Likelihood Ratios

Rice, Chapter 9 discuss hypothesis testing and likelihood ratios.

A hypothesis is a statement about the statistical model, and in the case of parametric models a statement about parameters. It is not a statement about random variables or a statistics. We have null hypotheses

$$H_0 : \theta \in \Theta_0$$

and alternative hypotheses

$$H_A : \theta \in \Theta_A$$

The sets Θ_0 and Θ_A are subsets of Θ, the parameter space. The hypothesis $H : \theta \in \Theta^*$ is a simple hypothesis if these θ uniquely determine the distribution of the observable r.v.s. When θ is identifiable, as is usually the case, then we need Θ^* to be a set of size 1, and hence knowing that θ satisfies the hypothesis then determines the distribution of the r.v.s X that are the observable data or r.v.s. When a hypothesis does not determine the distribution of the observable data then the hypothesis is called composite. This means that Θ^* must have more than one element.

Some examples of the use of likelihood ratios and translating or rewriting the rejection region into an nicer form are given here. The likelihood ratio LR used below is a function of the data $X_1, \ldots, n$. It is also sometimes a function of θ_0, a special value in the parameter space and the null hypothesis. Thus the student should keep in mind that it is a function of the n variables $\mathbf{x} = (x_1, \ldots, x_)$. To explicitly denote this we will sometimes write

$$LR(\mathbf{x}) = LR(x_1, x_2, \ldots, x_n)$$

to emphasis the dependence on the data $\mathbf{x}$. Sometimes we will also write $LR(\theta_0)$ when we wish to emphasize the dependence on θ_0. Sometimes we may need to emphasize the dependence on both these quantities and will write $LR(\theta_0, \mathbf{x})$ to denote this.

In the case of a simple hypothesis $H_0 : \theta = \theta_0$ versus an alternative $H_A : \theta = \theta_1$ (recall θ_0 and θ_1 are specific values in the parameter space) we have

$$LR(\mathbf{x}) = \frac{f(x_1, \ldots, x_n; \theta_0)}{f(x_1, \ldots, x_n; \theta_1)}$$

which is the ratio of the joint pdf (or pmf).

The same type of dependence also occurs for the generalized likelihood ratio (GLR) Λ (see Rice, section 9.5). Thus we will write in various places

$$\Lambda(\theta_0)\ ,\ \Lambda(\theta_0, \mathbf{x}),\ \text{or}\ \Lambda(\mathbf{x})$$

as needed. The GLR is by definition less than or equal to 1. This setting is used to test $H_0 : \theta \in \Theta_0$ versus $H_A : \theta \in \Theta_A$, where Θ_0 and Θ_A are specific subsets of Θ.

Aside : A GLR may be defined as

$$\Lambda^*(\mathbf{x}) = \frac{\max_{\theta \in \Theta_0} f(x_1, \ldots, x_n; \theta)}{\max_{\theta \in \Theta_A} f(x_1, \ldots, x_n; \theta)}$$

The rejection region is then of the form

$$RR = \{\mathbf{x} : \Lambda^*(\mathbf{x}) < c^*\}$$

for some constant c^*. We may also define

$$\Lambda(\mathbf{x}) = \frac{\max_{\theta \in \Theta_0} f(x_1, \ldots, x_n; \theta)}{\max_{\theta \in \Theta_0 \cup \Theta_A} f(x_1, \ldots, x_n; \theta)}$$

Calculus tool for solving maximization problems are much easier to apply to Λ than to Λ^* since typically the subsets of Θ in Λ^* have more boundaries. Notice

$$\Lambda(\mathbf{x}) = \min(\Lambda^*(\mathbf{x}), 1)$$

Thus rejection rules based on Λ^* are usually the same as those based on Λ and they are always the same whenever $c^* < 1$. The GLR in the form of $\Lambda(\mathbf{x})$ is normally used as opposed to any other form. In addition there is a nice limit theorem (see Theorem 9.5A p341 Rice) that makes it easy to obtain an approximate distribution under H_0.

