Unconditional Exact Tests in the exact2x2 R package

Michael P. Fay, Sally A. Hunsberger

August 2, 2020

Summary

These notes describe the calculations for the uncondExact2x2 function in the exact2x2 R package. This function does unconditional exact tests for the two sample binomial problem. It has options for serval different test statistics, mid p-value adjustments, and Berger and Boos adjustments.

1 Definition and Calculation of the Unconditional Exact Tests

1.1 Defining the General Method

Let $\mathbf{X} = [X_1, X_2]$ with $X_a \sim Binom(n_a, \theta_a)$ for $a = 1, 2$. Suppose we are interested in $\beta = b(\theta)$, where $b(\theta)$ is some function of θ_1 and θ_2. Common examples are the difference, $\beta_d = \theta_2 - \theta_1$, the ratio, $\beta_r = \theta_2/\theta_1$, and the odds ratio, $\beta_{or} = \{\theta_2(1-\theta_1)\} / \{\theta_1(1-\theta_2)\}$.

We want to test hypotheses of the form $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$, where Θ_0 and Θ_1 are the set of all possible values of $[\theta_1, \theta_2]$ under the null hypothesis or the alternative hypothesis, repspectively. It is convenient to write Θ_0 and Θ_1 in terms of β. For example,

$$\Theta_0 \quad = \quad \{\theta : b(\theta) = \beta_0\}$$

For this example, instead of writing the null hypothesis as $H_0 : \theta \in \Theta_0$, we write it in terms of $\beta = b(\theta)$ as $H_0 : \beta = \beta_0$. We are generally interested in three classes of hypotheses: two-sided hypotheses,

$$\begin{aligned} H_0 : &\quad \beta = \beta_0 \\ H_1 : &\quad \beta \neq \beta_0 \end{aligned}$$

or one of the one-sided hypotheses,

Alternative is Less	Alternative is Greater
$H_0 : \beta \geq \beta_0$	$H_0 : \beta \leq \beta_0$
$H_1 : \beta < \beta_0$	$H_1 : \beta > \beta_0$.

First consider parmtype="difference". Let $T(\mathbf{X})$ be some test statistic, where larger values suggest that θ_2 is larger than θ_1. Then a valid (i.e., exact) p-value for testing $H_0 : \beta \geq \beta_0$ is

$$p_U(\mathbf{x}, \beta_0) \quad = \quad \sup_{\theta : b(\theta) \geq \beta_0} Pr_\theta \left[T(\mathbf{X}) \leq T(\mathbf{x}) \right].$$

For testing $H_0 : \beta \leq \beta_0$ the p-value is

$$p_L(\mathbf{x}, \beta_0) \quad = \quad \sup_{\theta : b(\theta) \leq \beta_0} Pr_\theta \left[T(\mathbf{X}) \geq T(\mathbf{x}) \right].$$

When parmtype='ratio' then when $\mathbf{x} = [0,0]$ there is no information about the ratio and we define the p-value as 1. Similarly, when parmtype='oddsratio' and $\mathbf{x} = [0,0]$ or $\mathbf{x} = [n_1, n_2]$ there is no information about the odds ratio and we define the p-value as 1, and we do not calculate probabilities in p-value calculations over values with no information. Specifically, let $\mathcal{X}_I$ denote the set of $\mathbf{X}$ values with information about β. Then if $\mathbf{x} \notin \mathcal{X}_I$ set $p_U(\mathbf{x}, \beta_0)$ and $p_L(\mathbf{x}, \beta_0)$ to 1, otherwise let $p_U(\mathbf{x}, \beta_0)$ be

$$\sup_{\theta: b(\theta) \geq \beta_0} P_\theta \left[T(\mathbf{X}) \leq T(\mathbf{x}) | \mathbf{X} \in \mathcal{X}_I \right] P_\theta \left[\mathbf{X} \in \mathcal{X}_I \right]$$

and analogously, let $p_L(\mathbf{x}, \beta_0)$ be

$$\sup_{\theta: b(\theta) \leq \beta_0} P_\theta \left[T(\mathbf{X}) \geq T(\mathbf{x}) | \mathbf{X} \in \mathcal{X}_I \right] P_\theta \left[\mathbf{X} \in \mathcal{X}_I \right].$$

Since we never reject when $\mathbf{x} \notin \mathcal{X}_I$, these definitions give valid p-values, and additionally when $\mathbf{x} \notin \mathcal{X}_I$ we do not need to define $T(\mathbf{x})$.

