Chapter 3: Hypothesis Testing

This chapter consists of two parts: Part I: Likelihood ratio test, test evaluation and UMP test, and Part II: Pearson's chi-square test (also known as Pearson's Goodness of fit test).

1 INTRODUCTION

Previously, we introduced how to estimate the unknown parameter θ of our interest. More precisely, we provided some ways to find a point estimator of θ. Then we also discussed how to compare different point estimators under the criterion of MSE, and to find the UMVUE of a function of θ in some cases.

However, in practice, the problem confronting the scientist or engineer may not be so much the estimation of θ, but rather the formation of a data-based decision procedure that can produce a conclusion about some scientific system.

For instance,

- a medical researcher may decide on the basis of experimental evidence whether coffee drinking increases the risk of cancer in humans;
- a sociologist may wish to collect appropriate data to enable him or her to decide whether a person's blood type and eye color are independent variables.
- an engineer may decide on the basis of sample data whether the true average lifetime of a certain kind of tire is at least 22,000 miles;
- an agronomist may want to decide on the basis of experiments whether one kind of fertilizer produces a higher yield of soybeans than another;
- a manufacturer of pharmaceutical products may decide on the basis of samples whether 90% of all patients given a new medication will recover from a certain disease.

In each of these cases, the scientist or engineer conjectures something about a system. In addition, each must involve the use of experimental data and decision making that is based on the data. Formally, in each case, **the conjecture can be put in the form of a statistical test of hypothesis.**

Definition: A statistical hypothesis is an assertion or conjecture about the distribution of the random variable of our interest. If a parametric distribution is considered, then a statistical hypothesis becomes a conjecture about the true value of the unknown parameter(s) of the parametric distribution.

For instance, **in the third case**, we might say that the engineer has to test the hypothesis that θ, the parameter of an exponential distribution, is at least 22,000; **in the fourth case,** we might say that the agronomist has to decide whether $\mu_1 > \mu_2$ or not, where μ_1 and μ_2 are the means of two normal distributions; and **in the last case,** we might say that the manufacturer has to decide whether or not θ, the parameter of a binomial population, equals 0.90. **Note that for these three cases, parametric distributions are used.**

1.1 Null and Alternative Hypotheses

In hypothesis testing, the hypothesis of our interest is related to a particular class of θ, say Θ_0, and its complement Θ_1, where $\Theta = \Theta_0 \cup \Theta_1$ is the parameter space of θ. For instance, in our third example above, we have $\Theta_0 = [22000, \infty)$ and $\Theta_1 = [0, 22000)$, and in our last example we have $\Theta_0 = \{0.9\}$ and $\Theta_1 = [0, 0.9) \cup (0.9, 1]$.

Correspondingly, the hypothesis with $\theta \in \Theta_0$ is a ***null hypothesis*** H_0, and the one with $\theta \in \Theta_1$ is referred as an ***alternative hypothesis*** H_1. In particular, if Θ_0 (or Θ_1) contains only one element, then the null (or alternative) hypothesis is said to be ***simple***; otherwise, the hypothesis is said to be ***composite***. For instance, a test with $\Theta_0 = \{0.1\}$ and $\Theta_1 = \{0.5\}$ is a simple-simple test, or a simple test.

Note that **we always put the "equal" sign in H_0**, i.e. $=, \geq, \leq$. For a simple hypothesis, we can specify the functional form of the underlying parametric distribution, but a composite hypothesis does not assign a specific value to the parameter θ, like the third of the above examples.

In hypothesis testing, what we really want to do is to decide whether or not the null hypothesis H_0 should be rejected.

1.2 MAIN CONCEPT OF HYPOTHESIS TESTING

The basic idea of hypothesis testing is a "CONTRADICTION" with the following three steps:

Step 1:

Determine H_0 and H_1.

Step 2:

Under H_0, define a rare event --- the event which happens with a very small probability in one experiment of getting n data.

Step 3:

Collect data. If data contradicts H_0, then we can say that H_0 is false and reject H_0, while if data do not contradict H_0, then we canNOT say that H_0 is true and accept it, but we can say that we do NOT reject H_0.

Example: We want to know **whether or not a coin is fair**.

Consider a random experiment of flipping the coin, say 10 times.

Step 1: (H_0) The coin is fair, i.e. $P(\{H\})=P(\{T\})$ = 1/2, and (H_1) The coin is NOT fair.

Step 2: Under H_0, i.e. the coin is assumed to be fair, the probability of getting 10 tails in **ONE** experiment is $(1/2)^{10} \approx 0.00098$. So, we can define the event of getting 10 tails to be the rare event under H_0.

Step 3: Perform the experiment to **collect data**, i.e. we **now flip the coin 10 times**. If we get 10 tails, then (by the statistical belief that there is a highest probability of getting the collected data) the collected data tell us that getting 10 tails is NOT a rare event, i.e. they contradict H_0. Thus, we have evidence to suspect the reliability of H_0 and thus reject H_0 (= accept H_1).

In hypothesis testing, we only use **the data** to **see if there is enough evidence to reject** H_0.

- If we have enough evidence to reject H_0, we can have great confidence that H_0 is false and H_1 is true.
- However, if **we do not have enough evidence to reject** H_0, then *it* ***does not mean that*** *we have great confidence in the truth of* H_0. In this case, we should say **"do not reject** H_0**", instead of "accept** H_0**".**

1.3 TEST ERRORS AND ERROR PROBABILITIES

After we decide the null and alternative hypotheses, our next step is to determine a test statistic, i.e. the point estimator, to construct a test (procedure) of rejecting or not rejecting the null hypothesis H_0. Then, we draw a sample and compute the value of a test statistic to tell us what action we should take. Correspondingly, according to the possible value of the test statistic, we could define two complementary regions:

a ***rejection region*** C_1 of H_0: the set of random samples we would reject H_0, and

an ***non-rejection region*** C_0 of H_0: the set of random samples we would NOT reject H_0.

Note that there is no perfect test statement because of the randomness of the sample of data. Each test would lead to the following two kinds of errors.

	Not reject H_0	Reject H_0
If H_0 is true	No error	**TYPE I ERROR**
If H_0 is false	**TYPE II ERROR**	No error

Type I error: the error of rejecting H_0 when it is in fact true.
Type II error: the error of not rejecting H_0 when it is in fact false.

Correspondingly, we define

$$\gamma(\theta) = P(Type\ I\ error) \text{ for } \theta \in \Theta_0, \text{ and } \beta(\theta) = P(Type\ II\ error) \text{ for } \theta \in \Theta_1.$$

Ideally, we want to formulate a test such that these two error probabilities can be minimized. However, in general, we cannot control them simultaneously (when the sample size n is fixed). The following example illustrates this problem.

Example: Suppose we knew that the light bulbs produced from a standard manufacturing process have life times distributed as normal with a standard deviation σ = 300 hours. However, we did not know the mean lifetime θ. For simplicity, assume that we were sure that the mean lifetime should be either 1200 or 1240.

Then we may set up the following simple test:

$$\begin{cases} H_0: \theta = 1200 \\ H_1: \theta = 1240 \end{cases}$$

Suppose that we draw a sample of 100 light bulbs and measure their lifetimes. The sample mean $\bar{X}$ is used to estimate the true population mean θ.

Since the **hypothesized value in H_1 is larger than the hypothesized value in H_0**, intuitively, we can say that a **large value of $\bar{x}$ will lead to the rejection of H_0**, or we can set the test statement

Reject H_0 if $\bar{x} > c$.

We would later discuss how to determine the constant c, which is often called ***a critical value*** in practice. Note that in this example, $\Theta_0 = \{1200\}, \Theta_1 = \{1240\}$, and $\Theta = \{1200, 1240\}$, and the rejection region of H_0 is

$$C_1 = \{x_1, \dots, x_n : \bar{x} > c\}.$$

Thus, we have

$$\gamma = \gamma(1200) = P(reject\ H_0\ if\ H_0\ is\ true) = P(\bar{X} > c\ if\ \theta = 1200)$$

and

$$\beta = \beta(1240) = P(Not\ reject\ H_0\ if\ H_0\ is\ false) = P(\bar{X} \le c\ if\ \theta = 1240).$$

Note that (**Step 2)** under H_0 (i.e. when $\theta = 1200$), the event $\{\bar{X} > c\}$ occurs with a very small probability $\gamma(1200)$. Thus, $\{\bar{X} > c\}$ is a rare event under H_0. So, in **ONE** experiment of getting n data, we should **NOT get $\bar{x} > c$ IF H_0 is TRUE**.

