Notes: Multivariate Paired Permutation Tests

Kellie Ottoboni

July 11, 2016

1 Ordered variables

1.1 Binary paired outcomes

Suppose that we have two paired samples and binary outcomes, and we wish to test whether the frequency of outcomes is the same in the two samples. For example, we may be interested in whether the incidence of heart disease decreased after a cohort took some treatment. The McNemar test accomplishes this. We observe N pairs. The data may be written in a 2×2 contingency table, of the form

	Sample 1 +	Sample 1 -
Sample 2 +	a	b
Sample 2 -	c	d

where "+" and "-" denote the binary outcomes in each sample, a of the observations have "+" outcome in both samples, b of the observations have "+" outcome in sample 1 and "-" in sample 2, and so forth, with $a + b + c + d = N$.

Let (X_{i1}, X_{i2}) denote the ith pair of outcomes coming from sample 1 and sample 2, respectively. McNemar tests

$$H_0 : \mathbb{P}(X_{\cdot 1} = x) = \mathbb{P}(X_{\cdot 2} = x), x = 1, 2$$
$$H_1 : \mathbb{P}(X_{\cdot 1} = x) \neq \mathbb{P}(X_{\cdot 2} = x), \text{for some } x = 1, 2$$

The null further implies that $\mathbb{P}(X_{\cdot 1} = x \mid X_{\cdot 2} = y) = \mathbb{P}(X_{\cdot 2} = x \mid X_{\cdot 1} = y)$ Therefore, $\mathbb{P}((X_{i1}, X_{i2}) = (x, y)) = \mathbb{P}((X_{i1}, X_{i2}) = (y, x))$ for all $(x, y) \in \{+, -\}^2$. This means that within pairs, observations are exchangeable.

The test statistic is

$$\frac{(b - c)^2}{b + c}$$

Notice that the statistic doesn't depend on the row totals or a and d. To obtain the permutation distribution of the test statistic, we randomly exchange the order of observations within pairs, with probability 1/2, then count the numbers b^* and c^* in the permuted data. This is equivalent to doing a binomial test.

1.2 Ordered Paired Outcomes

The extension of McNemar's test to multiple outcomes is presented in Pesarin and Salmaso [2010], p. 87–88. Suppose that $X_{ij} \in \{1, 2, \dots, M\}$ for $j = 1, 2$ and we observe pairs $(X_{i1}, X_{i2}), i = 1, \dots, N$ as before. We can construct the same frequency table as above, except now it will have dimension $M \times M$ and elements $a_{k\ell} \equiv \sum_{i=1}^N \mathbb{I}(X_{i1} = k, X_{i2} = \ell)$.

There are several null hypotheses of interest:

- symmetry: if we define a suitable difference function $\phi(\cdot, \cdot)$, then we may wish to test whether $\phi(X_{i1}, X_{i2})$ is symmetric around 0. This requires us to define an ordering on the space of differences; then, we have a sort of binomial argument that conditional on the differences, they may take either sign with equal probability. TO DO: FLESH OUT

- marginal equality: under the null hypothesis that $X_1 \stackrel{d}{=} X_2$, we have $\mathbb{P}(X_{i1} = k) = \mathbb{P}(X_{i2} = k)$ for $i = 1, \dots, N$ and $k = 1, \dots, M$. Under this null, conditional on the observed data, $\mathbb{P}((X_{i1}, X_{i2})) = \mathbb{P}((X_{i2}, X_{i1}))$ for $i = 1, \dots, N$. Therefore, we can permute the order of observations within pairs to calibrate the null distribution of any test statistic. We may want to use a sum of McNemar-like statistics:

$$T = \sum_{k > \ell} \frac{(a_{k\ell} - a_{\ell k})^2}{a_{k\ell} + a_{\ell k}}$$

  or a sum of scaled binomials:

$$T = \sum_{k > \ell} \frac{a_{k\ell} - \nu_{k\ell}}{\sqrt{\nu_{k\ell}/4}}$$

  where $\nu_{k\ell} = \frac{a_{k\ell} + a_{\ell k}}{2}$ is fixed over permutations. The permutation statistic, T^*, is obtained by replacing the $a_{k\ell}$ and $a_{\ell k}$ by $a^*_{k\ell}$ and $a^*_{\ell k}$, which are computed after permuting within pairs.