The tsmethod option gives two ways to calculate the two-sided p-value. The default option is 'central' to give a central p-value, which is

$$\begin{aligned} p_{ts}(\mathbf{x}, \beta_0) &= p_{central}(\mathbf{x}, \beta_0) \\ &= \min \left\{ 1, 2p_U(\mathbf{x}, \beta_0), 2p_L(\mathbf{x}, \beta_0) \right\} \end{aligned}$$

The second options is tsmethod='square'. For this option, we square the test statistic, $T(\mathbf{x})$, defined in the next section, and define the p-value as

$$\begin{aligned} p_{ts}(\mathbf{x}, \beta_0) &= p_{square}(\mathbf{x}, \beta_0) \\ &= \begin{cases} \sup_{\theta \in \Theta_0} Pr_\theta \left[T^2(\mathbf{X}) \geq T^2(\mathbf{x}) \right] & \text{(for parmtype="difference")} \\ \sup_{\theta \in \Theta_0} Pr_\theta \left[T^2(\mathbf{X}) \geq T^2(\mathbf{x}) | X \in \mathcal{X}_I \right] Pr_\theta \left[X \in \mathcal{X}_I \right] & \text{(otherwise).} \end{cases} \end{aligned}$$

Since the probability expression only depends on the ordering, and since the ordering of the square of $T(\mathbf{X})$ is the same as the ordering of absolute value of $T(\mathbf{X})$, we can equivalently write p_{square} in terms of absolute values.

These exact p-values are necessarily conservative because for most $\theta \in \Theta_0$ we have

$$Pr_\theta \left[p_U(\mathbf{X}, \beta_0) \leq \alpha \right] < \alpha.$$

A less conservative approach, *but one that is no longer valid (i.e., no longer exact)*, is to use a mid-p value. For example, the mid-p value associated with p_U is

$$p_{Umid}(\mathbf{x}, \Theta_0) = \sup_{\theta: b(\theta) \geq \beta_0} \left\{ Pr_\theta \left[T(\mathbf{X}) < T(\mathbf{x}) \right] + \frac{1}{2} Pr_\theta \left[T(\mathbf{X}) = T(\mathbf{x}) \right] \right\}.$$

Other mid p-values are defined analogously.

1.2 Options for $T(\mathbf{x})$

1.2.1 Overview

We now give the $T(\mathbf{x})$ function that is defined by three options: parmtype, nullparm , and method. The option parmtype is one of 'difference', 'ratio' or 'odds ratio', and it determines the parameter associated with the confidence interval. The option nullparm defines β_0. The default for nullparm=NULL, which is replaced by $\beta_0 = 0$ for parmtype='difference' and $\beta_0 = 1$ for parmtype='ratio' or 'odds ratio'. Finally, method defines the type of T function:

simple: then T is an estimate of the parmtype using the estimates $\hat{\theta}_1 = x_1/n_1$ and $\hat{\theta}_2 = x_2/n_2$.

simpleTB: simple with a tie break. Ties in T using the simple method are broken based on variability, with larger variability further away from the null.

score: here T is based on a score statistic.

wald pooled: T is a Wald statistic on the difference in sample means using the pooled variance estimate.

wald unpooled: T is a Wald statistic on the difference in sample means using an **unpooled** variance estimate.

FisherAdj: T is a one-sided mid p-value using Fisher's exact test. Note that we create an exact unconditional test using the ordering by the mid p-value, so the test is valid (or exact), even though the mid p-values when used as p-values directly are not necessarily valid.

1.2.2 Simple: Difference

When method='simple' and parmtype='difference' we have,

$$T(\mathbf{x}) = T([x_1, x_2]) \quad = \quad \frac{x_2}{n_2} - \frac{x_1}{n_1} - \beta_0$$

The order does not change as β_0 changes.

1.2.3 Simple with Tie Break: Difference

When method='simpleTB' and parmtype='difference' and tsmethod='central' we use $T(\mathbf{x})$ from the previous subsection, then break ties by ordering by $T^*(\mathbf{x})$ *within* each tied value for $T(\mathbf{x})$, where

$$T^*(\mathbf{x}) \quad = \quad \frac{\hat{\theta}_2 - \hat{\theta}_1}{\sqrt{\frac{\hat{\theta}_1(1-\hat{\theta}_1)}{n_1} + \frac{\hat{\theta}_2(1-\hat{\theta}_2)}{n_2}}}$$

where $\hat{\theta}_1 = x_1/n_1$ and $\hat{\theta}_2 = x_2/n_2$. If T^* gives a ratio of 0/0 then it is set to 0.

The idea behind T^* is that with each $\hat{\beta}_d = \hat{\theta}_2 - \hat{\theta}_1$ value, values with lower variability are more extreme (i.e., ranked higher when $\hat{\beta}_d$ is positive and ranked lower when $\hat{\beta}_d$ is negative). We do not subtract β_0 from the numerator, because we do not want the order to change for different hypotheses, which makes calculations more difficult and could possibly lead to non-unified inferences (e.g., reject the null at level α but the $1-\alpha$ CI for β_d includes 0).

