Chapter 7: Hypothesis Testing and Confidence Intervals

By Joan Llull*

Probability and Statistics.
QEM Erasmus Mundus Master. Fall 2016

Main references:

— Mood: IX: 1, 2.1, 2.2, 3.1, 3.2, 3.3; VIIII: 1, 2.1-2.3

— Lindgren: 9.1-9.4, 9.8-9.12, 8.8

I. Hypothesis Testing

A ***statistical hypothesis*** is a hypothesis that is testable on the basis of observing a process that is modeled via a set of random variables. Consider a sample $(X_1, ..., X_N)$. What we want is to use this sample to see whether, with high enough chances of success, we can reject this hypothesis about the population that generated that sample.

The hypothesis that we are interested in testing is called the ***null hypothesis***, and is denoted by H_0. The null hypothesis describes a hypothetical data generating process. We call the ***alternative hypothesis***, denoted by H_1, or sometimes H_a, to the set of possible alternative hypothetical data generating processes that would be feasible if the null hypothesis was not true.

Statistical hypothesis testing is a method of statistical inference that compares our sample to a hypothetical sample obtained from an idealized model. The null hypothesis describes a specific statistical relationship between the two data sets. The comparison is deemed ***statistically significant*** if the relationship between the observed and hypothetical data sets would be an unlikely realization of the null hypothesis according to a threshold probability: the ***significance level***.

For example:

$$\begin{aligned} H_0: &\quad X_i \sim \mathcal{N}(0,1), \\ H_1: &\quad X_i \sim \mathcal{N}(\mu,1). \end{aligned} \tag{1}$$

In this example, we assume that the distribution of X_i is $\mathcal{N}(\cdot, 1)$, but we want to test whether the mean of the data generating process is equal to μ or equal to 0.

In this example, our hypothesis is called a ***simple hypothesis***, because we completely specified f_X (up parameter values). Alternatively, a ***composite hypothesis*** is any hypothesis that does not specify the distribution completely. The

* Departament d'Economia i Història Econòmica. Universitat Autònoma de Barcelona. Facultat d'Economia, Edifici B, Campus de Bellaterra, 08193, Cerdanyola del Vallès, Barcelona (Spain). E-mail: joan.llull[at]movebarcelona[dot]eu. URL: http://pareto.uab.cat/jllull.

key difference if we knew that a simple hypothesis is satisfied, we would know the entire distribution of X_i, because one and only one distribution satisfies it; instead, infinite distributions satisfy a composite hypothesis.

A ***test statistic*** $C(X)$ is a statistic that summarizes the comparison between the sample and the hypothetical sample obtained from the idealized model. A ***statistical test*** is a procedure to discern whether or not the test statistic *unlikely have been generated by the model described by the null hypothesis.* The ***critical region*** or region of rejection, denoted by R_C, is the set of values of the test statistic for which the null hypothesis is rejected. The set of values of the test statistic for which we fail to reject the null hypothesis is often called the ***acceptance region***. The ***critical value*** is the threshold value of $C(X)$ delimiting the regions of acceptance and rejection.

II. Type I and Type II Errors

As a combination of random variables, a test statistic is a random variable. As such, with certain probability it can lead us to take wrong decisions. The following table summarizes the possible situations:

$H_0 \backslash C(X)$	$C(X) \in R_C$	$C(X) \in R_C^c$
true	Type I error	Ok
false	Ok	Type II error

Therefore, we define as ***Type I error*** the situation in which we reject a true null hypothesis, and ***Type II error*** is the situation in which we do not reject the null hypothesis despite even though it was false. The probability that Types I and II errors occur are relevant to judge how *good* is the test. We define the ***size*** of a test as $\alpha \equiv P_{H_0}(C(X) \in R_C)$, the probability of rejecting a correct hypothesis, i.e. the false positive rate.[1] The ***power*** of a test is $(1-\beta) \equiv P_{H_1}(C(X) \in R_C)$, the probability of correctly rejecting the null hypothesis, i.e. the complement of the false negative rate, β. In the above expressions, P_{H_i} indicates that the probabilities are computed using the cdf described by the hypothesis H_i. In a parametric test, we can define $\pi(\theta)$ as the function that gives the power of the test for each possible value of θ. This function is called the ***power function***. If θ_0 is the parameter indicated in H_0, then $\pi(\theta_0) = \alpha$. Finally, the ***significance level*** of a test is the upper bound imposed on the size of the test, that is, the value

[1] For composite hypothesis, the size is the supremum of the probability of rejecting the null hypothesis over all cases covered by the null hypothesis.

chosen by the statistician that determines the maximum exposure to erroneously rejecting H_0 he/she is willing to accept.