We treat $\Lambda(\mathbf{x}) = \Lambda(x_1, \ldots, x_n)$ as a function mapping R^n to R^+. The rejection region is a set R (sometimes write as RR or A^c where A is the acceptance region) is a subset of the set of realizations of $X_1, \ldots, X_n$, so that $R = RR = A^c \subset R^n$. The decision rule is of the form : reject H_0 and hence accept H_A is $(X_1, \ldots, X_n) \in RR$, that is reject if the data falls into the rejection region. We then have

$$\begin{aligned} P(\text{Type I error}) &= P_{\theta_0}\left((X_1, \ldots, X_n) \in RR\right) \\ P(\text{Type II error}) &= P_{\theta}\left((X_1, \ldots, X_n) \in RR^c\right) \\ &= P_{\theta}\left(\text{Accept } H_0\right) \end{aligned}$$

The term $P(\text{Type II error})$ sometimes needs to be made a little more precise, since in the case of a composite hypothesis there are many possible θ. In this case we take the largest possible value of $P(\text{Type II error})$ over all relevant θ. The same thing is done for a composite null hypothesis. Most of the examples in this course are such that this level of detail is not needed, so we do not consider this further unless necessary.

There is however one topic where this is necessary, namely the power function. It is a function β given by

$$\beta(\theta) = P_{\theta}\left((X_1, \ldots, X_n) \in RR\right)$$

which is thus the probability of rejecting H_0 when θ is in the alternative set of θ. Notice that

$$1 - \beta(\theta) = P_{\theta}\left((X_1, \ldots, X_n) \in RR^c\right)$$

so that for θ satisfying the alternative hypothesis we have

$$1 - \beta(\theta) = P_{\theta}\left(\text{AcceptH}_0\right)$$

that is the probability of type II error for this particular $\theta \in \Theta_A$.

1 Normal Examples

X_i, $i = 1, \ldots, n$ is an iid sample from a $N(\mu, \sigma^2)$ distribution.

1.1 Normal Example, σ known. Simple null hypothesis $H_0 : \mu = \mu_0$ versus the simple alternative $H_A : \mu = \mu_A$

The likelihood ratio is

$$
\begin{aligned}
LR &= \frac{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-(X_i - \mu_0)^2/(2\sigma^2)\right\}}{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-(X_i - \mu_A)^2/(2\sigma^2)\right\}} \\
&= \exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(X_i - \mu_0)^2 - \sum_{i=1}^{n}(X_i - \mu_A)^2\right)\right\}
\end{aligned}
$$

Consider a rejection region of the form $R = \{\mathbf{x} : LR(\mathbf{x}) < c\}$, for some constant c, to be determined by the size of the test. Question for the student : Why does the rejection region have to be of this form?

The rejection region R is

$$
\begin{aligned}
R &= \{\mathbf{x} : LR(\mathbf{x}) < c\} \\
&= \{\mathbf{x} : \sum_{i=1}^{n}(x_i - \mu_0)^2 - \sum_{i=1}^{n}(x_i - \mu_A)^2 > c_1 = -2\sigma^2 \log(c)\} \\
&= \{\mathbf{x} : \sum_{i=1}^{n} x_i^2 - 2n\bar{x}\mu_0 + n\mu_0^2 - \sum_{i=1}^{n} x_i^2 + 2n\bar{x}\mu_A - n\mu_A^2 > c_1\} \\
&= \{\mathbf{x} : 2n\bar{x}(\mu_A - \mu_0) > c_2 = c_1 + n(\mu_A^2 - \mu_0^2)\}
\end{aligned}
$$

If $\mu_A > \mu_0$ then the rejection region is of the form

$$R = \{\mathbf{x} : \bar{x} > c_3\}$$

and if $\mu_A < \mu_0$ then the rejection region is of the form

$$R = \{\mathbf{x} : \bar{x} < c_3\}$$

where c_3 is a constant (that is it is not random).