1.2.4 Score:Difference

When method=‘score’ and parmtype=‘difference’ we have,

$$T([x_1, x_2]) \quad = \quad \frac{\frac{x_2}{n_2} - \frac{x_1}{n_1} - \beta_0}{\sqrt{\tilde{\theta}_1(1-\tilde{\theta}_1)/n_1 + \tilde{\theta}_2(1-\tilde{\theta}_2)/n_2}},$$

where $\tilde{\theta}_1$ and $\tilde{\theta}_2$ are the maximum likelihood estimates of θ_1 and θ_2 under the restriction that $b(\theta) = \beta_0$. See the code of constMLE.difference for the formula, or the Appendix of Farrington and Manning (1990).

1.2.5 Wald-Pooled: Difference

When method=‘wald-pooled’ and parmtype=‘difference’ we have,

$$T([x_1, x_2]) \quad = \quad \frac{\hat{\theta}_2 - \hat{\theta}_1 - \beta_0}{\sqrt{\hat{\theta}(1-\hat{\theta})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}},$$

where $\hat{\theta}_1 = x_1/n_1$ and $\hat{\theta}_2 = x_2/n_2$ and $\hat{\theta} = (x_1 + x_2)/(n_1 + n_2)$. If T gives a ratio of 0/0 then it is set to 0.

1.2.6 Wald-Unpooled: Difference

When method=‘wald-unpooled’ and parmtype=‘difference’ we have,

$$T([x_1, x_2]) \quad = \quad \frac{\hat{\theta}_2 - \hat{\theta}_1 - \beta_0}{\sqrt{\hat{\theta}_1(1-\hat{\theta}_1)/n_1 + \hat{\theta}_2(1-\hat{\theta}_2)/n_2}},$$

where $\hat{\theta}_1 = x_1/n_1$ and $\hat{\theta}_2 = x_2/n_2$. If T gives a ratio of 0/0 then it is set to 0.

1.2.7 Simple: Ratio

When method=‘simple’ and parmtype=‘ratio’ we have,

$$\begin{aligned} T(\mathbf{x}) = T([x_1, x_2]) \quad &= \quad \log\left(\frac{\hat{\theta}_2}{\beta_0 \hat{\theta}_1}\right) \\ &= \quad \log(\hat{\theta}_2) - \log(\hat{\theta}_1) - \log(\beta_0), \end{aligned}$$

where $\hat{\theta}_a = x_a/n_a$ for $a = 1, 2$. Note $\log(0) \equiv \infty$ and $log(0) - \log(0) \equiv NA$. We do not need to define NA values since $x = [0, 0]$ has no information (see Section 1.1).

1.2.8 Simple with Tie Break: Ratio

When method='simpleTB' and parmtype='ratio' we used $T(\mathbf{x})$ from the previous subsection, then break ties by ordering by $T^*(\mathbf{x})$ *within* each tied value for $T(\mathbf{x})$, where

$$T^*(\mathbf{x}) = \begin{cases} x_2 & \text{if } x_1 = 0 \text{ and } x_2 > 0 \\ 1/x_1 & \text{if } x_1 > 0 \text{ and } x_2 = 0 \\ 0 & \text{if } x_1 = n_1 \text{ and } x_2 = n_2 \\ \frac{\log(\hat{\theta}_2) - \log(\hat{\theta}_1)}{\sqrt{\frac{1}{x_1} - \frac{1}{n_1} + \frac{1}{x_2} - \frac{1}{n_2}}} & \text{if } x_1 > 0 \text{ and } x_2 > 0 \text{ and not}(x_1 = n_1 \text{ and } x_2 = n_2) \end{cases}$$

where $\hat{\theta}_1 = x_1/n_1$ and $\hat{\theta}_2 = x_2/n_2$.

In words, when $x_1/n_1 = \hat{\theta}_1 = 0$ and $x_2 > 0$ then $T(\mathbf{x}) = -\infty$ and we order by x_2; otherwise when we order $x_2/n_2 = \hat{\theta}_2 = 0$ and $x_1 > 0$ then $T(\mathbf{x}) = \infty$ and we order by $1/x_1$; otherwise when $\hat{\theta}_1 = \hat{\theta}_2 = 1$ we do not break the ties (by setting $T^*(\mathbf{x}) = 0$); otherwise for each $\log(\hat{\beta}_r) = \log(\hat{\theta}_2/\hat{\theta}_1)$ value, we rank values with lower variability are more extreme (i.e., ranked higher when $\hat{\beta}_r > 1$ and ranked lower when $\hat{\beta}_r < 1$ is negative). The variance formula comes from the variance estimate of the $\log(\hat{\beta}_r)$. Fleiss, Levin, and Paik (2003, p. 132, equation 6.112, except there is a typo) give the variance expression,

$$var(\log(\hat{\beta}_r)) \approx \sqrt{\frac{n_1 - x_1}{x_1 n_1} + \frac{n_2 - x_2}{x_2 n_2}} = \sqrt{\frac{1}{x_1} - \frac{1}{n_1} + \frac{1}{x_2} - \frac{1}{n_2}}.$$

We do not subtract $\log(\beta_0)$ from the numerator in the $T^*(|bfx)$ function to keep it simple.

