An adaptive and powerful randomization test

By MICHEL DONEGANI

Swiss Federal Institute of Technology, DMA, 1015 Lausanne, Switzerland

SUMMARY

Hogg & Lenth (1984) reviewed some adaptive tests. The main problems encountered while using such a procedure are the need for a good selector statistic and the size inflation due to the choice. At present, the solution to the latter problem is to construct a distribution-free procedure (Randles & Hogg, 1973). In this paper we propose a simple adaptive exact procedure. We then use it to construct an adaptive randomization test for matched pairs. This test has exactly size α, the actual size of test components and its power is asymptotically equal to the maximum power of the components. Finally, to observe the behaviour of the method for finite sample sizes we conduct a simulation study.

Some key words: Adaptive test; Asymptotic power; Pitman efficacy; Randomization; Robustness.

1. INTRODUCTION

Randomization tests can be used whenever a set of treatments has been randomly assigned to experimental units. After observation of the treatment response, the method consists in analyzing the given observations using all the different possible assignments of treatments to units. If we compare, for example, a treatment to a control using n and m units, there are $(n+m)!/(n!m!)$ different assignments of the treatment to the $(n+m)$ units. The randomization test computes a test statistic for each assignment and compares the value at the actual assignment to the randomization distribution. In the case of a comparison of a treatment and a control, a typical test statistic is Student's t.

Box & Andersen (1955) investigated the power of randomization tests. They showed that the power depends on the test statistic used, as well as on the underlying distribution of the observations. Moreover, an optimality theory for randomization would be problematic. For example, if we want to test the difference in mean of two normal populations, we should use the classical t distance. The optimal choice is therefore only possible if the underlying distribution is known, which is exactly the assumption we want to avoid. In practice we rarely know the distribution of the observations, in which case classical randomization, while being exact, is not necessarily powerful.

2. THE PROCEDURE AND ITS EXACTNESS

2·1. *Description of the adaptive test*

Let $X_n \in R^n$ be the outcome of a randomization experiment and let F_n be the empirical distribution of X_n. Let G_n be a group of transformations of R^n onto itself and let $\Omega_n = \{g_n X_n \mid g_n \in G_n\}$ be the set of all possible assignments. Note in particular that $X_n \in \Omega_n$. Let T_1 and T_2 be two different test statistics and let c_1 and c_2 be two positive real functions defined on the empirical distributions F_n^i ($i = 1, \ldots, \text{card}(\Omega_n)$) of the resamples $g_n X_n$. In a first step these two functions tell us which of the two tests is better for the data at hand. For example if $c_1(F_n) < c_2(F_n)$, the statistic T_1 is preferable. We suppose that big values of the test statistics T_1 and T_2 indicate

a departure from the null hypothesis. In this paper we study the following adaptive randomization test.

(i) Evaluate the criteria c_1 and c_2 on the observed outcome and let 'opt' refer to the index of the criterion having smaller value.

(ii) Partition the set Ω_n in two classes Ω_n^1 and Ω_n^2 defined by

$$\Omega_n^1 = \{g_n X_n : c_1(F_n^i) < c_2(F_n^i), g_n \in G_n\},$$

$$\Omega_n^2 = \{g_n X_n : c_1(F_n^i) > c_2(F_n^i), g_n \in G_n\}.$$

(iii) Let $\tilde{H}_{\text{opt}}$ be the randomization distribution obtained by evaluating the optimal test statistic T_{opt} on each element of the class that contains X_n. For example, if $X_n \in \Omega_n^1$, then the optimal test is T_1.

(iv) Reject the null hypothesis at the level α if $T_{\text{opt}}(F_n) > \tilde{H}_{\text{opt}}^{-1}(1 - \alpha)$.

2·2. *Exactness of the adaptive test*

Let $n_i = \text{card}(\Omega_n^i)$, let R_0 denote the rejection of H_0 and let C_i denote the choice of the test T_i. The size of the adaptive randomization test can be calculated as

$$\text{pr}_{H_0}(R_0) = \text{pr}_{H_0}(R_0 | C_1) \text{pr}(C_1) + \text{pr}_{H_0}(R_0 | C_2) \text{pr}(C_2)$$

$$= \alpha\{n_1/(n_1 + n_2)\} + \alpha\{n_2/(n_1 + n_2)\} = \alpha.$$

We have used the fact that, under the null hypothesis, the sample X_n is uniformly distributed in the set Ω_n. The exactness of the adaptive test is, therefore, independent of the choice of the criteria. Note that the crude strategy of using a criterion to choose the test and then applying it in the classical manner of randomization tests would lead to a size larger than α, unless the choice is independent of T_1 and T_2.

