1. Notation; Review

1.1. Treatment effects and treatment assignments. There are \( I \) pairs, \( i = 1, \ldots, I \), of two subjects, \( j = 1, 2 \), one treated, \( Z_{ij} = 1 \), the other control, \( Z_{ij} = 0 \), with \( Z_{i1} + Z_{i2} = 1 \), matched for \( x \), so \( x_{i1} = x_{i2} \) but possibly differing in an unmeasured covariate, \( u_{i1} \neq u_{i2} \). As in Neyman (1923) & Rubin (1973), subject \( ij \) has potential responses \( r_{Tij} \) if treated \( Z_{ij} = 1 \), or \( r_{Cij} \) if control, \( Z_{ij} = 0 \), so the observed response from \( ij \) is \( R_{ij} = Z_{ij} r_{Tij} + (1 - Z_{ij}) r_{Cij} \), and the treatment effect, \( r_{Tij} - r_{Cij} \), is not observed. Fisher’s (1935) sharp null hypothesis of no treatment effect is \( H_0 : r_{Tij} = r_{Cij}, \forall ij \). Write \( \mathcal{F} = \{ (r_{Tij}, r_{Cij}, x_{ij}, u_{ij}), i = 1, \ldots, I, j = 1, 2 \} \). If there is an additive effect, \( r_{Tij} - r_{Cij} = \tau, \forall ij \), then the \( i \text{th} \) treated-minus-control difference in observed responses, \( Y_i = (Z_{i1} - Z_{i2}) (R_{i1} - R_{i2}) \), is

\[
(1.1) \quad Y_i = (Z_{i1} - Z_{i2}) (r_{C1} + Z_{i1} \tau - r_{C2} - Z_{i2} \tau) = \tau + \epsilon_i \quad \text{where} \quad \epsilon_i = (Z_{i1} - Z_{i2}) (r_{C1} - r_{C2})
\]

Write \( \Omega \) for the set of possible values of \( Z = (Z_{11}, Z_{12}, \ldots, Z_{I2})^T \), so \( z \in \Omega \) if \( z = (z_{11}, z_{12}, \ldots, z_{I2})^T \) with \( z_{ij} = 0 \) or \( z_{ij} = 1 \) and \( z_{i1} + z_{i2} = 1 \) for every \( i \). Write \( \mathcal{Z} \) for the event \( Z \in \Omega \).

1.2. General signed rank statistics testing no effect in a randomized experiment. In a randomized paired experiment, one subject in each pair is picked at random to receive treatment, the other receiving control, with independent assignments in distinct pairs, so \( \Pr (Z_{ij} = 1 \mid \mathcal{F}, Z) = 1/2, \forall ij \), and \( \Pr (Z = z \mid \mathcal{F}, Z) = 2^{-I} \) for \( z \in \Omega \). If Fisher’s \( H_0 \) were true, then \( Y_i = Y_{Ci} = (Z_{i1} - Z_{i2}) (r_{C1} - r_{C2}) \). Let \( q_i \geq 0 \) be a function of the \( |Y_i| \)’s such that \( q_i = 0 \) if \( |Y_i| = 0 \). Let \( \text{sgn} (y) = 1 \) or \( 0 \) for, respectively \( y > 0 \) or \( y \leq 0 \). A general signed rank statistic is \( T = \sum_{i=1}^{I} \text{sgn} (Y_i) q_i \). Wilcoxon’s signed rank statistic takes \( q_i \) equal to the rank of \( |Y_i| \) when \( |Y_i| > 0 \). The sign test takes \( q_i = 1 \) when \( |Y_i| > 0 \). Randomization creates the null distribution \( \Pr (T \mid \mathcal{F}, Z) \) of \( T \). Under \( H_0 \), the absolute difference \( |Y_i| = |Y_{Ci}| = |r_{C1} - r_{C2}| \) is fixed by conditioning on \( \mathcal{F} \), so \( q_i \) is also fixed, and \( \text{sgn} (Y_i) = 0 \) or \( 1 \) each with equal probability \( 1/2 \) if \( |Y_i| > 0 \), or \( \text{sgn} (Y_i) = 0 \) if \( |Y_i| = 0 \); therefore, \( \Pr (T \mid \mathcal{F}, Z) \) is the distribution of the sum of the \( I \) independent discrete random variables \( \text{sgn} \{ (Z_{i1} - Z_{i2}) (r_{C1} - r_{C2}) \} q_i \), taking values \( q_i \) or \( 0 \) with equal probabilities.