1.2.9 Score: Ratio

When method='score' and parmtype='ratio' we have,

$$T([x_1, x_2]) = \frac{\hat{\theta}_2 - \hat{\theta}_1 \beta_0}{\sqrt{\beta_0 \tilde{\theta}_1 (1 - \tilde{\theta}_1)/n_1 + \tilde{\theta}_2 (1 - \tilde{\theta}_2)/n_2}},$$

where $\tilde{\theta}_1$ and $\tilde{\theta}_2$ are the maximum likelihood estimates of θ_1 and θ_2 under the restriction that $\beta_r = b(\theta) = \beta_0$; for the formula for $\tilde{\theta}_a$ for $a = 1, 2$, see either the constrMLE.ratio, Miettinen and Nurminen (1985).

1.2.10 Simple: Odds Ratio

When method='simple' and parmtype='odds ratio' we have,

$$T(\mathbf{x}) = T([x_1, x_2]) = \log\left(\frac{\hat{\theta}_2 (1 - \hat{\theta}_1)}{\beta_0 \hat{\theta}_1 (1 - \hat{\theta}_2)}\right),$$

where $\hat{\theta}_a = x_a/n_a$ for $a = 1, 2$.

1.2.11 Simple with Tie Break: Odds Ratio

When method=‘simpleTB’ and parmtype=‘oddsratio’ we used $T(\mathbf{x})$ from the previous subsection, then break ties by ordering by $T^*(\mathbf{x})$ *within* each tied value for $T(\mathbf{x})$, where

$$T^*(\mathbf{x}) = \begin{cases} x_2 & \text{if } x_1 = 0 \text{ or } x_2 = n_2 \\ 1/x_1 & \text{if } x_1 = n_1 \text{ or } x_2 = 0 \\ \frac{\log(x_2) - \log(n_2 - x_2) - \log(x_1) + \log(n_1 - x_1)}{\sqrt{\frac{1}{x_1} + \frac{1}{n_1 - x_1} + \frac{1}{x_2} + \frac{1}{n_2 - x_2}}} & \text{otherwise} \end{cases}$$

where $\hat{\theta}_1 = x_1/n_1$ and $\hat{\theta}_2 = x_2/n_2$.

In words, when $\hat{\beta}_{or} = \infty$ then we order by x_2; otherwise when $\hat{\beta}_{or} = -\infty$ then we order by $1/x_1$; otherwise for each $\log(\hat{\beta}_{or})$ value, we rank values with lower variability are more extreme (i.e., ranked higher when $\hat{\beta}_r > 1$ and ranked lower when $\hat{\beta}_r < 1$ is negative). The variance formula comes from the variance estimate of the $\log(\hat{\beta}_{or})$. Fleiss, Levin, and Paik (2003, p. 102, equation 6.19) give the variance estimate for $var(\hat{\beta}_{or})$, and using the delta method, the estimate for $var(\log(\hat{\beta}_{or}))$ is

$$var(\log(\hat{\beta}_{or})) \approx \sqrt{\frac{1}{x_1} + \frac{1}{n_1 - x_1} + \frac{1}{x_2} + \frac{1}{n_2 - x_2}}.$$

We do not subtract $\log(\beta_0)$ from the numerator to keep it simple.

1.2.12 Score: Odds Ratio

When method=‘score’ and parmtype=‘oddsratio’ we use (see Agresti and Min, 2002, p. 381, except we do not square the statistic because we want to allow one-sided inferences),

$$T([x_1, x_2]) = \left\{ n_2 \left(\frac{x_2}{n_2} - \tilde{\theta}_2 \right) \right\} \sqrt{\frac{1}{n_1 \tilde{\theta}_1 (1 - \tilde{\theta}_1)} + \frac{1}{n_2 \tilde{\theta}_2 (1 - \tilde{\theta}_2)}},$$

where $\tilde{\theta}_1$ and $\tilde{\theta}_2$ are the maximum likelihood estimates of θ_1 and θ_2 under the restriction that

$$\tilde{\beta}_{or} = \frac{\tilde{\theta}_2 (1 - \tilde{\theta}_1)}{\tilde{\theta}_1 (1 - \tilde{\theta}_2)} = \beta_0.$$

For the formula for $\tilde{\theta}_a$ for $a = 1, 2$, see either the function constrMLE.oddsratio or Miettinen and Nurminen (1985).

1.2.13 FisherAdj: Difference, Ratio, or Odds Ratio

When method=‘FisherAdj’ we order by the mid p-value from a one-sided Fisher’s exact test. We do not change the ordering as the β_0 changes, so it can be used with any parmtype.