- TO DO: TEST STATISTIC FROM STUART [1955]

TO DO: SHOW UNBIASED AND CONSISTENT

1.3 Multiple Binary Paired Outcomes

The extension of McNemar's test from a single measurement to multiple measurements is presented in Pesarin and Salmaso [2010], p. 289–292. Now instead of observing pairs of univariate observations (X_{i1}, X_{i2}), we observe K-vectors $(\mathbf{X}_{i1}, \mathbf{X}_{i2})$ with $\mathbf{X}_{ij} = (X_{ij1}, \dots, X_{ijK})$ for $j = 1, 2$. The idea is to test

$$H_0 : \mathbf{X}_{i1} \stackrel{d}{=} \mathbf{X}_{i2}$$

or equivalently

$$H_0 : \bigcap_{k=1}^K (X_{i1k} \stackrel{d}{=} X_{i2k}).$$

This reduces to K univariate problems, as above. We may conduct a partial test on each of the K components of the outcome and combine them with NPC.

1.4 Multiple Ordered Paired Outcomes

The extension to multiple ordered outcomes is presented in Pesarin and Salmaso [2010], p. 292–293. As in the previous section, we observe pairs of K-vectors $(\mathbf{X}_{i1}, \mathbf{X}_{i2})$ with $\mathbf{X}_{ij} = (X_{ij1}, \dots, X_{ijK})$ for $j = 1, 2$. Essentially, we combine the ideas behind the test for ordered paired outcomes and the test for multiple binary paired outcomes.

Consider the kth component of $\mathbf{X}_{\cdot}$, and suppose that it has V_k possible values. We may define a suitable difference function $\phi(\cdot, \cdot)$ and test for symmetry of $\phi(X_{i1k}, X_{i2k})$ around zero using the procedure described above. Alternatively, we can test the null hypothesis $X_{\cdot 1k} \overset{d}{=} X_{\cdot 2k}$ using the procedure from the previous section. Either way, we will compute a statistic T_k on the kth component.

This defines a partial test on each of the k components of the vectors $\mathbf{X}_{\cdot\cdot}$. NPC gives a test of the global null hypothesis $\cap_{k=1}^K H_{0k}$ against the global alternative $\cup_{k=1}^K H_{1k}$, where the null and alternative hypotheses may be specified as desired for each component. Without loss of generality, assume that we wish to test one-sided alternatives, with the convention that large test statistics are more extreme. The procedure is as follows:

1. Compute statistics $T_1, \dots, T_K$ on the observed dataset.
2. Repeat, for $b = 1, \dots, B$:
   (a) Randomly swap the order of $(\mathbf{X}_{i1}, \mathbf{X}_{i2})$ with probability 1/2 within pairs, independently across observations i, to get new data $\{(\mathbf{X}_{i1}^{b*}, \mathbf{X}_{i2}^{b*})\}_{i=1}^N$.
   (b) Compute test statistics $T_1^{b*}, \dots, T_K^{b*}$ from the permuted data.
3. Compute the empirical p-values for the observed partial tests: $p_k = \frac{1}{B}\sum_{d=1}^B \mathbb{I}(T_k \geq T_k^{d*}), k = 1, \dots, K$. Apply a suitable combining function Φ to obtain an overall test statistic $\mathbf{T} = \Phi(p_1, \dots, p_K)$.
4. Compute p-values for each permuted partial test statistic $p_k^{b*} = \frac{1}{B}\sum_{d=1}^B \mathbb{I}(T_k^{b*} \geq T_k^{d*}), k = 1, \dots, K$ and combine them as $\mathbf{T}^{b*} = \Phi(p_1^{b*}, \dots, p_K^{b*})$, for $b = 1, \dots, B$.
5. The overall p-value of the test is $\mathbf{P} = \frac{1}{B}\sum_{b=1}^B \mathbb{I}(\mathbf{T} \geq \mathbf{T}^{b*})$.

TO DO: If each partial test of H_{0k} against H_{1k} is unbiased and consistent for $k = 1, \dots, K$, then the global test is also unbiased and consistent.

