In this lecture we show how to construct optimal tests comparing two simple hypotheses.

### 16.1 Two simple hypotheses

Remember that a simple hypothesis asserts that $\theta$ takes a single, specified value.

We now consider the case where we wish to conduct a hypothesis test when both hypotheses are simple.

$$ H_0 : \theta = \theta_0 , \quad H_1 : \theta = \theta_1 . $$

In effect, we are supposing that $\theta$ can only have two possible values, $\theta_0$ or $\theta_1$.

**Remark 16.1** It is really superfluous now to allow $\theta$ to be a vector, but we stick with the general notation.

The power function for a test defined by critical region $C$ therefore also takes only two values:

$$ \pi_C(\theta_0) = P(X \in C | \theta = \theta_0) , \quad \pi_C(\theta_1) = P(X \in C | \theta = \theta_1) . $$

The first of these is the size of the test $C$, $\alpha_C = \pi_C(\theta_0)$. It is also referred to as the probability of the *first kind of error*.

In hypothesis testing, there are two kinds of error, to reject $H_0$ when it is true and to fail to reject it when it is false. So the ‘first kind of error’ is to reject $H_0$ when it is true, and the probability of this is $\alpha_C$.

The probability of the *second kind of error* is $\beta_C = P(X \notin C | \theta = \theta_1) = 1 - \pi_C(\theta_1)$.

Our goal is to make both probabilities small, and the ideal would be $\alpha_C = \beta_C = 0$.

Remember that in practice we wish to maximise power subject to fixing the test size. In the case of two simple hypotheses this reduces to fixing $\alpha_C = \alpha$ and finding the test with minimum possible value of $\beta_C$. This corresponds to controlling the probability of the first kind of error and then minimising the probability of the second kind of error.
Remark 16.2 Notice the asymmetry between the two kinds of error. Making the first kind of error is thought to be more serious and so must be controlled. This corresponds to the asymmetry in the hypotheses: the null hypothesis is the one to stick with unless we are suitably convinced that the alternative is more plausible.

16.2 The Neyman-Pearson Lemma

A famous result called the Neyman-Pearson (N-P) Lemma identifies the most powerful test of any given size for two simple hypotheses.

Definition 16.1 (Likelihood ratio) The likelihood ratio (LR) for comparing two simple hypotheses is

\[ \lambda(x) = \frac{L(\theta_1; x)}{L(\theta_0; x)} = \frac{f(x | \theta = \theta_1)}{f(x | \theta = \theta_0)}. \]

Remark 16.3 The unspecified proportionality constant in likelihoods cancels out when we take the ratio. Therefore the value of the likelihood ratio is unambiguous, and has absolute meaning.

Definition 16.2 (LR test) A likelihood ratio test is of the form

\[ C_k = \{x : \lambda(x) \geq k\} = \{x : L(\theta_1; x) \geq k L(\theta_0; x)\} \]

for some \( k \).

So the critical region includes all \( x \) for which \( \lambda \) is sufficiently large.

Remark 16.4 As we vary \( k \) we get different tests. Notice that if \( k' > k \) then \( C_{k'} \subseteq C_k \). So increasing \( k \) gives tests with decreasing size \( \alpha_{C_k} \), and also decreasing power \( 1 - \beta_{C_k} \).

Remark 16.5 In this ratio, remember that the alternative is on top.

The N-P Lemma says that the LR test \( C_k \) is the most powerful among all tests of size \( \alpha = \pi_k(\theta_0) = P(L(\theta_1; x) \geq k L(\theta_0; x) | \theta = \theta_0) \).

Its proof is not particularly difficult, but it is not particularly interesting either.
Theorem 1 (Neyman-Pearson Lemma) Let $C_k$ be the Likelihood Ratio test of $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$ defined by
\[ C_k = \left\{ x : \frac{L(\theta_1; x)}{L(\theta_0; x)} \geq k \right\}, \]
and with power function $\pi_k(\theta)$. Let $C$ be any other test such that $\pi_C(\theta_0) \leq \alpha_k = \pi_k(\theta_0)$, where $\pi_C(\theta)$ is the power function of $C$. Then $\pi_k(\theta_1) \geq \pi_C(\theta_1)$.