Using the phyper and dhyper functions for the hypergeometric distribution, this becomes:

$$T([x_1, x_2]) = \text{phyper}(x2, n2, n1, x2 + x1) - 0.5 * \text{dhyper}(x2, n2, n1, x1 + x2)$$

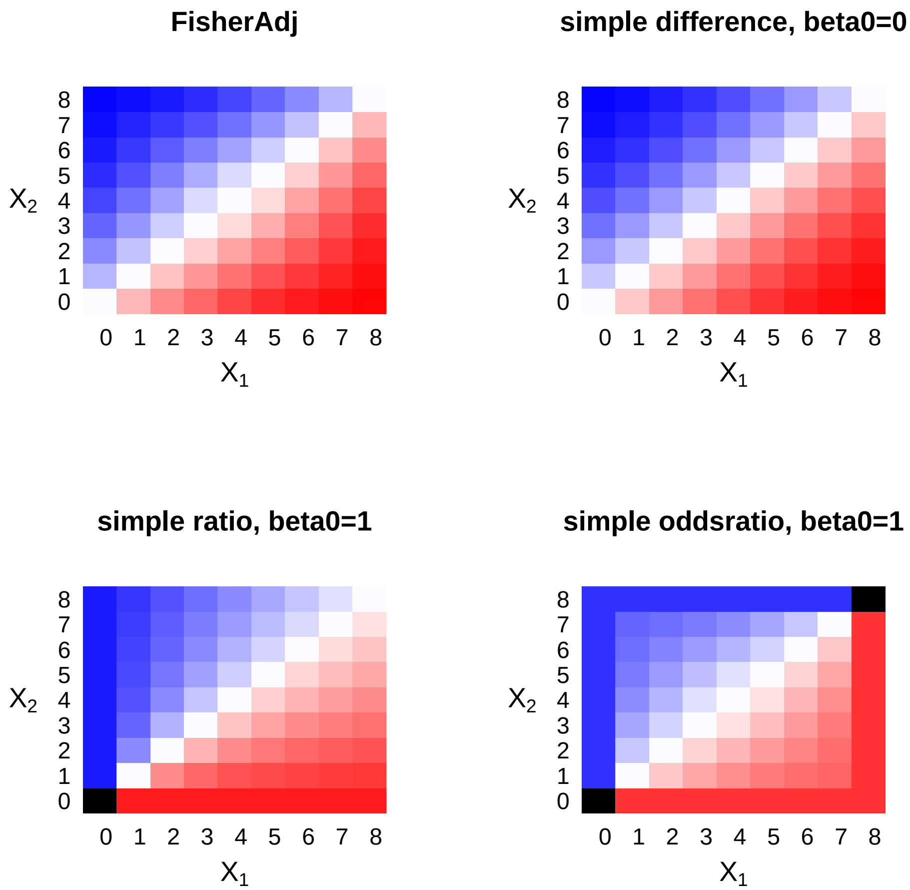

Figure 1: Plots of the orderings using plotT. Dark blue is highest, dard red is lowest, white is the middle, and black is no information. The default is method="FisherAdj" (same for all parmtypes), the method="simple" order by the plug-in estimates with sample proportions.

2 Comparing Orderings

In Figure 1 we show the default orderings and the method="simple" orderings for different values of parmtype.

In Figure 2 we show the similarity of several of the parmtype="difference" orderings.

The wald method gives a strange ordering at $x = (0, 0)$ and $x = (n_1, n_2)$ when β_0 is close to zero (see Figure 3).

When tsmethod="square" then a small difference in β_0 can make a big difference in the p-value (see Figure 4 for ordering difference, Figure 5 for a p-value example).

