Section 12.4: Neyman-Pearson Lemma

- **RECAP:**
  - **In this section, we only consider Simple Null Hypothesis against Simple Alternative Hypothesis testing based on a random Sample of n observations from the population with** $X \sim f(x;\theta)$.
  - $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$.
  - **Definition: Power of the TEST Procedure =**
    - $1 - P[\text{Type II error}] = P[\text{Reject } H_0 \mid H_0 \text{ False}] = 1 - \beta$
  - **There can be many critical regions for a given** α **.**
  - **Power Function: Relationship Between** α and $1-\beta$ **for all possible Critical Regions**
- **Definition: The Most Powerful Test (Best Critical Region) for a given** α **is the test with the largest power, at** $\theta = \theta_1$,
  - ***to detect the False NULL Hypothesis***
- **How does one find the Most Powerful (MP) Test?**
- **Use Neyman-Pearson Lemma:**
  - **For a fixed value of** α **, it characterizes the CR with smallest possible value of** β **.**
  - **Test procedure with the largest power** $= (1-\beta)$ **to detect the False Null.**
- **In this situation (simple null and alternative hypotheses), there are only two possible values** $\{\theta_0, \theta_1\}$ **, of the parameter** θ **.**
- **The likelihood of the two parameter values based on the observations** $\mathbf{x} = x_1, \ldots, x_n$ **, in a random sample of size n, are given by**

$$L_0 = f_0(\mathbf{x}) = \prod_{i=1}^{n} f(x_i \mid \theta_0), \text{ and } L_1 = f_1(\mathbf{x}) = \prod_{i=1}^{n} f(x_i \mid \theta_1).$$

- **Intuitive reasoning: Recall that the MLE of** θ **corresponds to the parameter value at which the likelihood takes the maximum value.**

- **Thus, for data** $\mathbf{x} = x_1, \ldots, x_n$ **for which**
  - $L_1 > L_0$, **the MLE is** θ_1,
  - $L_1 < L_0$, **the MLE is** θ_0, **and**
  - $L_1 = L_0$, **both** $\{\theta_0, \theta_1\}$ are the **MLE.**
- **So it makes sense to Reject the Null whenever** L_1 / L_0 is large.
  - **The larger this ratio, the more likely** θ_1 **is!**
- **Of course, the threshold for the rejection region is chosen so as to satisfy the criterion,** $P[\mathit{Type\ I\ error}] \le \alpha.$

- **Before stating and proving the Neyman-Pearson Lemma, we prove a more general result for finding a test procedure that minimizes a linear function** $a\alpha + b\beta$ **of** α and β, **for specified positive constants *a* and *b*.**

Theorem 1: [DeGroot, M.H. Probability and Statistics, 2nd Edition]

Let δ^* **denote a test procedure satisfying:**

(i) Don't Reject H_0 [Accept H_0], if $af_0(\mathbf{x}) > bf_1(\mathbf{x})$,

(ii) Reject H_0 [Accept H_1], if $af_0(\mathbf{x}) < bf_1(\mathbf{x})$, and

(iii) Either H_0 or H_1 may be accepted, if $af_0(\mathbf{x}) = bf_1(\mathbf{x})$.

Then, for any other test procedure δ,

$$a\alpha(\delta) + b\beta(\delta) \ge a\alpha(\delta^*) + b\beta(\delta^*).$$

Proof: **Let** R **denote the critical region of a test procedure** δ, **then** R **consists of all sample outcomes for which** δ **specifies to Reject H$_0$ and** R' **consists of all sample outcomes for which** δ **specifies to "Accept H$_0$." Therefore,**

$$a\alpha(\delta) + b\beta(\delta) = a\int\int_R \cdots \int f_0(\mathbf{x})d\mathbf{x} + b\int\int_{R'} \cdots \int f_1(\mathbf{x})d\mathbf{x}$$

$$= a\int\int_R \cdots \int f_0(\mathbf{x})d\mathbf{x} + b\{1 - \int\int_R \cdots \int f_1(\mathbf{x})d\mathbf{x}\}$$

$$= b + \int\int_R \cdots \int \{af_0(\mathbf{x}) - bf_1(\mathbf{x})\}d\mathbf{x} \qquad (1)$$

It follows from (1) that the value of $a\alpha(\delta) + b\beta(\delta)$ will be a minimum if the critical region R is chosen so that the value of the integral on the right hand side of (1) is minimum.

