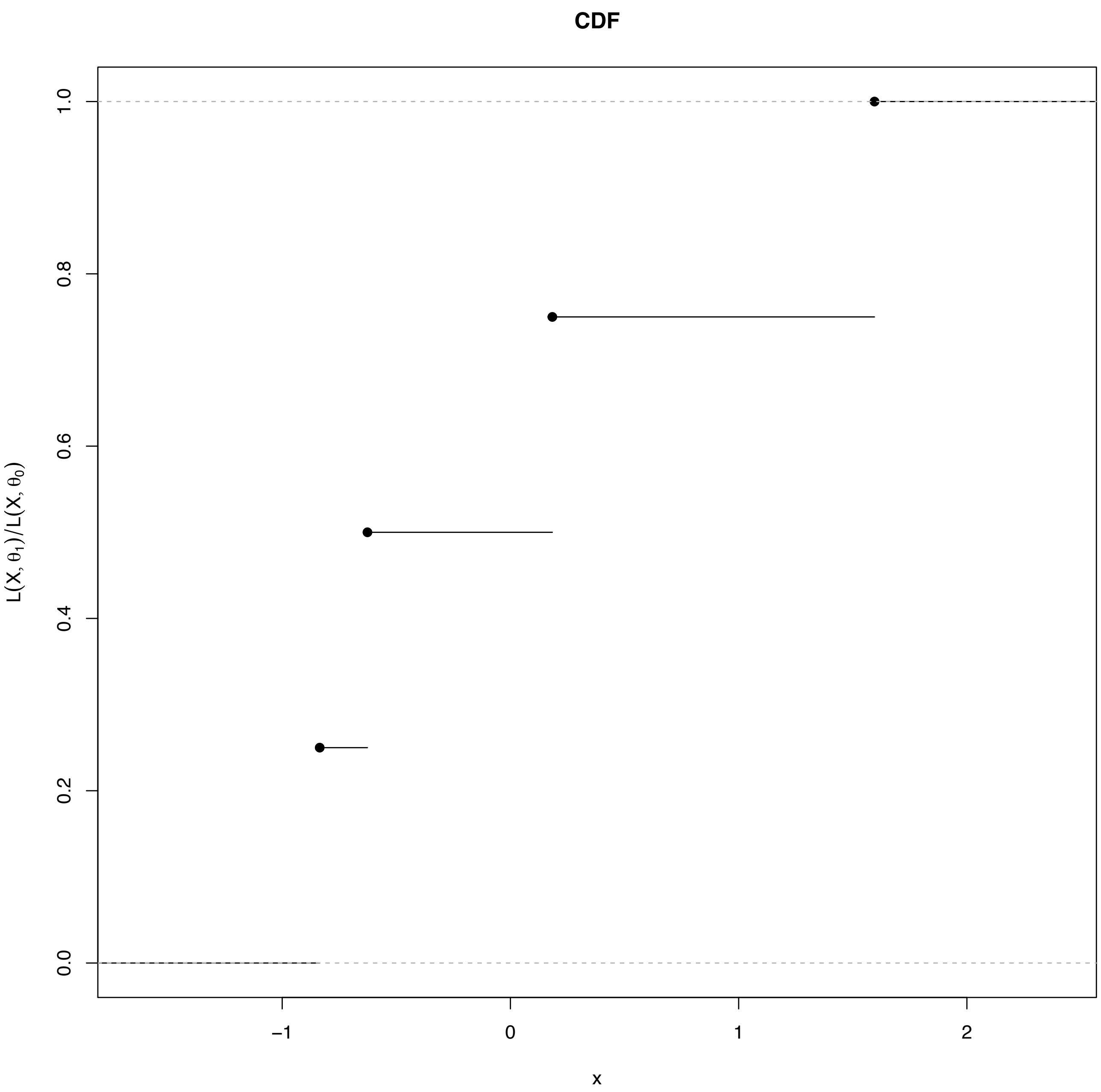
$$P_{\theta_0}\left\{\frac{L(X,\theta_1)}{L(X,\theta_0)} > C\right\}$$

is still $\leq \alpha$. In the continuous case we would actually have precisely

$$P_{\theta_0}\left\{\frac{L(X,\theta_1)}{L(X,\theta_0)} > C\right\} = \alpha$$

but equality might not be possible in the discrete case.


CDF
$L(X, \theta_1)/L(X, \theta_0)$
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5
x


CDF
1.0
0.8
0.6
0.4
0.2
0.0
L(X, θ1)/L(X, θ0)
−1
0
1
2
x

The constant C, which we choose in this manner, has a specific name. It is called the upper $\alpha \times 100\%$ point of the distribution of $\frac{L(X,\theta_1)}{L(X,\theta_0)}$ when θ_0 is the true parameter.

Then, looking at the definition of φ^* and using the definition of γ we see that

$$\begin{aligned}\mathbb{E}_{\theta_0}(\varphi^*) &= 1 \times P_{\theta_0}(\mathbf{X} \in S) + \gamma \times P_{\theta_0}(\mathbf{X} \in R)\\ &= P_{\theta_0}(\mathbf{X} \in S) + \alpha - P_{\theta_0}(\mathbf{X} \in S)\\ &= \alpha.\end{aligned}$$

since

$$\gamma = \frac{\alpha - P_{\theta_0}(\mathbf{X} \in S)}{P_{\theta_0}(\mathbf{X} \in R)}.$$

ii) Take any other α-test φ and divide the sample space $\mathcal{X}$ into

$$\mathcal{X} = \mathcal{X}^{+} \bigcup \mathcal{X}^{-} \bigcup \mathcal{X}^{=}$$

with:

$$\begin{aligned}
\mathcal{X}^{+} &= \{\mathbf{X} : \varphi^{*}(\mathbf{X}) - \varphi(\mathbf{X}) > 0\}, \\
\mathcal{X}^{-} &= \{\mathbf{X} : \varphi^{*}(\mathbf{X}) - \varphi(\mathbf{X}) < 0\}, \\
\mathcal{X}^{=} &= \{\mathbf{X} : \varphi^{*}(\mathbf{X}) - \varphi(\mathbf{X}) = 0\}.
\end{aligned}$$

Analysing the expression

$$Z(X) = (\varphi^{*}(X) - \varphi(X))(L(X, \theta_1) - CL(X, \theta_0))$$

separately for values of $X \in \mathcal{X}^{+}, X \in \mathcal{X}^{-}$ and $X \in \mathcal{X}^{=}$, we see that always $Z(X) \geq 0$ holds since, in all cases, the right-hand side is a product of two multipliers that are either both non-negative or both non-positive!

To see why, lets consider when $\varphi^*(X) > 0$ then

$$\frac{L(X,\theta_1)}{L(X,\theta_2)} > C \quad \text{which implies that} \quad L(X,\theta_1) - CL(X,\theta_2) > 0.$$

When $\varphi^*(X) < 1$ then

$$\frac{L(X,\theta_1)}{L(X,\theta_2)} \leq C \quad \text{which implies that} \quad L(X,\theta_1) - CL(X,\theta_2) \leq 0.$$

In all cases it implies that $Z(X) \geq 0$.

But then, of course,

$$\int_{\mathcal{X}} Z(X)\mathrm{d}X \geq 0 \qquad (*) \tag{11}$$

holds. Substituting back the value of $Z(X)$ in (11) we get:

$$\begin{aligned} 0 \leq & \int \varphi^*(X)L(X,\theta_1)dX - \int \varphi(X)L(X,\theta_1)dX \\ & - C\int \varphi^*(X)L(X,\theta_0)dX + C\int \varphi(X)L(X,\theta_0)dX \\ = & \, \mathbb{E}_{\theta_1}(\varphi^*) - \mathbb{E}_{\theta_1}(\varphi) - C\mathbb{E}_{\theta_0}(\varphi^*) + C\mathbb{E}_{\theta_0}(\varphi) \\ = & \, \mathbb{E}_{\theta_1}(\varphi^*) - \mathbb{E}_{\theta_1}(\varphi) - C\alpha + C\alpha \end{aligned}$$

so that

$$\mathbb{E}_{\theta_1}(\varphi^*) \geq \mathbb{E}_{\theta_1}(\varphi).$$

Since φ was arbitrarily chosen in the set of α-tests, this implies that φ^* cannot be improved with respect to power, that is, it is the best α-size test.

iii) Suppose $\bar{\varphi}$ is another best α test (that is $\mathbb{E}_{\theta_1}(\varphi^*) = \mathbb{E}_{\theta_1}(\bar{\varphi})$ holds), then according to our discussion in ii), we necessarily need to have $Z(X) \equiv 0$.