1.3. Sensitivity analysis in an observational study. A sensitivity analysis asks about the magnitude of departure from \( \Pr (Z_{ij} = 1 \mid \mathcal{F}, Z) = 1/2 \) that would need to be present to alter the qualitative conclusions of a study. A simple model for sensitivity analysis begins by assuming that in the population prior to matching, subjects have independent treatment assignments with
unknown probabilities, \(\pi_{ij} = \Pr (Z_{ij} = 1 \mid \mathcal{F})\), such that two subjects, say \(ij\) and \(ij'\), with the same observed covariates, \(x_{ij} = x_{ij'}\), may differ in their odds of treatment, \(\pi_{ij} / (1 - \pi_{ij})\) and \(\pi_{ij'} / (1 - \pi_{ij'})\), by at most a factor of \(\Gamma \geq 1\), and then restricts the distribution of \(Z\) to \(\Omega\) by conditioning on the event \(Z\); see Rosenbaum (2002,§4; 2011). This is the same as assuming

\[
\Pr (Z = z \mid \mathcal{F}, Z) = \frac{\exp (\gamma z^T u)}{\sum_{b \in \Omega} \exp (\gamma b^T u)} = \prod_{i=1}^{I} \frac{\exp (\gamma z_{i1} u_{11} + \gamma z_{i2} u_{22})}{\exp (\gamma u_{11}) + \exp (\gamma u_{22})}, \quad u \in [0, 1]^{2I},
\]

for \(z \in \Omega\), where \(\gamma = \log (\Gamma) \geq 0\), so that the \(I\) terms in the product in (1.2), namely \(\Pr (Z_{ij} = 1 \mid \mathcal{F}, Z) = \exp (\gamma u_{ij}) / \{\exp (\gamma u_{11}) + \exp (\gamma u_{22})\}\), are bounded below by \(1 / (1 + \Gamma)\) and above by \(\Gamma / (1 + \Gamma)\). For \(\Gamma = 1\) and \(\gamma = 0\), (1.2) equals the randomization distribution, \(\Pr (Z = z \mid \mathcal{F}, Z) = 2^{-I}\). Let \(\overline{T}_\Gamma\) be the sum of \(I\) independent random variables where the \(i\)th random variable takes the value \(q_i\) with probability \(\Gamma / (1 + \Gamma)\) and the value 0 with probability \(1 / (1 + \Gamma)\), and let \(\overline{T}_\Gamma\) be defined in the same way except with the roles of \(\Gamma / (1 + \Gamma)\) and \(1 / (1 + \Gamma)\) interchanged. It is straightforward to show (Rosenbaum 1987) that, under Fisher’s \(H_0\) and (1.2), the null distribution of \(\overline{T}_\Gamma\) satisfies

\[
\Pr (\overline{T}_\Gamma \geq k \mid \mathcal{F}, Z) \leq \Pr (T \geq k \mid \mathcal{F}, Z) \leq \Pr (\overline{T}_\Gamma \geq k \mid \mathcal{F}, Z) \quad \text{for all} \quad u \in [0, 1]^{2I},
\]

and the bounds are sharp, being attained for particular \(u \in [0, 1]^{2I}\), so that the bounds cannot be improved without further information about \(u\). Under mild conditions on the score function \(q_i\), as \(I \to \infty\), the probability \(\Pr (\overline{T}_\Gamma \geq k \mid \mathcal{F}, Z)\) may be approximated using a Normal approximation to the distribution of \(\overline{T}_\Gamma\) with \(E (\overline{T}_\Gamma \mid \mathcal{F}, Z) = \Gamma \sum_{i=1}^{I} q_i\) and \(\text{var} (\overline{T}_\Gamma \mid \mathcal{F}, Z) = \frac{\Gamma}{(1+\Gamma)^2} \sum_{i=1}^{I} q_i^2\) with an analogous approximation for \(\overline{T}_\Gamma\).