There is an interesting property in this simple alternative example. The constant c_3 is chosen to make the size of the test some specified value α. Since the r.v $\bar{X}_n$ has a normal distribution, when the null hypothesis is true, that is

$$\bar{X}_n \sim N(\mu_0, \frac{\sigma^2}{n})$$

we can then determine c_3 by solving (in the case that $\mu_A > \mu_0$)

$$P_{\mu_0}(\bar{X}_n > c_3) = \alpha$$

Writing z_α to denote the upper α critical value, or equivalently the $1-\alpha$ quantile, we obtain

$$\frac{\sqrt{n}(c_3-\mu_0)}{\sigma} = z_\alpha$$

or equivalently

$$c_3 = z_\alpha \frac{\sigma}{\sqrt{n}} + \mu_0$$

Notice this critical value, and hence the rejection region, is the same no matter what is the value of μ_A, subject to $\mu_A > \mu_0$. In this sense the rejection region is the same for any alternative $H_A : \mu = \mu_A$, subject to $\mu_A > \mu_0$.

Since this test is most powerful, we can also say it is *uniformly most powerful* for all alternatives $H_A : \mu = \mu_A$, subject to $\mu_A > \mu_0$.

1.2 Simple null hypothesis $H_0 : \mu = \mu_0$ versus the composite alternative $H_A : \mu > \mu_0$ and σ known

Here we need to use the generalized likelihood ratio statistic. There is only one parameter μ to be estimated. Thus

$$\begin{aligned}
\Theta_0 &= \{\mu : \mu = \mu_0\} \\
\Theta_A &= \{\mu : \mu > \mu_0\} \\
\Theta_0 \cup \Theta_A &= \{\mu : \mu \geq \mu_0\}
\end{aligned}$$

The generalized likelihood ratio is

$$\Lambda(\mathbf{X}) = \frac{\max_{\mu \in \Theta_0} \ell(\mu)}{\max_{\mu \in \Theta_0 \cup \Theta_A} \ell(\mu)} \tag{1}$$

where ℓ is the likelihood function

$$\ell(\mu) = \left(2\pi\sigma^2\right)^{-n/2} \exp\left\{-\frac{\sum_{i=1}^n (X_i-\mu)^2}{2\sigma^2}\right\}$$

In the numerator of $\Lambda = \Lambda(\mathbf{X})$ we have $\ell(\mu_0)$, and in the denominator we have

$$\begin{aligned}
&\ell(\mu_0) \quad \text{if} \quad \bar{X} \leq \mu_0 \\
&\ell(\bar{X}) \quad \text{if} \quad \bar{X} > \mu_0
\end{aligned}$$

The student should think about why this is so.

Thus the rejection region is (recall $c < 1$)

$$\begin{aligned}
R &= \{\mathbf{x} : \Lambda < c\} \\
&= \{\mathbf{x} : \sum_{i=1}^n (x_i-\mu_0)^2 - \sum_{i=1}^n (x_i-\bar{x})^2 > c_1 = -2\sigma^2 \log(c)\ ,\ \text{and } \bar{x} > \mu_0\} \\
&= \{\mathbf{x} : \sum_{i=1}^n (x_i-\bar{x}+\bar{x}-\mu_0)^2 - \sum_{i=1}^n (x_i-\bar{x})^2 > c_1\ ,\ \text{and } \bar{x} > \mu_0\}
\end{aligned}$$

$$
\begin{aligned}
&= \{\mathbf{x} : \sum_{i=1}^{n}(x_i - \bar{x})^2 + n(\bar{x} - \mu_0)^2 - \sum_{i=1}^{n}(x_i - \bar{x})^2 > c_1 \ , \ \text{and } \bar{x} > \mu_0\} \\
&= \{\mathbf{x} : n(\bar{x} - \mu_0)^2 > c_1 \ , \ \text{and } \bar{x} > \mu_0\} \\
&= \{\mathbf{x} : \sqrt{n}(\bar{x} - \mu_0) > c_2 = \sqrt{c_1} \ , \ \text{and } \bar{x} > \mu_0\}
\end{aligned}
$$

The constant c_2 (or c_1 or c) is then chosen so the size of the test is α.

$$
\begin{aligned}
\alpha &= P_0\left(\sqrt{n}(\bar{X} - \mu_0) > c_2\right) \\
&= P_0\left(\sqrt{n}(\bar{X} - \mu_0)/\sigma > \frac{c_2}{\sigma}\right)
\end{aligned}
$$

Therefore

$$c_2 = z_\alpha \sigma$$

where z_α is the upper α quantile, or equivalently the $1 - \alpha$-quantile of the standard normal distribution.