Since $Z(X)$ is a product of two factors, we either have one of the factors being zero: $\varphi^*(X) = \bar{\varphi}(X)$ or, if not, then the other one must be zero (which means $X \in R$).

Hence, always when X is not in R but in S or in A, we must have $\varphi^*(X) = \bar{\varphi}(X)$. $\square$

6.4 Comments related to the Neyman-Pearson Lemma

Any test φ with $\mathbb{E}_{\theta_0}(\varphi) = \alpha$ is called an α-test (equivalently, an α-size test).

The region S is called a rejection region (critical region).

The optimal test φ^* has the highest power among all tests of size $\leq \alpha$.

Looking at the structure of φ^* given in i) of the Neyman–Pearson Lemma we see that it has a very simple and intuitively appealing interpretation: we look at the ratio of likelihoods for the sample under the alternative and under the null hypothesis.

If the ratio is large enough, the alternative is more likely to have generated the sample and we “vote” for the alternative.

If the ratio is small enough, the hypothesis is more likely to have generated the sample and we “vote” for the hypothesis.

In the intermediate case of the ratio being equal to C, we are in doubt and this is why our decision is random and we decide for the alternative with a probability $\gamma \in (0, 1)$.

The choice of C and γ is tailored to make the test have exactly a size equal to α, since **by exhausting the allowed level** given to us we are hoping to maximise the power. All this simple reasoning finds its rigorous support in the Fundamental Lemma given above.

Example 6.47

Let X be a single observation from the density

$$f(x;\theta) = \theta x^{\theta-1}, \quad 0 \le x \le 1, \quad \theta > 0.$$

Find the most powerful test of size $\alpha = 0.05$ of

$$H_0 : \theta = 2 \quad \text{versus} \quad H_1 : \theta = 1.$$

Solution:

By the Neyman–Pearson lemma, for $H_0 : \theta = 2$ versus $H_1 : \theta = 1$, the best α-test is the one with a rejection region in the form

$$\left\{ \frac{L(\mathbf{X};1)}{L(\mathbf{X};2)} \ge k \right\}.$$

That is, the sample size is $n = 1$ and we have

$$\frac{L(\mathbf{X};1)}{L(\mathbf{X};2)} = \frac{1x^0}{2x} = \frac{1}{2x} \geq k.$$

Equivalently: $x \leq \frac{1}{2k} = k'$. To make it an α-test, we need to exhaust the level:

$$\mathbb{E}_{\theta_0}(\varphi^*) = \alpha = 0.05 = P(X \leq k' \mid \theta = 2) = \int_0^{k'} 2x\mathrm{d}x = (k')^2.$$

This implies that $k' = \sqrt{0.05} \approx 0.2236$. Hence, the best 0.05-size test of $H_0 : \theta = 2$ versus $H_1 : \theta = 1$ is

$$\varphi^* = \begin{cases} 1 & \text{if} \quad x \leq 0.2236 \\ 0 & \text{if} \quad x > 0.2236. \end{cases}$$

Exercise 6.26 (at lecture)

Let X be a single observation from the density

$$f(x;\theta) = \frac{e^{(x-\theta)}}{(1+e^{(x-\theta)})^2}, \quad -\infty < x < \infty \quad -\infty < \theta < \infty$$

Find the most powerful test of size $\alpha = 0.1$ of

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

Hint: Show that the ratio $\frac{L(\mathbf{X};1)}{L(\mathbf{X};0)}$ is an increasing function of x.

Hence the inequality $\frac{L(\mathbf{X};1)}{L(\mathbf{X};0)} \geq k$ is equivalent to rejecting when $x \geq k'$.

6.5 Simple $\mathbf{H}_0$ versus composite $\mathbf{H}_1$ - the simple case

We consider now the more realistic situation where we have a *collection* of alternatives instead of a simple one. This is a new situation, not covered by the Neyman–Pearson Lemma.

Imagine, however, that we obtain the same size α best critical region for all θ-values in the alternative set. Then the optimal Neyman–Pearson test φ^* that one can construct for *one concrete alternative value* θ_1 will be optimal no matter what the alternative θ value may be.

Then we can call such a test a uniformly most powerful (UMP) α-size test for the simple hypothesis versus the *collection* of alternatives.

Example 6.48

Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be i.i.d. $N(\theta, 1)$ and consider

$$H_0 : \theta = \theta_0 \in \mathbb{R}^1 \quad \text{versus} \quad H_1 : \theta > \theta_0.$$

We want to find the UMP α-test of H_0 versus H_1, i.e. we want to find a test φ^* which is such that for any other test $\varphi \in \Phi_\alpha$:

$$\mathbb{E}_\theta(\varphi^*) \geq \mathbb{E}_\theta(\varphi)$$

for all $\theta > \theta_0$ holds.

Solution: We are looking for UMP α-test φ^* which means that if we take *any* competitor

$$\varphi \in \Phi_\alpha = \{ \text{ set of all tests } \varphi \text{ such that } \mathbb{E}_{\theta_0}(\varphi) \leq \alpha\},$$

then we claim that $\mathbb{E}_\theta(\varphi^*) \geq \mathbb{E}_\theta(\varphi)$ *for all* $\theta > \theta_0$. We first simplify the problem by considering testing a simple $H_0 : \theta = \theta_0$ versus the simple alternative $\bar{H}_1 : \theta = \theta_1$ for a fixed $\theta 1 > \theta_0$

Since this is a Neymann-Pearson (NP) lemma type problem, for it we have the *most-powerful* α-test and it is given by

$$\varphi^* = \begin{cases} 1 & \text{if} \quad \dfrac{L(X, \theta_1)}{L(X, \theta_0)} > C \\ 0 & \text{if} \quad \dfrac{L(X, \theta_1)}{L(X, \theta_0)} \leq C. \end{cases}$$

Notice that

$$\frac{L(\mathbf{X};\theta_1)}{L(\mathbf{X};\theta_0)} = \frac{(2\pi)^{-\frac{n}{2}}\exp\{-\sum_{i=1}^n(x_i-\theta_1)^2/2\}}{(2\pi)^{-\frac{n}{2}}\exp\{-\sum_{i=1}^n(x_i-\theta_0)^2/2\}}$$
$$= \exp\left\{\left(-\sum_{i=1}^n(x_i-\theta_1)^2+\sum_{i=1}^n(x_i-\theta_0)^2\right)\Big/2\right\}$$

and after simple transformations

$$\frac{L(\mathbf{X};\theta_1)}{L(\mathbf{X};\theta_0)} = \exp\left\{\left(2(\theta_1-\theta_0)\sum_{i=1}^n x_i-\sum_{i=1}^n(\theta_1^2-\theta_0^2)\right)\Big/2\right\} \geq k$$

After some algebra this is equivalent to rejecting if

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i > \frac{2\log k + \sum_{i=1}^{n}(\theta_1^2 - \theta_0^2)}{2n(\theta_1 - \theta_0)} = k',$$

because $\theta_1 > \theta_0$.

Or notice that since $\theta_1 > \theta_0$ the ratio $\dfrac{L(X,\theta_1)}{L(X,\theta_0)}$ is monotonically increasing in $T = \sum_{i=1}^{n} X_i$ and $\dfrac{L(X,\theta_1)}{L(X,\theta_0)} > C$ is equivalent to $\displaystyle\sum_{i=1}^{n} X_i > C_1$, or by rearranging constants, to $\bar{X} > k'$.