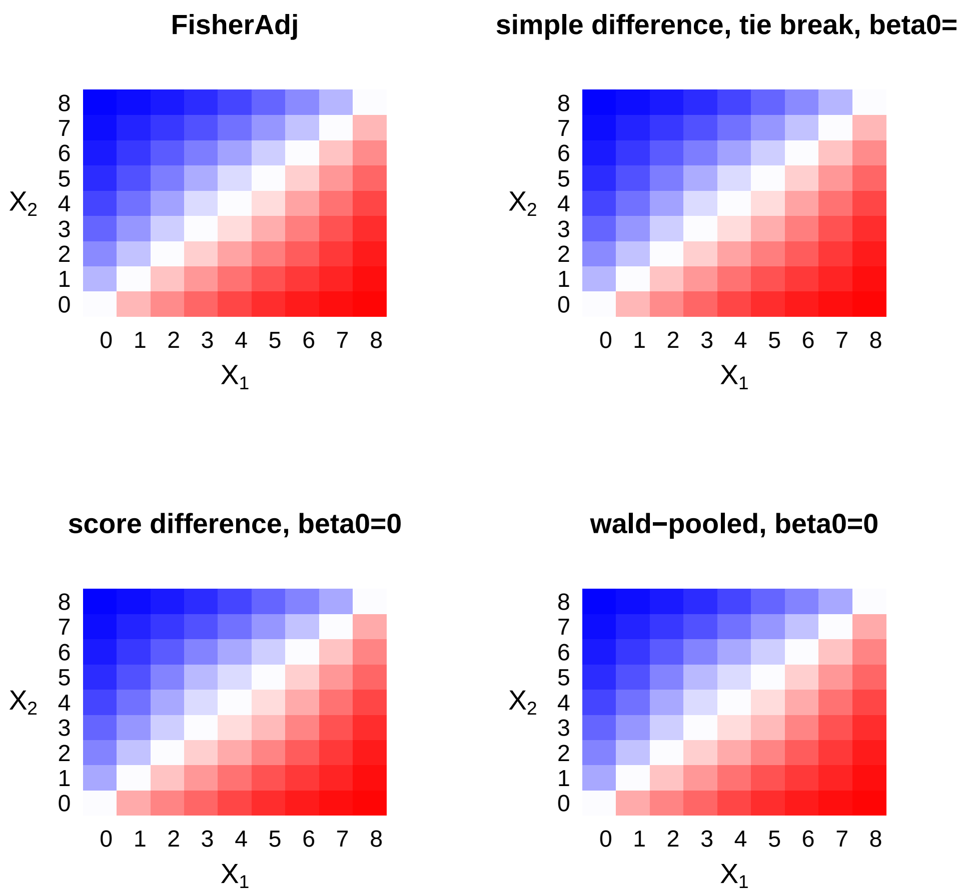

Figure 2: Plots of the orderings using plotT. Notice how the orderings are nearly the same for the 4 methods. The FisherAdj method has the advantage that it does not change with parmtype or β_0.

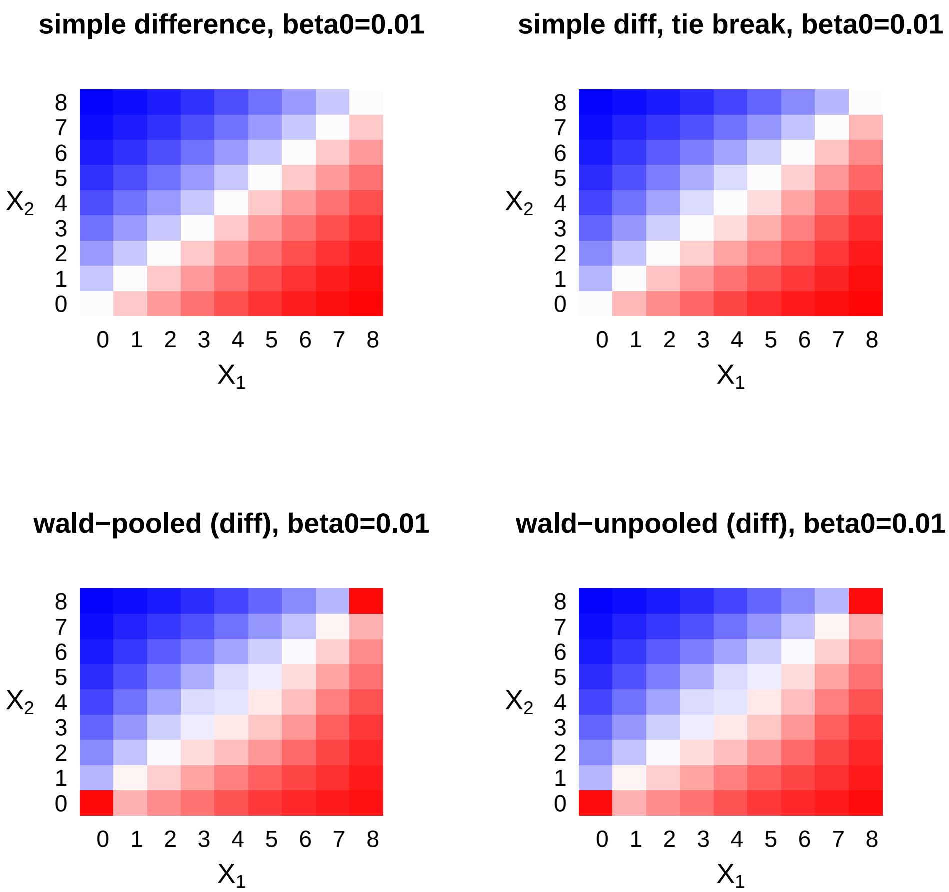

Figure 3: Plots of the orderings using `plotT`. Since we define $0/0 = 0$, when we have $\hat{\theta}_1 = \hat{\theta}_2$ and $\beta_0 = 0$ then the Wald methods give 0 (see Figure 1). But when $\beta_0 = 0.01$ these values at $x = (0, 0)$ and $x = (n_1, n_2)$ go to $-\infty$.