Notice that we could unravel this series of constants to obtain c_1 and then c. We could obtain

$$c = \exp\{-\frac{1}{2} z_\alpha^2\} \ .$$

For example if $\alpha = .05$ then $z_\alpha = 1.645$.

1.3 Composite null hypothesis $H_0 : \mu = \mu_0$ versus the composite alternative $H_A : \mu > \mu_0$ and σ unknown

In this case both the the null hypothesis and alternative are composite. The null hypothesis is

$$H_0 : \theta = (\mu, \sigma^2) \in \Theta_0 = \{(\mu, \sigma^2) : \mu = \mu_0, \sigma > 0\}$$

The alternative hypothesis is

$$H_A : \theta = (\mu, \sigma^2) \in \Theta_A = \{(\mu, \sigma^2) : \mu > \mu_0, \sigma > 0\}$$

The generalized likelihood ratio is

$$\Lambda = \frac{\max_{\mu \in \Theta_0} \ell(\mu, \sigma^2)}{\max_{\mu \in \Theta_0 \cup \Theta_A} \ell(\mu, \sigma^2)} \tag{2}$$

where ℓ is the likelihood function

$$\ell(\mu, \sigma^2) = \left(2\pi\sigma^2\right)^{-n/2} \exp\left\{-\frac{\sum_{i=1}^n (X_i - \mu)^2}{2\sigma^2}\right\}$$

The numerator and denominator are maximized separately. In the numerator the maximization is over the set Θ_0, which has one free variable σ^2. The denominator is maximized over $\Theta_0 \cup \Theta_A$, which is a proper subset of Θ. Thus when we use the calculus tool to do this maximization we need to pay attention to whether or not the solution of

$$\frac{\partial \ell(\mu, \sigma^2)}{\partial(\mu, \sigma^2)} = 0$$

falls in the interior of $\Theta_0 \cup \Theta_A$ or it does not fall in the interior, that is if $\bar{X} > \mu_0$ (the interior) or $\bar{X} = \mu_0$ (the boundary of $\Theta_0 \cup \Theta_A$) or $\bar{X} < \mu_0$ (the exterior of $\Theta_0 \cup \Theta_A$).

For the numerator of (2) we are maximizing over the set of possible values of σ^2. For this we obtain the argmax as

$$\hat{\sigma}_0^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu_0)^2$$

Thus the numerator is

$$\ell(\mu_0, \hat{\sigma}_0^2) = \left(2\pi\hat{\sigma}_0^2\right)^{-n/2} \exp\left\{-\frac{\sum_{i=1}^n (X_i - \mu_0)^2}{2\hat{\sigma}_0^2}\right\} = \left(2\pi\hat{\sigma}_0^2\right)^{-n/2} e^{-\frac{n}{2}}$$

For the denominator we have various cases to deal with. Let $(\hat{\mu}_1, \hat{\sigma}_1^2)$ be the argmax over the set $\Theta_0 \cup \Theta_A$.

- If $\bar{X} > \mu_0$ then

  $$\hat{\mu}_1 = \bar{X} \ , \ \hat{\sigma}_1^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

  The student should think about why this is so.

  Therefore

  $$\ell(\hat{\mu}_1, \hat{\sigma}_1^2) = \left(2\pi\hat{\sigma}_1^2\right)^{-n/2} \exp\left\{-\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{2\hat{\sigma}_1^2}\right\} = \left(2\pi\hat{\sigma}_1^2\right)^{-n/2} e^{-\frac{n}{2}}$$

  The student should also verify that $\hat{\sigma}_1^2 < \hat{\sigma}_0^2$. *Hint* : In the formula for $\hat{\sigma}_0^2$ add and subtract $\bar{X}_n$ and expand.

- If $\bar{X} \le \mu_0$ then

$$\hat{\mu}_1 = \mu_0 \ , \ \hat{\sigma}_1^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_0)^2 = \hat{\sigma}_0^2$$

The student should think about why this is so.

