INTRODUCTION TO INTERSECTION-UNION TESTS

Jimmy A. Doi, Cal Poly State University San Luis Obispo
Department of Statistics
(jdoi@calpoly.edu)

Key Words: Intersection-Union Tests; Multiple Comparisons; Acceptance Sampling; Bioequivalence.

Abstract

The Intersection-Union Test (IUT) has become increasingly popular, especially through its application in bioequivalence testing. Here we will provide a basic introduction and discuss some properties of the IUT. Since this method is based upon a combination of tests, there may be a concern for the need of a multiplicity adjustment to adequately control the overall type-I error rate. We will address this issue by considering two theorems. The first describes how the IUT can be level-α and the second shows under what conditions the test is size-α. Both results do not require any multiplicity adjustment. A simple example from acceptance sampling will be used to apply these theorems, and we will examine Monte Carlo simulations to verify the expected results.

1 Introduction

1.1 Multiple Comparisons

Consider an experiment consisting of 5 treatments. We would then have a total of 10 possible pairwise comparisons or *contrasts*. Let $t_1, t_2, \ldots, t_{10}$ represent 5% level t-tests corresponding to the respective contrasts. Suppose we combine all tests into $\boldsymbol{\phi_t}$, where $\boldsymbol{\phi_t}$ rejects if *any of the t_i rejects*. Then, the familywise error rate $\alpha^* \neq 0.05$ despite the fact that each of the individual t_i is a 0.05 level test. In fact, the correct value for α^* would be about 40%. Looking at Table 1 it is evident that the overall type I error rate quickly grows beyond 0.05 as the number of comparisons increases. In order to adequately control the overall type I error rate we would need to utilize a *multiplicity adjustment* procedure. Examples of such procedures are described by Steel

Table 1: Familywise error rates corresponding to levels of n, the total number of comparisons.

	n				
	3	5	10	20	45
α^*	14%	23%	40%	64%	90%

et al. (1997) in their chapter of multiple comparisons. The key issue to keep in mind here is the fact that $\boldsymbol{\phi_t}$ rejects if at least one of the individual tests t_i rejects. In a moment, we will see how this compares to the intersection-union test and its basis for rejection.

1.2 Model Definition

Given $X \sim f(x|\theta)$, suppose H_0 is expressed as a union of k sets (*note: the index set need not be finite*):

$$H_0 : \theta \in \Theta_0 = \bigcup_{i=1}^{k} \Theta_i \text{ vs. } H_A : \theta \in \Theta_0^c = \bigcap_{i=1}^{k} \Theta_i^c \tag{1}$$

For each i, let R_i be the rejection region for a test of

$$H_{0i} : \theta \in \Theta_i \text{ versus } H_{Ai} : \theta \in \Theta_i^c.$$

Then the rejection region for the IUT of H_0 versus H_A is $R = \bigcap_{i=1}^{k} R_i$. In other words, the IUT rejects only if *all of the tests reject*.

The rationale behind R is easy to understand since the null hypothesis (H_0) is false if and only if each of the individual null hypotheses (H_{0i}) is false. Note that the IUT is quite different from $\boldsymbol{\phi_t}$ in the manner in which rejection is determined. Will this difference have any effect upon the familywise error rate of the IUT? Is there a multiplicity adjustment need for the IUT as well? In order to address this issue, let us discuss two important theorems that deal with the level and size of the IUT.

2 Two Theorems

The following theorems are from Berger (1997).