Furthermore, the value of this integral will be a minimum:

- **if the set R includes every x such that $af_0(\mathbf{x}) - bf_1(\mathbf{x}) < 0,$ and**
- **doesn't include any point x for which $af_0(\mathbf{x}) - bf_1(\mathbf{x}) > 0.$**
- **If for some x for which $af_0(\mathbf{x}) - bf_1(\mathbf{x}) = 0,$ it doesn't matter whether this point is in R or $R',$ since such points contribute nothing to the integral in (1).**

It is easy to see that the critical region described here is exactly the same as that of the procedure δ^*, given in the statement of the Theorem 1.

- **Minimizing $\beta(\delta)$ subject to $\alpha(\delta) \le \alpha$ for a given value of α**
  - **Suppose that $\alpha(\delta) \le \alpha$, and it is desired to find a procedure δ^* for which $\beta(\delta^*)$ is the minimum.**
  - **We can apply the following result, which is closely related to Theorem 1, and is known as the Neyman-Pearson Lemma.**
    - **In honor of statisticians J. Neyman and E.S. Pearson, who developed these ideas in 1933.**

- **Neyman-Pearson Lemma. Suppose that δ^* is a test procedure which has the following form for some constant *k* > 0:**

  (i) Don't Reject H_0 [Accept H_0], if $f_0(\mathbf{x}) > kf_1(\mathbf{x})$,

  (ii) Reject H_0 [Accept H_1], if $f_0(\mathbf{x}) < kf_1(\mathbf{x})$, and

  (iii) Either H_0 or H_1 may be accepted, if $f_0(\mathbf{x}) = kf_1(\mathbf{x})$.

  If δ is any other test procedure such that $\alpha(\delta) \le \alpha(\delta^*),$ then it follows that $\beta(\delta) \ge \beta(\delta^*).$ Furthermore, if $\alpha(\delta) < \alpha(\delta^*),$ then it follows that $\beta(\delta) > \beta(\delta^*).$

  Proof: From the description of the procedure δ^* and Theorem 1 (with $a = 1, b = k$), it follows that for any other procedure δ,

  $$\alpha(\delta^*) + k\beta(\delta^*) \le \alpha(\delta) + k\beta(\delta) \qquad (2)$$

Now, if $\alpha(\delta) \le \alpha(\delta^*)$, **the right hand side of (2) satisfies**

$$\alpha(\delta) + k\beta(\delta) \le \alpha(\delta^*) + k\beta(\delta) \qquad (3)$$

Combining (2) with (3), it follows that,

$$\alpha(\delta^*) + k\beta(\delta^*) \le \alpha(\delta) + k\beta(\delta) \le \alpha(\delta^*) + k\beta(\delta).$$

Ignoring the middle term in the above inequalities and simplifying leads to $\beta(\delta) \ge \beta(\delta^*)$. **Furthermore, if** $\alpha(\delta) < \alpha(\delta^*)$, **(3) will have a strict inequality. Following the same line of arguments, we get** $\beta(\delta) > \beta(\delta^*)$.

- **To illustrate the use of Neyman-Pearson Lemma, if an experimenter wishes to use a test procedure for which** $\alpha(\delta) = .05$, **and** $\beta(\delta)$ **is minimum, then according to N-P lemma, he/she should try to find a *k* for which** $\alpha(\delta^*) = .05$ (say). **The test procedure** δ^* **will have a minimum possible value of** $\beta(\delta)$.
- **The key question that remains: how does find the value of *k?***
  - **If the population distribution is continuous, then it is usually possible to find such a *k*.**
  - **However, if the population distribution is discrete, then it is typically not possible to find such a *k*. We will discuss by Examples, as to what to do in such cases.**
  - **Examples:**
    1. **Hypergeometric distribution with unknown M**
    2. **Normal Populations for Yield of Corn**
    3. **Bernoulli samples with n=20**
    4. **Ex. 12.10**
    5. **Ex. 12.13**
    6. **Ex. 12.15**

- **Toy Example 1: Find the most powerful critical region for a given P[Type I error]**
  - Ex. 12.3 X~ Hypergeometric dist. with N=7, n=2
  - Urn has N=7 Balls [M Red, N-M Blue]
  - M not known.
  - Null hypothesis H_0: M=2 against Alternative H_1: M=4
  - Experiment: Draw 2 balls, Let X= # of red balls
  - Sample Space S={0, 1, 2}
  - It is easy to calculate the prob. mass functions of X under these two hypotheses.
  - The probability distributions under H_0: and H_1: and the likelihood ratio values are given in the table below:
  -