Therefore in this case

$$\ell(\hat{\mu}_1, \hat{\sigma}_1^2) = \left(2\pi\hat{\sigma}_0^2\right)^{-n/2} \exp\left\{-\frac{\sum_{i=1}^{n}(X_i - \mu_0)^2}{2\hat{\sigma}_0^2}\right\} = \left(2\pi\hat{\sigma}_0^2\right)^{-n/2} e^{-\frac{n}{2}}$$

Thus we find

$$\Lambda = \begin{cases} 1 & \text{if} \quad \bar{X}_n \le \mu_0 \\ \left(\frac{\hat{\sigma}_1^2}{\hat{\sigma}_0^2}\right)^{n/2} & \text{if} \quad \bar{X}_n > \mu_0 \end{cases}$$

Recall the rejection region is of the form

$$R = \{\mathbf{x} : \Lambda(\mathbf{x}) < c\}$$

and that $c < 1$. Recall also this requires in this particular problem that $\bar{x}_n > \mu_0$.

Aside : Why can we not have $c = 1$? (or $c > 1$?).

In the calculation below the sample variance S_n^2 is also given by

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = \frac{n-1}{n}\hat{\sigma}_1^2 \ .$$

$$\begin{aligned}
\mathbf{x} \in R \quad &\Leftrightarrow \quad \frac{\hat{\sigma}_1^2}{\hat{\sigma}_0^2} < c_1 = c^{2/n} \text{ and } \bar{x} > \mu_0 \\
&\Leftrightarrow \quad \frac{\hat{\sigma}_1^2}{\hat{\sigma}_1^2 + (\bar{x}_n - \mu_0)^2} < c_1 \text{ and } \bar{x} > \mu_0 \\
&\Leftrightarrow \quad \frac{1}{1 + \frac{(\bar{x}_n - \mu_0)^2}{\hat{\sigma}_1^2}} < c_1 \text{ and } \bar{x} > \mu_0 \\
&\Leftrightarrow \quad \frac{(\bar{x}_n - \mu_0)^2}{\hat{\sigma}_1^2} > c'_{2,1} = \frac{1 - c_1}{c_1} \quad \text{and } \bar{x} > \mu_0 \text{ student : why is } c_2 \text{ positive} \\
&\Leftrightarrow \quad \frac{(\bar{x}_n - \mu_0)^2}{\frac{n}{n-1}\hat{\sigma}_1^2} > c_2 = \frac{n-1}{n}c'_{2,1} \quad \text{and } \bar{x} > \mu_0 \text{ student : why is } c_2 \text{ positive} \\
&\Leftrightarrow \quad \frac{(\bar{x}_n - \mu_0)}{S_n} > c_3 = \sqrt{c_2} \quad \text{student : why positive square root?} \\
&\Leftrightarrow \quad \frac{(\bar{x}_n - \mu_0)}{S_n} > c_3 = \sqrt{c_2} \quad \text{student : why positive square root?} \\
&\Leftrightarrow \quad \frac{\sqrt{n}(\bar{x}_n - \mu_0)}{S_n} > c_4 = \sqrt{n}c_3
\end{aligned}$$

Notice this means that the test statistic is

$$t = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{S_n}$$

and decision rule is to reject H_0 in favour of H_A iff

$$\frac{\sqrt{n}(\bar{X}_n - \mu_0)}{S_n} > c_4$$

Moreover, under the null hypothesis, the test statistic has a Student's t distribution with $n-1$ degrees of freedom. The critical value c_4 is thus calculated using the Student's t distribution with df $n-1$.

The the generalized likelihood ratio test is equivalent to the Student's t test. This is the justification for using the Student's t statistic in this one sided alternative hypothesis test problem.

1.4 Composite null hypothesis $H_0 : \mu = \mu_0$ versus the composite alternative $H_A : \mu \neq \mu_0$ and σ unknown

The student should follow the previous section and develop the generalized likelihood ratio test for this problem. In particular show that it is equivalent to the usual two sided t test.

For a different type (non normal) of example see handout EpnonetialLikelihoodRatio.pdf

2 Multinomial GLR

See the handout *MultinomialLikelihoodRatio* for details of this example.

One very useful limit theorem, that is as the sample size $n \to \infty$, is given by Theorem 9.5A Rice, Third Edition, p 341. The proof is beyond the mathematical tools we study up to this level of course. The conditions for this Theorem are that Assumptions I and II for the regularity conditions for MLE are satisfied. Thus this Theorem will apply whenever we have MLE for which the score function and Fisher's information are also applicable or valid.