(Step 3) In other words, getting $\bar{x} > c$ in one experiment would contradict H_0, and then we would reject H_0.

It follows that

$$z_\gamma + z_\beta = \frac{4}{3},$$

where z_k satisfies $P(Z > z_k) = k$ and $Z \sim N(0, 1)$. This result indicates **a trade-off between the two types of error: making γ smaller will lead to a larger β, and vice versa.**

1.4 SIGNIFICANCE LEVEL AND CRITICAL VALUE

Since we cannot control $\gamma(\theta)$ and $\beta(\theta)$ at the same time (when n is fixed.), it is conventional practice to assign an upper bound to the Type I error probability $\gamma(\theta)$ over Θ_0 and find a test with $\beta(\theta)$ as small as possible.

In general, we want to find a test (by specifying C_1) which minimizes

$$\beta(\theta) = P(\{\boldsymbol{X} \in C_0\} \mid \theta), \quad \forall\theta \in \Theta_1 ,$$

subject to

$$\sup_{\theta\in\Theta_0} \gamma(\theta) = \alpha \in (0, 1),$$

where α is called ***the significance level***, i.e. it is the largest value of the Type I error probability over Θ_0. In practice, α is user-specified and it should be taken to be small.

Remarks that (1) $\sup\limits_{\theta\in\Theta_0}$ means the supremum over Θ_0.

Definitions. For a nonempty subset S of $\mathbb{R}$, S is *bounded above* iff there is some $M \in \mathbb{R}$ such that $x \leq M$ for all $x \in S$. Such an M is called an *upper bound* of S. The *supremum* or *least upper bound* (denoted by $\sup S$ or lub S) of S is an upper bound $\tilde{M}$ of S such that $\tilde{M} \leq M$ for all upper bounds M of S.

Examples. (1) For $S = \left\{\frac{1}{n} : n \in \mathbb{N}\right\} = \left\{1, \frac{1}{2}, \frac{1}{3}, \ldots\right\}$, the upper bounds of S are all $M \geq 1$. So $\sup S = 1 \in S$.
(2) For $S = \{x \in \mathbb{R} : x < 0\}$, the upper bounds of S are all $M \geq 0$. So $\sup S = 0 \notin S$.

(Details can be found in MATH2033.) Note that the maximum of S is the largest element of S, so it must be in S, while the supremum of S is just the least upper bound of S, so it is NOT necessary in S. In case 1, 1 is also the maximum of S, but in case 2, 0 is not the maximum of S because 0 is not in S.

(2) When H_0 is simple, say $\Theta_0 = \{\theta_0\}$, we have $\gamma(\theta_0) = \alpha$.

How do we formulate a test with a given α?

With reference to the previous example, we now want to test

$$\begin{cases} H_0: \theta = 1200 \\ H_1: \theta = 1240 \end{cases}$$

at a significance level $\alpha = 0.05$.

Since

$$0.05 = \alpha = \sup_{\theta \in \Theta_0 = \{1200\}} \gamma(\theta) = \gamma(1200) = P(\bar{X} > c \mid \theta = 1200),$$

we have the critical value

$$c = 1249.35$$

($z_{0.05} = 1.645$ is used). Thus, the rejection region of H_0 is

$$C_1 = \{x_1, \dots, x_n : \bar{x} > 1249.35\}.$$

and the test is

Reject H_0 at a significance level of 0.05 if $\bar{x} > 1249.35$.

Suppose that if the observed value of the sample mean is $\bar{x} = 1237$, then we could conclude that we DO NOT HAVE ENOUGH EVIDENCE TO REJECT H_0 at a level $\alpha = 0.05$.

2 LIKELIHOOD RATIO TEST

In Chapter 2 for parameter estimation, we studied two general methods of finding a point estimator of θ. Here we would also study a general method --- ***likelihood ratio test*** --- to construct a test/ a rejection region for a more general hypothesis about θ.

Definition: A likelihood ratio test (LRT) for tesing $H_0: \theta \in \Theta_0$ against $H_1: \theta \in \Theta_1$ at a significance level of α is a test with a rejection region

$$C_1 = \{x_1, \dots x_n : \lambda(\boldsymbol{x}) \leq k\},$$

where $k \in (0, 1)$ satisfies $\sup_{\theta \in \Theta_0} P(\lambda(\boldsymbol{X}) \leq k \mid \theta) = \alpha$, and the LRT statistic

$$\lambda(\boldsymbol{x}) = \frac{L(\hat{\theta}_0)}{L(\hat{\theta})}$$

with an MLE $\hat{\theta}_0$ of θ over Θ_0 and an MLE $\hat{\theta}$ over $\Theta = \Theta_0 \cup \Theta_1$.

Remark that (1) $0 < \lambda(\boldsymbol{x}) \leq 1$. (2) The value of $\lambda(\boldsymbol{x})$ close to 0 suggests that the data are NOT compatible with H_0, so H_0 should be rejected. (3) If the hypothesis is simple, then the hypothesized value of θ instead of the MLE should be used.

Example: Consider a rs $\{X_i: i = 1, \dots, n\}$ of size n from $Exp(\theta)$.

Construct a LRT for testing $\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta > \theta_0 \end{cases}$ at a significance level of α, where θ_0 is known and positive.

Note that $\Theta_0 = \{\theta_0\}$, $\Theta_1 = (\theta_0, \infty)$ and $\Theta = [\theta_0, \infty)$. That is, the parameter space is restricted!!

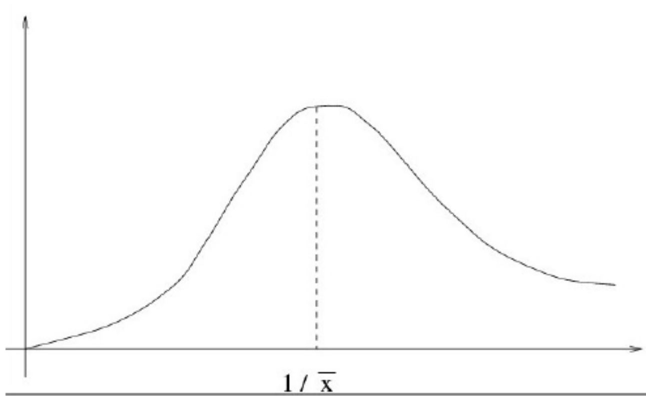
The likelihood function is

$$L(\theta) = \prod_{i=1}^{n} f(x_i|\theta) = \theta^n e^{-\theta \sum x_i}.$$

Thus, the log-likelihood function is

$$\begin{aligned} \log\, L(\theta) &= n \log\, \theta - \theta \sum x_i \\ \Longrightarrow \qquad \frac{\partial}{\partial \theta} \log\, L(\theta) &= \frac{n}{\theta} - n\bar{x} \\ \Longrightarrow \qquad \text{the critical point is } \hat{\theta} &= \frac{1}{\bar{x}} \end{aligned}$$

Note that it is a unique critical point, and the second order derivative of the log-likelihood at this point is negative. So, it is a maximum point.



The MLE of θ is

$$\begin{cases} \dfrac{1}{\bar{x}} \; if \; \dfrac{1}{\bar{x}} > \theta_0 \\ \theta_0 \; if \; \dfrac{1}{\bar{x}} \le \theta_0 \end{cases}$$

The likelihood ratio test statistic is

$$\lambda(\mathrm{x}) = \begin{cases} \dfrac{\theta_0^n e^{-n\theta_0 \bar{x}}}{(\bar{x})^{-n} e^{-n}}, & \text{if } \dfrac{1}{\bar{x}} > \theta_0 \\ 1, & \text{if } \dfrac{1}{\bar{x}} \leq \theta_0 \end{cases}$$

$$= \begin{cases} (\theta_0 \bar{x})^n e^{-n(\theta_0 \bar{x} - 1)}, & \text{if } \dfrac{1}{\bar{x}} > \theta_0 \\ 1, & \text{if } \dfrac{1}{\bar{x}} \leq \theta_0 \end{cases}$$

Thus, reject H_0 if

$$\frac{1}{\bar{x}} > \theta_0 \text{ AND } (\theta_0 \bar{x})^n e^{-n(\theta_0 \bar{x} - 1)} \leq k.$$

However, it is hard to use the above result to determine the value of k. So, what should we do?

If the term $(\theta_0 \bar{x})^n e^{-n(\theta_0 \bar{x} - 1)}$ is a monotonic function of some quantity, say y, where the distribution of Y (the random counterpart of y) can be easily determined, then the test based on y will be equivalent to the original test.

