Statistics and Applications
Volume 6, Nos.1 & 2, 2008 (New Series), pp. 25-34

Partition testing: A generalization of hypothesis testing

Daniel T. Voss
Wright State University

Abstract

We consider a variation on partition testing as introduced by Finner and Strassburger (2002), demonstrating this variation to be a useful generalization of traditional hypothesis testing. By choosing an appropriate partition of the parameter space and testing each set in the partition with a size-α test, more specific inferences are obtained compared to those provided by traditional methods of hypothesis testing, while still controlling the probability of making any false assertions to be at most α. Generalizing the traditional two-sided test of a point null hypothesis, one choice of the partition results in rejection of all values except those in the traditional two-sided confidence interval, and another choice of the partition provides better directional inference at no cost whenever the traditional point null hypothesis is rejected. Generalizing the traditional one-sided test, an appropriate choice of the partition results in rejection of all values except those 'accepted' in view of the corresponding one-sided confidence bound. In short, the proposed partition test generalizes traditional testing so as to make inferences for the partition test exactly match those obtained by the corresponding confidence estimation approach.

Key words: Partition principle; Partition test; Directional inference; Confidence set.

1 Introduction

A standard, fundamental procedure in statistics is the two-sided test of a point null hypothesis, testing the null hypothesis $H_o : \theta = \theta_o$ versus the two-sided alternative hypothesis $H_a : \theta \neq \theta_o$. Consider for example a t-test, where under suitable assumptions $(\hat{\theta} - \theta)/s(\hat{\theta}) \sim t_\nu$. The standard two-sided t-test rejects the null hypothesis at significance level α if $|t| > t_{\alpha/2,\nu}$, using the test statistic $t = (\hat{\theta} - \theta_o)/s(\hat{\theta})$. If the test statistic is in the rejection region—namely, if $|t| > t_{\alpha/2,\nu}$—and if α is small, then one has evidence against the null hypothesis and so in favor of the alternative hypothesis. Any more specific conclusions about θ within H_a generally depend on the operating characteristics of the test, but little or no attention is generally given to this, even though the broad conclusion $\theta \neq \theta_o$ is often too general to be of much use. If the null hypothesis is rejected, one would at least like to be able to make a directional inference, such as to conclude that $\theta > \theta_o$ if $\hat{\theta} > \theta_o$—see Hsu (1996, p. 38) for related discussion. Two options for directional inference come to mind.

As a first option, consider the $100(1-\alpha)\%$ confidence interval for θ—namely, $\hat{\theta} \pm t_{\alpha/2,\nu}\, s(\hat{\theta})$. As is well known, the above two-tailed test fails to reject $H_o : \theta = \theta_o$ if and only if θ_o is in this confidence interval. If for example the null hypothesis is rejected at level α and the confidence interval only contains values $\theta > \theta_o$, it seems reasonable to conclude that $\theta > \theta_o$. Likewise, if the null hypothesis is rejected at level α and the confidence interval only contains values $\theta < \theta_o$, it seems reasonable to conclude that $\theta < \theta_o$. One may wonder if this is rigorously justified.

As a second option, one can recognize that directional inference is available at no additional cost, (see for example Pfizer, 1967). In particular, if the null hypothesis is rejected—namely, if $|t| > t_{\alpha/2,\nu}$—suppose one asserts $\theta > \theta_o$ if $\hat{\theta} > \theta_o$ and one asserts $\theta < \theta_o$ if $\hat{\theta} < \theta_o$. If in fact $\theta \leq \theta_o$, then one only erroneously asserts $\theta > \theta_o$ if $t = (\hat{\theta} - \theta_o)/s(\hat{\theta}) > t_{\alpha/2,\nu}$, but the probability of this is at most $\alpha/2$:

$$P_\theta[(\hat{\theta} - \theta_o)/s(\hat{\theta}) > t_{\alpha/2,\nu} \mid \theta \leq \theta_o] \leq \alpha/2.$$

Similarly, if in fact $\theta \geq \theta_o$, then one only erroneously asserts $\theta < \theta_o$ if $t = (\hat{\theta} - \theta_o)/s(\hat{\theta}) < -t_{\alpha/2,\nu}$, but again the probability of this is at most $\alpha/2$:

$$P_\theta[(\hat{\theta} - \theta_o)/s(\hat{\theta}) < -t_{\alpha/2,\nu} \mid \theta \geq \theta_o] \leq \alpha/2.$$

Furthermore, when the null hypothesis is true, the probability of erroneously rejecting H_o and so necessarily making an incorrect directional inference is α. Hence, for any parameter value θ, the probability of making an error in directional inference is at most α.