2.1 Theorem 1

Theorem 1 *If R_i is a level-α test of H_{0i}, for $i = 1, \ldots, k$, then the IUT with rejection region $R = \cap_{i=1}^{k} R_i$ is a* <u>*level-α test*</u> *of H_0 versus H_A in 1.*

Proof. Let $\theta \in \Theta_0 = \cup_{i=1}^{k} \Theta_i$. So, for some $i = 1, \ldots, k$ (say $i = i'$), $\theta \in \Theta_{i'}$. Thus,

$$P_\theta(\cap R_i) \leq P_\theta(R_{i'}) \leq \alpha$$

Since $\theta \in \Theta_0$ was arbitrarily chosen, the IUT is level-α as

$$\sup_{\theta \in \Theta_0} P_\theta(\cap R_i) \leq \alpha \qquad \blacksquare$$

The result provides an overall type I error rate of α without the need for a multiplicity adjustment. It is made possible through the special way in which the individual tests are combined for the IUT. This illustrates the usefulness of the IUT as we can rely upon the simpler tests of the individual hypotheses and not worry about the formulation of the underlying joint multivariate distribution. This benefit is especially apparent when considering the situation where observations are believed to be correlated.

It is important to note, however, that the IUT constructed via Theorem 1 can be quite conservative. In order to find when the IUT is size-α, let us appeal to the following theorem.

2.2 Theorem 2

Theorem 2 *For some $i = 1, \ldots, k$, suppose R_i is a* <u>*size-α*</u> *rejection region for testing H_{0i} vs. H_{Ai}. For every $j = 1, \ldots, k, j \neq i$, suppose R_j is a* <u>*level-α*</u> *rejection region for testing H_{0j} vs. H_{Aj}. Suppose there exists a sequence of parameter points θ_l, $l = 1, 2, \ldots,$ in Θ_i such that*

$$\lim_{l \to \infty} P_{\theta_l}(R_i) = \alpha,$$

and, for every $j = i, \ldots, k, j \neq i$,

$$\lim_{l \to \infty} P_{\theta_l}(R_j) = 1.$$

Then the IUT with rejection region $R = \cap_{i=1}^{k} R_i$ is a <u>*size α test*</u> *of H_0 vs. H_A.*

Proof.

$$\lim_{l\to\infty} P_{\theta_l}\left(\bigcap_{m=1}^{k} R_m\right) \geq$$
$$\lim_{l\to\infty} \sum_{m=1}^{k} P_{\theta_l}(R_m) - (k-1) =$$
$$\alpha + (k-1) - (k-1) = \alpha.$$

The inequality above is based upon the Bonferroni inequality. Applying Theorem 1 we have

$$\alpha \geq \sup_{\theta\in\Theta_0} P_\theta\left(\bigcap_{m=1}^{k} R_m\right)$$

So, we have

$$\alpha \geq \sup_{\theta\in\Theta_0} P_\theta\left(\bigcap_{m=1}^{k} R_m\right) \geq \lim_{l\to\infty} P_{\theta_l}\left(\bigcap_{m=1}^{k} R_m\right) \geq \alpha$$

$$\text{Hence, } \sup_{\theta\in\Theta_0} P_\theta\left(\bigcap_{m=1}^{k} R_m\right) = \alpha. \qquad \blacksquare$$

Also, refer to the appendix for an alternative proof of Theorem 2. The key component of Theorem 2 is the use of the sequence of parameter points $\theta_l \in \Theta_0$. To gain a better understanding of these two theorems, let us consider the following application of an example in acceptance sampling.

3 Example and Simulation

Example. A sample of upholstery fabric is considered to be acceptable when it meets or exceeds certain criteria. Let θ_1 be the mean breaking strength of the fabric and θ_2 be the probability of the fabric passing a flammability test. Suppose the fabric is deemed to be acceptable when the following conditions are met: $\theta_1 > 50$ and $\theta_2 > 0.95$. Placing this under the hyposthesis testing framework, let us define

$$H_0 : \{(\theta_1, \theta_2)\colon \theta_1 \leq 50 \text{ or } \theta_2 \leq 0.95\} \text{ and}$$
$$H_A : \{(\theta_1, \theta_2)\colon \theta_1 > 50 \text{ and } \theta_2 > 0.95\}$$

