Lecture Notes 16
36-705

Today we will switch gears and talk about hypothesis testing. But before we do, there is one last, important fact about point estimation: the optimality of the mle. It's complicated to make this precise. (See *Asymptotic Statistics* by van der Vaart for a good treatment.)

1 The MLE is Optimal

Roughly, it goes like this. We know that mle satisfies

$$\sqrt{n}(\widehat{\theta} - \theta) \rightsquigarrow N\left(0, \frac{1}{I(\theta)}\right).$$

If $\widetilde{\theta}$ is any other well-behaved estimators, then

$$\sqrt{n}(\widetilde{\theta} - \theta) \rightsquigarrow N(0, \sigma^2)$$

where $\sigma^2 \geq 1/I(\theta)$. The phrase "well-behaved" refers to some desirable technical conditions on the estimator. Thus, the mle is the most precise estimator. Similarly, it can be shown that the mle is asymptotically minimax under a large class of loss functions. Unfortunately, making all this precise takes machinery that is beyond the scope of the course. But the message is just that there are good reasons for using the mle in parametric problems. In nonparametric problems, we shall see that the situation is quite different.

1.1 Notation

We will need the following notation. Let Φ be the cdf of a standard Normal random variable Z. For $0 < \alpha < 1$, let

$$z_\alpha = \Phi^{-1}(1 - \alpha).$$

Hence,

$$P(Z > z_\alpha) = \alpha \quad \text{and} \quad P(Z < -z_\alpha) = \alpha.$$

Sometimes we will write X^n to mean $(X_1, \ldots, X_n)$.

2 Hypothesis Testing

The classical statistical hypothesis testing framework (as with much of statistics) originated with Fisher.

Example 1: The story goes that a colleague of Fisher claimed to be able to distinguish if in an English tea, milk was added before water (or the other way around). Fisher proposed to give her 8 cups of tea, 4 of which had milk first, and 4 of which had tea first in a random order. The point was roughly, that if she was "labeling" at random then she would have a small chance $1/\binom{8}{4} = 1/70 = 0.014$ of getting every cup right. In his description, the null hypothesis was that she had no ability to distinguish. She actually got them all correct, which would have happened by chance with probability 0.014. He concluded that since this probability was less that 0.05 that it was "statistically significant". Notice the asymmetry in this description: only a null hypothesis is actually specified (i.e. there is no alternative hypothesis – it is in some sense implicit), i.e. the null hypothesis is often special. Furthermore, there is an arbitrary choice of a cut-off 0.05 below which we declare something is significant.

Hypothesis testing is really everywhere. It would probably alarm you to know how many policy decisions, nutrition decisions, scientific results live or die on the basis of hypothesis tests.

Exaample 2. In July, Castillo et al reported on a study about treating Covid with Vitamin D. 76 patients were randomized to treatment or no treatment. The outcome was whether they needed to be put in the ICU or not. The null hypothesis that there is no value in using Vitamin D had a p-value less than .001.

Example 3: A couple of typical examples to emphasize again why the null might really be special. A common example is in forensics. Things like fingerprint matches, DNA matches, deciding whether pieces of glass match in their chemical composition etc. are actually problems of a statistical nature. Here perhaps following the "innocent till proven guilty" adage, the null hypothesis is that the defendant is innocent. We then need to review evidence and choose to either reject or fail to reject (i.e. acquit) the defendant. It is perhaps clear that there in many cases is a heavier price for false convictions and so it makes sense to control this error. Indeed, deciding how to choose a significance level in this context is a huge debate.

3 The formal framework

Let $X_1, \ldots, X_n \sim p(x; \theta)$. Suppose we want to know if $\theta = \theta_0$ or not, where θ_0 is a specific value of θ. For example, if we are flipping a coin, we may want to know if the coin is fair; this corresponds to $\theta = 1/2$. If we are testing the effect of two drugs — whose means effects are θ_1 and θ_2 — we may be interested to know if there is no difference, which corresponds to $\theta_1 - \theta_2 = 0$.