In this case, we let $y = \theta_0 \bar{x}$. Note that $y^n e^{-n(y-1)}$ has its maximum at $y = 1$. Hence, we have

$$\frac{1}{\bar{x}} > \theta_0 \text{ AND } (\theta_0 \bar{x})^n e^{-n(\theta_0 \bar{x} - 1)} \leq k$$

$$\Updownarrow$$

$$y < 1 \text{ AND } y^n e^{-n(y-1)} \leq k$$

$$\Updownarrow$$

$$y \leq K \in (0, 1)$$

$$\Updownarrow$$

$$\sum_{i=1}^{n} x_i \leq \frac{nK}{\theta_0} = K'$$

Hence, the LRT reduces to "reject H_0 at a significance level of α when $\sum_{i=1}^{n} x_i \leq K'$ ", where $\sum_{i=1}^{n} X_i \sim Gamma(n, \theta)$ and K' can be determined by $P(\sum_{i=1}^{n} X_i \leq K' | \theta_0) = \alpha$, i.e. K' is the α^{th} quantile of Gamma(n, θ_0).

Example: Consider a rs $\{X_i : i = 1, \ldots, n\}$ of size n from $N(\theta, \sigma^2)$, where θ and σ^2 are unknown. Construct a LRT for testing $\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases}$ at a significance level of α.

Note that

$$\Theta_0 = \{(\theta, \sigma^2) : \theta = \theta_0, \sigma^2 \in R^+\}$$

$$\Theta = \{(\theta, \sigma^2) : \theta \in R, \sigma^2 \in R^+\}$$

and the likelihood function is

$$L(\theta, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \theta)^2\}$$

We have found in Chapter 2 that over Θ, the MLEs of θ and σ^2 are $\hat{\theta} = \bar{x}$ and $\hat{\sigma}^2 = s_n^2$, respectively. Over Θ_0, $\theta = \theta_0$ is specified, so we only have to estimate σ^2 based on θ_0. It follows that the MLE $\hat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \theta_0)^2$.

Therefore,

$$\lambda(\boldsymbol{x}) = \left[1 + \frac{n(\bar{x} - \theta_0)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]^{-n/2}.$$

2.1 LARGE-SAMPLE TESTS

Except for a few cases, in general, the distribution of the LRT statistic $\lambda(x)$ cannot be found easily, but we can simplify it by using asymptotic results.

In this section, we would study three large-sample tests that can be shown to be equivalent asymptotically. Let us suppose that we wish to test the null hypothesis $H_0\colon \theta = \theta_0$ against the alternative hypothesis $H_1\colon \theta \neq \theta_0$ (i.e. two-sided test) at a significance level of α.

Recall that according to the asymptotic properties of the MLE $\hat{\theta}_n(\boldsymbol{X})$, we have known that

$$\sqrt{n}\, I_X^{1/2}(\theta)\left(\hat{\theta}_n(\boldsymbol{X}) - \theta\right) \xrightarrow{d} N(0,1).$$

Note that $I_X(\theta)$ cannot be easily determined. Thus, we replace it by an observed Fisher Information $-\frac{1}{n} l''\left(\hat{\theta}_n\right)$, where $-\frac{1}{n} l''(\hat{\theta}_n) = -\frac{1}{n}\sum_{i=1}^{n} \frac{d^2}{d\theta^2} \ln f_X\left(X_i \middle| \hat{\theta}_n\right)$. Thus, under $H_0\colon \theta = \theta_0$, we have

$$T_1 = \sqrt{-l''(\hat{\theta}_n)}\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right) \xrightarrow{d} N(0,1).$$

Wald test, Score test, and LRT test

[One-parameter cases]

1. The **Wald test** statistic is defined by

$$X_W = T_1^2 = -l''(\hat{\theta}_n)(\hat{\theta}_n(\boldsymbol{X}) - \theta_0)^2.$$

2. The **Score test** statistic is defined by

$$X_S = [-l''(\theta_0)]^{-1}\left(l'(\theta_0)\right)^2.$$

3. The (large-sample) **likelihood ratio test** statistic is defined by

$$X_L = -2\ln\lambda(\boldsymbol{X}) = 2\left[\, l\left(\hat{\theta}_n(\boldsymbol{X})\right) - l(\theta_0)\,\right].$$

Theorem 1: Under H_0, these three test statistics are asymptotically equivalent, and follow an asymptotic chi-square distribution with 1 degree of freedom. Thus, we reject H_0 at a significance level of α when the actual value of the above test statistic is greater than $\chi_\alpha^2(1)$, the $(1-\alpha)^{th}$ quantile of chi-square distribution with 1 degree of freedom.

Remark that the above result is for two-sided tests only. For one-sided test, since the parameter space is restricted to $\theta \geq \theta_0$ (or $\theta \leq \theta_0$), the result for the likelihood ratio test has to be adjusted. We omit it here because of the complexity. **However,** the issue on one-sided tests is not a problem for Wald test and Score test, because

(i) for the Wald test, we have

$$T_1 = \sqrt{-l''(\hat{\theta}_n)}\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right) \xrightarrow{d} N(0, 1).$$

It follows that for $H_1: \theta > \theta_0$ $(or\ < \theta_0)$, we reject H_0 at a significance level of α if the actual value of $T_1 > z_\alpha$ $(or < z_\alpha)$.

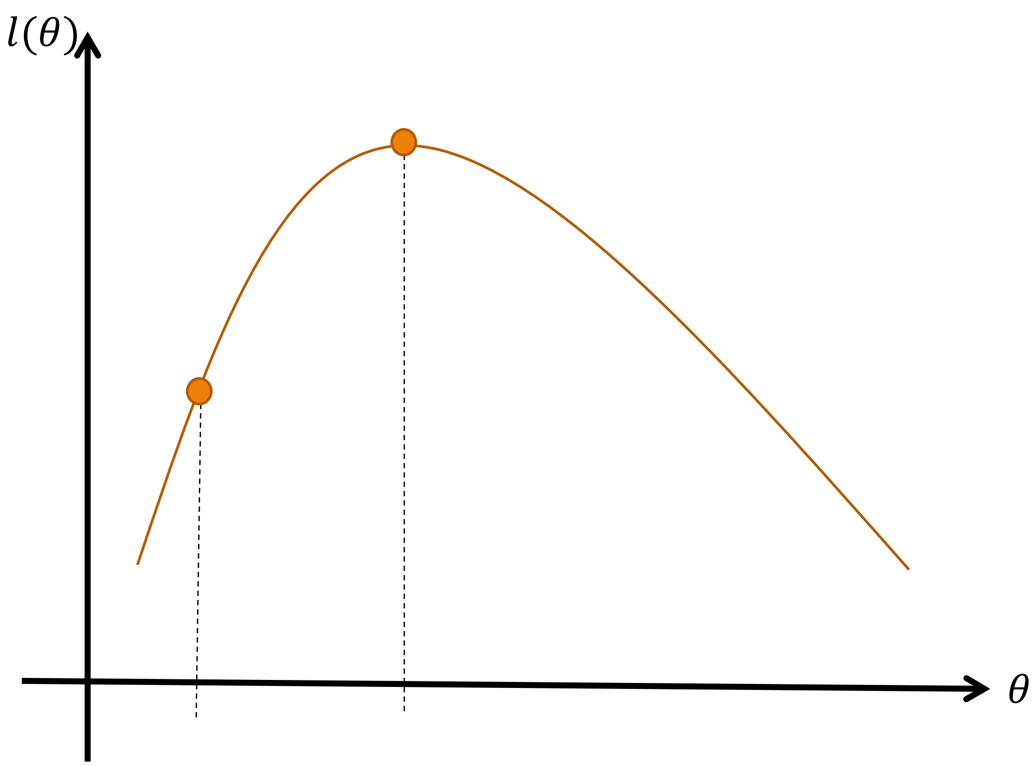
(ii) for the Score test, we have

$$T_2 = [-l''(\theta_0)]^{-1/2}\, l'(\theta_0) = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{d}{d\theta} \ln f_X(X_i|\theta_0)}{\sqrt{-\frac{1}{n} l''(\theta_0)}} \xrightarrow{d} N(0, 1)$$

It follows that for $H_1: \theta > \theta_0$ $(or\ < \theta_0)$, we reject H_0 at a significance level of α if the actual value of $T_2 > z_\alpha$ $(or < z_\alpha)$.

Proof of Theorem 1:

We plot a graph of the log-likelihood function against θ for a particular set of observations.