where the fabric is deemed to be acceptable when the null hypothesis is rejected. Suppose we model the data where $X_1, \ldots, X_n \overset{iid}{\sim} \mathrm{N}(\theta_1, \sigma^2)$ and $Y_1, \ldots, Y_m \overset{iid}{\sim} \mathrm{Bern}(\theta_2)$. For the X_i, let us use the corresponding likelihood ratio test (LRT) of $H_{01} : \theta_1 \leq 50$ which has the form $(\overline{x} - 50)/(s/\sqrt{n}) > t$. Likewise, for the Y_i, let us use the corresponding LRT of $H_{02} : \theta_2 \leq 0.95$ which has the form $\sum_{i=1}^{m} y_i > b$. Then, combining these two components, the *IUT* rejection region is

$$R = \bigcap_{i=1}^{k} R_i = \left\{ (\boldsymbol{x}, \boldsymbol{y}) : \frac{\overline{x} - 50}{s/\sqrt{n}} > t \text{ and } \sum_{i=1}^{m} y_i > b \right\}.$$

For the Monte Carlo simulation, let $n = m = 58$. Then, setting $t = 1.672$ and $b = 57$, the likelihood ratio tests above are approximate size-α tests.

In this example, since the individual tests are (approx.) size-α, applying Theorem 1 we have that the *IUT* is level-α. We can in fact go further by applying Theorem 2 and state that the *IUT* is size-α. To see this, let us construct a sequence of parameter points $\theta_l \in \Theta_0$ that will satisfy the conditions of Theorem 2. Figure 3 illustrates the graphical representation of the null and alternative hypothesis spaces. Notice that, for our particular example, the boundary of the two regions is the union of the sets $\{(\theta_1, \theta_2) : \theta_2 = 0.95,\ \theta_1 \geq 50\}$ and $\{(\theta_1, \theta_2) : \theta_1 = 50,\ \theta_2 \geq 0.95\}$. If we choose a sequence of parameter points from either set above (representing the horizontal and vertical boundary lines respectively) it can be easily shown that the *IUT* attains size-α. For verification, a Monte Carlo simulation was performed to calculate estimates of the type I error rate, α. In Table 2, estimates were calculated where θ_2 is fixed at 0.95 and θ_1 is increasing (selecting points along the horizontal boundary). In Table 3, estimates were calculated where θ_1 is fixed at 50 and θ_2 is increasing towards 1(selecting points along the vertical boundary). For each simulation, a total of 10,000 runs were performed. As expected, the results indicate that the *IUT* does attain a size of 0.05.

Finally, to determine Monte Carlo estimates of power, we chose 9 specific points from H_A (see Figure 2). The values 60, 65, and 70 for θ_1 were chosen since they seemed to generate fairly large differences among the corresponding LRT statistics. The values .975, .990, and .999 for θ_2 were likewise selected for the same reason. This was the basis for generating the 9 specific points and the results of the power estimates are summarized in Table 4.

Table 2: Monte Carlo Estimates of α, Fixed θ_2 and Increasing θ_1.

	θ_1=*50*	$\theta_1 = 60$	$\theta_1 = 75$	$\theta_1 = 100$
θ_2=.95	.002	.041	.049	.051

NOTE: Each estimate based on 10,000 MC runs.

Table 3: Monte Carlo Estimates of α, Fixed θ_1 and Increasing θ_2.

	θ_2=*.95*	$\theta_2 = .99$	$\theta_2 = .999$	$\theta_2 = .9999$
θ_1=*50*	.002	.015	.049	.049

NOTE: Each estimate based on 10,000 MC runs.

Table 4: *MC Estimates of Power.*

		θ_2		
		.975	*.99*	*.999*
	60	.185	.445	.757
θ_1	*65*	.226	.544	.927
	70	.230	.553	.944

NOTE: Each estimate based on 10,000 MC runs.

4 Summary

We have defined the *IUT* model and discussed two theorems that describe how the *IUT* can attain level-α and size-α. Through these theorems we found that an overall type I error rate could be attained without the need for a multiplicity adjustment. This was made possible by the special way in which the individual tests are combined.

