Statistical Science
2017, Vol. 32, No. 3, 346–348
DOI: 10.1214/16-STS582
Main article DOI: 10.1214/16-STS571
© Institute of Mathematical Statistics, 2017

Understanding Ding's Apparent Paradox

Peter M. Aronow and Molly R. Offer-Westort

1. INTRODUCTION

We are grateful for the opportunity to comment on "A Paradox From Randomization-Based Causal Inference" (Ding, 2017), an interesting discussion of the properties of Fisher's randomization test (FRT) with a comparison to a Wald-type test based on a variance estimator originally proposed by Neyman (1990). The article illustrates that the use of a Wald-type test using Neyman's variance estimator and the FRT (with the difference in means as a test statistic) can lead to situations where the null of zero average causal effect (Neyman's null) is rejected, while the sharp null of zero individual causal effects (Fisher's null) is not. The article in large part attributes this apparent paradox, which persists asymptotically, to a difference in the implied variances of the reference distributions used to construct tests. In this comment, we seek to situate Ding's (2017) findings in established statistical results, and to explain how the apparent paradox is a direct consequence of two well-known results.

We summarize our contribution as follows: (i) Rao-type tests may be sub-optimal under nonlocal alternatives (Engle, 1984), and as Ding (2017) shows, the FRT is asymptotically equivalent to a Rao-type test assuming constant effects. Rather than the FRT, a Wald-type analogue to the FRT using the difference in means exists (considered by Freedman, 2008, Samii and Aronow, 2012, Gerber and Green, 2012, and Lin, 2013) is a valid test of Fisher's null, and does not suffer from the potential pathology of Rao-type tests. Furthermore, this Wald-type test—equivalent to a pooled-variance two-sample z-test—is asymptotically equivalent to the FRT under local alternatives. Reichardt and Gollob (1999) discuss this point, with reference to Mosteller and Rourke (1973). (ii) As highlighted by Pratt (1964), Romano (1990) and Freedman (2008), the behavior of tests under incorrect working assumptions may depend on the joint distribution of the data. Analogous to the Behrens–Fisher problem, a test of the null of no average effect that assumes there is no effect on the variance may be more or less powerful than a test of no average effect that makes no such additional assumption. We illustrate this by comparing the Wald-type analogue to the FRT to the standard Wald-type test of Neyman's null. Combining (i) and (ii), we see that Ding's (2017) apparent paradox follows from the use of a suboptimal test under nonlocal alternatives and the well-known behavior of tests of joint hypotheses when one of the constituent hypotheses is false. (iii) We also note an error in Ding's (2017) Theorem 7 and suggest a refinement that is correct at full generality.

Note, we consider only asymptotic results, and do not discuss the finite N differences between exact tests and tests that are only known to be asymptotically valid. Finite N differences may account in part for Ding's (2017) simulation results; while perhaps of practical importance, such differences are in our view theoretically uninteresting as the source of a potential paradox.

2. A WALD-TYPE ANALOGUE TO THE FISHER RANDOMIZATION TEST

Ding (2017) demonstrates the asymptotic equivalence of the FRT and Rao's score test as applied to a linear model assuming homoskedasticity. This result builds on prior findings from Romano (1990), Freedman (2008) and Samii and Aronow (2012). Freedman (2008) considers the operating characteristics of a Wald test assuming a homoskedastic linear model; Samii and Aronow (2012) give the implied variance of this test a randomization basis, by establishing that the implied variance is equivalent to the randomization distribution of the difference-in-means estimator if the treatment effect is assumed to be constant and equal to the observed estimate. Gerber and Green (2012) also advocate for this Wald-type approach in practice. Ding (2017) further reproduces Lin (2013) and Samii and Aronow's (2012) result, showing

Peter M. Aronow is Associate Professor, Departments of Political Science and Biostatistics, Yale University, 77 Prospect Street, New Haven, Connecticut 06520, USA (e-mail: peter.aronow@yale.edu). Molly R. Offer-Westort is a graduate student, Departments of Political Science and Statistics and Data Science, Yale University, 77 Prospect Street, New Haven, Connecticut 06520, USA (e-mail: molly.offer-westort@yale.edu).