Example 16.1 (Normal sample, known variance) Let $X_1, X_2, \ldots, X_n$ be a sample from the $N(\mu, \sigma^2)$ distribution and suppose that $\sigma^2$ is known. Therefore $\theta = \mu$. Consider two simple hypotheses
\[ H_0 : \mu = \mu_0, \quad H_1 : \mu = \mu_1. \]

From previous examples, the LR is
\[ \lambda(x) = \frac{L(\mu_1; x)}{L(\mu_0; x)} = \exp\left(-\frac{n}{2\sigma^2}(\bar{x} - \mu_1)^2\right) = \exp\left(-\frac{n}{2\sigma^2}Q\right), \]
where
\[ Q = (\bar{x} - \mu_1)^2 - (\bar{x} - \mu_0)^2 = 2\bar{x}(\mu_0 - \mu_1) + (\mu_1^2 - \mu_0^2). \]
Therefore
\[ C_k = \left\{ x : \exp\left(-\frac{n}{2\sigma^2}Q\right) \geq k \right\} = \left\{ x : -\frac{n}{2\sigma^2}Q \geq \log k \right\} = \left\{ x : 2\bar{x}(\mu_0 - \mu_1) + (\mu_1^2 - \mu_0^2) \leq -\frac{2\sigma^2}{n} \log k \right\} = \left\{ x : \bar{x}(\mu_0 - \mu_1) \leq k^* \right\} \]
where $k^* = -\frac{n}{\pi} \log k - \frac{1}{2}(\mu_0^2 - \mu_1^2)$. Now we are going to divide by $(\mu_0 - \mu_1)$, but if this is negative we must change the direction of the inequality.

Therefore, if we now define $k^{**} = k^*/(\mu_0 - \mu_1)$, we have the following form for the LR test $C_k$:

- if $\mu_0 > \mu_1$, we reject $H_0$ if $\bar{x} \leq k^{**}$,
- if $\mu_0 < \mu_1$, we reject $H_0$ if $\bar{x} \geq k^{**}$.

So the LR criterion of ‘sufficiently large $\lambda$’ translates into ‘sufficiently small $\bar{x}$’ or ‘sufficiently large $\bar{x}$’ depending on whether $\mu_0$ is greater or less than $\mu_1$.\footnote{The footnote marker.}
Remark 16.6 You should always be very careful when dealing with inequalities! ♦

Remark 16.7 Notice that the lines leading to (16.1) are all working towards getting as simple as possible a function of the data on the left hand side of the inequality and a constant on the right. The only way the data appear in the LR is in the form of the sample mean $\bar{x}$, so it is this that we are trying to isolate on the left hand side, and the example ends by finishing this task. It is important to recognise that on the right hand side we then just have a constant, $k^{**}$. ♦

16.3 Test size

Having found the general form of the optimal tests, we have a whole family of tests. By varying $k$, we can get a test of any desired size from $\alpha = 1$ ($k = 0$, always reject $H_0$) to $\alpha = 0$ ($k = \infty$, never reject $H_0$). To complete the analysis, we need to be able to find the value of $k$ that makes $C_k$ have the desired size $\alpha_k = \alpha$, or else find the $P$-value $\alpha$ corresponding to the observed data.

Example 16.2 Let us continue the preceding example. The calculation will again depend on whether $\mu_0$ is greater or less than $\mu_1$, and we will deal with the case $\mu_0 < \mu_1$. Then

$$\alpha_k = P(\bar{X} \geq k^{**} | \mu = \mu_0).$$

To derive this probability, we will use the fact that the distribution of $\bar{X}$ is $N(\mu, \sigma^2/n)$, and we will standardise so that we can work in terms of standard normal probabilities.