Example: Suppose the following data are from an exponential distribution:

1, 3, 5, 8, 10, 15, 18, 19, 22, 25

Do a likelihood ratio test with

$$H_0: \lambda = 0.06 \;\; VS \;\; H_1: \lambda \neq 0.06$$

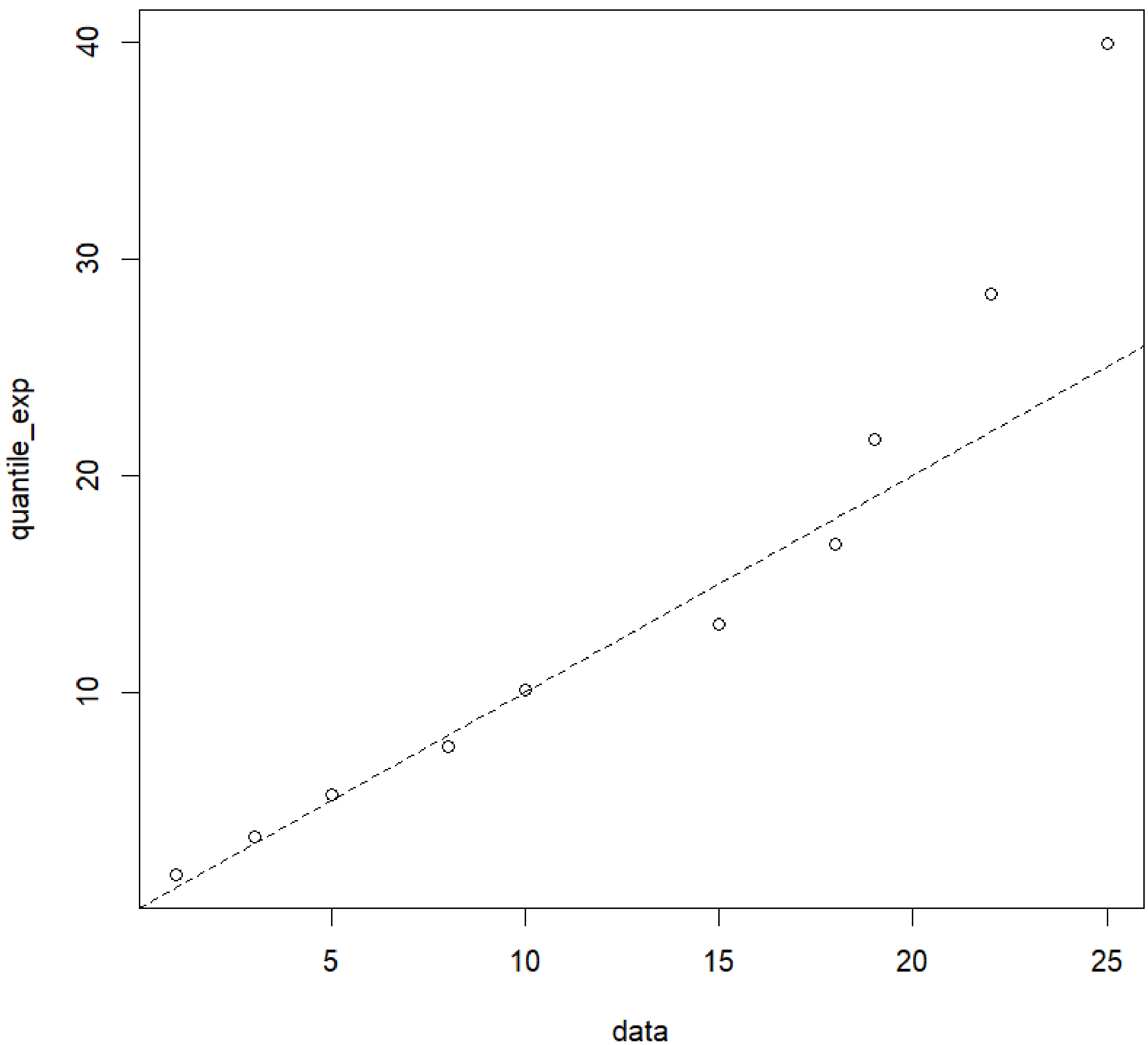
and draw a conclusion at a significance level of 0.05. Given that

```
> qchisq(0.95, df = 1)
[1] 3.841459
```

R corner

Check with Q-Q plot to see if data are from $Exp(\lambda = 0.06) = Gamma(\alpha = 1, \beta = 0.06)$.

```
> data = c(1, 3, 5, 8, 10, 15, 18, 19, 22, 25)
>
> quantile_exp = qgamma((1:10)/11, shape = 1, rate = 0.06)
>
> plot(data, quantile_exp)
> abline(a = 0, b = 1, lty = 2)
```



Recall that we have used the R package ***maxLik*** to maximize (log) $L(\theta)$ to find MLE before.

Now, we can use it to calculate $l\left(\hat{\theta}_n(\boldsymbol{X})\right)$ and $l(\theta_0)$.

```
> # install.packages("maxLik")
> library("maxLik")
>
> n= length(data)
>
> llik = function(par)
+ {
+ lambda = par
+
+ ll = n*log(lambda) - lambda*sum(data)
+
+ return(ll)
+ }
>
> # To obtain l(MLE)
> l_MLE = maxLik(logLik = llik, start = c(lambda = 0.05), method = "NR")
> summary(l_MLE)
--------------------------------------------
Maximum Likelihood estimation
Newton-Raphson maximisation, 5 iterations
Return code 1: gradient close to zero
Log-Likelihood: -35.33697
1  free parameters
Estimates:
       Estimate Std. error t value Pr(> t)
lambda  0.07937    0.02510   3.162 0.00157 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
--------------------------------------------
>
> # To obtain l(lambda_0)
> llik(0.06)
[1] -35.69411
```

[Multi-parameter cases]

Suppose that $\theta = \begin{pmatrix}\theta_r \\ \theta_q\end{pmatrix} \in \Theta \subset R^k$, where $\theta_r \in R^r$ and $\theta_q \in R^q$ with $r + q = k \geq 2$, $1 \leq r \leq k$, and $q \geq 0$. Consider $H_0: \theta_r = \theta_{r0}$ against the alternative hypothesis $H_1: \theta_r \neq \theta_{r0}$ at a significance level of α. Note that θ_q is called a nuisance parameter --- the parameter which is not of interest, but it is in θ.

Since the forms of **Wald test and Score test are a bit complicated,** we only consider the **likelihood ratio test** statistic defined by

$$X_L = 2\left[l(\hat{\theta}) - l(\theta_{r0}, \tilde{\theta}_{q0})\right],$$

where $\tilde{\theta}_{q0}$ is the MLE of θ_q under $\theta_r = \theta_{r0}$, and $\hat{\theta}$ is the MLE of θ in Θ.

Theorem 2: Under H_0, X_L follows an asymptotic chi-square distribution with r degrees of freedom.

Proof: Omitted.

EXAMPLE: Suppose the following data are from $N(\mu, \sigma^2)$:

-10.85, -8.81, -7.94, -3.40, 3.52, 4.31, 9.31, 9.61, 16.19, 24.11

Do a likelihood ratio test with

$$H_0: \mu = 3.5 \;\; VS \;\; H_1: \mu \neq 3.5$$

and draw a conclusion at a significance level of 0.05. Given that

```
> qchisq(0.95, df = 1)
[1] 3.841459
```

```
> data = c(-10.85,  -8.81,  -7.94,  -3.40,   3.52,   4.31,   9.31,   9.61,  16.19,  24.11)
>
> n = length(data)
>
> llik = function(par)
+ {
+ mu = par[1]
+ sigma = par[2]
+
+ ll = -0.5*n*log(2*pi) - n*log(sigma) - sum(0.5*(data-mu)^2/sigma^2)
+
+ return(ll)
+ }
>
> # To obtain l(MLE)
> l_MLE = maxLik(logLik = llik, start = c(mu = 2, sigma = 4), method = "NR")
> summary(l_MLE)
--------------------------------------------
Maximum Likelihood estimation
Newton-Raphson maximisation, 8 iterations
Return code 2: successive function values within tolerance limit
Log-Likelihood: -38.0992
2  free parameters
Estimates:
      Estimate Std. error t value  Pr(> t)
mu       3.605      3.295   1.094    0.274
sigma   10.924      2.376   4.597 4.28e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
--------------------------------------------
>
> # To obtain l(sigma2)
> l_sigma2 = maxLik(logLik = llik, start = c(mu = 3.5, sigma = 4), fixed = "mu", method = "NR")
> summary(l_sigma2)
--------------------------------------------
Maximum Likelihood estimation
Newton-Raphson maximisation, 8 iterations
Return code 2: successive function values within tolerance limit
Log-Likelihood: -38.09967
1  free parameters
Estimates:
      Estimate Std. error t value Pr(> t)
mu       3.500      0.000      NA      NA
sigma   10.925      2.474   4.416   1e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
--------------------------------------------
```

(Optional) More general cases with multiple-parameter cases

1.2.4 Test of a composite hypothesis

Under the same setup as in Section 1.2.1, let the hypothesis to be tested be $H_0 : h(\theta) = c$, where h is an $r \times 1$ vector function of the p-vector θ with $p \geq r$ and c is a given r-vector of constants. The corresponding *Holy Trinity* is as follows:

1. Likelihood ratio test [Neyman and Pearson (1928)]

$$LR = 2\left[l(\hat{\theta}|X) - l(\tilde{\theta}|X)\right] \tag{1.13}$$

where $\tilde{\theta}$ is the ml of θ under the restriction $h(\theta) = c$.