We formalize this by stating a *null hypothesis* H_0 and an alternative hypothesis H_1. For

example:

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

More generally, consider a parameter space Θ. We consider

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

where $\Theta_0 \cap \Theta_1 = \emptyset$. If Θ_0 consists of a single point, we call this a *simple null hypothesis*. If Θ_0 consists of more than one point, we call this a *composite null hypothesis*.

Example 1 $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$.

$$H_0 : p = \frac{1}{2} \qquad H_1 : p \neq \frac{1}{2}. \quad \square$$

The question is not whether H_0 is true or false. The question is whether there is sufficient evidence to reject H_0, much like a court case. Our possible actions are: reject H_0 or retain (don't reject) H_0.

	Decision	
	Retain H_0	Reject H_0
H_0 true	$\surd$	Type I error (false positive)
H_1 true	Type II error (false negative)	$\surd$

4 Constructing Tests

Hypothesis testing involves the following steps:

1. Choose a *test statistic* $T_n = T_n(X_1, \ldots, X_n)$.
2. Choose a rejection region $R \subset \mathcal{X}^n$. Often this has the form

   $$R = \Big\{(x_1, \ldots, x_n) : \ T_n(x_1, \ldots, x_n) > t\Big\}$$

   for some t.

3. If $(X_1, \ldots, X_n) \in R$ we reject H_0 otherwise we retain H_0.

Although you can define the rejection region without an associated test statistic, often it will be the case that R will be defined in terms of the test statistic, i.e. we simply reject if the test statistic takes an "extreme value". We define the *test function* ϕ by:

$$\phi(x_1, \ldots, x_n) = \begin{cases} 1 & \text{if } (x_1, \ldots, x_n) \in R \\ 0 & \text{otherwise.} \end{cases}$$

Example 2 *Let* $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$. *Suppose we test*

$$H_0 : p = \frac{1}{2} \qquad H_1 : p \neq \frac{1}{2}.$$

Let $T_n = n^{-1} \sum_{i=1}^n X_i$ *and*

$$R = \left\{ (x_1, \ldots, x_n) : \ |T_n(x_1, \ldots, x_n) - 1/2| > \delta \right\}.$$

So we reject H_0 *if* $|T_n - 1/2| > \delta$.

We need to choose T and R so that the test has good statistical properties. We will consider the following tests:

1. The Neyman-Pearson Test
2. The Wald test
3. The Likelihood Ratio Test (LRT)
4. The permutation test.

Before we discuss these methods, we first need to talk about how we evaluate tests.

5 Error Rates and Power

Suppose we reject H_0 when $(X_1, \ldots, X_n) \in R$. Define the *power function* by

$$\beta(\theta) = P_\theta((X_1, \ldots, X_n) \in R).$$

We want $\beta(\theta)$ to be small when $\theta \in \Theta_0$ and we want $\beta(\theta)$ to be large when $\theta \in \Theta_1$. The general strategy is:

1. Fix $\alpha \in [0,1]$.
2. Now try to maximize $\beta(\theta)$ for $\theta \in \Theta_1$ subject to $\beta(\theta) \leq \alpha$ for $\theta \in \Theta_0$.

Notice the asymmetry that we always favor the null hypothesis and only consider tests that control the Type-I error.

We need the following definitions. A test is *size* α if

$$\sup_{\theta \in \Theta_0} \beta(\theta) = \alpha.$$

A test is *level* α if

$$\sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha.$$

A size α test and a level α test are almost the same thing. The distinction is made bcause sometimes we want a size α test and we cannot construct a test with exact size α but we can construct one with a smaller error rate.