Theorem 1 (Rice Theorem 9.5A) *Suppose the regularity assumptions I and II for the regular MLE case are satisfied. Consider the hypothesis $H_0 : \theta \in \Theta_0$ versus the hypothesis $H_A : \theta \in \Theta_A$ and consider the Generlized Likelihood Ratio (GLR) test Λ. Under the assumption that $\theta \in \Theta_0$ (that is the null hypothesis is true) then*

$$-2\log(\Lambda(\mathbf{X})) \Rightarrow \chi^2_{(d)}$$

where the convergence is in distribution as $n \to \infty$ and

$$d = \dim(\Theta_0 \cup \Theta_A) - \dim(\Theta_0)$$

This Theorem makes it easy to obtain an approximation to the critical values for the GLR, the test statistic, when the assumptions are satisfied and the sample size n is large. We use this critical value to obtain the rejection region.

3 Hypothesis Testing and Confidence Intervals or Confidence Sets

There is a one to one correspondence between confidence intervals and hypothesis testing with two sided or general alternatives.

Consider the hypothesis $H_0 : \theta = \theta_0$ versus the alternative $H_A : \theta \neq \theta_0$. Let $T_{n,\theta_0}(\mathbf{X})$ be a test statistic and consider a size α test. Consider the following set

$$A = \{\theta_0 : \theta_0 \text{ is not rejected at size } \alpha\}$$

The set A is a 100(1-α) confidence set for θ, that is the set of possible values that are not rejected at level α. We may also consider as the set of possible parameter values that *consistent* with the observed data, in the sense these are the parameter values that are not rejected by our hypothesis test. The confidence set is then the possible values of θ_0 which fall into the acceptance regions for each possible null hypothesis above. One may construct these sets for a parameter $\theta \in \Theta \subset R^d$ or one may construct these for a component of a parameter, for example μ being one of the components of (μ, σ^2) in the normal case, or one of the components in the parameter for a Gamma distribution.

Often in the case of real parameters these confidence sets are intervals and we call the result a confidence interval.

Remark: The confidence set (or interval) is a random set, in that it is a set valued function of the random variables that are observable. As usual there are two forms that are of interest to us (i) as a random object so that we can calculate probabilities of some functions of the confidence set A and (ii) as an observed confidence set (or interval) after data has been observed from an experiment. The later is useful as it then tells us which possible parameters are *consistent* with the observed data. In particular if some special value θ^* is not in the observed confidence interval then at level α this value θ^* is rejected and so there is evidence (at this significance level) that θ^* is not a reasonable choice for θ.

End of Remark

Typically our test statistic will be of the form

$$T_n = \frac{\sqrt{n}(\hat{\theta}_n - \theta_0)}{v_n}$$

where $\hat{\theta}_n$ is an estimator of θ and v_n is either a population variance of an estimate of the standard deviation of $\sqrt{n}(\hat{\theta}_n - \theta_0)$. For the hypothesis test the size α rejection region is the complement of the *acceptance* region which is

$$c_L \leq \frac{\sqrt{n}(\hat{\theta}_n - \theta_0)}{v_n} \leq c_U$$

where c_L and c_U are the appropriate lower and upper quantiles or critical values for T_n as calculated under the null hypothesis. After solving for θ_0 satisfying these inequalities we obtain a formula for the confidence interval for θ_0 as

$$\left\{\theta_0 : \hat{\theta}_n - c_U \frac{v_n}{\sqrt{n}} \leq \theta_0 \leq \hat{\theta}_n - c_L \frac{v_n}{\sqrt{n}}\right\}$$

We often write this in a simpler or more convenient notation as

$$\left[\hat{\theta}_n - c_U \frac{v_n}{\sqrt{n}} \ , \ \hat{\theta}_n - c_L \frac{v_n}{\sqrt{n}}\right] \ .$$

In some cases we cannot easily calculate the critical values c_L and c_U since the distribution of T_n may not be easy to find. The we may be able to calculate c_L and c_U from an approximation such as normal approximation.

