〈Research Paper〉

Journal of the Korean
Statistical Society
Vol. 15, No. 1, 1986

A Randomization Test for Weak Notion of Equality in Paired Experiments

Myung-Hoe Huh*

ABSTRACT

Basu(1980) examined Fisher Randomization Test(FRT) of matched pair experimental data with critical point of view. Additionally, Lane(1980) pointed out that "the experimenter may be interested in a weaker notion of equality between two treatments," than the notion of equality which FRT relies on. In this study, a randomization test is developed so that it can test a weaker hypothesis of equalty.

1. Introduction

Suppose that there are $2n$ experimental units available for the comparative study of two treatments, say T_1 and T_2. For better precision of experimental results, experimenter carefully matches(blocks) units into n pairs so that there are as little differences as possible within a pair of units. Treatment assignment is done by physical randomization such as independent tossing of a fair coin. Then the experiment is performed according to the protocole and responses are observed for each unit. Let x_i and y_i be observed responses from the i-th pair of units receiving T_1 and T_2 respectively. Enperimenter wants to test to the null hypothesis that T_1 and T_2 are equally effective against the alternative hypothesis that T_2 is more effective than T_1. It is thought that $T=\sum_{i=1}^{n} D_i$ is appropriate for the test statistic, where $D_i=Y_i-X_i$ $(i=1,\cdots,n)$. Fisher randomization test(FRT) of paired observations is based on the observed significance level (SL) $P_0(T\geq t)$, where P_0 is the probability distribution which assigns equal probabil-

* Department of Statistics, Korea University, Seoul 132, Korea.

ity mass to 2^n sample points $(\pm|d_1|, \cdots, \pm|d_n|)$, and t is the observed value of T. An interesting fact is that the usual t test, heavily dependent on the normality assumption, gives the same significance level as FRT in the asymptotic case. For the full description of FRT in the same context as in this study, see Fisher(1960, Chapter Ⅲ) or Cox and Hinkley(1974, Section 6.4). See also Kempthorne and Folks(1971, Section 12.11) in which FRT is formulated on different conceptual basis. Since R.A. Fisher invented FRT, it received strong support from a group of statisticians because FRT is not dependent on the usual assumption of normality and because FRT fully exploits the device of physical randomization. R.A. Fisher(1960, p.48), however, said that "somewhat extravagant claims have often been made on their (FRT) behalf." Basu(1980) examined Fisher randomization test (FRT) of matched pair experimental data with critical point of view. His main point is that FRT suffers from logical weakness under restricted and unequal probability randomization. Also Lane(1980) pointed out that "the experimenter may be interested in a weaker notion of equality between two treatments," than the notion of equality which FRT relies on.

In this study, R.A. Fisher is supported in a way that a randomization test for weak notion of equality is devised as an extension of FRT.

2. A Randomization Test for Weak Notion of Equality

Suppose that μ_{ij} be the response of the j-th unit $(j=1,2)$ of the i-th pair$(i=1, \cdots, n)$ when it receives T_1. Similarly, denote ν_{ij} to be the response of same unit when it receives T_2. In the usual experiments, each experimental unit receives only one treatment, and, hence, either μ_{ij} or ν_{ij} is observable. Let $(a_1, a_2, \cdots, a_n)$ be the actual design outcome, where

$a_i=1$ if the first unit of the i-th pair recieves T_1 and the second unit receives T_2,

$=2$ if the first unit of the i-th pair receives T_2 and the second unit receives T_1,

for $i=1, \cdots, n$. As a consequence, μ_{i,a_i} and $\nu_{i,3-a_i}$ are observed and equal to x_i and y_i, respectively. On the other hand, $\mu_{i,3-a_i}$ and ν_{i,a_i} are not available to the experimenter. Note that the average pairwise observed difference can be written as

$$\bar{d}=(1/n)\sum_{i=1}^{n} d_i=(1/n)\sum_{i=1}^{n}(y_i-x_i)=(1/n)\sum_{i=1}^{n}(\nu_{i,3-a_i}-\mu_{i,a_i})$$

Often the experimenter is interested in the quantity

$$\delta=(1/2n)\sum_{i=1}^{n}\{(\nu_{i1}-\mu_{i1})+(\nu_{i2}-\mu_{i2})\}$$

$$=(1/2)\left\{(1/n)\sum_{i=1}^{n}(\nu_{i,3-a_i}-\mu_{i,a_i})+(1/n)\sum_{i=1}^{n}(\nu_{i,a_i}-\mu_{i,3-a_i})\right\}$$

which is the unobserved average difference.