Example 3 *$X_1, \ldots, X_n \sim N(\theta, \sigma^2)$ with σ^2 known. Suppose we test*

$$H_0 : \theta = \theta_0, \qquad H_1 : \theta > \theta_0.$$

This is called a **one-sided alternative**. *Suppose we reject H_0 if $T_n > c$ where*

$$T_n = \frac{\overline{X}_n - \theta_0}{\sigma/\sqrt{n}}.$$

Then

$$\begin{aligned} \beta(\theta) &= P_\theta\left(\frac{\overline{X}_n - \theta_0}{\sigma/\sqrt{n}} > c\right) = P_\theta\left(\frac{\overline{X}_n - \theta}{\sigma/\sqrt{n}} > c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}}\right) \\ &= P\left(Z > c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}}\right) \\ &= 1 - \Phi\left(c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}}\right), \end{aligned}$$

where Φ is the cdf of a standard Normal and $Z \sim \Phi$. Now

$$\sup_{\theta \in \Theta_0} \beta(\theta) = \beta(\theta_0) = 1 - \Phi(c).$$

To get a size α test, set $1 - \Phi(c) = \alpha$ so that

$$c = z_\alpha$$

where $z_\alpha = \Phi^{-1}(1-\alpha)$. Our test is: reject H_0 when

$$T_n = \frac{\overline{X}_n - \theta_0}{\sigma/\sqrt{n}} > z_\alpha.$$

Example 4 *$X_1, \ldots, X_n \sim N(\theta, \sigma^2)$ with σ^2 known. Suppose*

$$H_0 : \theta = \theta_0, \qquad H_1 : \theta \neq \theta_0.$$

This is called a **two-sided** *alternative. We will reject H_0 if $|T_n| > c$ where T_n is defined as before. Now*

$$\begin{aligned}
\beta(\theta) &= P_\theta(T_n < -c) + P_\theta(T_n > c) \\
&= P_\theta\left(\frac{\overline{X}_n - \theta_0}{\sigma/\sqrt{n}} < -c\right) + P_\theta\left(\frac{\overline{X}_n - \theta_0}{\sigma/\sqrt{n}} > c\right) \\
&= P\left(Z < -c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}}\right) + P\left(Z > c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}}\right) \\
&= \Phi\left(-c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}}\right) + 1 - \Phi\left(c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}}\right) \\
&= \Phi\left(-c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}}\right) + \Phi\left(-c - \frac{\theta_0 - \theta}{\sigma/\sqrt{n}}\right)
\end{aligned}$$

since $\Phi(-x) = 1 - \Phi(x)$. The size is

$$\beta(\theta_0) = 2\Phi(-c).$$

To get a size α test we set $2\Phi(-c) = \alpha$ so that $c = -\Phi^{-1}(\alpha/2) = \Phi^{-1}(1 - \alpha/2) = z_{\alpha/2}$. The test is: reject H_0 when

$$|T| = \left|\frac{\overline{X}_n - \theta_0}{\sigma/\sqrt{n}}\right| > z_{\alpha/2}.$$

When $\alpha = .05$, $z_{\alpha/2} = 1.96 \approx 2$. In this case we reject when $|T| > 2$.

6 The Neyman-Pearson Test

Let $\mathcal{C}_\alpha$ denote all level α tests. A test in $\mathcal{C}_\alpha$ with power function β is **uniformly most powerful (UMP)** if the following holds: if β' is the power function of any other test in $\mathcal{C}_\alpha$ then $\beta(\theta) \geq \beta'(\theta)$ for all $\theta \in \Theta_1$.

Consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$. (Simple null and simple alternative.)

Theorem 5 *Let $L(\theta) = p(X_1, \ldots, X_n; \theta)$ and*

$$T_n = \frac{L(\theta_1)}{L(\theta_0)}.$$

Suppose we reject H_0 if $T_n > k$ where k is chosen so that

$$P_{\theta_0}(X^n \in R) = \alpha.$$

This test is a UMP level α test.

One nice thing about this is that it is a "general recipe" for doing a hypothesis test. The drawback of course is that it only applies to the restricted class of simple versus simple tests. The Neyman-Pearson test, despite its restricted applicability is a very important conceptual contribution. When it is applicable it is an optimal test. This is often called the Neyman-Pearson Lemma.