To find k' we must **"exhaust the given level** α" which means

$$E_{\theta_0}(\varphi^*) = 1 \cdot P(\bar{X} \geq k' | \theta = \theta_0) = \alpha$$

must hold. To determine k' to obtain a specified α, we use the fact that

$$\bar{X} \sim N\Big(\theta_0, \frac{1}{n}\Big).$$

Therefore

$$\begin{aligned} E_{\theta_0}(\varphi^*) &= P(\bar{X} \geq k' | \theta = \theta_0) \\ &= P\left(\sqrt{n}\frac{\bar{X} - \theta_0}{1} > \sqrt{n}\frac{k' - \theta_0}{1} \right) \\ &= P(Z \geq (k' - \theta_0)\sqrt{n}) \\ &= \alpha, \end{aligned}$$

where $Z \sim N(0,1)$. Hence,

$$(k' - \theta_0)\sqrt{n} = z_\alpha,$$

where z_α is the upper $\alpha \times 100\%$ point of the standard normal distribution (that is, for $Z \sim N(0,1) : P(Z > z_\alpha) = \alpha$.)

This equality results in

$$k' = \theta_0 + z_\alpha / \sqrt{n}$$

and we get the rejection region S of the best (NP) test for H_0 versus $\bar{H}_1$ in the form

$$S = \left\{ \mathbf{x} : \bar{x} \geq \theta_0 + \frac{z_\alpha}{\sqrt{n}} \right\}.$$

Notice that this rejection region obviously **does not** depend on the specific θ_1 chosen in the alternative. Then

$$\varphi^* = \begin{cases} 1 & \text{if} \quad \bar{x} \geq \theta_0 + (z_\alpha / \sqrt{n}) \\ 0 & \text{if} \quad \bar{x} < \theta_0 + (z_\alpha / \sqrt{n}) \end{cases}$$

will be the most powerful α-test of H_0 versus $\bar{H}_1$.

Now we notice that the resulting $\varphi^*(X)$ above, although having been constructed for a particular $\bar{H}_1 : \theta = \theta_1$, **does not** involve this $\theta_1 > \theta_0$ in its shape.

Hence the **same test** φ^* will be the most powerful α-test for *any* chosen $\theta_1 > \theta_0$!

Hence, φ^* will also be the **uniformly** most powerful (UMP) for testing also $H_0 : \theta = \theta_0$ versus $H_1 : \theta > \theta_0$.

Notice also that we used the monotonicity of the likelihood ratio in our arguments. This example was *generalized* in the Blackwell & Girshick Theorem, which we will explore next after another exercise.

Exercise 6.27 (at lecture)

Show that for a random sample $X_1, \dots, X_n$ from $N(0, \sigma^2)$ population, the most powerful test of

$$H_0 : \sigma = \sigma_0 \quad \text{versus} \quad H_1 : \sigma > \sigma_0$$

is given by

$$\varphi^* = \begin{cases} 1 & \text{if} \quad \sum X_i^2 \geq c \\ 0 & \text{if} \quad \sum X_i^2 < c. \end{cases}$$

For a given value α, show how the value of c is explicitly determined by using the Neyman–Pearson lemma.

Hint: $\sum_{i=1}^{n} X_i^2/\sigma^2 \sim \chi_n^2$ where χ_n^2 denotes the chi-square distribution with n degrees of freedom.

6.6 Composite $\mathbf{H}_0$ versus composite $\mathbf{H}_1$

In general, for such types of hypothesis testing problems, there is no UMP α-size test. On the other hand, we have been able to construct some UMP directly in two specific examples already.

As a generalisation of these examples, we consider one common practical situation in which UMP α-tests exist. This is the scenario in which a one-dimensional parameter of a Monotone Likelihood Ratio (MLR) family of distributions is tested.

6.6.1 MLR family of distributions

The family $L(\mathbf{x}, \theta), \theta \in R$ has a monotone likelihood ratio (MLR) in the statistic $T(\mathbf{X})$ if for any fixed θ' and θ'' such that $\theta' < \theta''$, it holds that

$$\frac{L(\mathbf{x}, \theta'')}{L(\mathbf{x}, \theta')}$$

is a *non-decreasing function* of $T(\mathbf{x}) = T(x_1, x_2, .., x_n)$.

Remark 6.15

Typical examples are from the one-parameter exponential family: if

$$f(x,\theta) = a(\theta)b(x)\exp\Big(c(\theta)d(x)\Big)$$

and $c(\theta)$ is strictly monotone increasing then

$$\frac{L(\mathbf{x},\theta'')}{L(\mathbf{x},\theta')} = \frac{a^n(\theta'')}{a^n(\theta')}\exp\left\{\Big[(c(\theta'')-c(\theta')\Big]\sum_{i=1}^{n} d(x_i)\right\}$$

and this family has a MLR in $T(\mathbf{X}) = \sum_{i=1}^{n} d(X_i)$.

The nice thing about the MLR family of distributions is they allow for UMP tests to be constructed for specific composite hypotheses against composite alternatives. The precise statement is given by the Theorem of Blackwell and Girshick.

6.6.2 Theorem of Blackwell & Girshick

Theorem 6.23

Suppose $\mathbf{X} \sim L(\mathbf{x}, \theta)$ and the family is with MLR in $T(\mathbf{X})$. Then for testing

$$H_0 : \theta \leq \theta_0 \quad \text{versus} \quad H_1 : \theta > \theta_0$$

the α-test φ^* with the structure:

$$\varphi^*(\mathbf{x}) = \begin{cases} 1 & \text{if} \quad T(\mathbf{x}) > k \\ \gamma & \text{if} \quad T(\mathbf{x}) = k \\ 0 & \text{if} \quad T(\mathbf{x}) < k \end{cases}$$

where k is the upper $\alpha \times 100\%$ point of the P_{θ_0}-distribution of $T(\mathbf{X})$ and $\gamma = \frac{\alpha - P_{\theta_0}(T>k)}{P_{\theta_0}(T=k)}$) has an increasing power function $E_\theta(\varphi^*)$ (i.e., its power as a function of θ is increasing) and the test is a UMP α-test.

Remark 6.16

There is an obvious variant of the theorem: under the same conditions on the family, the test that rejects $H_0 : \theta \geq \theta_0$ in favour of $H_1 : \theta < \theta_0$ when $T(\mathbf{x}) < k$ with k chosen such that the level α is exhausted, i.e.,

$$\alpha = P_{\theta_0}(T < k) + \gamma P_{\theta_0}(T = k)$$

is the UMP α-test.

6.6.3 UMP α-test examples

Example 6.49

Let X be a single observation from the density

$$f(x;\theta) = \theta x^{\theta-1}, \quad 0 \le x \le 1, \quad \theta > 0.$$

Find the UMP size α test of $H_0 : \theta \ge 2$ versus $H_1 : \theta < 2$ and calculate its power function.

Solution:

Since $f(x;\theta) = \theta e^{(\theta-1)\ln x}$ is a member of the one-parameter exponential family with $d(x) = \ln x$ and a monotonically increasing $c(\theta) = \theta - 1$.

Hence, according to the note to the Blackwell–Girshick theorem, the UMP-α test exists and has a rejection region $S = \{x : \ln x \le k\}$.

But

$$\ln x \le k \iff x \le k'.$$

For $\alpha = 0.05$ we get:

$$\alpha = 0.05 = P(x \le k'|\theta = 2) = \int_0^{k'} 2x^{2-1}\mathrm{d}x = x^2\Big|_{x=0}^{x=k'} = (k')^2$$

and we again get $k' = \sqrt{0.05} = 0.2236.$

This means that the same test as in Example 6.47 is UMP α-size test of $H_0 : \theta \geq 2$ versus $H_1 : \theta < 2$.