In original FRT, the null hypothesis can be formulated as either

(1) $H_0: \mu_{i1}=\nu_{i1}, \mu_{i2}=\nu_{i2}$ for all i,

or, in a slightly weaker form,

(2) $H_0: \mu_{i1}+\mu_{i2}=\nu_{i1}+\nu_{i2}$ for all i.

Using observed differnces $d_1, \cdots, d_n$, (2) is equivalent to

(2′) $H_0: \nu_{i,a_i}-\mu_{i,3-a_i}=-d_i$, for all i.

Either (1) or (2) implies $\delta=0$ but not vice versa. Note that $\delta=0$ if and only if

(3) $H_0: \sum_{i=1}^{n}(\mu_{i1}+\mu_{i2})=\sum_{i=1}^{n}(\nu_{i1}+\nu_{i2})$.

Noting that

$$\delta=(1/2)\left\{\bar{d}+(1/n)\sum_{i=1}^{n}(\nu_{i,a_i}-\mu_{i,3-a_i})\right\},$$

it is easy to see that (3) is equivalent to

(3′) $H_0: \sum_{i=1}^{n}(\nu_{i,a_i}-\mu_{i,3-a_i})=-n\bar{d}$,

once $d_1, \cdots, d_n$ are observed. Note that (3) or (3′) is weaker than (2) or (2′), and that, in this sense, (3) is more appropriate as a null hypothesis of no treatment difference than (2). Although (3) is more appealing as a null hypothesis, we cannot induce finite sample space with discrete probability measure even under the null hypothesis, which is key to the main idea of FRT. Therefore, for the practical purpose, we need to consider a hypothesis which is stronger than (3) but weaker than (2). Now, let

(4) $H_0: \nu_{i,a_i}-\mu_{i,3-a_i}=e_i$, $\underset{\sim}{e} \in S_{\underset{\sim}{d}}$ for all i,

where $\underset{\sim}{e}=(e_1, \cdots, e_n)$ denote a permutation of elements in $-\underset{\sim}{d}=(-d_1, \cdots, -d_n)$ and $S_{\underset{\sim}{d}}$ is the collection of such permutations. We may call e_i's "assumed hidden differences" under the null hypothesis (4), noting that d_i's are observed differences. At this point, it becomes clear that (4) is a composite hypothesis.

The null distribution of $T=\sum_{i=1}^{n}D_i$ can be constructed, for each given $(e_1, \cdots, e_n)$, by generating random variables D_i independently in such a way that

$D_i=d_i$ with probability 1/2

$=e_i$ with probability 1/2,

for $i=1,\cdots,n$. Such null distribution is needed for each member of $S_{\underline{d}}$. Thus, in order to specify a class of null distributions, the test statistic should be evaluated at $n!\times 2^n$ sample points, where $n!$ corresponds to the number of all permutations of negative observed differences and 2^n corresponds to the number of all possible combinations within pairs of observed and "assumed hidden differences". Note that $n!\times 2^n$ is a very large number even for moderate values of n:

n	2	3	4	5	10	15
$n!\times 2^n$	8	48	384	3840	3715891200	4.2850×10^{16}

In order to obtain the significance level, individual SL's are computed first with each null distribution of T since the null hypothesis contemplated is composite. And, then, the maximum of individual SL's is the SL associated with the observed value of T under (4). That is,

$$SL=\sup_{\underline{e}\in S_{\underline{d}}} P_0(T\geq t : e_1,\cdots,e_n),$$

where $P(\cdot : e_1,\cdots,e_n)$ is appropriately defined null distribution associated with $(e_1,\cdots,e_n)$. Note that the significance level of this test is, of course, larger than that of FRT. That makes sense since this test is for a weaker hypothesis.

Since computing all $n!\times 2^n$ possible values of T is an enormous job for moderately large n, even with a modern computer, it may not be practical to carry out this test routinely. However, FRT can be used limitedly in some cases: if FRT does not reject the "equality" of two treatments, then this test should not reject it.

ACKNOWLEDGEMENTS

The author is grateful to reviewers of the previous manuscript for clarifying many parts of presentation and especially for pointing out a critical logical error.

References

(1) Basu, D. (1980) "Randomization Analysis of Experimental Data: The Fisher Randomization Test." *J. Amer. Statist. Assoc.*, 75, 575~582.

(2) Cox, D.R., and Hinkley, D.V. (1974) *Theoretical Statistics.* London: Chapman and Hall.

(3) Fisher, R.A. (1960) *The Design of Experiments* (7th ed.). Edinburgh: Oliver and Boyd.

(4) Kempthorne, O., and Folks, J.L. (1971) *Probability, Statistics, and Data Analysis.* Ames: Iowa State University Press.

(5) Lane, D.A. (1980) "Comment"of Basu (1980). *J. Amer. Statist. Assoc.*, 75, 587~589.