X	$f_0(x)$	$f_1(x)$	$LR = \frac{f_0(x)}{f_1(x)}$
0	**10/21**	**3/21**	**10/3**
1	**10/21**	**12/21**	**10/12**
2	**1/21**	**6/21**	**1/6**

  - From the 4th column of the above table, it is clear the likelihood ratio is smallest at *x=2*, 2nd smallest at *x=1*, and largest at *x=0*. Depending on the threshold value of *k*, the rejection regions are as follows:

K	$R = \{x : LR < k\}$	$\alpha(\delta^*) = P\{R \mid f_0\}$	$\beta(\delta^*) = P\{R' \mid f_1\}$
$0 < k < 1/6$	$\varnothing$	0	1
$\frac{1}{6} \le k < \frac{10}{12}$	{2}	$\frac{1}{21}$	$\frac{3}{21} + \frac{12}{21} = \frac{15}{21}$
$\frac{10}{12} \le k < \frac{10}{3}$	{2,1}	$\frac{1}{21} + \frac{10}{21} = \frac{11}{21}$	$\frac{3}{21}$
$\frac{10}{3} \le k$	{2,1,0}=S	1	0

- Note that there is no set R in this example for which $\alpha(\delta^*) = .05$

- Example: n independent Bernoulli trials
- Notation: $\mathrm{P}[Success] = \theta$
- $H_0 : \theta = \theta_0$ Against $H_1 : \theta = \theta_1$, where $\theta_1 > \theta_0$.
  The likelihood function:
-
$$L(\theta) = \prod_i f(x_i \mid \theta) = \prod_i \theta^{x_i} (1-\theta)^{1-x_i}$$
$$= \theta^{\sum x_i} (1-\theta)^{(n-\sum x_i)} = \theta^y (1-\theta)^{(n-y)}, \text{ where } \mathrm{y} = \sum x_i.$$

- Note that the likelihood ratio depends only on the sufficient statistics:
$$LR(y) = \frac{L(\theta_0)}{L(\theta_1)} = \frac{\theta_0{}^y (1-\theta_0)^{(n-y)}}{\theta_1{}^y (1-\theta_1)^{(n-y)}} = \left(\frac{(1-\theta_0)}{(1-\theta_1)} \right)^n \left[\frac{\{\theta_0 / (1-\theta_0)\}}{\{\theta_1 / (1-\theta_1)\}} \right]^y .$$
- Given the values of θ_0 and θ_1, the first term in the above expression is some constant. So it can be ignored for the purpose of finding the MP region, i.e.,
- $LR(y) < k \Leftrightarrow \left[\frac{\{\theta_0 / (1-\theta_0)\}}{\{\theta_1 / (1-\theta_1)\}} \right]^y < k_1$, for some constant k_1.
- Taking natural log on both sides, we get $y \ln \left(\left[\frac{\{\theta_0 / (1-\theta_0)\}}{\{\theta_1 / (1-\theta_1)\}} \right] \right) < k_2$,
- Now, since $\theta_0 < \theta_1$, $\left(\left[\frac{\{\theta_0 / (1-\theta_0)\}}{\{\theta_1 / (1-\theta_1)\}} \right] \right) < 1$, hence its logarithm is negative.
- The MP rejection region is: $y \geq c$, where c is *chosen so that*
  $P[Y \geq c \mid f_0] \leq \alpha$. *(*)*
- Note that, under H_0, the distribution of $Y = \sum X_i \sim B(n, \theta_0)$ with
  $g(y \mid \theta_0) = \binom{n}{y} \theta_0^y (1-\theta_0)^{n-y}$
- For any given α, we need to use the Table of Binomial probabilities for the sample size *n*, in Appendix B.2 to find the largest CR (in this case, smallest c) such that * is satisfied. [Since the Binomial distribution is discrete, we can't find a cut-off point corresponding to some exact values of α.]
- For example, let n=20, and θ_0=.5 and $\theta_1 = .9$. Now for
  $\alpha = .01, .05, .10$, c=16,15, and 14 respectively.

