1. Introduction

Welcome to the series of E-learning modules on Neyman Pearson Lemma, MP test & UMP test. In this module we are going cover the statement of Neyman Pearson Lemma and understand its application. We shall also cover the concepts of Best Critical Region, Most Powerful tests and Uniformly Most Powerful tests.

By the end of this session, you will be able to:

- Describe the Best critical region
- Explain Most Powerful and Uniformly Most Powerful tests
- Explain Neyman- Pearson Fundamental lemma
- Understand the application of the lemma

As we learned from our work in the previous lessons, whenever we perform a hypothesis test, we should make sure that the test we are conducting has sufficient power to detect a meaningful difference from the null hypothesis.
That said, how can we be sure that the **T**-*test* for a mean **mu *is*** the "most powerful" test we could use?
Is there instead a **D**-test or a Y-test or some other test that would provide us with more power?

A very important result, known as the Neyman Pearson Lemma, will reassure us that each of the tests is the most powerful test for testing statistical hypotheses about the parameter under the assumed probability distribution.

Before we can present the lemma, however, we need to
- Define some notation,
- Recollect the distinction between simple and composite hypotheses, and
- Define what it means to have a best critical region of size **alpha**

Notation.
If x one*,* x two*,* etc*,* X n is a random sample of size *n* from a distribution with probability density (or mass) function f *of (*x *, theta),* then the joint probability density (or mass) function *of* x one*,* x two*, etc,* x n is denoted by the likelihood function L *of* (theta).
That is, the joint probability density function or probability mass function is:
L of (theta) equal to L of (theta; x one, x two, etc till x n) which is equal to f of (x one; theta) into f of (x2;theta) till into f of (x n; theta)

Note that for the sake of ease, we drop the reference to the sample x one*,* x two*,* till x n in using L *of* (theta) as the notation for the likelihood function. We will want to keep in mind though that the likelihood L *of (*theta) still depends on the sample data.

Simple and Composite Hypothesis
If a random sample is taken from a distribution with parameter **theta**, a hypothesis is said to be a simple hypothesis if the hypothesis **uniquely specifies** the distribution of the population from which the sample is taken. Any hypothesis that is not a simple hypothesis is called

a composite hypothesis.

Let a simple null hypothesis H naught: theta is equal to theta naught be tested against the simple alternative hypothesis H one: theta is equal to theta one.
Let C be a critical region for testing H naught of size alpha against H one.
C is called the best critical region of size alpha for testing H naught against H one if C has at least the same power as any other critical region of size alpha.
That is, if C is a critical region of size alpha and C dash is any other critical region of size alpha then Power of C is greater than or equal to Power of C dash.

In testing of hypothesis we keep alpha, the level of significance at a fixed level (say zero point zero five or zero point zero one) and try to minimize the Type two error.
The sample space may be partitioned into several ways so that each critical region, w has the same size alpha. Of all these critical regions choose that which has least type two error.
This is called the best critical region of size alpha.

2. Most Powerful Test Neyman-Pearson Lemma for Finding the Best Test

Most Powerful Test:
Among the critical regions of the same size alpha, that which renders the minimum Type two error is called the most powerful critical region.
The test based on the most powerful critical region is called the most powerful test.
Therefore among all tests possessing the same size of Type one error, choose one for which the size of the Type two error is small as possible. This test is called the Most Powerful test.

Hence if two tests have the same level of significance, then the test with a smaller-size type two error is the most powerful test of the two at that significance level.

Now that we have clearly defined what we mean for a critical region **C** to be "best," we're ready to turn to the Neyman Pearson Lemma, to learn what form, a hypothesis test must take in order for it to be the best, that is, to be the most powerful test.

Neyman- Pearson Lemma for finding out the best test
Let the probability density function of the population be f of (x, theta) where theta represents the parameters.
Draw a sample (x one, x two, etc x n) from this population.
Let L be the likelihood function.
Then L is equal to f of (x one, x two, etc till x n).
Since the sample is drawn independently f of (x one, x two, etc till x n) is equal to f of (x one) into f of (x two) etc till into f of (x n)

Let L one represent the likelihood function when H one is true and L naught represent the likelihood function when H naught is true.
H naught and H one are simple hypotheses.
L naught is equal to f of (x one given H naught) into f of (x two given H naught) into etc till into f of (x n given H naught)
L one is equal to f of (x one given H one) into f of (x two given H one) into etc till into f of (x n given H one).