The asymptotic power property of the adaptive test depends on the criteria and the test statistics used. In the next paragraph we develop an example of an adaptive randomization test with good power properties.

3. An adaptive test for matched pairs

3·1. *Description of the test*

The simplest situation in which a randomization test is possible occurs if we compare one treatment to a control, using matched pairs. Let $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ be the actual observations and let

$$Z_n = \{z_1 = x_1 - y_1, \ldots, z_n = x_n - y_n\}.$$

There are $2^n = N$ possible assignments of treatment and control, since we can switch for each of the n pairs. The set $\Omega_n = \{g_n Z_n | g_n \in G_n\}$ consists of all possible samples of the form $\{\pm z_1, \ldots, \pm z_n\}$. The null hypothesis is H_0: the treatment has no effect, which is often interpreted in a more restricted manner as H_0: the centre of the distribution of Z is equal to 0. Without loss of generality we consider alternatives of the form H_1: the treatment has a positive effect. Let $c_i(F)$ be the Pitman efficacy of the test T_i when F is the underlying distribution and let $c_i(F_n)$ be consistent estimates of $c_i(F)$.

We consider the following two test statistics.

(i) The classical t of Student: $T_1(F_n) = n^{\frac{1}{2}}\bar{z}/s$, where $\bar{z} = \Sigma z_i/n$ and $s^2 = (n-1)^{-1} \Sigma (z_i - \bar{z})^2$, associated with the criterion $c_1(F_n) = n^{-\frac{1}{2}}s$.

(ii) A robust test based on a M-estimate of location: $T_2(F_n) = \hat{\theta}\{\text{var}(\hat{\theta})\}^{-\frac{1}{2}}$, where

$$\hat{\theta} = \arg\min \frac{1}{n} \sum_{i=1}^{n} \rho\left(\frac{z_i - t}{km}\right),$$

$$\text{var}(\hat{\theta}) = \frac{\Sigma_i \psi^2\{(z_i - \hat{\theta})/(km)\}(k^2 m^2)}{[\Sigma_i \psi'\{(z_i - \hat{\theta})/(km)\}]^2},$$

with $\psi(u) = \rho'(u)$. We work with Huber's function (Huber, 1981, p. 146)

$$\rho(u) = \begin{cases} \frac{1}{2}u^2 & \text{if } |u| \le 1, \\ |u| - \frac{1}{2} & \text{otherwise.} \end{cases}$$

The mean absolute deviation m is defined as the median of

$$\{|z_1 - \bar{z}|, \ldots, |z_n - \bar{z}|\},$$

with $\bar{z} = \text{med} \{z_1, \ldots, z_n\}$. The parameter k is a number close to 1·5. Associated with this robust statistic we use the criterion $c_2(F_n) = \{\text{var}(\hat{\theta})\}^{\frac{1}{2}}$, which is then the estimated asymptotic standard error of the robust M-estimate. Note that the criteria $c_1(F_n)$ and $c_2(F_n)$ are straightforward estimates of the Pitman efficacy of the tests T_1 and T_2.

3·2. Asymptotic power of the adaptive test

The power of our adaptive test depends on the ability of the criteria to select the better test, at least asymptotically, for the sample at hand. In other words, we would like the criterion to recognize certain attributes of the true distribution which generated a given sample. To see why the criteria lead to an adaptive test having good asymptotic power properties, we use the asymptotic power theory developed by Pitman and presented in the context of permutation theory by Albers, Bickel & Van Zwet (1976) or Puri & Sen (1971, p. 70). Our purpose is to show that the proposed adaptive test has asymptotic power equal to the power of its better component. This is shown in the following proposition.

PROPOSITION. *If the underlying distribution F is symmetric under the null hypothesis and with the notation of § 2, we have* $\text{pr}(\Omega_n^2) \to 1$ *if* $c_1(F) > c_2(F)$, *and* $\text{pr}(\Omega_n^2) \to 0$ *if* $c_1(F) < c_2(F)$.
The asymptotic power of the adaptive test at alternatives of the form $\theta n^{-\frac{1}{2}}$ *is*

$$\max_{i=1,2} \left[1 - \Phi \left\{ \Phi^{-1}(1 - \alpha) - \frac{\theta}{c_i(F)} \right\} \right] = 1 - \Phi \left[\Phi^{-1}(1 - \alpha) - \theta \Big/ \left\{ \min_{i=1,2} c_i(F) \right\} \right]. \quad (3 \cdot 1)$$

A short proof of these results can be found in the Appendix.