- You are supposed to find the shape of the M-P Rejection region in Problem 12.12 (assigned in HW#5) when $\theta_0 > \theta_1$.
- [Hint: In this case, you need to show that the MP Critical Region is of the form $y \le c$.]
- ***Note: The shape of the Critical Region depends on whether*** $\theta_1 > \theta_0$ or $\theta_0 > \theta_1$. **But the cut-off point for the *CR* depends only on the value of** θ_0 **in the Null Distribution.**
- In the Toy example with 20 independent Bernoulli trials, we had $\theta_0 > \theta_1$. We showed how to calculate the values of α and β, given some cut-off points c. When we are given a value of α, we need to find the largest CR [in this case, largest c] such that, $P\left[Y \le c \mid f_0\right] \le \alpha$.
- Recap for given rejection regions:
- $\cdot$Rejection region : $\{X \le 14\} \Leftrightarrow P[\text{Type I error}] = P[X \le 14 \mid \theta_0 = 0.9]$
  $= P[\# \text{ Failure} \ge 6 \mid P(Fail) = 0.1] = .0114$ [Table I, p. 567]
- $\cdot$Rejection region : $P[\text{Type I error}] = P[X \le 15 \mid \theta = 0.9]$
  $= P[\# \text{ Failure} \ge 5 \mid P(Fail) = 0.1] = .0114 + .0319 = .0433$ [Table I, p. 567]
- $\cdot$Critical region : $\{X \le 16\} \Leftrightarrow [\text{Type I error}] = P[X \le 16 \mid \theta = 0.9]$
  $= P[\# \text{ Failure} \ge 4 \mid P(Fail) = 0.1] = .0114 + .0319 + .0898 = .1331$ [Table I, p. 567]

- From these calculations, it follows that if
- $\alpha = .025$, then c=14.
- $\alpha = .05$, then c=15.
- $\alpha = .10$, then c=15.
- $\alpha = .15$, then c=16.

- Example 3: Data Model - Normal distribution
  - Simple (Null) vs. Simple (Alternative)
  - Corn Yield Application:
    $\mathbf{H_0}: \mu = 120$ against $\mu = 125$
  - $\sigma^2 = 36$
    $n = 4\ farms$
  - M-P Rejection Region:
    - RS of size n from Normal population with unknown mean and *known* variance σ^2.
- The likelihood function for n independent $N(\theta, \sigma^2)$ random variables is

$$L(\theta) = \prod_i f(x_i \mid \theta) = \prod_i \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2}(x_i-\theta)^2} = \frac{1}{(2\pi\sigma)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum_i (x_i-\theta)^2}.$$

- The LR is $\dfrac{L(\theta_0)}{L(\theta_1)} = \dfrac{\frac{1}{(2\pi\sigma)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum_i (x_i-\theta_0)^2}}{\frac{1}{(2\pi\sigma)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum_i (x_i-\theta_1)^2}} = e^{-\frac{1}{2\sigma^2}[\sum_i (x_i-\theta_0)^2 - \sum_i (x_i-\theta_1)^2]}$
  - However, it was shown earlier that $\sum_i (x_i-\theta)^2 = \sum_i (x_i-\bar{x})^2 + n(\bar{x}-\theta)^2$.
  - Therefore, the LR ratio can be simplified to

$$\frac{L(\theta_0)}{L(\theta_1)} = e^{-\frac{n}{2\sigma^2}[(\bar{x}-\theta_0)^2-(\bar{x}-\theta_1)^2]} = e^{\frac{n}{2\sigma^2}[2\bar{x}(\theta_0-\theta_1)-(\theta_1^2-\theta_0^2)]}.$$

  - Now, taking log on both sides of the inequality LR< k, we get
  - $\frac{n}{2\sigma^2}[2\bar{x}(\theta_0-\theta_1)-(\theta_1^2-\theta_0^2)] < k_1 \Leftrightarrow \bar{x}(\theta_0-\theta_1) < c$
  - Now, if $\frac{n}{2\sigma^2}[2\bar{x}(\theta_0-\theta_1)-(\theta_1^2-\theta_0^2)] < k_1 \Leftrightarrow \bar{x}(\theta_0-\theta_1) < c$
  - Therefore, when
    - $(\theta_0-\theta_1) > 0$ the MP critical Region is $\bar{X} < c$
    - $(\theta_0-\theta_1) < 0$ the MP critical Region is $\bar{X} > c$

- Depending on the relationship between the null and alternative models, the value of c in each case is found by solving for
- $P(\bar{X} > c \mid \theta = \theta_0) = \alpha$ or $[P(\bar{X} < c \mid \theta = \theta_0)]$
- We know that, when H_0 is true, $\bar{X} \sim N(\theta_0, \frac{\sigma^2}{n})$. Therefore, for $\theta_0 < \theta_1$

  $$P(\bar{X} > c \mid \theta = \theta_0) = \alpha \Leftrightarrow P\left[\frac{\bar{X} - \theta_0}{\sigma / \sqrt{n}} > \frac{c - \theta_0}{\sigma / \sqrt{n}}\right] = \alpha$$