Let alpha be the level of significance
Let w be the critical region. So w is a subset of the sample space.
Then the theorem states that we can determine w in such a way that L one by L naught is greater than k.
Where k is the value determined on the basis of the level of significance or size of the test.

Then w is called the most powerful critical region of the significant level alpha for testing H naught against H one.
The test based on the most powerful critical region is called the most powerful test.

Hence Neyman Pearson lemma gives Best Critical Region for testing a simple null hypothesis that theta is equal to theta naught against H one that theta is equal to theta one which is a sufficient condition for a best critical region of size **alpha**.

3. Uniformly Most Powerful Test

Uniformly Most Powerful Test:
Suppose T is a test of size alpha for testing the null hypothesis that theta belongs to the parameter space under h naught, against an alternative hypothesis that theta belongs to the parameter space under H one.
T is said to a uniformly most powerful test (UMPT) of size alpha for testing H naught against H one, if for any other test T dash of size alpha for testing H naught against H one, P t of theta is greater than P t dash of theta for all theta belongs to the parameter space under H one where PT of theta denotes the power of the test.

The Neyman-Pearson Lemma can be used in certain cases to derive optimal tests of a simple null versus a composite alternative.
The region C is a uniformly most powerful critical region of size alpha for testing the simple hypothesis H against a composite alternative hypothesis H one if C is a best critical region of size alpha for testing H naught against each simple hypothesis in H one. The resulting test is said to be uniformly most powerful.

In statistical hypothesis testing, a uniformly most powerful (UMP) test is a hypothesis test which has the greatest power one minus beta among all possible tests of a given size alpha.
For example, according to the Neyman–Pearson lemma, the likelihood-ratio test is UMP for testing simple (point) hypotheses.
Uniformly most powerful tests don't always exist, but when they do, the Neyman Pearson Theorem provides a technique for finding them.

For simple H naught and composite H one, the critical region C is a uniformly most powerful critical region of size alpha if C is the most powerful (best) critical region for testing H naught against every simple hypotheses in H one.

The corresponding test is a uniformly most powerful (UMP) test with level of significance alpha for testing the simple H naught versus the composite (or simple) H one.

Comments:
- A UMP test may not exist. (There is usually trouble with the two-sided alternative hypothesis, theta not equal to theta naught)
- H 1 could be simple. Then the most powerful or best test described above (for simple versus simple) is, by default, uniformly most powerful.
- A UMP test may be easily defined for the composite versus composite case

Neyman-Pearson Framework:
1. Fix a significance level alpha for the test
2. Among all rules respecting the significance level, pick the one that uniformly maximizes power

When do UMP tests exist?
Insight on which composite pairs typically admit UMP tests:

- Hypothesis pair concerns a single real-valued parameter
- Hypothesis pair is “one-sided”

However existence of UMP test not only depends on hypothesis but also depends on specific model.

4. Unbiasedness and Consistency of a Test Procedure

Unbiasedness and consistency of a test procedure
The test procedure for which the power is more than its size is called an unbiased test.
That is if one minus beta is greater than alpha then it is an unbiased test procedure.
Similarly a critical region whose power is more than its size is called unbiased critical region.

In a test procedure if the power tends to one as n tends to infinity then the test procedure is called consistent test procedure
That is one minus beta tends to one as n tends to infinity.
Similarly a critical region whose power tends to one as n tends to infinity then it is said to be a consistent critical region.

Result:
In testing a simple null hypothesis against a simple alternative hypothesis the power of the Best Critical Region cannot be less than its size or Best Critical Region obtained by NP lemma is unbiased.

Proof:
Let x one, x two, etc till x n be a sample of size n taken from the population f of (x, theta).
Using a NP lemma the BCR is given by
C equal to { (x one, x two, etc till x n): L one of (x i , theta) by L naught of (x I , theta) greater than or equal to k.
Therefore Acceptance region, A is equal to (x one, x two, etc till x n): L one of (x I , theta) by L naught of (x i , theta) is less than k.
This implies L one of (x i , theta) less than k into L naught of (x i , theta)
in the acceptance region and L one of (x i, theta) greater than or equal to k into L naught of (x i, theta) in the rejection region.