One goal of this paper is to make directional inference a more natural consequence of testing. Another goal is to tighten the connection between testing and confidence interval estimation. To this end, in the next section we present a generalization of hypothesis testing based on ideas borrowed from partition tests and the partitioning principle, introduced by Finner and Strassburger (2002) in the context of multiple inference. Partition testing is a refinement of closed testing introduced by Marcus, Peritz and Gabriel (1976), each providing a means toward more powerful multiple tests strongly controlling familywise error rates, though in this paper we focus on individual tests. Our end result in some cases hinges upon the notion that confidence intervals result from testing each parameter value, a notion exploited by Stefansson, Kim and Hsu (1988) in the context of multiple estimation.

2 Partition testing

Partition testing was introduced by Finner and Strassburger (2002), and in its true sense it involves the testing of multiple hypotheses, partitioning the subset of the parameter space corresponding to the union of these multiple null hypotheses, testing each set of this partition with a size-α test, and using the fact that testing disjoint hypotheses using a size-α test for each provides a simultaneous size-α test. Unlike Finner and Strassburger (2002), here we consider testing just a single null hypothesis, so this is not a multiple testing setting per se. As another distinction, rather than partitioning the subset of the parameter space corresponding to the null hypothesis, we instead partition the entire parameter space. However, like Finner and Strassburger (2002), we also test each set of the partition with a size-α test as if testing multiple hypotheses, and it likewise follows that this provides a size-α test—namely, the probability of making any false assertions is at most α. So, adapting the approach of Finner and Strassburger (2002) as just described, we propose and utilize the following variation on their partition principle.

Partition Principle: Consider a partition $\Omega = \cup_i \Omega_i$ of the parameter space $\Omega = \{\theta : \theta \in \Re\}$. For each set Ω_i in the partition, conduct a size-α test of the null hypothesis $H_{oi} : \theta \in \Omega_i$, and conclude $\theta \notin \Omega_i$ for each i for which $H_{oi} : \theta \in \Omega_i$ is rejected. Then the probability of making any false assertions is at most α.

Such a test is referred to as a *partition test.* The following example illustrates simple directional inference as a consequence of a partition test.

Example 1. Instead of testing $H_o : \theta = \theta_o$ versus $H_a : \theta \neq \theta_o$, consider the partition of $\Omega = \{\theta : \theta \in \Re\}$ consisting of $\Omega_1 = \{\theta : \theta < \theta_o\}$, $\Omega_2 = \{\theta_o\}$, and $\Omega_3 = \{\theta : \theta > \theta_o\}$. For the corresponding size-α tests, reject Ω_1 if $t > t_{\alpha,\nu}$, reject Ω_2 if $|t| > t_{\alpha/2,\nu}$, and reject Ω_3 if $t < -t_{\alpha,\nu}$, where $t = (\hat{\theta} - \theta_o)/s(\hat{\theta})$. The possible conclusions are as follows.

(i)	If $t < -t_{\alpha/2,\nu}$,	reject Ω_2 and Ω_3	and conclude $\theta \in \Omega_1$, i.e. $\theta < \theta_o$;
(ii)	If $-t_{\alpha/2,\nu} < t < -t_{\alpha,\nu}$,	reject Ω_3	and conclude $\theta \in \Omega_1 \cup \Omega_2$, i.e. $\theta \leq \theta_o$;
(iii)	If $-t_{\alpha,\nu} < t < t_{\alpha,\nu}$,	reject no Ω_i,	so conclude $\theta \in \Omega_1 \cup \Omega_2 \cup \Omega_3$, i.e. $\theta \in \Re$;
(iv)	If $t_{\alpha,\nu} < t < t_{\alpha/2,\nu}$,	reject Ω_1	and conclude $\theta \in \Omega_2 \cup \Omega_3$, i.e. $\theta \geq \theta_o$;
(v)	If $t_{\alpha/2,\nu} < t$,	reject Ω_1 and Ω_2	and conclude $\theta \in \Omega_3$, i.e. $\theta > \theta_o$.