Note that there exists a tension between size and power. In the classical method of hypothesis testing, also known as Neyman-Pearson method, we give most of the importance to minimize the size of the test. But note that if we pick a critical value that takes α to zero, then the power of the test would be also zero.

For example, consider the following (one-sided) tests for the mean of a normal distribution. Assume that $X_i \sim \mathcal{N}(\mu, \sigma^2)$. The hypothesis we want to test is:

$$\begin{aligned} H_0 : &\quad \mu = \mu_0, \\ H_1 : &\quad \mu > \mu_0. \end{aligned} \tag{2}$$

We consider two situations, depending on whether σ^2 is known or not. If σ^2 is known, we know from Chapter 4 that:

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{N}} \sim \mathcal{N}(0, 1). \tag{3}$$

Hence, define the following statistic:

$$C \equiv \frac{\bar{X} - \mu_0}{\sigma/\sqrt{N}}. \tag{4}$$

Under H_0, $\mu = \mu_0$, and, hence, $C \underset{H_0}{\sim} \mathcal{N}(0, 1)$, but under H_1, $C \sim \mathcal{N}(\theta, 1)$, where $\theta \equiv \frac{\sqrt{N}(\mu - \mu_0)}{\sigma}$, since C is a linear transformation of the statistic defined in Equation (7).

Now we consider a critical region: $R_\alpha = \{C > C_\alpha\}$. As we know that $\mu \geq \mu_0$ (these are all the possible cases included in the null and alternative hypotheses), the critical region is defined by the set of values that are so large that are unlikely to be obtained if $C \sim \mathcal{N}(0, 1)$, and hence constitute evidence against the null hypothesis, and in favor of the alternative hypothesis that we defined.

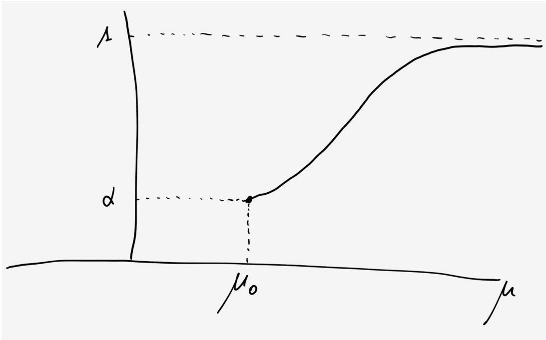
The critical value is the value C_α that satisfies:

$$P_{\mu_0}(C > C_\alpha) = \alpha = 1 - \Phi(C_\alpha) \quad \Rightarrow C_\alpha = \Phi^{-1}(1 - \alpha). \tag{5}$$

The power function is:

$$\pi(\mu) = P_\mu(C > C_\alpha) = 1 - \Phi(C_\alpha - \theta) = \Phi(\theta - C_\alpha). \tag{6}$$

Hence, the power function has the following shape:



If σ^2 is unknown, we also know from Chapter 4 that:

$$\frac{\bar{X}-\mu}{s/\sqrt{N}} \sim t_{N-1}. \tag{7}$$

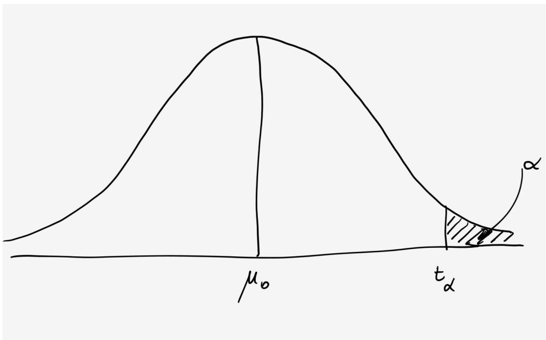
Hence, define the following statistic:

$$t \equiv \frac{\bar{X}-\mu_0}{s/\sqrt{N}}. \tag{8}$$

Under H_0, $\mu = \mu_0$, and, hence, $t \underset{H_0}{\sim} t_{N-1}$, independently of the value of σ^2. Now the critical value is the value t_α such that:

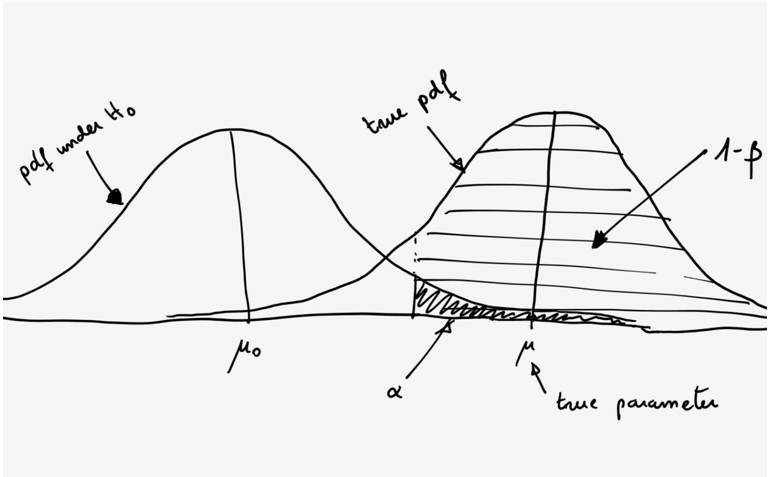
$$P_{\mu_0}(t > t_\alpha) = \alpha = 1 - F_t(t_\alpha). \tag{9}$$

Hence, we need the distribution tables for the Student-t distribution. Graphically, if H_0 is true:



III. Likelihood Ratio Test

Note that, in the previous example, we could define infinite different regions of size α (any region that contains 5% of the area below the pdf if H_0 is true). Then, the question is: why do we tend to consider the critical region in the tail(s)? The answer has to do with maximizing the power of the test. Intuitively, consider the following graphical representation:



By choosing the critical region at the tails, we are increasing the power of the test. In this Section, we formally prove that, by choosing the critical region at the tails (one tail for one-sided alternative hypotheses, the two tails for two-sided hypotheses, e.g. $\mu \neq \mu_0$) we maximize the power function). More generally, we analyze what is the best possible test in different situations.

First, we consider the case in which null and alternative hypotheses are simple:

$$\begin{aligned} H_0 : \quad & C(X) \sim F_0(\cdot), \\ H_1 : \quad & C(X) \sim F_1(\cdot). \end{aligned} \tag{10}$$

Let R_α and R'_α be two critical regions of size α:

$$P_{H_0}(C \in R_\alpha) = P_{H_0}(C \in R'_\alpha) = \alpha. \tag{11}$$

We say that R_α is ***preferred*** to R'_α for the alternative H_1 if:

$$P_{H_1}(C \in R_\alpha) > P_{H_1}(C \in R'_\alpha). \tag{12}$$

Therefore, among two tests with the same size, the one that has more power is preferred. More formally, the ***Neyman-Pearson lemma*** states that in the test of $F_0(\cdot)$ vs $F_1(\cdot)$ (or, equivalently, $f_0(\cdot)$ vs $f_1(\cdot)$), if a size α critical region, R_α, and a constant $k > 0$ exist, such that:

$$R_\alpha = \left\{ X : \lambda(X) = \frac{f_0(X)}{f_1(X)} < k \right\}. \tag{13}$$

then R_α is the most powerful critical region for any size α test of H_0 vs H_1. This lemma is very strong, even though it is also quite restrictive, as it requires both hypotheses to be simple (e.g. it is not applicable to $\mu = \mu_0$ vs $\mu > \mu_0$). $\lambda(X)$ is know as the ***likelihood ratio***. Small values of $\lambda(X)$ indicate small likelihood of H_0 and large likelihood of H_1, and, hence, we want those cases in the critical region. The critical region R_α defined in Equation (13) is known as the ***size*** α ***critical region of likelihood ratio***.