The usefulness of the *IUT* is that there is no need to postulate a multivariate model since we only need to make use of the individual hypothesis tests and these are usually much easier to construct. In addition to the field of acceptance sampling, the *IUT* has been used for the comparison of regression functions (see Berger (1984)) and testing for contingency tables (see Cohen et al. (1983)). But, the field that has placed most attention to the *IUT* is the growing area of bioequivalence (see Berger and Hsu (1996)).

APPENDIX

Alternate Proof of Theorem 2
Recall that $\bigcap_{m=1}^{k} R_m$ is the rejection region of the IUT.

$$\lim_{l\to\infty} P_{\theta_l}\left(\bigcap_{m=1}^{k} R_m\right) = \lim_{l\to\infty}\left(1 - P_{\theta_l}\left(\bigcup_{m=1}^{k} R_m^c\right)\right)$$

$$\geq 1 - \lim_{l\to\infty}\left(\sum_{m=1}^{k} P_{\theta_l}\left(R_m^c\right)\right)$$

The last inequality above follows from the fact that $P\left(\bigcup_m A_m\right) \leq \sum_m P\left(A_m\right)$.

$$Note: \sum_{m=1}^{k} P_{\theta_l}\left(R_m^c\right) = P_{\theta_l}\left(R_i^c\right) + \sum_{m\neq i} P_{\theta_l}\left(R_m^c\right)$$

Now, we have $\lim_{l\to\infty} P_{\theta_l}(R_i^c) \longrightarrow 1-\alpha$.

Also, for all $m \neq i$ we have $\lim_{l\to\infty} P_{\theta_l}(R_m^c) \longrightarrow 0$. So,

$$1 - \lim_{l\to\infty}\left(\sum_{m=1}^{k} P_{\theta_l}\left(R_m^c\right)\right) = 1 - \left((1-\alpha) + \sum_{m\neq i} 0\right) = \alpha$$

So, we have

$$\lim_{l\to\infty} P_{\theta_l}\left(\bigcap_{m=1}^{k} R_m\right) \geq \alpha \tag{2}$$

Applying Theorem 1 we also have

$$\alpha \geq \sup_{\theta\in\Theta_0} P_{\theta}\left(\bigcap_{m=1}^{k} R_m\right) \tag{3}$$

Putting 2 and 3 together we finally have

$$\alpha \geq \sup_{\theta\in\Theta_0} P_{\theta}\left(\bigcap_{m=1}^{k} R_m\right) \geq \lim_{l\to\infty} P_{\theta_l}\left(\bigcap_{m=1}^{k} R_m\right) \geq \alpha$$

showing that the IUT is indeed a size α test! ■

REFERENCES

BERGER, R.L. (1984), "Testing whether one regression function is larger than another", *Communications in Statistics, Theory and Methods* 13, 1793–1810.

BERGER, R.L. (1997), "Likelihood Ratio Tests and Intersection-Union Tests", *Advances in Statistical Decision Theory and Applications*, Boston: Birkhauser. pp. 225–237.

BERGER, R.L. and Hsu, J.C. (1996), "Bioequivalence trials, intersection-union tests, and quivalence confidence sets." With discussion. *Statistical Science*, 11, 283–319.

COHEN, A., GATSONIS, C. and MARDEN, J.I. (1983), "Hypothesis testing for marginal probabilities in a $2 \times 2 \times 2$ contingency table with conditional independence", *Journal of the American Statistical Association*, 78, 920–929.

STEEL, R.G.D., TORRIE, J.H., DICKEY, D.A. (1997), *Principles and Procedures of Statistics A Biometrical Approach*, New York: McGraw-Hill.