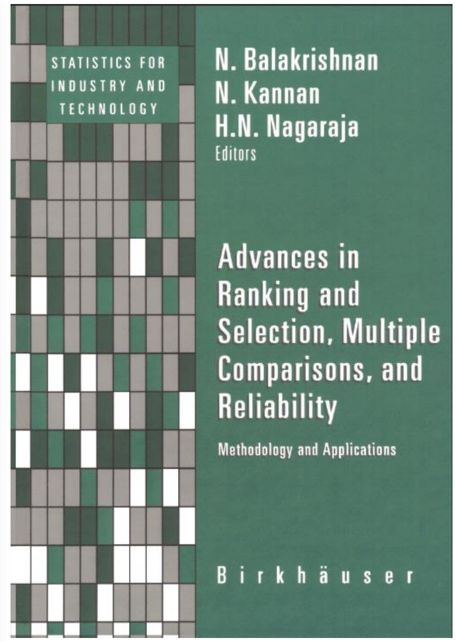
2. Wald test [Wald (1943)]

$$W = \left[h(\hat{\theta}) - c\right]' \left[A(\hat{\theta})\right]^{-1} \left[h(\hat{\theta}) - c\right] \tag{1.14}$$

where

$$A(\theta) = \{H(\theta)][I(\theta)]^{-1}[H(\theta)]',$$
$$H(\theta) = (\partial h_i(\theta)/\partial \theta_j)\,, h(\theta) = (h_1(\theta), \ldots, h_r(\theta))'\,,$$

and $I(\theta)$ is as defined in (1.2).



The Fisher information matrix of order $p \times p$ is defined by

$$ni(\theta) = I(\theta) = E[s(\theta)s'(\theta)] = (i_{rs}(\theta)) \tag{1.2}$$

where $i_{rs}(\theta) = E\left[s_r(\theta)s_s(\theta)\right]$. The maximum likelihood estimate of θ is obtained as a solution of the p equations

$$s_i(\theta) = 0, \; i = 1, \ldots, p \tag{1.3}$$

which we represent by $\hat{\theta}$. Under suitable regularity conditions [Lehmann (1999, pp. 499–501)], using the multivariate cental limit theorem

$$n^{-1/2}s(\theta_0) \sim N_p\left(0, i(\theta_0)\right) \tag{1.4}$$

3. Rao Score test [Rao (1948)]

$$RS = [s(\tilde{\theta})]'[I(\tilde{\theta})]^{-1}[s(\tilde{\theta})]. \tag{1.15}$$

All the three statistics have an asymptotic chi-square distribution on r degrees of freedom.

An alternative way of expressing the RS statistic is as follows. Note that $\tilde{\theta}$, the restricted ml of θ, is a solution of the equation

$$s(\theta) + [H(\theta)]'\lambda = 0,\ h(\theta) = c$$

where λ is an r-vector of the Lagrangian Multiplier so that $[s(\tilde{\theta})]' = -\lambda' H(\tilde{\theta})$. Substituting in (1.15) we have

$$\mathrm{RS} = \lambda' H(\tilde{\theta})[I(\tilde{\theta})]^{-1}[H(\tilde{\theta})]'\lambda = \lambda'[A(\tilde{\theta})]\lambda \tag{1.16}$$

where $A(\theta)$ is as defined in (1.14). Silvey (1959) expressed the RS statistic (1.15) in the form (1.16) and called it the Lagrangian Multiplier (LM) test. (In econometric literature, the RS test is generally referred to as the LM test.)

1.2.5 Special form of composite hypothesis

In many problems, the p-vector parameter θ consists of two parts, θ_1 an r vector and θ_2 a $(p-r)$ vector and the null hypothesis is of the form $H_0 : \theta_1 = \theta_{10}$ (a specified vector) and θ_2 (known as a nuisance parameter) is arbitrary. This becomes a special case of the composite hypothesis considered in Subsection 1.2.4 if we take $h(\theta) = \theta_1$. Denote the unrestricted ml of (θ_1, θ_2) by $(\hat{\theta}_1, \hat{\theta}_2)$ and its asymptotic covariance matrix by

$$\begin{aligned} cov(\hat{\theta}, \hat{\theta}) &= [I(\theta)]^{-1} \\ &= \begin{pmatrix} I_{11}(\theta) & I_{12}(\theta) \\ I_{21}(\theta) & I_{22}(\theta) \end{pmatrix}^{-1} = \begin{pmatrix} A & B \\ B' & C \end{pmatrix} \end{aligned}$$

where the partitions of the information matrix, I_{11}, I_{12}, and I_{22} are matrices of orders $r \times r$, $r \times (p-r)$ and $(p-r) \times (p-r)$, respectively, The Wald statistic can be written as

$$\begin{aligned} W &= (\hat{\theta}_1 - \theta_{10})'\hat{A}^{-1}(\hat{\theta}_1 - \theta_{10}),\ \hat{A} = A(\hat{\theta}) \\ &= (\hat{\theta}_1 - \theta_{10})' I_{1.2}(\hat{\theta})(\hat{\theta}_1 - \hat{\theta}_{10}) \end{aligned} \tag{1.17}$$

where

$$I_{1.2} = I_{11} - I_{12} I_{22}^{-1} I_{21}$$

the Schur complement of I_{22}.

To compute LR and RS statistics, we need to find the restricted ml estimates of θ_1, θ_2 under the restriction $\theta_1 = \theta_{10}$. Using the Lagrangian multiplier we have to maximize

$$L(\theta|x) - \lambda(\theta_1 - \theta_{10})$$

with respect to θ. The estimating equations are

$$s_1(\tilde{\theta}) = \lambda,\ s_2(\tilde{\theta}) = 0,\ \tilde{\theta}_1 = \theta_{10}.$$

The Rao score statistic is

$$\begin{aligned} \text{RS} &= [s_1(\tilde{\theta})', 0']'[I(\tilde{\theta})]^{-1}[s_1(\tilde{\theta})', 0'] \\ &= [s_1(\tilde{\theta})]'[I_{1.2}(\tilde{\theta})]^{-1}[s_1(\tilde{\theta})] \\ &= \lambda'[I_{1.2}(\tilde{\theta})]^{-1}\lambda. \end{aligned} \tag{1.18}$$

The LR statistic is

$$\text{LR} = 2\left[L(\hat{\theta}) - L(\tilde{\theta})\right] \tag{1.19}$$

All the three statistics have asymptotically chi-square distribution on r d.f.

3 POWER FUNCTION AND POWER OF A TEST STATEMENT

Analogous to the fact that in parameter estimation we have many point estimators to estimate an unknown parameter, in hypothesis testing, we can also use different test statistics (i.e. point estimators) to construct tests for testing H0 against H1. Then, which one of them is the best? We first need a quantity (like unbiasedness or MSE) such that we can compare different tests.

3.1 POWER FUNCTION

Let $\boldsymbol{X} = \{X_i : i = 1, \dots, n\}$. For a test, we define its ***power function*** Q mapping from Θ to [0,1] such that for all $\theta \in \Theta$,

$$Q(\theta) = \begin{cases} \displaystyle\sum_{\boldsymbol{x} \in C_1} p_{\boldsymbol{X}}(\boldsymbol{x}|\theta) & (discrete) \\ \\ \displaystyle\int_{C_1} f_{\boldsymbol{X}}(\boldsymbol{x}|\theta)\, dx & (continuous) \end{cases}.$$

Indeed, the power function of a test is the probability of rejecting H0. In particular, for $\theta \in \Theta_1$, $Q(\theta) = 1 - \beta(\theta)$ is called ***the power of the test at*** θ ***--- the probability of rejecting H0 at*** $\theta \in \Theta_1$ ***(i.e. H0 is false).***

Example: The average length of time for students to register for classes at a certain college has been 46 minutes. A new registration procedure using modern computing machines is being tried. If a random sample of 12 students had an average registration time of 42 minutes with a standard deviation of 11.9 minutes under the new system. Test the hypothesis that the population mean length θ of time under the new system is now less than 46. Use a 0.05 level of significance with the assumption that the data are from a normal distribution.