For example, consider a random sample obtained from a normal distribution with known variance σ^2. Consider the following hypotheses:

$$\begin{aligned} H_0: &\quad \mu = \mu_0, \\ H_1: &\quad \mu = \mu_1, \end{aligned} \tag{14}$$

with $\mu_1 > \mu_0$. This is the example drawn in the figure above. These hypotheses are simple, because the distributions under the null and alternative hypotheses are completely specified. The likelihood ratio is:

$$\lambda(X) = \frac{\prod_{i=1}^N \frac{1}{\sigma}\phi\left(\frac{X_i - \mu_0}{\sigma}\right)}{\prod_{i=1}^N \frac{1}{\sigma}\phi\left(\frac{X_i - \mu_1}{\sigma}\right)} = \exp\left(\frac{N}{2\sigma^2}\left[\mu_1^2 - \mu_0^2 - 2\bar{X}(\mu_1 - \mu_0)\right]\right). \tag{15}$$

Now we need to find critical regions of size α for this test statistic. Since we do not directly know the distribution of $\lambda(X)$, we can transform it so that we have an expression in terms of something for which we can compute probabilities:

$$\begin{aligned} \lambda(X) < k \Leftrightarrow &\ln \lambda(X) < \ln k \\ &\Leftrightarrow -\frac{N}{\sigma^2}(\mu_1 - \mu_0)\bar{X} < \ln k - \frac{N}{2\sigma^2}(\mu_1^2 - \mu_0^2) \\ &\Leftrightarrow \bar{X} > \frac{\mu_1 + \mu_0}{2} - \frac{\sigma^2 \ln k}{N(\mu_1 - \mu_0)}, \end{aligned} \tag{16}$$

where in the last step we used the fact that $\mu_1^2 - \mu_0^2 = (\mu_1 - \mu_0)(\mu_1 + \mu_0)$. Therefore, $\lambda(X) < k$ is equivalent to $C > C_\alpha$, which is what we did above, where, in this case, since we know σ^2, C is defined by Equation (8). This explains why we pick the critical region in the tail of the distribution: the statistic C does not depend on μ_1. Therefore, the critical region will be the same regardless of the alternative hypothesis.

The case of composite hypothesis is much more useful in practice. The simple hypotheses case, nonetheless, is useful because it allows us to implement the Neyman-Pearson lemma. It will also serve as a key ingredient in the implementation of the Neyman-Pearson lemma to the composite case.

Let H_0 and H_1 be composite hypotheses (the case in which only one of them is composite is a special case):

$$\begin{aligned} H_0: &\quad \theta \in \Theta_0, \\ H_1: &\quad \theta \in \Theta_0^c = \Theta \backslash \Theta_0, \end{aligned} \tag{17}$$

where Θ is the set of all possible parameters. A test with critical region R_α and power function $\pi(\theta)$ is ***uniformly more powerful*** for a size α if:

1) $\max_{\theta\in\Theta_0} \pi(\theta) = \alpha$, that is, it is of size α.

2) $\pi(\theta) \geq \pi'(\theta)$ for any $\theta \in \Theta$, and any test of size α and power function $\pi'(\cdot)$.

In general, uniformly more powerful tests do not exist because it is difficult that the second condition is satisfied in general. Therefore, there is no equivalent of the Neyman-Pearson lemma for the composite case. However, we proceed with an alternative: the ***generalized likelihood ratio test***, which defines the likelihood ratio as:

$$\lambda(X) = \frac{\max_{\theta\in\Theta_0} L(X;\theta)}{\max_{\theta\in\Theta} L(X;\theta)} = \frac{L(X;\hat{\theta}_0)}{L(X;\hat{\theta}_1)}. \tag{18}$$

This test statistic is very useful to test equality restrictions on the parameters. In this case, $\lambda = \frac{L(\hat{\theta}_r)}{L(\hat{\theta}_u)}$, where $\hat{\theta}_r$ and $\hat{\theta}_u$ indicate, respectively, the estimated coefficients for the restricted and unrestricted models.

To build the test, we need to know the distribution of $\lambda(X)$ or of a transformation of it. Interestingly, if the samples are large (see Chapter 8 for a reference) the distribution of $-2\ln\lambda$ is approximately χ^2.