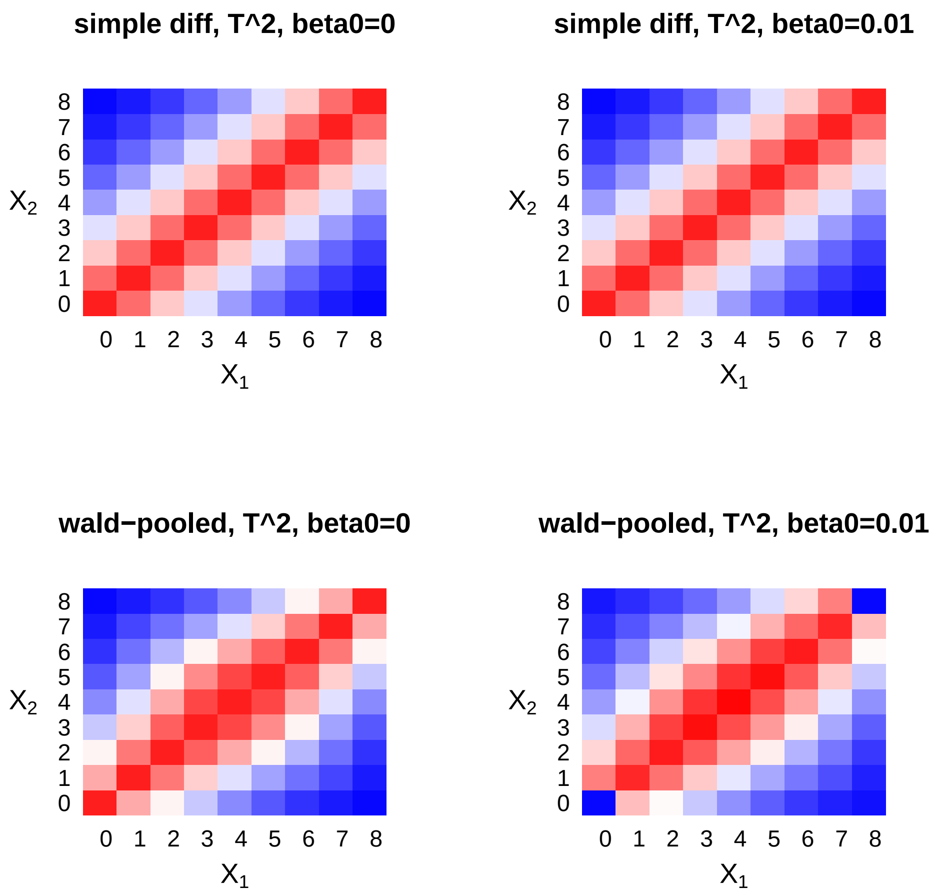

Figure 4: Plots of the orderings using plotT. Small changes in β_0 can have large changes in the ordering, because of the denominators equalling 0 at $x = (0, 0)$ and $x = (n_1, n_2)$.

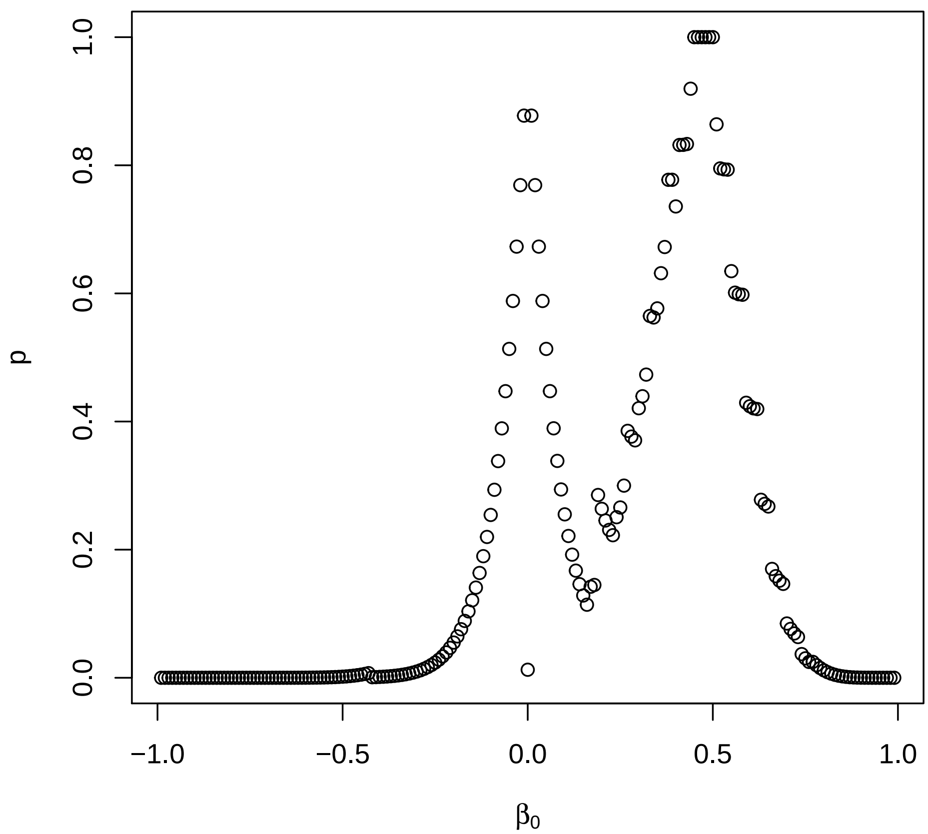

Figure 5: P-values from method="wald-pooled", tsmethod="square", and parmtype="difference" for the data $x_1/n_1 = 5/13$ and $x_2/n_2 = 12/14$. Notice the strange behaviour of the p-value at $\beta_0 = 0$. This is because the denominator at $x = (0, 0)$ and $x = (n_1, n_2)$ is 0 and 0/0 is defined as zero, and the p-value is defined as the sup over the sample space which can give very large probability mass at $x = (0, 0)$ or $x = (n_1, n_2)$.