Given that the (one-sided left) test is to

$$\text{reject H0 at 0.05 level of significance if } \bar{x} < 46 - t_{11,\,0.05}\frac{11.9}{\sqrt{12}} = 39.83072..$$

Thus, the power of this test at $\theta \in \Theta_1$, i.e. $\theta < 46$, is

$$Q(\theta) = P\left(T_{n-1} < \frac{39.83072 - \theta}{11.9/\sqrt{12}}\right),$$

where $T_{n-1} \sim t(n-1)$. Note that the power is a decreasing function of θ in Θ_1.

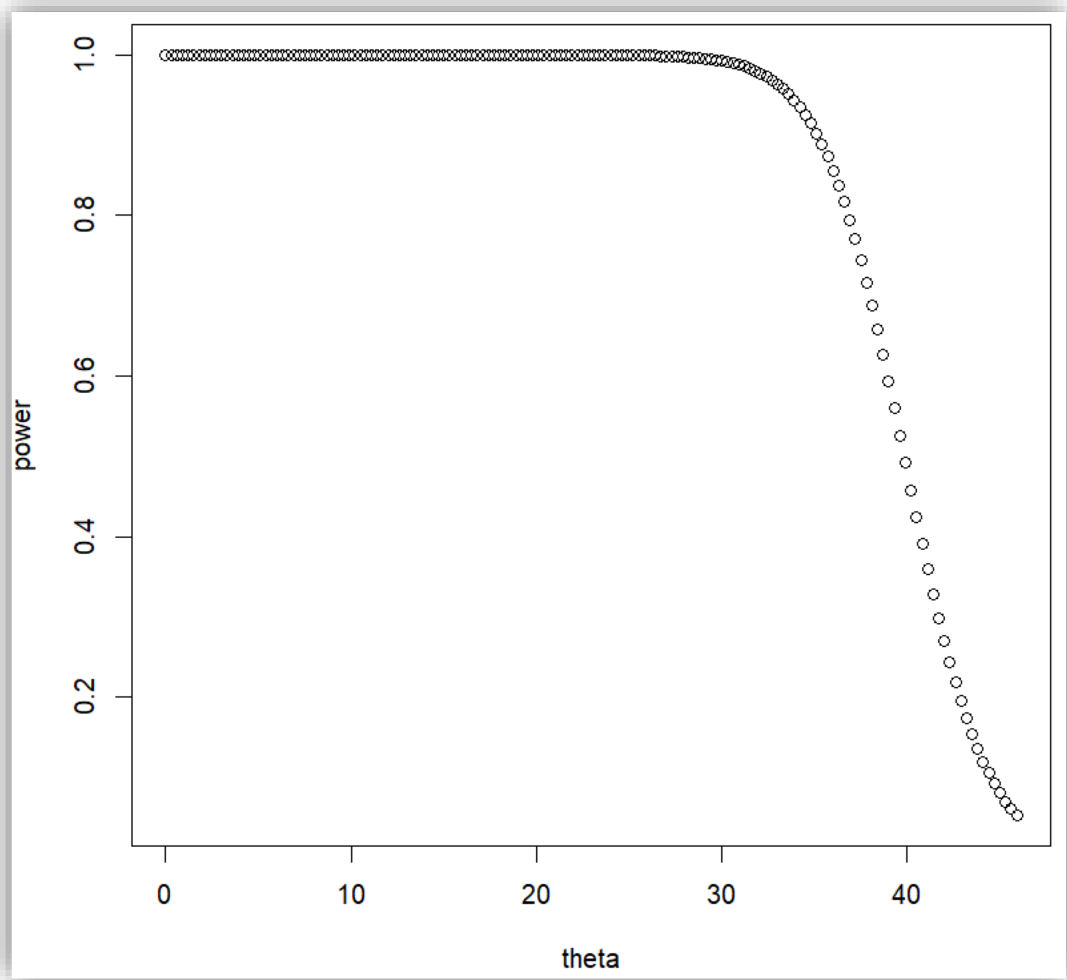
R corner:

```
theta = seq(0, 45.9, by = 0.3)

t = (39.83072 - theta)/(11.9/sqrt(12))

power = pt(t, df = 11)

plot(theta, power)
```



Note that, *in term of the power function*, our goal now is to find a test for which the value of the power at $\theta \in \Theta_1$ is as large as possible, subject to the condition that

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

For the comparison of two tests, first we require them **at the same significance level** α, and then the test is said to be *more powerful* at a value θ^* of $\theta \in \Theta_1$ if it has a higher power at θ^*. A test is said to be *the most powerful* at θ^* if it is more powerful than any other tests at θ^*.

3.2 Uniformly Most Powerful Test

A test is said to be a ***uniformly most powerful*** (in short, UMP) test, if it is the most powerful for all $\theta \in \Theta_1$. To be more precise, the UMP test at a significance level α is the test with a power function $Q(\theta)$ satisfying

1. $\sup_{\theta \in \Theta_0} Q(\theta) = \alpha$
2. $Q(\theta) \geq Q^*(\theta)$ for all $\theta \in \Theta_1$

where $Q^*(\theta)$ is the power of any other test at the significance level α.

Note that the UMP test in hypothesis testing has a similar role to the best estimator under the criterion of MSE in parameter estimation, but unlike the best MSE estimator that does not exist in general, the UMP test ALWAYS exists for simple tests and one-sided tests.

3.2.1 Simple test

In this section, we will study one important result which tells us how to construct the UMP test for a simple test, i.e. $\Theta_0 = \{\theta_0\}$, $\Theta_1 = \{\theta_1\}$, and $\Theta = \{\theta_0, \theta_1\}$.

For testing $H_0: \theta = \theta_0$ vs $H_1: \theta = \theta_1$ at a significance level $\alpha = \gamma(\theta_0) = Q(\theta_0)$ and its power $Q(\theta_1) = 1 - \beta(\theta_1)$, we have the following Neyman-Pearson lemma to construct a UMP test.

Lemma 1 (Neyman-Pearson Lemma):

Jerzy Neyman
(1894 - 1981)

Let $\boldsymbol{X} = \{X_i: i = 1, \dots, n\}$ be a r.s. from pdf $f(\cdot\,|\theta)$ or pmf $p(\cdot\,|\theta)$, where $\theta \in \Theta = \{\theta_0, \theta_1\}$, and $\boldsymbol{x} = \{x_i: i = 1, \dots, n\}$ is its realization, i.e. our collected data. Then, at a significance level α, a test with a rejection region

$$C_1 = \left\{\boldsymbol{x}: \frac{L(\theta_0)}{L(\theta_1)} \le k\right\}$$

is the UMP test for testing $H_0: \theta = \theta_0$ against $H_1: \theta = \theta_1$ at a significance level α, where $k > 0$.

Egon Pearson
(1895 - 1980)

Proof omitted.

Theorem 3: The likelihood ratio test (LRT) for a simple test is a UMP test.

Proof: For a simple test, the LRT at a significance level of α is a test with a rejection region

$$C_1 = \{\boldsymbol{x}:\ \lambda(\boldsymbol{x}) \le k < 1\},$$

where $P(\lambda(\boldsymbol{X}) \le k\,|\theta_0) = \alpha$ and

$$\lambda(\boldsymbol{x}) = \frac{L(\theta_0)}{\max\{L(\theta_0), L(\theta_1)\}} = \begin{cases} 1 & \text{if } L(\theta_0) \ge L(\theta_1) \\ \dfrac{L(\theta_0)}{L(\theta_1)} & \text{if } L(\theta_0) < L(\theta_1) \end{cases}.$$

Thus, we have

$$C_1 = \{\boldsymbol{x}:\ \lambda(\boldsymbol{x}) \le k < 1\} = \left\{\boldsymbol{x}: \frac{L(\theta_0)}{L(\theta_1)} \le k < 1\right\},$$

and hence it follows by Neyman-Pearson lemma that the LRT is a UMP test.

Example:

Consider a rs $\{X_i: i = 1, \dots, n\}$ of size n from $N(\theta, \sigma_0^2)$, where θ is unknown but σ_0^2 is known. Construct a UMP test for testing $\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta = \theta_1 \end{cases}$ at a significance level of α, where $\theta_0 < \theta_1$.