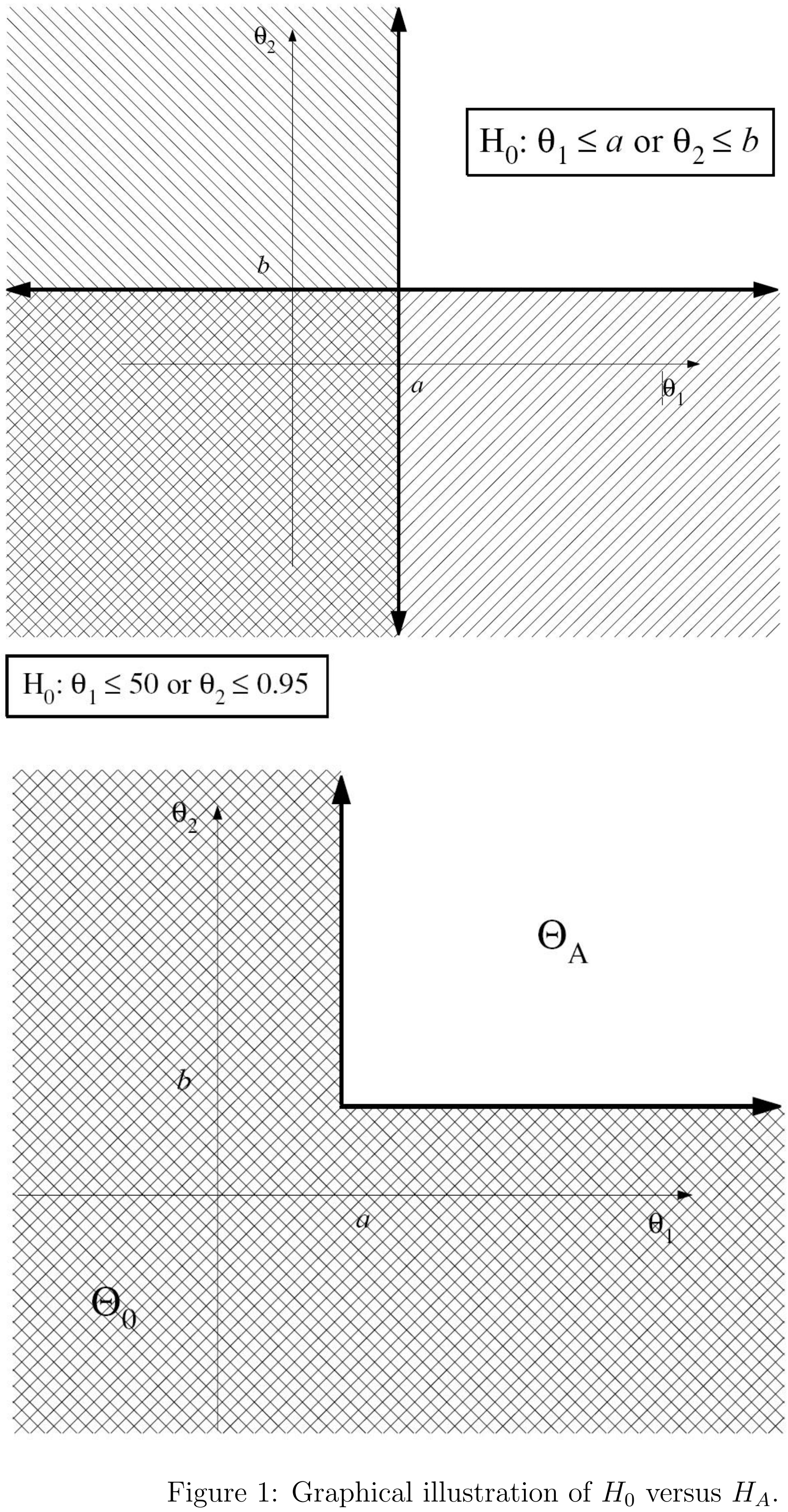

Figure 1: Graphical illustration of H_0 versus H_A.

Figure 2: Selected points in H_A for determination of power estimates