We could have also argued about this by noticing that in Example 6.47 the rejection region did not depend on the θ value under the alternative, hence the same test as in Example 6.47 will be a UMP-0.05 test.

Exercise 6.28 (at lecture)

Apply the Blackwell & Girshick Theorem directly to find the UMP α-test of $H_0 : \theta = \theta_0$ versus $H_1 : \theta > \theta_0$ when $X_1, \ldots, X_n$ are i.i.d. $N(\theta, 1)$

Exercise 6.29 (at lecture)

Assume that $X = (X_1, \ldots, X_n)$ are i.i.d. from

$$f(x,\theta) = \begin{cases} \frac{2x}{\theta^2} & \text{if} \quad 0 < x < \theta \\ 0 & \text{else} \end{cases}.$$

Construct the UMP α-test of $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$

All the optimal tests we have considered so far did not need any randomisation and the values of the test function were just 0 or 1.

In fact, randomisation is only needed in the Fundamental Lemma and in the Blackwell–Girshick Theorem when the data comes from a discrete distribution. Indeed, only then can the likelihood ratio or the test statistic attend the borderline value with a non-zero probability!

For all data from continuous distributions, the probability of such an event is equal to zero and, as a consequence, we will only reject or accept the null hypothesis.

The next example shows when randomisation is needed and how is it implemented.

Example 6.50

For a sample of size $n = 10$ from a Poisson(λ) family construct the UMP $\alpha = 0.10$ size test of

$$H_0 : \lambda \leq 1 \quad \text{versus} \quad H_1 : \lambda > 1.$$

You may utilise the following extract of a table of Poisson(10) probabilities:

x	12	13	14	15	16
$P(X \leq x)$	0.7915	0.8644	0.9165	0.9512	0.9729

Solution:

We have an MLR property in $T(X) = \sum_{i=1}^{10} X_i$ since the Poisson distribution belongs to the one-parameter exponential family with $d(x) = x$.

Moreover, T is the sum of n i.i.d. Poisson distributed random variables, each with parameter λ and, as such, is known to be $T \sim \text{Poisson}(10\lambda)$.

For $\lambda_0 = 1$ this is the Poisson distribution with parameter 10. The Blackwell–Girshick theorem tells us that a UMP-α test ($\alpha = 0.1$) exists and is in the form

$$\varphi^* = \begin{cases} 1 & \text{if} \quad T(X) > 14 \\ \gamma & \text{if} \quad T(X) = 14 \\ 0 & \text{if} \quad T(X) < 14 \end{cases}$$

Indeed, 14 is the smallest integer value (a possible realisation of the statistic T) for which the probability that $T \sim$ Poisson(10) exceeds it, is still smaller than the chosen $\alpha = 0.1$

The value of γ is then (compare with the Fundamental Lemma)

$$\gamma = \frac{\alpha - P_{\theta_0}(S)}{P_{\theta_0}(R)} = \frac{0.1 - 0.0835}{0.9165 - 0.8644} = \frac{0.0165}{0.0521} = 0.317.$$

You may argue that the randomisation is an artificial procedure, and indeed it is. It is only needed in the discrete case to of precisely exhaust the (given in advance) confidence level α since only then can we achieve optimality.

In practice, if your client is unhappy with randomisation, you could ask them if they would be happy with a test of size 0.06, say, instead of the given $\alpha = 0.05$. Then you could construct an optimal non-randomied test for a slightly changed confidence level.

Exercise 6.30 (at lecture)

Assume that you have a sample of $n = 25$ i.i.d observations from Bernoulli (θ) distribution and that $\alpha = 0.01$ is given as a level of significance. You are asked to construct the UMP α test of

$$H_0 : \theta \leq \theta_0 = 0.15 \quad \text{versus} \quad H_1 : \theta > 0.15.$$

Exercise 6.31 (at lecture)

Find the UMP size α test of

$$H_0 : \sigma \leq \sigma_0 \quad \text{versus} \quad H_1 : \sigma > \sigma_0$$

based on n i.i.d. observations from $N(0, \sigma^2)$ population. Sketch a graph of its power function. Try to answer the same question in the case that

$$H_0 : \sigma \geq \sigma_0 \quad \text{versus} \quad H_1 : \sigma < \sigma_0$$

was to be tested.

Exercise 6.32 (at lecture)

Find the form of the rejection region of the UMP test of

$$H_0 : \theta \le \theta_0 \quad \text{versus} \quad H_1 : \theta > \theta_0$$

based on independent random variables $X_1, X_2, \dots, X_n$ each with density

$$f(x;\theta) = \frac{1}{\theta} e^{-x/\theta}, x > 0, \theta > 0.$$

Use the Central Limit Theorem (CLT) to determine approximately the constant specifying the rejection region for a size α test. Hence, find an appropriate expression for the power function.

6.7 Unbiasedness and UMPU α-tests

We have already seen that for a one-parameter exponential family, for example, a UMP α-test for composite alternatives exists.

If no UMP test exists, we should think about another criterion to make the optimal choice among possible tests. A further restriction of the set of α-tests is necessary in order to find an optimal solution in a smaller set of competing tests.

For example, one can:

- choose a 'typical' alternative in the set of alternatives and use the most powerful test for that alternative;
- maximize the power locally, by considering only the θ-values from the alternative set that are close to the hypothetical θ-values. This leads to the notion of a locally most powerful test;
- maximize some weighted average of power for the different alternatives.

6.7.1 General discussion and definition

One more solution is mathematically attractive and leads to reasonable tests. We will restrict ourselves to unbiased tests.

Definition 6.9

A test φ of $H_0 : \theta \in \Theta_0$ $(\Theta_0 \subset \Theta)$ versus $H_1 : \theta \in \Theta \backslash \Theta_0$ is an unbiased size α-test if $E_\theta(\varphi) \leq \alpha$ for all $\theta \in \Theta_0$ and $E_\theta(\varphi) \geq \alpha$ for all $\theta \in \Theta \backslash \Theta_0$.

The definition ensures that there exist no alternatives for which acceptance of the hypothesis is more probable than in cases when the null hypothesis is true. This is a very reasonable requirement! (the majority of reasonable tests of size α would still be allowed to compete).

6.7.2 Basic results

Theorem 6.24

Suppose $\mathbf{X} \sim L(\mathbf{x}, \theta)$ with

$$L(\mathbf{x}, \theta) = (a(\theta))^n \prod_{i=1}^{n} b(x_i) \exp\left[c(\theta) \sum_{i=1}^{n} d(x_i)\right]$$

and $T(\mathbf{X}) = \sum_{i=1}^{n} d(x_i)$. Then, for testing $H_0 : \theta_1 \leq \theta \leq \theta_2$ versus $H_1 : \theta < \theta_1$ or $\theta > \theta_2$, the test φ^* is:

$$\varphi^*(\mathbf{x}) = \begin{cases} 1 & \text{if} \quad T(\mathbf{x}) \notin [c_1, c_2] \\ \gamma_i & \text{if} \quad T(\mathbf{x}) = c_i, i = 1, 2 \\ 0 & \text{if} \quad c_1 < T < c_2 \end{cases}$$

where $c_1, \gamma_1, c_2, \gamma_2$ are determined by the conditions $E_{\theta_1}(\varphi^*) = E_{\theta_2}(\varphi^*) = \alpha$. Moreover,the power function has a minimum somewhere within (θ_1, θ_2) and is monotone outside (θ_1, θ_2).