Depending on the assumptions we typically calculate the critical values c as follows

- iid normal sampling, σ^2 known : $v_n = \sigma$ and c is calculated from a standard normal distribution
- iid normal sampling, σ^2 not known, $v_n = \sqrt{S_n^2}$ and c calculated from student's t
- most other cases :
  - sampling distribution of T is approximated by a normal distribution. If v_n^2 is the the limiting variance of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ under the null hypothesis assumption, then T has an approximate standard normal distribution under H_0 and c is calculated from the standard normal distribution
  - the sampling distribution of T may be obtained numerically by a Monte Carlo method, such as parameter bootstrap or another useful method called the non parametric bootstrap
- there are a few other special cases where the sampling distribution of T may be known exactly (that is without approximation) and we may discuss some of these if time permits in the course

Example : iid normal. The test statistic of $H_0 : \mu = \mu_0$ versus $H_A : \mu \neq \mu_0$ is

$$T_n = \frac{\sqrt{n}(\hat{\mu}_n - \mu_0)}{S_n}$$

where $\hat{\mu}_n = \bar{X}_n$ and S_n^2 is the sample variance. T_n has, under the null hypothesis assumption, a student's t distribution with degrees of freedom $n-1$. Then

$$c_U = t_{(n-1),\frac{\alpha}{2}}, c_L = -t_{(n-1),\frac{\alpha}{2}}$$

where $t_{(n-1),\frac{\alpha}{2}}$ is the upper $\frac{\alpha}{2}$ critical value, that is the $1 - \frac{\alpha}{2}$ quantile.

Example : In a regular model the MLE has a normal limit distribution, or asymptotic normal distribution

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \approx N(0, \frac{1}{I(\theta_0)})$$

and after some further manipulation we obtain

$$\sqrt{nI(\hat{\theta}_0)}\left(\hat{\theta}_n - \theta_0\right) \approx N(0,1)$$

where $I(\hat{\theta}_n)$ is the observed Fisher's information. Our upper and lower critical values are then obtained from the standard normal distribution, that is for a 100(1 - α) confidence interval we use

$$c_U = z_{\frac{\alpha}{2}}, c_L = -z_{\frac{\alpha}{2}}$$

where $z_{\frac{\alpha}{2}}$ is the upper $\frac{\alpha}{2}$ critical value of the standard normal distribution.

In some cases, but not too many, the test statistic is of the form

$$T = \frac{\hat{\theta}_n}{\theta_0}$$

One example of this type is in the case of iid normal sampling where we have (see Rice Chapter 6, or Stat 3657 examples from Chapter 3)

$$T = \frac{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}{\sigma_0^2} = \frac{(n-1)S_n^2}{\sigma_0^2} \sim \chi^2_{(n-1)}$$

In this case we obtain the confidence interval in the form

$$\left\{\sigma_0^2 : c_L \leq \frac{(n-1)S_n^2}{\sigma_0^2} \leq c_U\right\}$$

where c_L and c_U are the corresponding $\frac{\alpha}{2}$ and $1 - \frac{\alpha}{2}$ quantiles from a $\chi^2_{(n-1)}$ distribution. This confidence interval is usually written as

$$\left[\frac{(n-1)S_n^2}{c_U} \, , \, \frac{(n-1)S_n^2}{c_L}\right] \, .$$

In our examples above often we can rewrite the rejection region (the subset of the sample space for which we reject the null hypothesis in favour of the alternative) in order to find a simpler or more intuitive form of the rejection region. Sometimes this is not possible.

In this case we can find, in the regular models case, an approximation theorem to the sampling distribution of the test statistic. Instead of working with the r.v. $\Lambda(\underline{X})$ we instead work with the r.v.

$$-2\log(\Lambda(\underline{X}))\ .$$

Theorem 9.4A (Rice) : Suppose the statistical model satisfies the regularity (smoothness) Assumptions I and II. Then under the assumption that H_0 holds then

$$-2\log(\Lambda(\underline{X})) \Rightarrow \chi^2_{(d)}$$

(converges in distribution) as $n \to \infty$, where the degrees of freedom d for the limiting chi-square distribution is given by

$$d = \dim(\Theta_0 \cup \Theta_A) - \dim(\Theta_0)\ .$$

end of theorem

We will not prove this Theorem. However in the case of the iid exponential statistical model, we will use a Taylor's expansion or order 1 to examine $-2\log(\Lambda(\underline{X}))$. See that handout.