Directional inference follows immediately from the above reformulation of the standard hypothesis test as a partition test, using a three-set partition of the parameter space that is finer than the two-set partition induced by the null and alternative hypotheses, and testing each of the three sets instead of just the one corresponding to the null hypothesis. This partition test is somewhat more powerful than the "directional inference at no additional cost" obtained as option 2 in the Introduction, since here either direction can be rejected even if the null hypothesis is not rejected—potentially useful information. For example, for an experimental factor at two levels, in testing the null hypothesis that the effect is zero, if one fails to reject the null hypothesis, it may still be useful to conclude that the effect is not negative, say.

Result (Finner and Strassburger, 2002). The partition principle works for the following simple reason. Only one of the sets Ω_i in the partition will contain the true value of θ, and one can only make an error if that set Ω_i (and correspondingly the true value of θ) is rejected. Since that Ω_i is tested using a size-α test, the probability of making a false assertion is at most α. It does not matter that one does not know which set Ω_i contains the true but unknown value of θ.

The next example provides a true equivalence between the two-sided test and the two-sided confidence interval.

Example 2. Instead of testing $H_o : \theta = \theta_o$ versus $H_a : \theta \neq \theta_o$, consider conducting a partition test, partitioning $\Omega = \{\theta : \theta \in \Re\}$ as $\Omega = \cup_{\theta_1} \Omega_{\theta_1}$, with a single-element set $\Omega_{\theta_1} = \{\theta_1\}$ for each $\theta_1 \in \Re$. For each θ_1, reject Ω_{θ_1} at level α if $|t| > t_{\alpha/2,\nu}$, where $t = (\hat{\theta} - \theta_1)/s(\hat{\theta})$. For each value θ_1, observe that θ_1 is not rejected if and only if θ_1 is contained in the two-sided confidence $\hat{\theta} \pm t_{\alpha/2,\nu}\, s(\hat{\theta})$. To reiterate, this partition test rejects with significance level α exactly those values of θ not in the $100(1-\alpha)\%$ two-sided confidence interval for θ, and the probability of erroneously rejecting the true value of θ is at most α.

From the perspective of directional inference, this partition test is equivalent to the first option considered in the Introduction—namely, asserting a direction based on whether θ is larger or smaller than θ_o any time the null hypothesis $H_o : \theta = \theta_o$ is rejected at level α. Furthermore, this partition test further pins down the value of θ, providing exactly the same information as provided by the two-sided confidence interval. However, this partition test based on the partition of Ω into sets of individual real values is less powerful for directional inference than the three-set partition test of Example 1, since this test fails to eliminate a direction when either $-t_{\alpha/2,\nu} < t < -t_{\alpha,\nu}$ or $t_{\alpha,\nu} < t < t_{\alpha/2,\nu}$.

Partition testing as a generalization of standard hypothesis testing. The above discussion and examples suggest the following generalization of the standard formulation of hypothesis testing. In short, instead of formulating hypothesis testing as a test of a null hypothesis versus an alternative hypothesis, corresponding to a partition of the parameter space into two sets, and testing only the null hypothesis, one should view hypothesis testing as a partition test, partitioning the parameter space into more than two sets if that is beneficial, and testing each set in the partition.

Directional inference. For the case of the standard two-sided test considered above, if directional information is more important than pinning down the value of θ, then one may prefer the three-set partition test of Example 1 over the confidence-interval-equivalent test of Example 2 based on the finest possible partition, since the former may eliminate a direction even when the value θ_o is not rejected. However, if directional inference is not the top priority, then the approach of Example 2 seems preferable, as it provides more information about the value of θ whether or not the value θ_o is rejected. Both of these partition tests rejects the value θ_o whenever the traditional test rejects $H_o : \theta = \theta_o$.

In view of the above, we have established the following fundamental result.

Lemma 1: Equivalence of two-sided tests and two-sided confidence intervals. The two-sided size-α partition test of Example 2, based on the finest possible partition of Ω, rejects exactly those values θ not contained in the standard $100(1-\alpha)\%$ two-sided confidence interval for θ.

In the next section, we show analogously that one-sided partition tests are equivalent to one-sided confidence bounds.