Theorem 6.25

Consider the same family like in the previous Theorem 6.24. Then, for testing $H_0 : \theta = \theta_0$ versus $H_2 : \theta \neq \theta_0$ an UMPU α-test exists with the structure:

$$\varphi^*(\mathbf{x}) = \begin{cases} 1 & \text{if} \quad T(\mathbf{x}) < c_1 \text{ or } T(\mathbf{x}) > c_2 \\ \gamma_i & \text{if} \quad T(\mathbf{x}) = c_i, \quad i = 1, 2 \\ 0 & \text{if} \quad c_1 < T < c_2 \end{cases}$$

The constants c_i, γ_i satisfy:

$$\text{Power}(\theta_0) = \alpha = E_{\theta_0}(\varphi^*)$$

$$\frac{\partial}{\partial\theta} Power(\theta_0) = 0 = \frac{\partial}{\partial\theta} E_\theta(\varphi^*)\Big|_{\theta=\theta_0} .$$

Remark 6.17

The latter Theorem can be justified as a limiting case of previous Theorem when the interval $[\theta_1, \theta_2]$ collapses to a single point θ_0.

See plot in lecture!

6.8 UMPU α-tests examples

Example 6.51

Assume, a sample of one observation $(n = 1)$ from an exponential family with density

$$f(x, \theta) = \frac{1}{\theta} \exp\left(-\frac{x}{\theta}\right), \quad x > 0$$

is available. The parameter $\theta > 0$ is to be tested. Construct an UMP unbiased α-test of:

i)
$$H_0 : \theta \in [1, 2] \quad \text{versus} \quad H_1 : \theta \notin [1, 2].$$

ii)
$$H_0 : \theta = 1 \quad \text{versus} \quad H_1 : \theta \neq 1.$$

Solution:

Case (i) is covered by Theorem 6.22 and here we have

$$\varphi^*(x) = \begin{cases} 1 & \text{if} \quad T < C_1 \quad \text{or} \quad T > C_2 \\ 0 & \text{if} \quad C_1 \leq T \leq C_2 \end{cases}$$

where $T(x) = x$ in this case (one-parameter exponential family with $d(x) = x, n = 1$). We only need to find C_1 and C_2 in order to uniquely specify the above test.

To determine C_1 and C_2 we need to satisfy:

$$\begin{cases} E_{\theta=1}(\varphi^*) = 1 - e^{-C_1} + e^{-C_2} = \alpha \\ E_{\theta=2}(\varphi^*) = 1 - e^{-C_1/2} + e^{-C_2/2} = \alpha \end{cases}$$

This system can be solved by substituion. Set $k_1 = e^{-C_1/2}$ and $k_2 = e^{-C_2/2}$ then we have

$$\begin{cases} k_1^2 - k_2^2 = 1 - \alpha \\ k_1 - k_2 = 1 - \alpha \end{cases}$$

This implies

$$\begin{cases} k_1 + k_2 = 1 \\ k_1 - k_2 = 1 - \alpha \end{cases} \quad \text{or} \quad \begin{cases} k_1 = 1 - \alpha/2 \\ k_2 = \alpha/2 \end{cases}$$

Hence,

$$C_1 = -2\log(1-\alpha/2) \quad \text{and} \quad C_2 = -2\log(\alpha/2).$$

Which implies that

$$E_\theta(\varphi^*) = 1 - e^{-C_1/2} + e^{-C_2/2} = 1 - (1-\alpha/2)^{2/\theta} + (\alpha/2)^{2/\theta}.$$

According to the Theorem 6.23, the UMPU α-test φ^* for (ii) has the same structure as part (i):

$$\varphi^*(x) = \begin{cases} 1 & \text{if} \quad T < C_1 \quad \text{or} \quad T > C_2 \\ 0 & \text{if} \quad C_1 \le T \le C_2 \end{cases}$$

where $T(x) = x$ in this case (one-parameter exponential family with $d(x) = x, n = 1$).We only need to find C_1 and C_2 in order to uniquely specify the above test.

Note that

$$E_\theta(\varphi^*) = P_\theta(x \notin (C_1, C_2)) = 1 - \exp(-C_1/\theta) + \exp(-C_2/\theta)$$

since the cdf is $F(x,\theta) = 1 - \exp(-x/\theta), x > 0)$.

The two conditions on $E_\theta(\varphi^*)$ are:

- $$E_\theta(\varphi^*)\Big|_{\theta=1} = \alpha = 1 - \exp(-C_1) + \exp(-C_2)$$
- $$\begin{aligned}\frac{\partial}{\partial\theta}E_\theta(\varphi^*)\Big|_{\theta=1} &= -\frac{C_1}{\theta^2}\exp(-\frac{C_1}{\theta}) + \frac{C_2}{\theta^2}\exp(-\frac{C_2}{\theta})\Big|_{\theta=1} \\ &= -C_1\exp(-C_1) + C_2\exp(-C_2) = 0\end{aligned}$$

We get a system of two equations with respect to C_1 and C_2. It can be solved numerically (iteratively) given the level α and hence the UMPU α-test will be completely specified. We will skip over these details but the key is a solution can be found from here.

Example 6.52

If $X_1, X_2, \ldots, X_n$ are i.i.d. $N(\theta, \sigma^2)$ with σ^2 known. Then for testing

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0$$

there exists an UMPU α-test (according to Theorem 6.23). We want to show that it coincides with the well-known test (the so-called Z-test)

$$\varphi^*(\mathbf{x}) = \begin{cases} 1 & \text{if} \quad \dfrac{\sqrt{n}|\bar{\mathbf{x}} - \theta_0|}{\sigma} \geq z_{\alpha/2} \\ 0 & \text{if} \quad \dfrac{\sqrt{n}|\bar{\mathbf{x}} - \theta_0|}{\sigma} < z_{\alpha/2}. \end{cases}$$

Solution:

We have that

$$
\begin{aligned}
E_\theta(\varphi^*) &= P_\theta(\bar{\mathbf{x}} \le C_1 \quad \text{or} \quad \bar{\mathbf{x}} \ge C_2) \\
&= P_\theta\left\{\frac{\sqrt{n}(\bar{\mathbf{x}} - \theta)}{\sigma} \le \frac{\sqrt{n}(C_1 - \theta)}{\sigma} \quad \text{or} \quad \frac{\sqrt{n}(\bar{\mathbf{x}} - \theta)}{\sigma} \ge \frac{\sqrt{n}(C_2 - \theta)}{\sigma}\right\} \\
&= P\left\{N(0,1) \le \frac{\sqrt{n}(C_1 - \theta)}{\sigma}\right\} + P\left\{N(0,1) \ge \frac{\sqrt{n}(C_2 - \theta)}{\sigma}\right\} \\
&= \Phi\left(\frac{\sqrt{n}(C_1 - \theta)}{\sigma}\right) + 1 - \Phi\left(\frac{\sqrt{n}(C_2 - \theta)}{\sigma}\right).
\end{aligned}
$$

where Φ is the cdf of the standard normal distribution.

The two equations that have to be satisfied, are:

$$E_{\theta_0}(\varphi^*) = \alpha$$

and

$$\frac{\partial}{\partial\theta} E_\theta(\varphi^*)\Big|_{\theta=\theta_0} = 0.$$

This leads us to the following two equations:

$$\Phi\Big(\frac{\sqrt{n}(C_1-\theta_0)}{\sigma}\Big) + 1 - \Phi\Big(\frac{\sqrt{n}(C_2-\theta_0)}{\sigma}\Big) = \alpha$$

$$\phi\Big(\frac{\sqrt{n}(C_1-\theta_0)}{\sigma}\Big) = \phi\Big(\frac{\sqrt{n}(C_2-\theta_0)}{\sigma}\Big)$$

where ϕ denotes the density of the standard normal distribution.

From the second equation we get (since $C_1 \neq C_2$ and the standard normal density is symmetric around zero): $C_1 + C_2 = 2\theta_0$.