- $$\Leftrightarrow P\left[Z > \frac{c - \theta_0}{\sigma / \sqrt{n}}\right] = \alpha \Leftrightarrow \frac{c - \theta_0}{\sigma / \sqrt{n}} = z_\alpha$$

  $$\Leftrightarrow c = \theta_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$$

- Similarly, for $\theta_0 > \theta_1$

  $$P(\bar{X} < c \mid \theta = \theta_0) = \alpha \Leftrightarrow P\left[\frac{\bar{X} - \theta_0}{\sigma / \sqrt{n}} < \frac{c - \theta_0}{\sigma / \sqrt{n}}\right] = \alpha$$

- $$\Leftrightarrow P\left[Z < \frac{c - \theta_0}{\sigma / \sqrt{n}}\right] = \alpha \Leftrightarrow \frac{c - \theta_0}{\sigma / \sqrt{n}} = z_{1-\alpha} \Leftrightarrow \frac{c - \theta_0}{\sigma / \sqrt{n}} = -z_\alpha$$

  $$\Leftrightarrow c = \theta_0 - z_\alpha \frac{\sigma}{\sqrt{n}}$$

- Again, note that the shape of the Critical Region depends on whether $\theta_1 > \theta_0$ or $\theta_0 > \theta_1$.
  - But the cut-off point for the CR depends only on the value of θ_0 in the Null Distribution.

- ## Section 12.5 Power Function of a Test

  - Given the MP Critical Region for a simple against simple hypothesis, we can find the power (to reject the false Null) $(= 1 - \beta)$ of the test by calculating the $P(\bar{X} \in CR \mid \theta = \theta_1)$.

- For example, when $\theta_0 < \theta_1$, the power $\pi(\theta_1)$ of the test for any value of θ_1 in the alternative region is given by

$$\pi(\theta_1) = P(\bar{X} > c \mid \theta = \theta_1) = P(\bar{X} > \theta_0 + z_\alpha \frac{\sigma}{\sqrt{n}} \mid \theta = \theta_1)$$

$$= P\left[\frac{\bar{X} - \theta_1}{\sigma / \sqrt{n}} > \frac{\theta_0 + z_\alpha \frac{\sigma}{\sqrt{n}} - \theta_1}{\sigma / \sqrt{n}} \right]$$

$$= P\left[Z > \frac{\theta_0 - \theta_1}{\sigma / \sqrt{n}} + z_\alpha \right]$$

- Since the shape of the CR of the MP test depends only on the ordering relationship between the null and alternative parameter values, and the actual cut-off point for the CR depends only on the null,
  - it follows that if either or both of these are ***one sided composite*** hypothesis, the same test procedure would be usually good.
- For example,
  - In the test for the effectiveness of the new medicine, we could consider the null and the alternative Hypotheses as $H_0 : \theta \geq .90 \ against \ H_0 : \theta < .90$
  - And in the corn yield example, $H_0 : \mu \leq 120$ against $H_1 : \mu > 120$

- **Note that in the above expression, the power to detect the false null depends on the difference between the values of the null and the alternative parameters.**

- Thus we can actually calculate the power of a given test, with a specified value of α at the boundary point of NULL and Alternative Parameters Set, for detecting the false null for any point in the alternative set, as well as the probability of rejecting the true null for any point in the Null set. This is called the Power function of the test as a function of θ.

- **Definition: The Power Function of a given test procedure (Critical Region) is defined by**

$$\pi(\theta) = \begin{cases} P(\text{Reject } \mathrm{H}_0|\theta \text{ under } \mathrm{H}_0) = \alpha(\theta) \\ P(\text{Reject } \mathrm{H}_0|\theta \text{ under } \mathrm{H}_1) = 1 - \beta(\theta) \end{cases}$$

- For θ under the alternative, the power function is same as the probability of not committing the Type II error.
- For one sided composite Null and Alternative hypotheses, the MP test is sometimes Uniformly Most Powerful (UMP) test for all θ under the alternative.
- For simple Null $\theta = \theta_0$ against two sided alternatives $\theta \neq \theta_0$, there is no UMP test.
- An alternative to power function is to calculate probability of (Accepting) Not-Rejecting H_0 for all θ. It is called the Operating Characteristic (OC) Curve in Engineering and Medical science literature.
- IGNORE the Last paragraph of Section 12.5
- Next Lecture: Likelihood Ratio Tests for composite against composite hypotheses.