2 Missing data

2.1 Problem Set-up

Suppose, as in Section 1.4, we observe pairs of K-vectors $(\mathbf{X}_{i1}, \mathbf{X}_{i2})$ with $\mathbf{X}_{ij} = (X_{ij1}, \dots, X_{ijK})$ for $j = 1, 2$. Now, some of the data are missing. For each individual, define the inclusion indicators $(\mathbf{O}_{i1}, \mathbf{O}_{i2})$ with $O_{ijk} = \mathbb{I}(X_{ink} \text{ is observed})$, for $i = 1, \dots, N$, $j = 1, 2$, and $k = 1, \dots, K$. That is, O_{ijk} equals 1 if the outcome is measured for variable k in sample j for individual i and equals 0 if the measurement is missing. Hence, the complete set of observed data is represented by $\{(\mathbf{X}_{i1}, \mathbf{X}_{i2}, \mathbf{O}_{i1}, \mathbf{O}_{i2})\}_{i=1}^N$.

Assume that we are interested in testing for effects in the K components. The null hypothesis is that there is no difference between the samples due to treatment. Then, we must not only test for differences in $\mathbf{X}$, but also in $\mathbf{O}$. The null hypothesis can be expressed

$$H_0 : \{(\mathbf{X}_1, \mathbf{O}_1) \overset{d}{=} (\mathbf{X}_2, \mathbf{O}_2)\}$$

against the alternative hypothesis $H_1 : \{H_0 \text{ is not true}\}$. We may rewrite the null hypothesis as

$$\begin{aligned} H_0 : &\left\{ \left(\mathbf{O}_1 \overset{d}{=} \mathbf{O}_2\right) \bigcap \left[\left(\mathbf{X}_1 \overset{d}{=} \mathbf{X}_2\right) |(\mathbf{O}_1, \mathbf{O}_2) \right] \right\} \\ &= \left\{ \left[\bigcap_{k=1}^{K} \left(O_{\cdot 1k} \overset{d}{=} O_{\cdot 2k} \right) \right] \bigcap \left[\bigcap_{k=1}^{K} \left(X_{\cdot 1k} \overset{d}{=} X_{\cdot 2k} |(\mathbf{O}_1, \mathbf{O}_2) \right) \right] \right\} \end{aligned}$$

2.2 Missing Completely at Random

Data is said to be missing at random (MAR) if the missingness pattern does not depend on the unobserved values, conditional on the observed data. Similarly, data is said to be observed at random if the missingness pattern does not depend on the observed values, conditional on the unobserved data. Data is missing completely at random (MCAR) if the data are both missing at random and observed at random (Rubin [1976]). This is a special case of a missing data process.

In MCAR data, the observed values may be considered a random subset of the complete data set. The missingness pattern does not vary by sample, by definition, which implies that $\mathbf{O}_1 \overset{d}{=} \mathbf{O}_2$. Thus, we may omit testing for differences in the pattern of missingness, but we must still condition on it when testing for differences in measurements. The test reduces to

$$\begin{aligned} H_0 : &\left\{ \left(\mathbf{X}_1 \overset{d}{=} \mathbf{X}_2\right) |(\mathbf{O}_1, \mathbf{O}_2) \right\} \\ &= \left\{ \bigcap_{k=1}^{K} \left(X_{\cdot 1k} \overset{d}{=} X_{\cdot 2k} |(\mathbf{O}_1, \mathbf{O}_2) \right) \right\} \end{aligned}$$

against the alternative hypothesis $H_1 : \{H_0 \text{ is not true}\}$.

2.3 Exact tests

TO DO: exact results vs approximate results (see Bertacche and Pesarin [1997])

References

Rosita Bertacche and Fortunato Pesarin. Treatments of missing data in multidimensional testing problems for categorical variables. *Metron*, 55:135–149, 1997.

Fortunato Pesarin and Luigi Salmaso. *Permutation tests for complex data: theory, applications and software.* John Wiley & Sons, Inc., 2010.

Donald B. Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, December 1976. ISSN 0006-3444, 1464-3510. doi: 10.1093/biomet/63.3.581. URL `http://biomet.oxfordjournals.org/content/63/3/581`.

Alan Stuart. A test for homogeneity of the marginal distributions in a two-way classification. *Biometrika*, 42(3/4):412–416, 1955. ISSN 0006-3444. doi: 10.2307/2333387. URL `http://www.jstor.org/stable/2333387`.