Substituting into the first equation, we get:

$$2\left[1 - \Phi\left(\frac{\sqrt{n}(C_2 - \theta_0)}{\sigma}\right)\right] = \alpha.$$

The latter relation means that

$$C_2 = \theta_0 + \frac{\sigma z_{\alpha/2}}{\sqrt{n}}$$

and

$$C_1 = 2\theta_0 - C_2 = \theta_0 - \frac{\sigma z_{\alpha/2}}{\sqrt{n}}.$$

Hence, the form of φ^* is indeed

$$\varphi^*(\mathbf{x}) = \begin{cases} 1 & \text{if} \quad \dfrac{\sqrt{n}|\bar{\mathbf{x}} - \theta_0|}{\sigma} \geq z_{\alpha/2} \\ 0 & \text{if} \quad \dfrac{\sqrt{n}|\bar{\mathbf{x}} - \theta_0|}{\sigma} < z_{\alpha/2}. \end{cases}$$

6.9 Locally most powerful tests

Another way to handle the situation in which no UMP test exists is to look at tests which have high power at some particular alternatives. That is, consider the behaviour of the power for alternative parameter values that are *close* to the null hypothesis (since when hypothesis and alternative are relatively far from each other, many tests would do a good job).

Definition 6.10 (locally most powerful test)

Test φ^* with power function $E_\theta(\varphi^*)$ is (at its size) locally most powerful (LMP) for $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$ if for any other test φ with

$$E_{\theta_0}(\varphi) = E_{\theta_0}(\varphi^*),$$

there exists $\Delta > 0$ such that

$$E_\theta(\varphi^*) \geq E_\theta(\varphi)$$

for every $\theta \in (\theta_0, \theta_0 + \Delta)$.

Remark 6.18

In most practical situations, the tests have differentiable power functions. In such cases, a LMP test will maximize:

$$\frac{\partial}{\partial\theta}E_\theta(\varphi)\bigg|_{\theta=\theta_0} = \int..\int \varphi(\mathbf{x})\frac{\partial}{\partial\theta}L(\mathbf{x},\theta)\bigg|_{\theta=\theta_0}\mathrm{d}\mathbf{x}$$

under the constraint

$$\int \varphi(\mathbf{x})L(\mathbf{x},\theta_0)\mathrm{d}\mathbf{x} = \alpha = E_{\theta_0}(\varphi).$$

The structure of this optimization problem is the same as in the NP Lemma and leads to the optimal solution:

$$\varphi^*(\mathbf{x}) = \begin{cases} 1 & \text{if} \quad \frac{\partial}{\partial\theta} L(\mathbf{x},\theta)\big|_{\theta=\theta_0} > kL(\mathbf{x},\theta_0) \\ 0 & \text{if} \quad \frac{\partial}{\partial\theta} L(\mathbf{x},\theta)\big|_{\theta=\theta_0} \leq kL(\mathbf{x},\theta_0) \end{cases}$$

and the above optimal test is unique.

If we denote, as usual, $V(\mathbf{x},\theta)$ to be the score function then φ^* has the following simple and appealing structure:

$$\varphi^*(\mathbf{x}) = \begin{cases} 1 & \text{if} \quad V(\mathbf{x},\theta_0) > k \\ 0 & \text{if} \quad V(\mathbf{x},\theta_0) \leq k \end{cases}$$

Exercise 6.33 (at lecture)

Suppose $\mathbf{X} = (X_1, X_2, .., X_n)$ are i.i.d. Cauchy$(\theta, 1)$ variables with density

$$f(x, \theta) = \frac{1}{\pi} \cdot \frac{1}{1 + (x - \theta)^2}, \quad x \in R,$$

with θ being an unknown parameter. Test the hypothesis

$$H_0 : \theta \leq 0 \quad \text{versus} \quad H_1 : \theta > 0.$$

6.10 Likelihood ratio tests (LRT)

Using uniformly most powerful (UMP) tests is the ideal goal in hypothesis testing. However, we have seen that, except in very special circumstances (like for monotone likelihood ratio families), such tests do not exist.

The uniformly most powerful unbiased (UMPU) tests can also be derived for specific families of distributions (one-parameter exponential).

Other criteria, not discussed here, have also been proposed by mathematical statisticians to solve the optimal problem under reasonable restrictions but they all have a limited applicability for specific families of distributions only.

But if we agree with the likelihood principle and believe the set of models to be best represented by its most likely member given the observed data, we can arrive at a relatively simple and *universal* procedure also in a testing context.

When no points in the parameter space specified by H_0 are preferred to others, the likelihood function can be maximised under the null and alternative hypotheses.

6.10.1 General formulation

Assume that

$$\Theta \subseteq \mathbb{R}^{r+s} = \mathbb{R}^k, \quad \Theta_0 = \{\theta \in \Theta \mid \theta_1 = \theta_{10}, \ldots, \theta_r = \theta_{r0}\},$$

and

$$H_0 : \theta \in \Theta_0 \quad \text{and} \quad H_1 : \theta \in \Theta \backslash \Theta_0.$$

Define the statistic

$$\lambda(\mathbf{X}) = \frac{\sup\limits_{\theta \in \Theta_0} L(\mathbf{X}, \theta)}{\sup\limits_{\theta \in \Theta} L(\mathbf{X}, \theta)}.$$

A likelihood ratio test (LRT) is any test that has a rejection region of the form $S = \{\mathbf{x} \mid \lambda(\mathbf{x}) \leq C\}$, where C is any number satisfying $0 \leq C \leq 1$.

Intuitively, it makes sense to define the rejection region as

$$S = \{\mathbf{x} \mid \lambda(\mathbf{x}) \leq C\}$$

for a certain constant C. However, the optimum properties of likelihood ratios for simple hypotheses, as discussed in the NP lemma, no longer apply, except asymptotically.

In addition, the exact distribution of $\lambda(\mathbf{X})$, that is needed to determine the constant C, is also difficult to obtain without using asymptotic approximations.

Thus, the fact that the deviance statistic has an asymptotic distribution that is well known and is generally used to obtain significance levels or to get the constant C for α given in advance.

Bearing in mind the derivations related to the deviance in Chapter 5, we can formulate a slightly more general statement:

Theorem 6.26

Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ be a random sample from $f(x, \theta), \theta \in \mathbb{R}^{r+s}$. Suppose the regularity conditions for consistency and asymptotic normality of MLE under H_0 and H_1 hold.

Then under H_0:

$$-2 \ln \lambda(\mathbf{X}) \xrightarrow{d} \chi^2_r$$

where r is interpreted as the difference between the number of free parameters specified by $\theta \in \Theta$ and the number of free parameters specified by $\theta \in \Theta_0$.

Example 6.53

For testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ for a sample of n i.i.d. $N(\mu, \sigma^2)$, where σ^2 is known one has

$$-2\ln\lambda(\mathbf{x}) = \frac{n(\bar{\mathbf{x}} - \mu_0)^2}{\sigma^2} = D(\mu_0) \sim \chi_1^2$$

and the result is exact.

Solution:

Let $X_1, \ldots, X_n$ be a random sample from a $N(\mu, \sigma^2)$ population where σ^2 is assumed known. Consider

$$H_0 : \mu = \mu_0 \quad \text{versus} \quad H_1 : \mu \neq \mu_0$$

where μ_0 is a number fixed by the experimenter prior to the experiment. Since there is only one value of μ specified by H_0, the numerator of $\lambda(\mathbf{X})$ is $L(\mathbf{X}, \mu_0)$.

We should also recall that the unrestricted MLE of μ is $\bar{X}$, the sample mean. Thus the denominator of $\lambda(\mathbf{X})$ is $L(\mathbf{X}, \bar{x})$.

Hence the LRT statistic is

$$\lambda(\mathbf{x}) = \frac{(2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu_0)^2\right\}}{(2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}}$$

$$= \exp\left\{\frac{1}{2\sigma^2}\left(-\sum_{i=1}^{n}(x_i-\mu_0)^2+\sum_{i=1}^{n}(x_i-\bar{x})^2\right)\right\}.$$

The expression for $\lambda(\mathbf{x})$ can be simplified by noting that

$$\sum_{i=1}^{n}(x_i-\mu_0)^2=\sum_{i=1}^{n}(x_i-\bar{x})^2+n(\bar{x}-\mu_0)^2.$$

Thus the LRT statistic is

$$\lambda(\mathbf{x})=\exp\left\{-n\frac{(\bar{x}-\mu_0)}{2\sigma^2}\right\}$$

and hence

$$-2\ln\lambda(\mathbf{x})=n\frac{(\bar{x}-\mu_0)^2}{\sigma^2}=D(\mu_0)\sim\chi_1^2$$

This should be $\sim\chi^2$ asymptotically but in this case, because of dealing with normal, the result is precise (not only asymptotically.)