2. POWER OF A SENSITIVITY ANALYSIS: DESIGN SENSITIVITY

For fixed \(\Gamma \geq 1\), (1.3) yields an upper bound on the one-sided significance level. For fixed \(\Gamma \geq 1\), the power of an \(\alpha\) level sensitivity analysis is the probability that this upper bound will be less than or equal to \(\alpha\); see Rosenbaum (2004). For \(\Gamma = 1\), this is the power of a randomization test. Power is computed under some model for the generation of \(\mathcal{F}\) and \(Z\). In the ‘favorable situation’ there is a treatment effect and no bias from unmeasured covariates, and we hope to report insensitivity to unmeasured bias. In the favorable situation, \(Z\) is randomized, \(Z_{i1} - Z_{i2} = \pm 1\) with equal conditional probabilities of \(\frac{1}{2}\) given \((\mathcal{F}, Z)\), and \(\mathcal{F}\) is produced under some model for treatment effects. In the discussion here, the \(Y_i\) in (1.1) are independent and identically distributed with a distribution \(G (\cdot)\) with density \(g (\cdot)\); e.g., \(Y_i \sim N (\tau, 1)\). Not knowing that we are in the favorable situation, we perform a sensitivity analysis hoping to report a high degree of insensitivity when the favorable situation does arise.

Given a test statistic and model generating \(\mathcal{F}\), there is a value \(\Gamma\), the design sensitivity, such that, as \(I \to \infty\), the power of the sensitivity analysis tends to 1 if performed with \(\Gamma < \Gamma\) and to 0 if performed with \(\Gamma > \Gamma\). In large sample sizes, this test statistic can distinguish this model for \(\mathcal{F}\) from all biases smaller than \(\Gamma\) but not from some biases larger than \(\Gamma\).

3. A NEW U-STATISTIC

Fix an integer \(m\) with \(1 \leq m \leq I\), write \(K\) for the set containing the \(I\) sequences \(\mathcal{T} = (i_1, \ldots, i_m)\) of \(m\) distinct integers \(1 \leq i_1 < \cdots < i_m \leq I\), and write \(Y_{\mathcal{T}} = (Y_{i_1}, \ldots, Y_{i_m})\). A U-statistic (Hoeffding 1948) has the form \(T = \frac{1}{m!} \sum_{\mathcal{T} \in K} h (Y_{\mathcal{T}})\) where the kernel, \(h (\cdot)\), is a symmetric
function of its $m$ arguments $(Y_{i_1}, \ldots, Y_{i_m})$. For $I = \langle i_1, \ldots, i_m \rangle \in K$, sort $Y_{i_1}, \ldots, Y_{i_m}$, into increasing order by their absolute values, $0 < |Y_{i_1}| < \cdots < |Y_{i_m}|$. Fix two integers $m_1, \overline{m}$ with $1 \leq m_1 \leq \overline{m} \leq m$. In the new u-statistic, $h(Y_I)$ is the number of positive differences among $Y_{i_1}, Y_{i_2}, \ldots, Y_{i_m}$, so $h(Y_I)$ is an integer in $\{0, 1, \ldots, \overline{m} - m_1 + 1\}$. If $m = \overline{m} = m_1 = 1$, then $h(Y_I) = \text{sgn}(Y_{i_1}) = \text{sgn}(Y_{i_1})$ and $T$ is the sign statistic, whereas if $m = \overline{m} = m_1 = 2$, then $h(Y_I) = \text{sgn}(Y_{i_1})$, and $T$ is the U-statistic that closely approximates Wilcoxon’s signed rank statistic. If $m = \overline{m} = m$, then $h(Y_I) = \text{sgn}(Y_{i_1})$ and $T$ is Stephenson’s (1981) statistic which has excellent power when only a subset of treated subjects respond to treatment; see Conover and Salsburg (1988) and Rosenbaum (2010, DOS, §16). With $m = 8$, the statistic $(m_1, m_1, m) = (8, 7, 8)$ has $h(Y_I) = \text{sgn}(Y_{i_1}) + \text{sgn}(Y_{i_2}) + \text{sgn}(Y_{i_3})$ with values 0, 1, 2. This U-statistic is a signed rank statistic with $q_i = \binom{I}{m}^{-1} \sum_{t=m}^{\overline{m}} \binom{a_t}{t} - \binom{I-a_t}{m-t}$ where $a_t$ is the rank of $|Y_i|$. 