Now

$$\alpha_k = P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \geq \frac{k^{**} - \mu_0}{\sigma/\sqrt{n}} | \mu = \mu_0\right),$$

and we know that if $\mu = \mu_0$ then $(\bar{X} - \mu_0)/(\sigma/\sqrt{n})$ has the standard normal distribution. So if we let $Z \sim N(0, 1)$ we have

$$\alpha_k = P\left(Z \geq \frac{k^{**} - \mu_0}{\sigma/\sqrt{n}}\right) = 1 - \Phi\left(\frac{k^{**} - \mu_0}{\sigma/\sqrt{n}}\right). \quad (16.2)$$

This equation links the test size $\alpha$ to the critical value $k^{**}$ for the test. It immediately enables us to find the $P$-value (observed significance) for
the observed data. We simply equate the critical value \( k^{**} \) to the observed value \( \bar{x} \) of \( X \), and obtain

\[
P = 1 - \Phi \left( \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \right) = \Phi \left( \frac{\mu_0 - \bar{x}}{\sigma / \sqrt{n}} \right),
\]

using the general result that \( 1 - \Phi(z) = \Phi(-z) \). Now suppose we wish to choose \( k \), or equivalently \( k^{**} \), so as to obtain a test of specified size \( \alpha \). This means solving (16.2):

\[
\frac{k^{**} - \mu_0}{\sigma / \sqrt{n}} = Z_\alpha.
\]

\[
\therefore k^{**} = \mu_0 + \frac{\sigma}{\sqrt{n}} Z_\alpha.
\]

So we reject \( H_0 \) if \( \bar{x} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} Z_\alpha \). Therefore the test of size \( \alpha \) formally has critical region \( C = \{ x : \bar{x} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} Z_\alpha \} \).

If \( \mu_0 > \mu_1 \), we end up with \( P = \Phi \left( \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \right) \) for observed significance, and for a fixed size test we have \( k^{**} = \mu_0 - \frac{\sigma}{\sqrt{n}} Z_\alpha \), and reject \( H_0 \) if \( \bar{x} \leq \mu_0 - \frac{\sigma}{\sqrt{n}} Z_\alpha \). [These are the one-sided tests for a normal mean that you will have met in Level 1 statistics.]

There are several features to notice about this example.

**Remark 16.8** First, note that we did not need to actually find \( k \). It was enough to find \( k^{**} \). In Example 16.1 we found the general form of the LR test. If \( \mu_0 < \mu_1 \) it is to reject \( H_0 \) if \( \bar{x} \) is sufficiently large, and if \( \mu_0 > \mu_1 \) we reject \( H_0 \) if \( \bar{x} \) is sufficiently small. In Example 16.2 we found just how large or small \( \bar{x} \) needed to be to get a test of a given size. We did not need to have recorded what function \( k^{**} \) was of \( k \), \( \sigma^2 \), \( \mu_0 \) and \( \mu_1 \), because we did not use the information again.

**Remark 16.9** Something that we did need to know is the distribution of \( \bar{X} \), at least in the case \( \mu = \mu_0 \) (also for \( \mu = \mu_1 \) if we want to know \( \beta \) or the power).

### 16.4 The LR test statistic

Examples 16.1 and 16.2 illustrate a common finding, at least in simple models, that the LR statistic \( \lambda(x) \) is a monotone function of some simpler test statistic \( T = T(x) \). Then the LR test reduces to either \( C_k = \{ x : T \geq k' \} \) or \( C_k = \{ x : T \leq k' \} \), for some \( k' \), depending on whether the LR is a monotone increasing or decreasing function of \( T \). We do not need in
general to know the relationship between $k$ and $k'$. It is sufficient to know
the form of the test, that it amounts to seeing if $T$ is sufficiently large or
sufficiently small.

Then to identify the most powerful test of a given size $\alpha$, we need to
be able to derive the distribution of $T(X)$ given $\theta = \theta_0$, and hence to find
$k'$ such that $\alpha = P(T(X) \geq k' \mid \theta = \theta_0)$ or $\alpha = P(T(X) \leq k' \mid \theta = \theta_0)$, as
appropriate.

This gives us a general procedure for finding optimal tests, in two steps.

1. Find the likelihood ratio $\lambda(x)$. Then manipulate the inequality $\lambda(x) \geq k$ so as to express the test in terms of as simple a test statistic $T(x)$
as possible. In doing so, the objective is to find the form of the LR
test in terms of $T$. If the LR is a monotone function of $T$, the form
of the test will be either to reject $H_0$ if $T$ is sufficiently large or to
reject if $T$ is sufficiently small.

2. Derive the distribution of $T(X)$ given $\theta = \theta_0$, and thereby

   (a) find the $P$-value corresponding to the observed value $t = T(x)$
of the test statistic, or

   (b) find the appropriate critical value of $T(x)$ to produce a test of
   the required size $\alpha$.