Indeed, we know that under $H_0 : \bar{X} \sim N(\mu_0, \frac{\sigma^2}{n})$ and so

$$\sqrt{n}\frac{(\bar{X}-\mu_0)}{\sigma} \sim N(0,1)$$

and hence

$$n\frac{(\bar{X}-\mu_0)^2}{\sigma^2} \sim \chi_1^2.$$

The generalized likelihood ratio test is then

$$\varphi^* = \begin{cases} 1 & \text{if} \quad n\frac{(\bar{X}-\mu_0)^2}{\sigma^2} > \chi^2_{1,\alpha} \\ 0 & \text{if} \quad n\frac{(\bar{X}-\mu_0)^2}{\sigma^2} \le \chi^2_{1,\alpha} \end{cases}$$

which is equivalent to the standard Z-test in this case.

Example 6.54

Consider a Normal sample with both μ and σ^2 unknown. We want to test

$$H_0 : \mu = \mu_0 \quad \text{versus} \quad H_1 : \mu \neq \mu_0.$$

The MLR test is equivalent to the classical t-test with a rejection region

$$S = \left\{ \mathbf{x} \left| \frac{(\bar{\mathbf{x}} - \mu_0)\sqrt{n}}{s} \right| \geq k \right\}$$

where

$$s = (s^2)^{1/2} = \left[\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{\mathbf{x}})^2 \right]^{1/2}.$$

Solution:

In this example we let $X_1, X_2, \ldots, X_n$ be i.i.d. $N(\mu, \sigma^2)$ with both μ and σ^2 unknown. We are interested in testing

$$H_0 : \mu = \mu_0 \quad \text{versus} \quad H_1 : \mu \neq \mu_0.$$

We will first show that the test that rejects H_0 when

$$|\bar{x} - \mu_0| \geq t_{n-1,\alpha/2} \sqrt{s^2/n}$$

is a test of size α, where

$$s = (s^2)^{1/2} = \left[\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \right]^{1/2}.$$

The size of the test is

$$
\begin{aligned}
Size &= P_{\mu_0}\left\{\left|\bar{X}-\mu_0\right| \geq t_{n-1,\alpha/2}\sqrt{s^2/n}\right\} \\
&= 1 - P_{\mu_0}\left\{-t_{n-1,\alpha/2}\sqrt{s^2/n} < \bar{X}-\mu_0 < t_{n-1,\alpha/2}\sqrt{s^2/n}\right\} \\
&= 1 - P_{\mu_0}\left\{-t_{n-1,\alpha/2} < \frac{\bar{X}-\mu_0}{\sqrt{s^2/n}} < t_{n-1,\alpha/2}\right\} \\
&= 1 - (1-\alpha) = \alpha,
\end{aligned}
$$

Since

$$
\frac{\bar{X}-\mu_0}{\sqrt{s^2/n}} \sim t_{n-1}
$$

under H_0.

Now we will show that this test, which is a classical t-test, can be derived as an LRT.

The unrestricted MLEs are $\hat{\mu} = \bar{X}$ and $\hat{\sigma}^2 = \sum_i (X_i - \bar{X})^2/n$. The restricted MLEs are $\hat{\mu}_0 = \mu_0$ and $\hat{\sigma}_0^2 = \sum_i (X_i - \mu_0)^2/n$. Then,

$$
\begin{aligned}
\lambda(\mathbf{x}) &= \frac{(2\pi\hat{\sigma}_0)^{-n/2}}{(2\pi\hat{\sigma})^{-n/2}} \frac{\exp\{-n\hat{\sigma}_0^2/(2\hat{\sigma}_0^2)\}}{\exp\{-n\hat{\sigma}^2/(2\hat{\sigma}^2)\}} \\
&= \left[\frac{\sum_i (x_i - \bar{x})^2}{\sum_i (x_i - \mu_0)^2} \right]^{n/2} \\
&= \left[\frac{\sum_i (x_i - \bar{x})^2}{\sum_i (x_i - \bar{x})^2 + n(\bar{x} - \mu_0)^2} \right]^{n/2}
\end{aligned}
$$

For a constant C we reject H_0 if

$$\left[\frac{\sum_i (x_i - \bar{x})^2}{\sum_i (x_i - \bar{x})^2 + n(\bar{x} - \mu_0)^2}\right]^{n/2} \leq C.$$

After some algebra we obtain

$$\left|\frac{(\bar{x} - \mu_0)\sqrt{n}}{s}\right| \geq \sqrt{(C^{-2/n} - 1)(n-1)} = C'$$

We choose $C' = t_{n-1,\alpha/2}$ the upper $(\alpha/2) \times 100\%$ point of the t distribution with $n-1$ degrees of freedom.

Thus, this popular test (known to be also the UMPU α-test for the above problem) could be constructed using the LRT construction.

We will only note here that the asymptotic distribution result in Theorem 4.22 is formulated under the hypothesis only but more sophisticated reasoning can also be used to derive the asymptotic distribution under alternative parameter values. This distribution is a non-central χ^2 and can be used for (approximate asymptotic) power computations. We omit the details.

Exercise 6.34 (challange for you to try)

Suppose $X_1, X_2, \ldots, X_m$ are independent $N(\mu_1, 1)$ random variables and $Y_1, Y_2, \ldots, Y_n$ is an independent set of independent $N(\mu_2, 1)$ random variables. Show that the generalised likelihood ratio statistic is

$$-2\ln\Lambda_{m,n} = \frac{mn(\bar{x}-\bar{y})^2}{m+n}$$

for testing

$$H_0 : \mu_1 = \mu_2 \quad \text{versus} \quad H_1 : \mu_1 \neq \mu_2$$

Find the distribution of $-2\log\Lambda_{m,n}$ under H_0.

Hints:

- When $\mu_1 = \mu_2 = \mu$ the MLE of μ is $\tilde{\mu} = \frac{m\bar{X}+n\bar{Y}}{m+n}$.
- The unrestricted MLEs of μ_1 and μ_2 are $\hat{\mu}_1 = \bar{X}$ and $\hat{\mu}_2 = \bar{Y}$.

Sketch solutions:

The joint likelihood function for $\mathbf{X}$ and $\mathbf{Y}$ is:

$$L(\mathbf{X},\mathbf{Y};\mu_1,\mu_2) = (2\pi)^{-\frac{m+n}{2}} \exp\left\{-\frac{1}{2}\Big[\sum_{i=1}^{m}(x_i-\mu_1)^2+\sum_{i=1}^{n}(y_i-\mu_2)^2\Big]\right\}$$

Unrestricted maximisation with respect to both μ_1 and μ_2 leads to $\bar{X},\bar{Y}$ as solutions. This can be shown by differentiation of the log-likelihood function as follows:

$$\frac{1}{2}\left(2\sum_{i=1}^{m}x_i-2m\mu_1\right)=0$$

and so

$$\hat{\mu}_1=\frac{1}{m}\sum_{i=1}^{m}x_i=\bar{x}.$$

Furthermore,

$$\frac{1}{2}\left(2\sum_{i=1}^{n} y_i - 2n\mu_2\right) = 0$$

and so

$$\hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y}$$

Now, restricted maximisation under the restriction $\mu_1 = \mu_2 = \mu$ leads to

$$\hat{\mu}_{\text{mle(restricted)}} = \frac{\sum_{i=1}^{m} x_i + \sum_{i=1}^{n} y_i}{m+n}.$$