4. NUMERICAL SIMULATION RESULTS

The asymptotic properties developed in § 3 do not ensure that the adaptive test has a good behaviour for small samples sizes. To compare the performances of the two individual tests and the adaptive test, we simulated samples of size $n = 7$, $n = 9$, $n = 11$ and $n = 13$. At $n = 13$ the power of the adaptive test should be close to the power of the better component. This case provides a check of our theoretical results. To simulate, we use a normal distribution and a Student's t distribution with 2 degrees of freedom. The estimates are based on 5000 simulated samples generated at alternative $\theta = n^{-\frac{1}{2}}$, and are given in Table 1. As our estimates are based on the percentage of rejection of the null hypothesis, the standard errors do not exceed 0·0071. Finally, to observe the convergence of the simulated power of the adaptive test, we add the asymptotic power obtained by formula (3·1).

Table 1. *Simulated power of the adaptive test,* $\theta = n^{-\frac{1}{2}}$, $\alpha = 0·05$

	t_2 distribution			Normal distribution		
n	Student's test	Robust test	Adaptive test	Student's test	Robust test	Adaptive test
7	0·183	0·178	0·160	0·220	0·190	0·210
9	0·149	0·159	0·158	0·231	0·201	0·223
11	0·163	0·180	0·171	0·236	0·202	0·228
13	0·155	0·180	0·179	0·239	0·209	0·234
Asymp.	0·050	0·186	0·186	0·256	0·243	0·256

5. COMMENTS

Even for these moderate sample sizes, the adaptive procedure behaves in the way we hoped, in the sense that the power of the adaptive test converges to the power of the better component. The advantages of this adaptive procedure lie in the two following properties. First, we obtain a test that is exact and secondly, we obtain a test which has asymptotically good power properties. This new method bridges robustness of the size inherent in all randomization techniques with robustness of efficiency. In our example, we wanted to prevent the use of the t test when the underlying distribution has heavy tails. It is possible to generalize the ideas explained here to more than two components and to other linear model testing problems.

APPENDIX

Proof of the Proposition

The result is straightforward under the null hypothesis. Let $g_\infty = \{\pm 1, \pm 1, \ldots\}$ be an infinite sequence of ± 1 and let g_n be the first n components of g_∞. Let $Z_\infty = \{z_1, z_2, \ldots\}$ be an infinite sequence of independent and identically distributed variables from the symmetric distribution F, and let Z_n be the first n components of Z_∞. Consider the resample $w_n = g_n Z_n$ with empirical distribution $\tilde{F}_n$. For each g_n, the convergence of the estimates $c_1(\tilde{F}_n)$ and $c_2(\tilde{F}_n)$ to $c_1(F)$ and $c_2(F)$ ensures the desired result.

Consider now alternatives of the form $\theta = n^{-\frac{1}{2}}$. Let $v_n = g_n(Z_n + n^{-\frac{1}{2}})$ be the perturbed sample based on Z_n and g_n so that v_n is a random sample from the distribution

$$H_n = \tfrac{1}{2} F(z + n^{-\frac{1}{2}}) + \tfrac{1}{2} F(z - n^{-\frac{1}{2}}).$$

Using the definitions

$$c_1^2(H_n) = \int z^2 \, dH_n(z), \quad c_2^2(H_n) = \int \Psi^2(z) \, dH_n(z) \Big/ \left\{ \int \Psi'(z) \, dH_n(z) \right\}^2,$$

we obtain

$$c_1^2(H_n) = c_1^2(F) + n^{-1}, \quad c_2^2(H_n) = c_2^2(F) + o_p(n^{-1}).$$

This proves $(3 \cdot 1)$.

Since the t test and the robust test have limiting normal distribution, $(3 \cdot 1)$ is a consequence of Theorem $5 \cdot 1$ of Albers, Bickel & van Zwet (1976, p. 131).

REFERENCES

ALBERS, W., BICKEL, P. J. & VAN ZWET, W. R. (1976). Asymptotic expansions for one-sample tests. *Ann. Statist.* **4**, 108–56.

BOX, G. E. P. & ANDERSEN, S. L. (1955). Permutation theory in the derivation of robust criterion and the study of departures from assumptions (with discussion). *J. R. Statist. Soc.* B **17**, 1–34.

HOGG, R. V. & LENTH, R. V. (1984). A review of some adaptive statistical techniques. *Comm. Statist.* **13**, 1551–79.

HUBER, P. J. (1981). *Robust Statistics.* New York: Wiley.

RANDLES, R. H. & HOGG, R. V. (1973). Adaptive distribution-free tests. *Comm. Statist.* A **2**, 337–56.

PURI, M. L. & SEN, P. K. (1971). *Nonparametric Methods in Multivariate Analysis.* New York: Wiley.