Consider:
L one of (x i, theta) greater than or equal to k into L naught of (x i, theta).
This implies L one greater than or equal to k into L naught where L one is equal to L one of (x i, theta) and L naught is equal to L naught of (x i, theta).
This implies integral over C L one dx greater than or equal to k into integral over C L naught dx.
Implies probability of reject H naught when H one is true is greater than or equal to k into probability of reject H naught when H naught is true.
This implies (one minus beta) is greater than or equal to k into alpha. Call this as equation one.

Consider L one of (x i , theta) by L naught of (x i , theta) less than k. this implies L one of (x i , theta) is less than k into L naught of (x i , theta). Which implies, L one is less than k into L naught.
Implies integral over A L one dx less than k into integral over A L naught dx.

Implies probability of Accept H naught when H one is true less than k into probability of Accept H naught when H naught is true.
Implies beta is less than k into (one minus alpha). Call this as equation two.

Equation one multiplied by (minus one) gives
Minus (one minus beta) less than k into alpha. Call this as (3)
Equation (3) by (2) gives
Minus (one minus beta) by beta is less than minus k into alpha by (k into (one minus alpha).
This implies, minus (one minus beta) into k into (one minus alpha) is less than minus k into alpha into beta.
This implies [alpha minus (one minus beta)] is less than zero implies one minus beta is greater than alpha.

Hence the power of the Best Critical Region obtained by NP lemma is greater than its size or BCR is unbiased.

In hypothesis testing situations, there are often several statistical tests from which to choose.
Ideally, we prefer tests with small probabilities of type one and type two errors and high power.
That is, we prefer to use the test with maximum power, often referred to as the most powerful test.

5. Examples

Example one
Find the BCR of size alpha for testing null hypothesis, lambda is equal to lambda naught against the alternative hypothesis, lambda is equal to lambda one when a sample of size n is drawn from the Poisson population with unknown parameter lambda.

Solution:
Let x one, x two, etc x n be a random sample of size n from Poisson population with parameter lambda. By NP Lemma a size alpha Best Critical Region is given is given by, C equal to { (x one, x two, etc till x n), L one of (x i , lambda) by L naught of (x i , lambda) is greater than k.
Where k is such that Probability of x belongs to C given that H naught is true is equal to alpha

L one of (x i, lambda) by L naught of (x i, lambda) is equal to 'e' to the power minus n into lambda one into lambda one to the power summation xi by product of x i factorial whole divided by 'e' to the power minus 'n' into lambda naught into lambda naught to the power summation xi by product of x i factorial.
Which is equal to 'e' to the power minus n into (lambda one minus lambda naught) into (lambda one by lambda naught) to the power summation x i.
Now L one of (x i , lambda) by L naught of (x i , lambda) greater than k implies 'e' to the power minus 'n' into (lambda one minus lambda naught) into (lambda one by lambda naught) to the power summation x i is greater than k.

This implies minus 'n' into (lambda one minus lambda naught) plus summation xi into Ln into (lambda one by lambda naught) is greater than Ln into k
Implies, summation x i into Ln (lambda one by lambda naught) is greater than Ln k plus 'n' into (lambda one minus lambda naught) which is equal to k one (say)

Case 1:
Let lambda one greater than lambda naught
L one of (x i , lambda) by L naught of (x i , lambda) greater than k implies
Summation x i greater than k one by Ln into (lambda one by lambda naught) which is equal to k two.
The BCR when lambda one is greater than lambda naught is given by,
C is equal to { (x one, x two, etc till x n): summation x i greater than k two}

Case 2:
Let lambda one less than lambda naught
L one of (x i , lambda) by L naught of (x i , lambda) less than k implies
Summation x i is less than k one by Ln into (lambda one by lambda naught) which is equal to k two.
The BCR when lambda one less than lambda naught is given by,
C is equal to { (x one, x two, etc till x n): summation x i less than k two}

Concluding, in hypothesis testing we prefer tests with small probabilities of type one and type two error and high power.

That is, we prefer to use the test with maximum power.
When we are testing a simple null hypothesis versus a simple alternative hypothesis we use most powerful test and in case of a simple null hypothesis versus a composite alternative hypothesis we apply Uniformly most powerful test.

Here’s a summary of our learning in this session where we have understood the following:

- Concept of Best Critical region
- Statement of Neyman Pearson Fundamental lemma for finding the best test
- Most Powerful Test and Uniformly Most Powerful Test
- Application of N P lemma to get Best Critical Region