Note that

$$\frac{L(\theta_0)}{L(\theta_1)} = exp\left\{\frac{n}{2\sigma_0^2}[\theta_1^2 - \theta_0^2 - 2\bar{x}(\theta_1 - \theta_0)]\right\}.$$

Thus, $\frac{L(\theta_0)}{L(\theta_1)} \le k$ iff $\bar{x} \ge \frac{1}{2}(\theta_0 + \theta_1) - \frac{\sigma_0^2 \log k}{n(\theta_1 - \theta_0)} = K$, where K is determined by

$$\alpha = P(\bar{X} \ge K | \theta_0) = P\left(Z \ge \frac{K - \theta_0}{\sigma_0/\sqrt{n}}\right).$$

That is, $K = \theta_0 + z_\alpha \frac{\sigma_0}{\sqrt{n}}$. Hence, by Neyman-Pearson lemma, the UMP test for testing $\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta = \theta_1 \end{cases}$ at a significance level of α is the test with a rejection region

$$C_1 = \left\{\boldsymbol{x}: \bar{x} \ge \theta_0 + z_\alpha \frac{\sigma_0}{\sqrt{n}}\right\},$$

where $\theta_0 < \theta_1$.

3.2.2 One-sided test

Although Neyman-Pearson lemma only provides us with a way of constructing a UMP test for a simple test, for some situations, it can help to find a UMP test for one-sided test. See the following example.

[Since we can reparametrize the distribution to convert the one-sided right test into one-sided left test and vice versa, without loss of generality, we only discuss how to find the UMP test for one-sided right test in the following.]

Referring to the previous example, we got a UMP test at a significance level of $\alpha = Q(\theta_0)$ with the rejection region

$$C_1 = \left\{\boldsymbol{x}: \ \bar{x} \geq \theta_0 + z_\alpha \frac{\sigma_0}{\sqrt{n}}\right\}$$

for a simple test $\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta = \theta_1 \end{cases}$, where $\theta_0 < \theta_1$. Note that C_1 **does NOT depend** on θ_1, i.e. the UMP test is independent of the value of θ_1. So, the test with this C_1 is the most powerful at a significance level of α **for any arbitrary** $\theta_1 > \theta_0$. Hence, we can conclude that the test at a significance level of $\alpha = Q(\theta_0)$ with the rejection region

$$C_1 = \left\{\boldsymbol{x}: \ \bar{x} \geq \theta_0 + z_\alpha \frac{\sigma_0}{\sqrt{n}}\right\}$$

is also a UMP test for a one-sided test $\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta > \theta_0 \end{cases}$. Indeed, we can ***further*** extend this result to be a UMP test for $\begin{cases} H_0: \theta \leq \theta_0 \\ H_1: \theta > \theta_0 \end{cases}$ at a significance level of $\alpha = \sup_{\theta \in \Theta_0} Q(\theta)$. Note that the power function of the test is

$$Q(\theta) = P\left(\bar{X} \geq \theta_0 + z_\alpha \frac{\sigma_0}{\sqrt{n}} \middle| \theta\right) = P\left(Z \geq \frac{\theta_0 - \theta}{\frac{\sigma_0}{\sqrt{n}}} + z_\alpha \middle| \theta\right).$$

Thus, it follows that

$$\sup_{\theta \in \Theta_0} Q(\theta) = \sup_{\theta \leq \theta_0} P\left(Z \geq \frac{\theta_0 - \theta}{\frac{\sigma_0}{\sqrt{n}}} + z_\alpha \middle| \theta\right) = Q(\theta_0) = \alpha$$

Remark that this approach based on Neymann-Pearson lemma to extend a UMP test from a simple test to a one-sided test does NOT always work. So, when does it work?
Before anwsering this question, let's first study a property called ***monotone likelihood ratio***.

Definition:

A distribution is said to have the property of *monotone likelihood ratio* *(MLR)* in T if the likelihood ratio $\dfrac{L(\theta')}{L(\theta'')}$ is non-decreasing in T for $\theta' > \theta''$, where at least one of $L(\theta')$ and $L(\theta'')$ is positive.

Example:

Consider a rs $\{X_i: i = 1, \dots, n\}$ of size n from a Bernoulli distribution with an unknown parameter $\theta \in (0, 1)$. Note that the likelihood ratio

$$\frac{L(\theta')}{L(\theta'')} = \left(\frac{1-\theta'}{1-\theta''}\right)^n \left[\frac{\theta'(1-\theta'')}{\theta''(1-\theta')}\right]^{\Sigma x_i}$$

is non-decreasing in $T = \sum_{i=1}^n x_i$ for $\theta' > \theta''$ becaue $\frac{\theta'(1-\theta'')}{\theta''(1-\theta')} > 1$ when $\theta' > \theta''$, so MLR holds for this Bernoulli distribution in $T = \sum_{i=1}^n x_i$.

Example (cont'):

Consider a rs $\{X_i: i = 1, \dots, n\}$ of size n from $N(\theta, \sigma_0^2)$ with a known σ_0^2 and an unknown parameter $\theta \in (-\infty, \infty)$. The likelihood ratio

$$\frac{L(\theta')}{L(\theta'')} = exp\left\{\frac{n}{2\sigma_0^2}[(\theta'')^2 - (\theta')^2 - 2\bar{x}(\theta'' - \theta')]\right\}.$$

is non-decreasing in $T = \bar{x}$ for $\theta' > \theta''$. That is, this normal distribution has MLR in $T = \bar{x}$.

Theorem 4a (Karlin-Rubin Theorem --- Version 1):

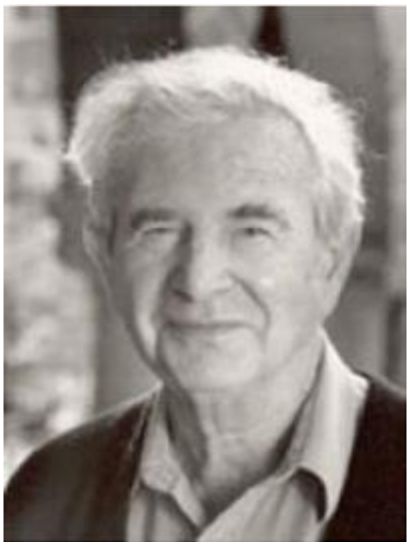
Let $\boldsymbol{X} = \{X_i: i = 1, \dots, n\}$ be a r.s. from a distribution with a parameter θ having MLR in $T(\boldsymbol{x})$, where $\boldsymbol{x}$ is the realization of $\boldsymbol{X}$. Then, at a significance level α, a test with a rejection region

$$C_1 = \{\boldsymbol{x}: T(\boldsymbol{x}) \geq K\}$$

is a UMP test for testing $H_0: \theta \leq \theta_0$ against $H_1: \theta > \theta_0$ at a significance level α for some K.

This test is also a UMP test for testing $H_0: \theta = \theta_0$ against $H_1: \theta > \theta_0$.

Proof omitted.



Samuel Karlin (1924-2007)

Herman Rubin (1926-)

Example:

Consider a rs $\{X_i: i = 1, \dots, n\}$ of size n from $N(\theta, \sigma_0^2)$, where θ is unknown but σ_0^2 is known.

(i) Construct a LRT for testing $\begin{cases} H_0: \theta \leq \theta_0 \\ H_1: \theta > \theta_0 \end{cases}$ at a significance level of α.

(ii) Show that the LRT in (i) is a UMP test.

For (i), note that over $\Theta_0 = \{\theta: \theta \leq \theta_0\}$, the MLE of θ is $\min\{\bar{x}, \theta_0\}$. Thus,

$$\lambda(\boldsymbol{x}) = \begin{cases} 1, & \bar{x} \leq \theta_0 \\ exp\left\{-\dfrac{n(\bar{x} - \theta_0)^2}{2\sigma_0^2}\right\}, & \bar{x} > \theta_0 \end{cases}.$$

It follows that for the LRT, the rejection region is $\left\{\boldsymbol{x}: exp\left\{-\frac{n(\bar{x}-\theta_0)^2}{2\sigma_0^2}\right\} \leq k \text{ } and \text{ } \bar{x} > \theta_0\right\}$, which is equivalent to $C_1' = \{\boldsymbol{x}: \bar{x} > K'\}$, where K' can be shown to be $\theta_0 + z_\alpha \frac{\sigma_0}{\sqrt{n}}$.