In practice, both steps can be mathematically tricky.

Here is another example to illustrate the process.

**Example 16.3 (Exponential sample)** Let $X_1, X_2, \ldots, X_n$ be a sample
from the exponential distribution $Ex(\lambda)$. We have simple hypotheses
$H_0 : \lambda = \lambda_0$ and $H_1 : \lambda = \lambda_1 > \lambda_0$. NB in this example $\lambda$ is the parameter!

**Step 1.** The LR statistic is

$$
\frac{L(\lambda_1; x)}{L(\lambda_0; x)} = \left( \frac{\lambda_1^n \exp(-n\bar{x} \lambda_1)}{\lambda_0^n \exp(-n\bar{x} \lambda_0)} \right) = \left( \frac{\lambda_1}{\lambda_0} \right)^n \exp(-n\bar{x}(\lambda_1 - \lambda_0)),
$$

which is a monotone decreasing function of $T(x) = \bar{x}$. Hence the LR test
is $C_k = \{ x : \bar{x} \leq k' \}$ for some $k'$ (and we do not need to know any more
about it, like the relationship between $k'$ and $k$).

**Step 2.** We know from distribution theory that the sum of independent
exponential random variables has a gamma distribution, and specifically
that $n\bar{X} = \sum_{i=1}^n X_i \sim Ga(n, \lambda)$. We also know that a multiple
of a gamma random variable has a gamma distribution, and we know a
relationship between the gamma and chi squared distributions, so that
\[ 2 \lambda n \bar{X} \sim \text{Ga}(n, \frac{1}{2}) = \chi^2_{2n}. \]
We now use this result as follows.

\[
\alpha = P(\bar{X} \leq k' | \lambda = \lambda_0) = P(2\lambda_0 n \bar{X} \leq 2\lambda_0 nk' | \lambda = \lambda_0) = P(Y \leq 2\lambda_0 nk'),
\]

(16.3)

where \( Y \) has the \( \chi^2_{2n} \) distribution. This gives us the observed significance
\[ P = P(Y \leq 2\lambda_0 n \bar{x}) \]
for the observed value \( \bar{x} \) of the test statistic \( \bar{X} \). However, this simple case illustrates why traditionally hypothesis testing was
done with certain conventional fixed values of \( \alpha \), because before the days
of computers it would have been impossible to calculate this probability
when required.

Consider, then, finding the critical value \( k' \) to obtain a test of size \( \alpha \). Denoting the upper
100\( p \)% point of the \( \chi^2_m \) distribution as in Lecture 2 by
\( \chi^2_{m,p} \) then we have

\[ 2\lambda_0 nk' = \chi^2_{2n,1-a} \quad \therefore \quad k' = \frac{\chi^2_{2n,1-a}}{2\lambda_0 n}. \]
So the test of size \( \alpha \) rejects \( H_0 \) if \( \bar{x} \geq \frac{\chi^2_{2n,1-a}}{2\lambda_0 n} \). We can get \( \chi^2_{2n,1-a} \) from
tables (e.g. Neave Table 3.2) for a range of values of \( \alpha \), including the usual
5\%, 1\% etc.

With modern computer software, of course, we can readily find the
\( P \)-value for any \( \bar{x} \), or the critical value \( k' \) for a test of any desired size.

\textbf{Remark 16.10} In this example, Step 1 was relatively straightforward. However,
to do Step 2 we needed to string together a rather complicated argument
around various results in distribution theory. Most of the facts you
are likely to need in arguments like this are on the handouts on distributions
and useful facts. When constructing hypothesis tests, it is very useful to
know your way around these handouts.

\textbf{Remark 16.11} In Level 1 statistics, you may have constructed various
tests using more intuitive arguments. In effect, you will have identified a
suitable test statistic \( T(x) \) and also the form of the test heuristically. The
new thing in this course is the formal Step 1, where \( T \) and the form of test
are derived from the likelihood ratio. Instead of guessing what would make
a good statistic and form of test, we have the N-P Lemma to tell us what
the optimal test looks like. But Step 2 is essentially the same, whichever
approach we use up to that point.