3 One-sided tests and confidence bounds

Suppose one is interested in conducting a one-sided test, say testing $H_o : \theta \leq \theta_o$ versus $H_a : \theta > \theta_o$. The standard size-α test rejects H_o and accepts H_a if $t = (\hat{\theta} - \theta_o)/s(\hat{\theta}) > t_{\alpha,\nu}$. If the test statistic is in the rejection region, one is able to conclude that $\theta > \theta_o$ with significance level α. Instead of conducting this one-sided test of hypotheses, some would prefer to use a $100(1-\alpha)\%$ lower confidence bound for θ, since then with $100(1-\alpha)\%$ confidence one can conclude that $\theta > \hat{\theta} - t_{\alpha,\nu}\, s(\hat{\theta})$. As is well know, the test results in a rejection of the null hypothesis if and only if the lower confidence bound is above θ_o. Hence, the lower confidence bound provides not only the information needed to determine whether or not to reject the null hypothesis—in cases where the test would reject the null hypothesis, the lower confidence bound provides additional information by virtue of providing a lower bound for θ that is strictly larger than θ_o.

Consider the following formulation of this testing problem using the partition principle. As in Example 2, instead of testing $H_o : \theta \leq \theta_o$ versus $H_a : \theta > \theta_o$, consider partitioning $\Omega = \{\theta : \theta \in \Re\}$ as $\Omega = \cup_{\theta_1} \Omega_{\theta_1}$, with a single-element set $\Omega_{\theta_1} = \{\theta_1\}$ for each $\theta_1 \in \Re$. For each $\theta_1 \in \Omega$, reject Ω_{θ_1} at level α if $t > t_{\alpha,\nu}$, where $t = (\hat{\theta} - \theta_1)/s(\hat{\theta})$. Observe that each value θ is not rejected if and only if θ is above the one-sided lower confidence bound, $\hat{\theta} - t_{\alpha,\nu}\, s(\hat{\theta})$. So, as in the two-sided testing problem, one again gets a perfect equivalence between the information provided by the partition test and the information provided by the lower confidence bound. In particular, one rejects at level α exactly those values θ below the $100(1-\alpha)\%$ lower confidence bound for θ.

In view of the above, we have established the following second fundamental result.

Lemma 2: Equivalence of one-sided tests and one-sided confidence bounds. Let $t = (\hat{\theta} - \theta_1)/s(\hat{\theta})$. For the one-sided size-α partition test that rejects Ω_{θ_1} at level α if $t > t_{\alpha,\nu}$, each value θ is not rejected if and only if θ is above the $100(1-\alpha)\%$ lower confidence bound, $\hat{\theta}-t_{\alpha,\nu}\, s(\hat{\theta})$. Analogously for one-sided inferences in the other direction, for the one-sided size-α partition test that rejects Ω_{θ_1} at level α if $t < -t_{\alpha,\nu}$, each value θ is not rejected if and only if θ is below the $100(1-\alpha)\%$ upper confidence bound, $\hat{\theta}+t_{\alpha,\nu}\, s(\hat{\theta})$.

4 Closing remarks

Partition testing as considered in this paper provides a useful generalization of traditional hypothesis testing, providing more specific conclusions without any loss of power. The standard approach to hypothesis testing involves partitioning the parameter space based on the null and alternative hypotheses and testing only the null hypothesis. The generalization considered here—a variation on partition testing as introduced by Finner and Strassburger (2002)—is to select a finer partition of the parameter space, and to test each set in the partition using a size-α test. The resulting partition test is a size-α test, controlling the probability of making any false assertions to be at most α. While an appropriate partition test rejects the traditional null hypothesis whenever the traditional test does, the clear benefit of the partition test is that, when the traditional null hypothesis is rejected, the partition test provides more useful, specific inferences than simply rejecting the null hypothesis in favor of the alternative as in the traditional approach.

For example, compared to the standard two-tailed test of a point null hypothesis, a partition test with the parameter space Ω partitioned as finely as possible—namely, with one set in the partition for each real value $\theta \in \Omega$—is exactly equivalent to constructing a $100(1-\alpha)\%$ two-sided confidence interval for θ. Alternatively, if directional inference is of primary interest in the two-tailed test setting, a partition test using a three-set partition has the same rejection region as the standard two-sided test but also provides powerful directional inference while still controlling the probability of any false assertions to be at most α. Compared to the standard one-tailed test, there is a partition test with Ω partitioned as finely as possible that is exactly equivalent to constructing a $100(1-\alpha)\%$ one-sided confidence bound for θ.