Proof of the Neyman-Pearson Lemma. Let us denote the test function of the NP test as ϕ_{NP} and the test function of any other test we want to compare against as ϕ_A. The test function simply takes the value 1 if the test rejects and 0 otherwise. To ease notation we will assume that $n = 1$. Let $f_0(x) = L(\theta_0; x)$ and $f_1(x) = L(\theta_1; x)$. So with this notation, we reject if:

$$\frac{f_1(x)}{f_0(x)} \geq k.$$

To prove the NP Lemma, we will first argue that the following is true:

$$\int_x \underbrace{(\phi_{NP}(x) - \phi_A(x))}_{U_1} \underbrace{(f_1(x) - kf_0(x))}_{U_2} dx \geq 0.$$

To see this we can just consider some cases:

1. If both tests reject or if both tests accept then the inequality is clearly true since the LHS is 0.
2. If NP rejects, and the test A accepts then $\phi_{NP}(x) = 1$, and $\phi_A(x) = 0$, so $U_1 \geq 0$. Since the NP test rejected the null we know that:
$$\frac{f_1(x)}{f_0(x)} \geq k,$$
so that $U_2 \geq 0$. So the inequality is true in this case.
3. If NP accepts and the test A rejects then both U_1 and U_2 are negative so the inequality is also true in this case.

So we can see that for every x, $U_1 \times U_2 \geq 0$ so it is true when we integrate over x. Now, we can rearrange this inequality to see that:

$$\begin{aligned}\int_x (\phi_{NP}(x) - \phi_A(x)) f_1(x) dx &\geq k \int_x (\phi_{NP}(x) - \phi_A(x)) f_0(x) dx \\ &= k \left(\underbrace{\int_x \phi_{NP}(x) f_0(x) dx}_{=\alpha} - \underbrace{\int_x \phi_A(x) f_0(x) dx}_{\leq \alpha} \right) \\ &\geq 0.\end{aligned}$$

This proves the NP lemma, i.e. that the power of the NP test is larger than the power of any other test. □

Now we develop some tests that are useful in other more complex settings.

7 The Wald Test

When we are testing a simple null hypothesis against a possibly composite alternative, the NP test is no longer applicable. A general method is the Wald test. We are interested in testing the hypotheses in a parametric model:

$$\begin{aligned} H_0 : &\quad \theta = \theta_0 \\ H_1 : &\quad \theta \neq \theta_0. \end{aligned}$$

The Wald test most generally is based on an asymptotically normal estimator, i.e. we suppose that we have access to an estimator $\widehat{\theta}$ which, under the null, satisfies the property that:

$$\frac{\widehat{\theta} - \theta_0}{se_0} \xrightarrow{d} N(0,1)$$

where $se_0 = \sqrt{\text{Var}}(\widehat{\theta})$ is the standard deviation of $\widehat{\theta}$ under the null. In this case, we could consider the statistic:

$$T_n = \frac{\widehat{\theta} - \theta_0}{se_0}$$

or, if se_0 is not known, we use

$$T_n = \frac{\widehat{\theta} - \theta_0}{\widehat{se_0}}.$$

Under the null $T_n \xrightarrow{d} N(0,1)$, so we simply reject the null if: $|T_n| \geq z_{\alpha/2}$. This controls the Type-I error only asymptotically (i.e. only if $n \to \infty$) but this is relatively standard in applications. That is

$$P_{\theta_0}(|T_n| \geq z_{\alpha/2}) \to \alpha.$$

It is also valid to use the statistic

$$T_n = \frac{\widehat{\theta} - \theta_0}{\widehat{se}}$$

where $\widehat{se}$ is any consistent estimate of the standard error; it's not necessary to assume H_0 is true when estimating the standard error. (This follows from Slutsky's theorem and the continuous mapping theorem.)

Example: Suppose that $X_1, \ldots, X_n \sim \text{Ber}(p)$, and the null is that $p = p_0$. Defining $\widehat{p} = \frac{1}{n}\sum_{i=1}^n X_i$. Let

$$T_n = \frac{\widehat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}},$$

which has an asymptotic $N(0,1)$ distribution. As mentioned above, we can also use

$$T_n = \frac{\widehat{p} - p_0}{\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}}.$$

Observe that this alternative test statistic also has an asymptotically standard normal distribution under the null. Its behaviour under the alternate is a bit more pleasant as we will see.