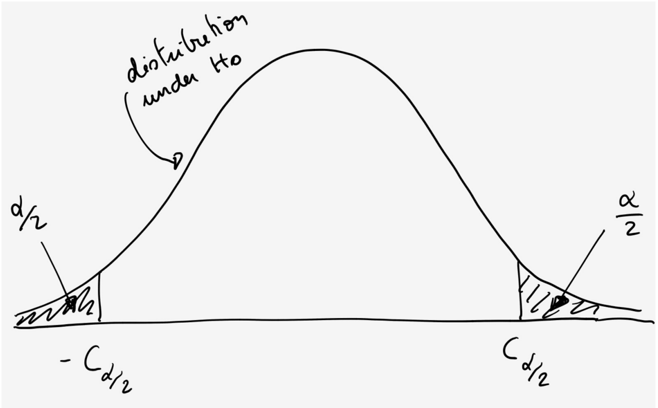
Finally, we say that a test is ***unbiased*** if the power under H_0 is always smaller that the power under H_1, and we say that the test if ***consistent*** if the power under H_1 tends to 1 when $N \to \infty$.

To illustrate all this, we retake some of the examples above, and we introduce some new examples. The first example is the one-sided test of the normal mean with known σ^2 described above. In this case, as the null hypothesis is simple, we can simply apply the Neyman-Pearson lemma for all the possible values of the alternative (i.e., the test $\mu = \mu_0$ vs $\mu > \mu_0$ is an infinite sequence of tests of the form $\mu = \mu_0$ vs $\mu = \mu_1$ for all $\mu_1 > \mu_0$). Therefore, $C = \frac{\sqrt{N}(\bar{X}-\mu_0)}{\sigma^2} > C_\alpha$ describes a test that is uniformly more powerful for a size α. The case of unknown σ^2 will be analogous, except that the test will be defined by $t > t_\alpha$, which will be distributed as a Student-t instead of a normal.

Consider, as a second example, the two tail test for the mean of a normal distribution with known variance, that is:

$$\begin{aligned} H_0 : &\quad \mu = \mu_0, \\ H_1 : &\quad \mu \neq \mu_0. \end{aligned} \tag{19}$$

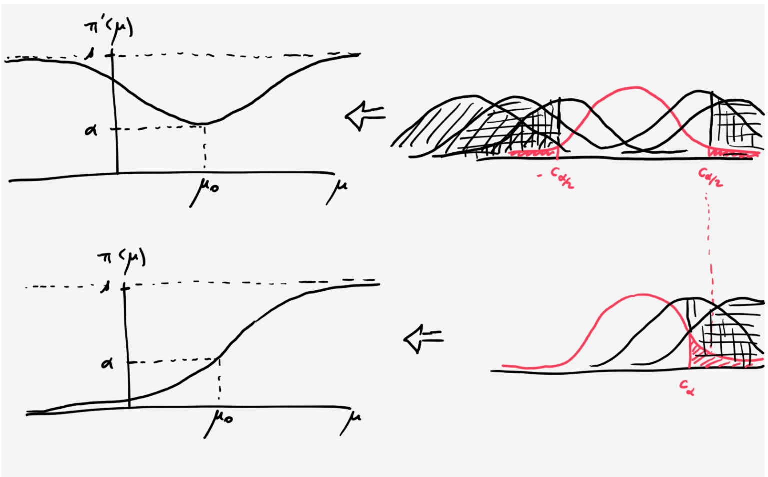
The critical region is still defined by the statistic C defined above, but now the critical region is defined by $|C| > C'_\alpha = C_{\alpha/2}$:



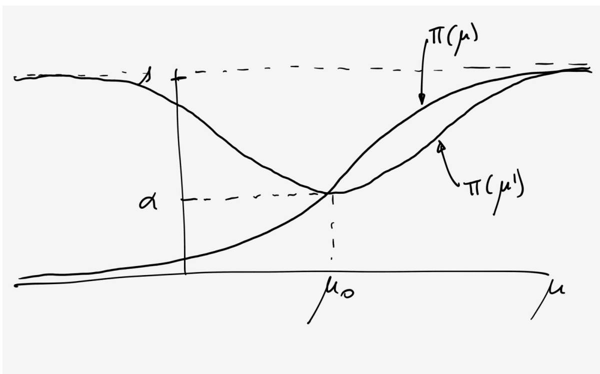
Note that, as before, the distribution under the alternative is given by $\mathcal{N}(\theta, 1)$ with $\theta \equiv \frac{\sqrt{N}}{\sigma}(\mu - \mu_0)$. Therefore, unlike in the previous case, the power function is:

$$\pi'(\mu) = P_\mu(|C| > C_{\alpha/2}) = 1 - [\Phi(C_{\alpha/2} - \theta) - \Phi(-C_{\alpha/2} - \theta)]. \tag{20}$$

To illustrate that a uniformly most powerful test does not exist, let us compare the $\pi'(\mu)$ with the power function $\pi(\mu)$ defined in Equation (6):



It is illustrative to put the two power functions in the same graph to compare:



Clearly, the one-sided test is preferred for the alternatives that imply $\mu > \mu_0$, and the two tail is preferred for $\mu < \mu_0$ (a one sided test of the type $\mu = \mu_0$ vs $\mu < \mu_0$ would be the most powerful one when $\mu < \mu_0$). The key here is that the one tail test is very good for the alternatives with $\mu > \mu_0$, but very bad for the alternatives with $\mu < \mu_0$. We would need to prove that $|C| > C_{\alpha/2}$ is the Neyman-Pearson critical region for the two-tail test, but we will not do it in class (it is strongly recommended as an exercise).

IV. Confidence Intervals

In Chapters 5 and 6, we provided very specific approximations to the parameter value of interest: ***point estimates***. However, it is useful to provide an approximation in the form of an interval. In this section, we provide an interval approximation to the true parameter value that we call ***confidence interval***. It is natural to study here, as they are closely related to hypothesis testing.

A confidence interval is defined by a pair of values $r_1(X)$ and $r_2(X)$ (or $r_1(\hat{\theta})$ and $r_2(\hat{\theta})$) such that $P(r_1(X) < \theta_0 < r_2(X)) = 1 - \alpha$, where α indicates the significance level as in the previous sections. In words, the confidence interval is a range of possible values for θ that, given the sample obtained, we infer contains the true parameter with probability $1 - \alpha$. Importantly, the functions that define the confidence intervals do not depend on the true parameter value.

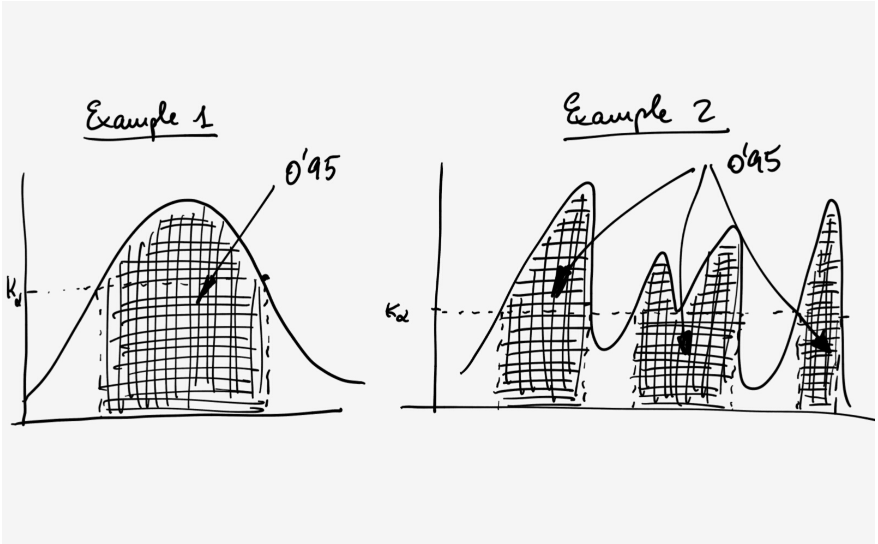
The confidence intervals are constructed in the exact same way that we find the critical value for a two-tail hypothesis test. If the distribution of $r(\hat{\theta})$ is symmetric and unimodal, the confidence intervals will typically be symmetric. One could also build one-sided confidence intervals (i.e., either $r_1(X) = -\infty$ or $r_2(X) = \infty$, but this practice is rather rare.