3 Confidence Intervals

Then we can create $100(1-\alpha)\%$ confidence regions as the set of β_0 value that fail to reject the associated null hypothesis. For example,

$$C_{ts}(\mathbf{x}, 1-\alpha) \quad = \quad \{\beta : p_{ts}(\mathbf{x}, \beta) > \alpha\}$$

gives a "two-sided" confidence region. The region may not be an interval if the p-value function is not unimodal. This problem occurs with Fisher's exact test (the Fisher-Irwin version, or 'minlike' version). For central confidence regions we take the union of the one-sided confidence regions, in other words,

$$C_c(\mathbf{x}, 1-\alpha) = C_L(\mathbf{x}, 1-\alpha/2) \cup C_U(\mathbf{x}, 1-\alpha/2),$$

where C_L and C_U are the one-sided confidence regions,

$$C_L(\mathbf{x}, 1-\alpha/2) \quad = \quad \{\beta : p_L(\mathbf{x}, \beta) > \alpha/2\}$$

and

$$C_U(\mathbf{x}, 1-\alpha/2) \quad = \quad \{\beta : p_U(\mathbf{x}, \beta) > \alpha/2\}.$$

If the regions are intervals, and we let $L(\mathbf{x}, 1-\alpha/2) = \min C_L(\mathbf{x}, 1-\alpha/2)$ and $U(\mathbf{x}, 1-\alpha/2) = \max C_U(\mathbf{x}, 1-\alpha/2)$, then the central interval is

$$C_c(\mathbf{x}, 1-\alpha) \quad = \quad \{L(\mathbf{x}, 1-\alpha/2), U(\mathbf{x}, 1-\alpha/2)\}.$$

For the mid-p confidence regions, we replace the p-values with the mid-p values.

4 Berger and Boos Adjustment

The Berger-Boos (1994) adjustment is as follows. Do the usual unconditional exact test, but instead of taking the supremum over the entire null parameter space, we calculate a $100(1-\gamma)\%$ confidence region over the null space, and only search within that. The $100(1-\gamma)\%$ confidence region is the union of the $100(1-\gamma/2)$ exact central two-sided confidence interval for θ_1 and the analogous $100(1-\gamma/2)$ interval for θ_2. This is the method used by StatXact. Searching over that space gives anti-conservative p-values, so we turn those anti-conservative p-values into valid p-values by adding γ to them. For details see Berger and Boos (1994) or the StatXact manual.

5 The E+M Adjustment

Lloyd (2008) proposed another adjustment called the estimated and maximized $(E+M)$ p-value that can be applied to any ordering and any parmtype. In this method, we replace an ordering statistic, T, with T^*, where T^* is an estimated p-value when testing $H_0 : \beta \leq \beta_0$ (or the negative estimated p-value when testing $H_0 : \beta \geq \beta_0$). We estimate the p-value by plugging in $\hat{\theta}_0 = [\hat{\theta}_{10}, \hat{\theta}_{20}]$ instead of taking the supremum of θ under the null, where $\hat{\theta}_0$ is the maximum likelihood estimator of θ under the null hypothesis. For example, the approximation for p_L uses $\hat{p}_L(\mathbf{x}, \beta_0) = P_{\hat{\theta}_0}[T(\mathbf{X}) \leq T(\mathbf{x})]$. Then we "maximize" using $T^*(\mathbf{x}) = \hat{p}_L(\mathbf{x}, \beta_0)$ instead of T as the ordering function. For details see Loyd (2008).

References

Agresti, A and Min, Y (2002). Biostatistics 3: 379-386.

Berger, RL, and Boos, DD (1994). JASA 89: 1012-1016.

Farrington and Manning (1990). Statistics in Medicine 1447-1454.

Fay, MP, and Brittain, EH (2016). "Finite sample pointwise confidence intervals for a survival distribution with right-censored data." *Statistics in Medicine* 35: 2726-2740.

Fleiss, JL, Levin, B, Paik, MC (2003). *Statistical Methods for Rates and Proportions, Third edition.* Wiley: New York.

Lloyd, CJ (2008). Exact p-values for discrete models obtained by estimation and maximization. Australian and New Zealand Journal of Statistics 50(4): 329-345.

Miettinen and Nurminen (1985). Statistics in Medicine 213-226.