7.1 Power of the Wald Test

To get some idea of what happens under the alternate, suppose we are in some situation where the MLE has "standard asymptotics", i.e. $\sqrt{n}(\widehat{\theta} - \theta) \rightsquigarrow N(0, 1/(I_1(\theta)))$. Suppose that we use the statistic:

$$T_n = \sqrt{nI_1(\widehat{\theta})}(\widehat{\theta} - \theta_0),$$

and that the true value of the parameter is $\theta_1 \neq \theta_0$. Let us define:

$$\Delta = \sqrt{nI_1(\theta_1)}(\theta_0 - \theta_1),$$

then the probability that the Wald test rejects the null hypothesis is asymptotically:

$$1 - \Phi\left(\Delta + z_{\alpha/2}\right) + \Phi\left(\Delta - z_{\alpha/2}\right).$$

You will prove this on your HW (it is some simple re-arrangement, similar to what we have done previously when computing the power function in a Gaussian model). There are some aspects to notice:

1. If the difference between θ_0 and θ_1 is very small the power will tend to α, i.e. if $\Delta \approx 0$ then the test will have trivial power.

2. As $n \to \infty$ the two Φ terms will approach either 0 or 1, and so the power will approach 1.

3. As a rule of thumb the Wald test will have non-trivial power if $|\theta_0 - \theta_1| \gg \frac{1}{\sqrt{nI_1(\theta_1)}}$.

8 Likelihood Ratio Test (LRT)

To test composite versus composite hypotheses the general method is to use something called the (generalized) likelihood ratio test. We want to test:

$$\begin{aligned} H_0 : &\quad \theta \in \Theta_0 \\ H_1 : &\quad \theta \notin \Theta_0. \end{aligned}$$

This test is simple: reject H_0 if $\lambda(X_1, \ldots, X_n) \leq c$ where

$$\lambda(X_1, \ldots, X_n) = \frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\theta \in \Theta} L(\theta)} = \frac{L(\widehat{\theta}_0)}{L(\widehat{\theta})}$$

where $\widehat{\theta}_0$ maximizes $L(\theta)$ subject to $\theta \in \Theta_0$.

We can simplify the LRT by using an asymptotic approximation. This fact that the LRT generally has a simple asymptotic approximation is known as *Wilks' phenomenon.* First, some notation:

Notation: Let $W \sim \chi^2_p$. Define $\chi^2_{p,\alpha}$ by

$$P(W > \chi^2_{p,\alpha}) = \alpha.$$

We let $\ell(\theta)$ denote the log-likelihood in what follows.

Theorem 6 *Consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ where $\theta \in \mathbb{R}$. Under H_0,*

$$-2 \log \lambda(X_1, \ldots, X_n) \rightsquigarrow \chi^2_1.$$

Hence, if we let $T_n = -2 \log \lambda(X^n)$ then

$$P_{\theta_0}(T_n > \chi^2_{1,\alpha}) \to \alpha$$

as $n \to \infty$.

Proof: Using a Taylor expansion:

$$\ell(\theta) \approx \ell(\widehat{\theta}) + \ell'(\widehat{\theta})(\theta - \widehat{\theta}) + \ell''(\widehat{\theta}) \frac{(\theta - \widehat{\theta})^2}{2} = \ell(\widehat{\theta}) + \ell''(\widehat{\theta}) \frac{(\theta - \widehat{\theta})^2}{2}$$

and so

$$\begin{aligned} -2\log\lambda(x_1,\ldots,x_n) &= 2\ell(\widehat{\theta}) - 2\ell(\theta_0) \\ &\approx 2\ell(\widehat{\theta}) - 2\ell(\widehat{\theta}) - \ell''(\widehat{\theta})(\theta_0-\widehat{\theta})^2 = -\ell''(\widehat{\theta})(\theta_0-\widehat{\theta})^2 \\ &= \frac{-\frac{1}{n}\ell''(\widehat{\theta})}{I_1(\theta_0)}(\sqrt{nI_1(\theta_0)}(\widehat{\theta}-\theta_0))^2 = A_n \times B_n. \end{aligned}$$