Therefore,

$$
\begin{aligned}
&2\ln\Lambda_{m,n}\\
&=\sum_{i=1}^{m}(x_i-\bar{x})^2+\sum_{i=1}^{n}(y_i-\bar{y})^2-\sum_{i=1}^{m}(x_i-\frac{m\bar{x}+n\bar{y}}{m+n})^2-\sum_{i=1}^{n}(y_i-\frac{m\bar{x}+n\bar{y}}{m+n})^2\\
&=-m\bar{x}^2-n\bar{y}^2+2\frac{m\bar{x}+n\bar{y}}{m+n}(\sum_{i=1}^{m}x_i+\sum_{i=1}^{n}y_i)-(m+n)(\frac{m\bar{x}+n\bar{y}}{m+n})^2\\
&=\dots\\
&=-\frac{mn(\bar{x}-\bar{y})^2}{m+n}
\end{aligned}
$$

Hence

$$-2\ln\Lambda_{m,n} = \frac{mn(\bar{x}-\bar{y})^2}{m+n} = T$$

which can be seen directly to be distributed as chi-square with one degree of freedom under the hypothesis of equal means. The LRT test is then:

$$\varphi = \begin{cases} 1 & \text{if} \quad T \geq \chi^2_{1,\alpha} \\ 0 & \text{if} \quad T < \chi^2_{1,\alpha} \end{cases}.$$

6.11 Alternatives to the GLRT

The GLRT is widely used but there are circumstances where other test procedures may be preferred. We define two of these test procedures for the case $s = 0$, that is, $k = r$.

Score test:

$$S = V(\mathbf{X}, \theta_0)' I_{\mathbf{X}}^{-1}(\theta_0) V(\mathbf{X}, \theta_0)$$

instead of $-2 \log \lambda(\mathbf{X})$.

Wald test:

$$(\hat{\theta} - \theta_0)' I_{\mathbf{X}}(\hat{\theta})(\hat{\theta} - \theta_0)$$

instead of $-2 \log \lambda(\mathbf{X})$ where θ_0 is the hypothetical vector and $\hat{\theta}$ is the MLE.

For both the Score test, and the Wald test, the asymptotic distribution of the test statistic under H_0 is the same as the distribution of the GLRT statistic (that is, chi-square with $r = k$ degrees of freedom).

Score tests have a numerical advantage in comparison to GLRT and the Wald test, that they do **not** require the MLE to be calculated!

Specifically in the econometrics literature, the Score test is known as Lagrange Multiplier Test. The name comes from its alternative derivation in which the Likelihood function is maximized subject to the restrictions of H_0 using Lagrange multipliers.

Much research has been devoted to selecting one of these three tests as a preferred test in a particular situation for relatively small sample sizes. We will only say that this is a difficult task and will not be discussed in our course.

Example 6.55

Let X be the number of successes in a binomial experiment with a probability of success p. We wish to test

$$H_0 : p = p_0 \quad \text{versus} \quad H_1 : p \neq p_0.$$

Denote $\hat{p} = \frac{X}{n}$.

Solution:

Then the Score statistic is

$$S = \frac{(X - np_0)^2}{np_0(1 - p_0)},$$

the Wald statistic is

$$W = \frac{(X - np_0)^2}{n\hat{p}(1 - \hat{p})}$$

and

$$-2\ln\lambda = 2\left\{X \ln\frac{\hat{p}(1 - p_0)}{p_0(1 - \hat{p})} + n\ln\frac{1 - \hat{p}}{1 - p_0}\right\}.$$

6.12 Confidence interval and confidence region construction

The main topic of concern during this chapter was hypothesis testing. Confidence interval construction (and confidence region construction for that matter) is very strongly related to this topic. However, we will only consider only hypothesis testing. The relation is expressed via the following reasoning:

Suppose that for a chosen confidence level $\alpha \in (0,1)$ an α-size test of the hypothesis $H_0 : \theta = \theta_0$ against a two-sided alternative $H_1 : \theta \neq \theta_0$ exists and can be constructed for any specific hypothetical value θ_0.

Having observed the sample $X_1, X_2, \dots, X_n$ of n i.i.d. observations from $f(X, \theta)$, we can take the set of the parameter values for which the test "answers" with an acceptance when the sample is substituted in the test statistic. This set of parameter values is then a confidence set at level $1 - \alpha$ for the unknown parameter θ.

Symbolically, we can say that the subset in Θ defined via:

$\{\theta' | H_0 : \theta = \theta$ is accepted given the realisation of $\mathbf{X} = \mathbf{x}$ of the sample$\}$

represents a confidence set at level $(1-\alpha)$ for the unknown parameter θ.

In other words, knowing how to construct tests, we basically also know how to construct confidence sets. Moreover, the usefulness of the relationship between testing hypotheses and confidence sets is further exemplified by the fact that optimality results carry over.

It can be shown quite generally that the above procedure of constructing confidence sets leads to confidence sets with optimality properties if the hypothesis test used in the construction was optimally designed.

Without going into further details, let us next illustrate the construction of confidence regions with a few examples.

Example 6.56

Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be a sample from $N(\theta, \sigma^2)$ and assume that σ^2 is known. Construct a $(1-\alpha)\times 100\%$ confidence interval for θ.

Solution:

We start with the α-size test of $H_0 : \theta = \theta_0$ against the two-sided alternative $H_1 : \theta \neq \theta_0$. The classical z-test then looks like

$$\varphi^* = \begin{cases} 1 & \text{if} \quad \sqrt{n}\dfrac{|\bar{x} - \theta_0|}{\sigma} \geq z_{\alpha/2}, \\ 0 & \text{if} \quad \sqrt{n}\dfrac{|\bar{x} - \theta_0|}{\sigma} < z_{\alpha/2}. \end{cases}$$

The lemma tells us to include in the confidence interval for θ all θ values for which the hypothesis is accepted given the sample.

Since the sample has given rise to a fixed $\bar{x}$ value, we will, therefore, include in the $(1-\alpha)\times 100\%$ confidence interval:

$$\left\{\theta : \sqrt{n}\frac{|\bar{x}-\theta|}{\sigma} < z_{\alpha/2}\right\}.$$

Solving this inequality for θ gives the confidence interval

$$\left(\bar{x} - \frac{\sigma z_{\alpha/2}}{\sqrt{n}}, \bar{x} + \frac{\sigma z_{\alpha/2}}{\sqrt{n}}\right).$$

Example 6.57

Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be a sample from $N(\mu, \sigma^2)$. Construct a $(1-\alpha) \times 100\%$ confidence interval for σ^2.

Solution:

We start with the α-size test of $H_0 : \sigma^2 = \sigma_0^2$ against the two-sided alternative $H_1 : \sigma^2 \neq \sigma_0^2$. The classical test then looks like

$$\varphi^* = \begin{cases} 1 & \text{if} \quad \frac{(n-1)s^2}{\sigma_0^2} \leq \chi^2_{n-1,1-\alpha/2} \quad \text{or} \quad \frac{(n-1)s^2}{\sigma_0^2} \geq \chi^2_{n-1,\alpha/2}, \\ 0 & \text{if} \quad \chi^2_{n-1,1-\alpha/2} \leq \frac{(n-1)s^2}{\sigma_0^2} \leq \chi^2_{n-1,\alpha/2}. \end{cases}$$

The lemma tells us to include in the confidence interval for σ^2 all σ^2 values for which the hypothesis is accepted given the sample.

We will, therefore, include in the $(1-\alpha)\times 100\%$ confidence interval:

$$\left\{\sigma^2 : \chi^2_{n-1,1-\alpha/2} \le \frac{(n-1)s^2}{\sigma^2} \le \chi^2_{n-1,\alpha/2}\right\}.$$

Solving this inequality for σ^2 gives the confidence interval

$$\left(\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}\right).$$

Exercise 6.35 (at lecture)

Let again $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be a sample from $N(\theta, \sigma^2)$ but now assume that σ^2 is unknown. Construct a $(1-\alpha)\times 100\%$ confidence interval for θ.