For example, in our example of the mean, if $C_{0.025} \approx 1.96$ (obtained from the tables of the normal distribution), the confidence interval for the mean would be $[\bar{X} - 1.96 * \frac{\sigma}{\sqrt{N}}, \bar{X} + 1.96 * \frac{\sigma}{\sqrt{N}}]$.

In Bayesian inference, we construct confidence intervals based on the posterior distribution. In that case, we define ***Bayesian confidence intervals*** grouping the regions of the posterior distribution that accumulate more density so that we accumulate density up to $1 - \alpha$. Thus:

$$R_{\theta,\alpha} \equiv \{\theta : h(\theta|X) > k_\alpha\}, \tag{21}$$

so that $P_h(\theta \in R_{\theta,\alpha} \geq 1 - \alpha)$. The interval does not need to be a contiguous set. In the following figure, you can see two examples, one in which the area is contiguous (Example 1), and one in which it is not (Example 2):



V. Hypothesis Testing in a Normal Linear Regression Model

A. Tests for Single Coefficient Hypotheses

Consider the normal linear regression model defined by Assumptions 1 through 3 in Chapter 6. Recall that:

$$\begin{aligned} &Y = W\delta + U, \\ &\hat{\delta} = (W'W)^{-1}W'y = \delta + (W'W)^{-1}W'U, \\ &Y|W \sim \mathcal{N}(W\delta, \sigma^2 I_N), \end{aligned} \tag{22}$$

which implies that:

$$(\hat{\delta} - \delta)|W \sim \mathcal{N}(0, \sigma(W'W)^{-1}). \tag{23}$$

For this model, we want to test hypotheses of the form:

$$\begin{aligned} H_0 :&\quad \delta_j = \delta_{j0}, \\ H_1 :&\quad \delta_j \neq \delta_{j0}, \end{aligned} \quad \text{or} \quad \begin{aligned} H_0 :&\quad \delta_j = \delta_{j0}, \\ H_1 :&\quad \delta_j > \delta_{j0}. \end{aligned} \tag{24}$$

Define the following statistic:

$$Z_j \equiv \frac{\hat{\delta}_j - \delta_j}{\sigma\sqrt{(W'W)^{-1}_{jj}}}, \tag{25}$$

where $(W'W)^{-1}_{jj}$ indicates the jjth element of the matrix $(W'W)^{-1}$. Cearly:

$$Z_j|W \sim \mathcal{N}(0, 1). \tag{26}$$

To derive the unconditional distribution of the statistic, we note that $f(Z_j|W)$ does not depend on W, and, thus, $f(Z_j|W) = f(Z_j)$. Hence, we can conclude that:

$$Z_j \sim \mathcal{N}(0,1). \tag{27}$$

If σ^2 is unknown, we follow the analogy for the sample mean derived in Chapter 4, and derive a t statistic:

$$t \equiv \frac{\hat{\delta}_j - \delta_j}{\widehat{s.e.}(\hat{\delta}_j)} \sim t_{N-K}, \tag{28}$$

where K is the size of the vector δ and $\widehat{s.e.}(\hat{\delta}_j) = s\sqrt{(W'W)^{-1}_{jj}}$ is the estimated standard error of the coefficient. To prove that $t \sim t_{N-K}$, we proceed as in Chapter 4. Dividing the numerator by the standard error, we obtain Z_j, which is distributed as a standard normal. More specifically:

$$Z_j = \frac{\hat{\delta}_j - \delta_j}{\sigma\sqrt{(W'W)^{-1}_{jj}}} = \frac{1}{\sqrt{(W'W)^{-1}_{jj}}}(W'W)^{-1}W'\tilde{U} \equiv P\tilde{U} \sim \mathcal{N}(0,1), \tag{29}$$

where $\tilde{U} \sim \mathcal{N}(0, I_N)$. Therefore, we can rewrite the t statistic as:

$$t = \frac{Z_j}{\sqrt{V/(N-K)}}, \tag{30}$$

where $V \equiv \frac{(N-K)s^2}{\sigma^2}$. Now:

$$V \equiv \frac{(N-K)s^2}{\sigma^2} = \frac{\hat{U}'\hat{U}}{\sigma^2} = \frac{U'MU}{\sigma^2} = \tilde{U}'M\tilde{U} \sim \chi^2_{N-K}, \tag{31}$$

where $M = (I_N - W(W'W)^{-1}W')$, which is symmetric, idempotent, and its rank is $N-K$ (the proof is exactly as in Chapter 4). Now we only need to prove that Z_j and V are independent, which we do by showing that $PM = 0$:

$$\begin{aligned}
PM &= \frac{1}{\sqrt{(W'W)^{-1}_{jj}}}(W'W)^{-1}W'[I_N - W(W'W)^{-1}W'] \\
&= \frac{1}{\sqrt{(W'W)^{-1}_{jj}}}[(W'W)^{-1}W' - (W'W)^{-1}W'W(W'W)^{-1}W'] \\
&= \frac{1}{\sqrt{(W'W)^{-1}_{jj}}}[(W'W)^{-1}W' - (W'W)^{-1}W'] \\
&= 0.
\end{aligned} \tag{32}$$

This completes the proof of $t|W \sim t_{N-K}$. As a final step, to derive the unconditional distribution, we note again that $f(t|W)$ does not depend on W, and, thus,

$f(t) = f(t|W)$, hence concluding that:

$$t \sim t_{N-K}. \tag{33}$$

Given all these test statistics and their distributions, we then proceed with inference in the same way as described in previous sections.

B. *Tests for Multiple Coefficients Hypotheses*

Consider the following test of linear restrictions:

$$\begin{aligned} H_0: &\quad R\delta = R\delta_0, \\ H_1: &\quad R\delta \neq R\delta_0. \end{aligned} \tag{34}$$

where R is a matrix of size $Q \times K$, where $Q \leq K$, and $\text{rank}(R) = Q$. This is general enough to test any linear combination of regressors. For notation compactness, we define $A \equiv (W'W)^{-1}$. We can write the following statistic:

$$F \equiv \frac{(\hat{\delta} - \delta)'R'[RAR']^{-1}R(\hat{\delta} - \delta)/Q}{s^2} \sim F_{Q,N-K}. \tag{35}$$

Note that the rank condition for R is necessary for RAR' to be invertible. The detailed proof is left as an exercise, but intuitively:

$$\begin{aligned} & \hat{\delta}|W \sim \mathcal{N}(\delta, \sigma^2(W'W)^{-1}) \\ \Rightarrow \quad & R\hat{\delta}|W \sim \mathcal{N}(R\delta, \sigma^2 R(W'W)^{-1}R') \\ \Rightarrow \quad & R(\hat{\delta} - \delta)|W \sim \mathcal{N}(0, \sigma^2 R(W'W)^{-1}R'). \end{aligned} \tag{36}$$

Now note that the numerator is a combination of Q squared standard normals (provided that we divide by σ), that we should prove that are independent (in which case is a χ^2_Q), divided by the degrees if freedom Q. The denominator (once divided by σ) is a χ^2_{N-K} divided by the degrees of freedom $N - K$, as we have been doing for the t-test several times. Thus, proving that the χ^2 variables in the numerator and denominator are independent, we would complete the proof, given the definition of the F-distribution in Chapter 4.

One particular application of this test is testing for the values of several coefficients simultaneously. For example, consider the following case for $\delta = (\alpha, \beta)'$:

$$\begin{aligned} H_0: &\quad \alpha = \alpha_0 \text{ and } \beta = \beta_0, \\ H_1: &\quad \alpha \neq \alpha_0 \text{ or } \beta \neq \beta_0. \end{aligned} \tag{37}$$

In this case, $R = I_2$. And, hence, the statistic boils down to:

$$F = \frac{(\hat{\delta} - \delta)'W'W(\hat{\delta}' - \delta)/2}{s^2} \sim F_{2,N-2}. \tag{38}$$

The following figure determines how we construct the confidence interval (given by $P[F < F^{\alpha}_{2,N-2}] = 1 - \alpha$):