Now $A_n \xrightarrow{p} 1$ by the WLLN and $\sqrt{B_n} \rightsquigarrow N(0,1)$. The result follows by Slutsky's theorem. ■

Example 7 *$X_1,\ldots,X_n \sim \text{Poisson}(\lambda)$. We want to test $H_0: \lambda = \lambda_0$ versus $H_1: \lambda \neq \lambda_0$. Then*

$$-2\log\lambda(x^n) = 2n[(\lambda_0 - \widehat{\lambda}) - \widehat{\lambda}\log(\lambda_0/\widehat{\lambda})].$$

We reject H_0 when $-2\log\lambda(x^n) > \chi^2_{1,\alpha}$.

Now suppose that $\theta = (\theta_1,\ldots,\theta_k)$. Suppose that $H_0: \theta \in \Theta_0$ fixes some of the parameters. Then, under conditions,

$$T_n = -2\log\lambda(X_1,\ldots,X_n) \rightsquigarrow \chi^2_\nu$$

where

$$\nu = \dim(\Theta) - \dim(\Theta_0).$$

Therefore, an asymptotic level α test is: reject H_0 when $T_n > \chi^2_{\nu,\alpha}$.

Example 8 *Consider a multinomial with $\theta = (p_1,\ldots,p_5)$. So*

$$L(\theta) = p_1^{y_1}\cdots p_5^{y_5}.$$

Suppose we want to test

$$H_0: p_1 = p_2 = p_3 \quad \text{and} \quad p_4 = p_5$$

versus the alternative that H_0 is false. In this case

$$\nu = 4 - 1 = 3.$$

The LRT test statistic is

$$\lambda(x_1,\ldots,x_n) = \frac{\prod_{j=1}^5 \widehat{p}_{0j}^{Y_j}}{\prod_{j=1}^5 \widehat{p}_j^{Y_j}}$$

where $\widehat{p}_j = Y_j/n$, $\widehat{p}_{01} = \widehat{p}_{02} = \widehat{p}_{03} = (Y_1+Y_2+Y_3)/n$, $\widehat{p}_{04} = \widehat{p}_{05} = (1-3\widehat{p}_{01})/2$. Now we reject H_0 if $-2\lambda(X_1,\ldots,X_n) > \chi^2_{3,\alpha}$. □

9 p-values

When we test at a given level α we will reject or not reject. It is useful to summarize what levels we would reject at and what levels we would not reject at.

The p-value is the smallest α at which we would reject H_0.

In other words, we reject at all $\alpha \geq p$. So, if the pvalue is 0.03, then we would reject at $\alpha = 0.05$ but not at $\alpha = 0.01$.

Hence, to test at level α, we reject when $p < \alpha$.

Theorem 9 *Suppose we have a test of the form: reject when $T(X_1, \ldots, X_n) > c$. Then the p-value is*

$$p = \sup_{\theta \in \Theta_0} P_\theta(T_n(X_1^*, \ldots, X_n^*) \geq T_n(x_1, \ldots, x_n))$$

where $x_1, \ldots, x_n$ are the observed data and $X_1^*, \ldots, X_n^* \sim p_{\theta_0}$.

Example 10 *$X_1, \ldots, X_n \sim N(\theta, 1)$. Test that $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. We reject when $|T_n|$ is large, where $T_n = \sqrt{n}(\overline{X}_n - \theta_0)$. Let t_n be the observed value of T_n. Let $Z \sim N(0, 1)$. Then,*

$$p = P_{\theta_0}\left(|\sqrt{n}(\overline{X}_n - \theta_0)| > t_n\right) = P(|Z| > t_n) = 2\Phi(-|t_n|).$$

The p-value is a random variable. Under some assumptions that you will see in your HW the p-value will be uniformly distributed on $[0, 1]$ under the null.

Important. Note that p is NOT equal to $\mathbb{P}(H_0|X_1, \ldots, X_n)$. The latter is a Bayesian quantity which we will discuss later.