### Table 2. Design sensitivities $\overline{\Gamma}$ with additive effect $\tau$. Errors are standard Normal, standard logistic or $t$-distributed.

<table>
<thead>
<tr>
<th>Errors Statistic</th>
<th>Normal $\tau = 1/2$</th>
<th>Logistic $\tau = 1$</th>
<th>$t$ with 4 df $\tau = 1$</th>
<th>$t$ with 3 df $\tau = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wilcoxon (8,7,8)</td>
<td>3.2 3.9 6.8 6.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(8,6,7)</td>
<td>3.5 4.5 9.0 7.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(20,16,19)</td>
<td>4.9 5.6 10.1 7.8</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.1. **A formula for the design sensitivity.** Assume $Y_i$ are iid from some distribution $G(\cdot)$ and there is no unobserved bias, $\text{Pr}(Z_{ij} | X, Z) = \frac{1}{2}$. Let $\theta = E(h(Y_I))$.

**Proposition:** The design sensitivity of the U-statistic $(m_1, m_1, \overline{m})$ is $\overline{\Gamma} = \theta / (\overline{m} - m_1 + 1 - \theta)$.

### 4. Testing one hypothesis twice

Suppose there are two tests of $H_0$ using the same $Y_i$ but different scores, $T = \sum_{i=1}^I \text{sgn}(Y_i) q_i$ and $T' = \sum_{i=1}^I \text{sgn}(Y_i) q_i'$, where $q_i \geq 0$ and $q_i' \geq 0$. It is important here that $T$ and $T'$ both receive a nonnegative contribution whenever $\text{sgn}(Y_i) = 1$ or $Y_i > 0$. In the sensitivity analysis, there are now two upper bound random variables, $\overline{T}_I$ and $\overline{T}_I'$, which are each the sum of $I$ independent random variables, both taking the value 0 with probability $1 / (1 + \Gamma)$ or else the values $q_i$ and $q_i'$ with probability $\Gamma / (1 + \Gamma)$. Under mild conditions on the scores, $q_i$ and $q_i'$, as $I \to \infty$, the
joint distribution of $\overline{T}$ and $\overline{T}$ tends to a bivariate Normal distribution with known, typically high correlation $\rho$. The bounding statistics $\left(\overline{T}, \overline{T}'\right)$ are jointly stochastically larger than $\left(\overline{T}, \overline{T}'\right)$. Hence, the required computations when you pick the least sensitive of two tests involve straightforward manipulations with the bivariate Normal distribution. With $L$ tests, $L \geq 2$, the computations involve an $L$-variate Normal distribution. Compute using the mvtnorm package in R. Joint method has design sensitivity equal to the maximum of the $L$ design sensitivities of the $L$ tests.

Related software: http://www-stat.wharton.upenn.edu/~rosenbap/software.html

5. References


Conover, W. J. & Salsburg, D. S. (1988), “Locally most powerful tests for treatment effects when only a subset can be expected to `respond’ to treatment,” Biometrics, 44, 189-196.


Department of Statistics, University of Pennsylvania, Philadelphia PA 19104