For (ii), we have already shown that this normal distribution has MLR in $T = \bar{x}$. By Karlin-Rubin Theorem, a test with a rejection region

$$C_1 = \{\boldsymbol{x}: \bar{x} \geq K\}$$

is a UMP test for testing $H_0: \theta \leq \theta_0$ against $H_1: \theta > \theta_0$ at a significance level α for some K

(Again, it is also a UMP test for testing $H_0: \theta = \theta_0$ against $H_1: \theta > \theta_0$). It is clear that $C_1' = C_1$. That is, the LRT for testing $\begin{cases} H_0: \theta \leq \theta_0 \\ H_1: \theta > \theta_0 \end{cases}$ at a significance level of α is a UMP test!

Theorem 4b (Karlin-Rubin Theorem --- Version 2):

Let $\boldsymbol{X} = \{X_i: i = 1, \dots, n\}$ be a r.s. from a distribution with a parameter θ and $S(\boldsymbol{X})$ is a sufficient statistic for θ. If the distribution of $S(\boldsymbol{X})$ has MLR in itself. Then, at a significance level α, a test with a rejection region

$$C_1 = \{\boldsymbol{x}: S(\boldsymbol{x}) \geq K\}$$

is a UMP test for testing $H_0: \theta \leq \theta_0$ against $H_1: \theta > \theta_0$ at a significance level α for some K.

This test is also a UMP test for testing $H_0: \theta = \theta_0$ against $H_1: \theta > \theta_0$.

Proof omitted.

Example:

Consider a rs $\{X_i: i = 1, \dots, n\}$ of size n from $U[0, \theta]$, where $\theta > 0$ is unknown.

Construct a UMP test for testing $\begin{cases} H_0: \theta \leq \theta_0 \\ H_1: \theta > \theta_0 \end{cases}$ at a significance level of α, where $\theta_0 > 0$.

Recall that in Chapter 2 we have shown that $X_{(n)}$ is sufficient for θ, and the pdf of $X_{(n)}$ is

$$f_{X_{(n)}}(y|\theta) = \frac{ny^{n-1}}{\theta^n} I_{\{y \leq \theta\}}$$

For checking MLR, we have for $\theta' > \theta''$,

$$\frac{f_{X_{(n)}}(y|\theta')}{f_{X_{(n)}}(y|\theta'')} = \begin{cases} \left(\frac{\theta''}{\theta'}\right)^n < 1, & if\ y < \theta'' \\ \infty\,, & if\ \theta'' < y \le \theta' \end{cases}$$

Hence, MLR holds in y (i.e. $X_{(n)}$) itself. Note that we only consider $y \le \theta'$ because both of $L(\theta')$ and $L(\theta'')$ are zero when $y > \theta'$. By Karlin-Rubin Theorem (Theorem 4b), a test with a rejection region

$$C_1 = \{\boldsymbol{x}: x_{(\boldsymbol{n})} \ge K\}$$

is a UMP test for testing $H_0: \theta \le \theta_0$ against $H_1: \theta > \theta_0$ at a significance level α, where K can be shown to be $\theta_0(1-\alpha)^{1/n}$.

Corollary 1: Let $\boldsymbol{X} = \{X_i : i = 1, \dots, n\}$ be a r.s. from a distribution belonging to a *one-parameter **exponential family*** in form of

$$exp\{a(\theta) + b(x) + c(\theta)d(x)\}.$$

Then, for the test of $H_0: \theta \le \theta_0$ against $H_1: \theta > \theta_0$ OR $H_0: \theta = \theta_0$ against $H_1: \theta > \theta_0$ at a significance level α, the test with a rejection region

(i) for an increasing function $c(\theta)$

$$C_1 = \left\{\boldsymbol{x}: \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} d(x_i) \ge K\right\}$$

(ii) for a decreasing function $c(\theta)$

$$C_1 = \left\{\boldsymbol{x}: \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} d(x_i) \le K\right\}$$

is the UMP test at a significance level α for some K.

3.2.3 Two-sided test

In general, a UMP test ***does not exist*** for two-sided tests because the UMP for H1: $\theta < \theta_0$ and the UMP test for H1: $\theta > \theta_0$ are often not the same, but a UMP test for two-sided tests must be the most powerful across all values of $\theta \neq \theta_0$.

The following example illustrates such a case.

Suppose that $\mathbf{X} = \{X_1, \ldots, X_n\}$ is a r.s. of size n from $N(\theta, \sigma^2)$ with unknown θ and known σ_0^2. We have already found that the test with

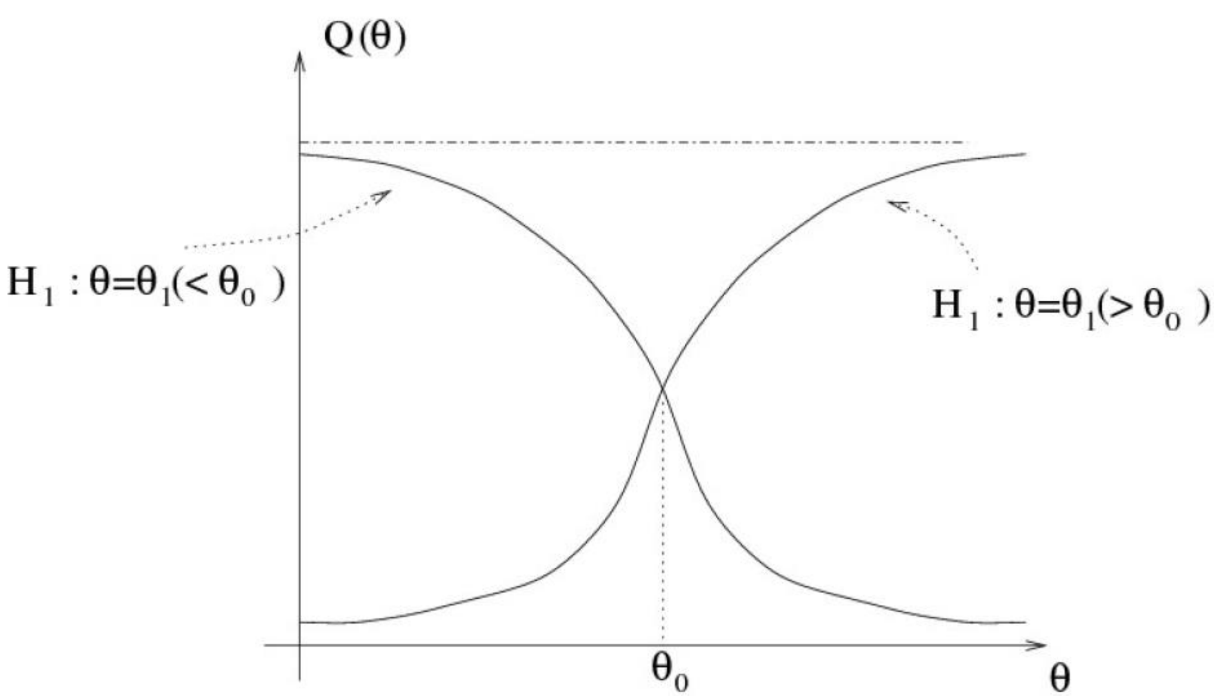
$$C_1 = \{\mathbf{x} : \bar{x} \geq \theta_0 + \frac{\sigma_0}{\sqrt{n}} z_\alpha\}$$

is the uniformly most powerful for $\begin{cases} H_0 : & \theta \leq \theta_0 \\ H_1 : & \theta > \theta_0 \end{cases}$ at a level $\alpha = Q(\theta_0)$. Similarly, we can also find that the test with

$$C_1 = \{\mathbf{x} : \bar{x} \leq \theta_0 - \frac{\sigma}{\sqrt{n}} z_\alpha\}$$

is the uniformly most powerful for $\begin{cases} H_0 : & \theta \geq \theta_0 \\ H_1 : & \theta < \theta_0 \end{cases}$ at a level $\alpha = Q(\theta_0)$.

Thus, from the picture below, it is easy to see that there does not exist a uniformly most powerful test for $\begin{cases} H_0 : & \theta = \theta_0 \\ H_1 : & \theta \neq \theta_0 \end{cases}$



Recall that when we looked for a uniformly "best" estimator in estimation problems, we also encountered the problem that the uniformly best estimator with minimum MSE does not exist. To fix this problem, we considered a smaller class — a class of unbiased estimators. Therefore, here we can also consider a criterion of the test such that we can find a UMP test on a class of tests satisfying a criterion.

Definition 1. *A test at significant level α for testing $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$ is said to be* **unbiased** *if*

$$Q(\theta) \geq \alpha \qquad \forall\, \theta \in \Theta_1.$$

So, a uniformly most powerful unbiased (UMPU) test is the UMP test among all unbiased tests.

Remark that the materials on a UMPU test are not included in the syllabus of this course. For those who are interested in the UMPU test, please take/audit MATH 5432.