To put the results of this paper in proper perspective, it is important to recognize that an equivalence between confidence sets and hypothesis testing is well known to exist—see for example Casella and Berger (1990, pp. 406–408) and Hochberg and Tamhane (1987, pp. 22–23)—though the equivalence illustrated here is somewhat different and, one might say, stronger in character. The standard notion of equivalence is as follows. As is well known, one can invert a family of tests to obtain a confidence set, the confidence set consisting of all values θ_o for which $H_o : \theta = \theta_o$ is not rejected; likewise, given a confidence set, one can obtain a corresponding test by rejecting $H_o : \theta = \theta_o$ if θ_o in not in the confidence set. Casella and Berger (1990) refer to this equivalence of testing and confidence sets as theoretically interesting, noting that it is most useful for constructing confidence sets, since it is relatively easy to construct tests and difficult to obtain confidence sets. This traditional sense of equivalence, however, maintains the traditional formulation of hypothesis testing, whereby the null hypothesis is either rejected or not—"a YES-NO test of hypothesis" in the words of Natrella (1960)—and rejection of the null hypothesis provides evidence in support of the alternative hypothesis but nothing more specific without consideration of the operating characteristics of the test. In particular, the test does not provide the same information as the confidence interval. In contrast, the equivalence considered here is based on use of partition testing, a true generalization of traditional hypothesis testing, and *the appropriate partition test provides exactly the same information about the parameter θ as the analogous confidence set, whether or not the null hypothesis is rejected.* Thus, the appropriate partition test and the analogous confidence set are perfectly equivalent.

In the context of the traditional two-tailed test of a point null hypothesis, one can perhaps do better still. Hayter and Hsu (1994) and Finner (1994) considered directional inference and confidence sets for θ if the traditional two-sided test of $H_o : \theta = \theta_o$ results in rejection of $H_o : \theta = \theta_o$. Interestingly, when H_o is rejected, besides directional inference based on whether $\hat{\theta}$ is larger or smaller than θ_o, Finner (1994) provides one-sided confidence bounds, bounding θ away from θ_o in the appropriate direction, with a bound that is not only stronger than the two-sided bound provided here but in some cases even approaching the traditional one-sided bound one would have if one chose from the outset the correct one-sided confidence bound—"correct" as if the direction of θ from θ_o were known in advance.

The goal of this paper has been accomplished. In teaching two-sided hypothesis testing, I have often lamented the need to resort to confidence intervals for directional inference when rejecting the point null hypothesis—that directional inference is not a direct consequence of the standard hypothesis testing formulation. The reformulation and generalization of traditional hypothesis testing using partition testing eliminates this problem, and furthermore provides a perfect equivalence between testing and confidence sets.

Acknowledgments : I am delighted to contribute to this special issue of *Statistics and Applications* in honour of Professor Aloke Dey.

References

Casella, G., and Berger, R. L. (1990). *Statistical Inference.* Duxbury Press, California.

Finner, H. (1994). Two-sided tests and one-sided confidence bounds. *The Annals of Statistics* **22**, 1502–1516.

Finner, H., and Strassburger, K. (2002). The partitioning principle: A powerful tool in multiple decision theory. *The Annals of Statistics* **30**, 1194–1213.

Hayter, A. J., and Hsu, J. C. (1994). On the relationship between stepwise decision procedures and confidence sets. *Journal of the American Statistical Association* **89**, 128–136.

Hochberg, Y., and Tamhane, A. C. (1987). *Multiple Comparison Procedures.* Wiley, New York.

Hsu, J. C. (1996). *Multiple Comparisons: Theory and Methods.* Chapman and Hall, London.

Marcus, R., Peritz, E., and Gabriel, K. R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* **63**, 655–660.

Natrella, M. G. (1960). The relation between confidence intervals and tests of significance-A teaching aid, *The American Statistician* **14**, (1), 20–23.

Pfizer, D. B. (1967). A note on directional inference. *Psychological Bulletin* **68**, 448.

Stefansson, G., Kim, W. C., and Hsu, J. C. (1988). On confidence sets in multiple comparisons. *Statistical Decision Theory and Related Topics IV* (S. S. Gupta and J. O. Berger, eds.) **2**, 89–104. Academic Press, New York.

Daniel T. Voss
Department of Mathematics and Statistics
Wright State University
Dayton, Ohio, 45435-001, USA
Email : dan.voss@wright.edu