Some connections between permutation tests and t-tests and their relevance for adaptive designs

Ekkehard Glimm¹, Michael Proschan², Martin Posch³
¹Novartis Pharma, Basel, Switzerland
²National Institute of Allergy and Infectious Diseases, Bethesda, Maryland
³Medical University of Vienna, Vienna, Austria

MRC Hub Network workshop "Adaptive designs for clinical trials"
Cambridge, 27 January 2014
Outline

- Permutation tests
  - one-sample case
  - two-sample case
- Connection with $t$-test and rotation tests
- Application to adaptive designs: sample size reestimation
Permutation test: one-sample case

Basic idea in the one-sample case:

- $H_0$: no difference between treatment and control
- $D_i$ denotes differences between a treatment and a control measurement on a patient
- Assign $Z_i = +1$ or $-1$ w.p. $\frac{1}{2}$ to each observation $d_i$ of $D_i$
- Go through all $2^n$ possibilities and calculate $T_i^* = \sum_{i=1}^{n} Z_i d_i$
- The $2^n$ values of $T_i^*$ constitute the conditional empirical null distribution $F(T_i)$ of $T_i = \sum_{i=1}^{n} Z_i D_i$ given $D_i = d_i$
- $p$-value of the test: percentage of $T_i^* \geq \sum_{i=1}^{n} d_i$.
- For $n$ large, approximate $F(T_i)$ by a random subsample.
One-sample permutation test asymptotics

- \( E(T_i) = 0; \) \( \text{var}(T_i) = \sum_{i=1}^{n} d_i^2 \) under \( H_0 \)
- By the Lindeberg-Feller-theorem,
  \[
  \tilde{T} = \frac{\sum_{i=1}^{n} Z_i D_i}{\sqrt{\sum_{i=1}^{n} D_i^2}}
  \]
  is asymptotically \( N(0,1) \)-distributed.
- \( \tilde{\sigma}^2 = \sum_{i=1}^{n} D_i^2 / n \) is the \textit{total variance} = variance under \( H_0 \)
- \( \tilde{\sigma}^2 \) and \( \tilde{T} \) are asymptotically stochastically independent.
One-sample permutation test with normal data

- If $D_i$ are independent, normally distributed, then

\[
\tilde{T} = \frac{\sum_{i=1}^{n} Z_i D_i}{\sqrt{\sum_{i=1}^{n} D_i^2}} \sim \pm \frac{1}{2} \sqrt{n \cdot \text{Beta}(\frac{1}{2}, \frac{n-1}{2})}.
\]

- Can be proved in many ways.

- Most intuitive proof is given on the next slide.

\[\pm \frac{1}{2} \sqrt{n \cdot \text{Beta}(\frac{1}{2}, \frac{n-1}{2})}\] is shorthand for the distribution where

\[\frac{\tilde{T}^2}{n} \sim \text{Beta}(\frac{1}{2}, \frac{n-1}{2})\]  

- For $\tilde{T}$ and $-\tilde{T}$, the density takes the same value.
One-sample permutation test with normal data

- \((D_1, \ldots, D_n)\) given \(\sum_{i=1}^{n} D_i^2 = r^2\) has a uniform distribution on the \(n\)-dimensional sphere with radius \(r\).

- Since this holds for all values of \(\sum_{i=1}^{n} D_i^2 = r^2\), and since \(Z_i D_i\) and \(D_i\) are equally distributed, \(\frac{\sum_{i=1}^{n} Z_i D_i}{\sqrt{\sum_{i=1}^{n} D_i^2}}\) and \(\sum_{i=1}^{n} D_i^2\) are stochastically independent.

- \(\tilde{T} = \frac{\sum_{i=1}^{n} Z_i D_i}{\sqrt{\sum_{i=1}^{n} D_i^2}}\) has the distribution of the sum of coordinates of vectors distributed uniformly on the \(n\)-dimensional unit sphere.

- Unique characterization of the distribution:
  \[
  \frac{1}{n} \tilde{T}^2 \sim Beta\left(\frac{1}{2}, \frac{n-1}{2}\right)
  \]
  and for \(+\tilde{T}\) and \(-\tilde{T}\) the density takes the same value.
One-sample permutation and t-test

- A more common representation is in terms of the well-known t-test statistic:

\[ t = \sqrt{n - 1} \left( n \cdot (\pm \bar{t}^{-2}) - 1 \right)^{-\frac{1}{2}} \sim t(n - 1) \]
Rotation test

- rotation test: approximate the distribution of $\tilde{T}$ by
  - fixing $\sum_{i=1}^{n} D_i^2 = r^2$ to its observed value
  - randomly rotate the vector $(d_1, \ldots, d_n)$
    =
    multiply $(d_1, \ldots, d_n)$ with an orthonormal matrix $Q$ where all orthonormal matrices have equal probability.
    For example for $n=2$:
    - generate $\theta$ from uniform distribution on $[0, 2\pi]$  
    - Let $d_1 = r \cdot \cos(\theta)$ and $d_2 = r \cdot \cos(\theta)$. 
Permutation test: two-sample case

- $H_0$: no difference between treatment and control
- $X_i$ observation from patient $i$ (in treatment or control group)
- For simplicity: $n/2$ patients per group
- To each observation $x_i$, assign $Z_i = +1$ or $-1$ w.p. $\frac{1}{2}$
- Restriction $\sum_{i=1}^{n} Z_i = 0$
- Go through all $\binom{n}{2}$ possibilities and calculate $T_i^* = \sum_{i=1}^{n} Z_i x_i$

⇒ Only difference in comparison with one-sample case: restriction $\sum_{i=1}^{n} Z_i = 0$
Two-sample permutation test asymptotics

- Conditional on observed $X_i = x_i$, $T_i = \sum_{i=1}^{n} Z_i X_i$ has
- $E(T_i) = 0; \text{var}(T_i) = \frac{n}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 = n\tilde{\sigma}^2$ under $H_0$
- Again,
  - $\tilde{T} = \frac{\sum_{i=1}^{n} Z_i X_i}{\sqrt{n\tilde{\sigma}^2}}$ is asymptotically $N(0,1)$-distributed
  - $\tilde{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$ is the "total" variance = variance under $H_0$
  - $\tilde{T}$ and $\tilde{\sigma}^2$ are asymptotically stochastically independent
Two-sample permutation test: normal data

By a similar argument as for the one-sample case, if $X_i \sim N(\mu, \sigma^2)$ i.i.d.

- $\tilde{T} = \frac{\sum_{i=1}^{n} Z_i X_i}{\sqrt{n\tilde{\sigma}^2}} \sim \pm \frac{1}{2} \sqrt{n \cdot \text{Beta}(\frac{1}{2}, \frac{n-1}{2})}$, i.e. corresponding to the uniform distribution on an $(n - 1)$-dimensional sphere
- $\tilde{T}$ and $\tilde{\sigma}^2$ are stochastically independent.

For the case of unequal sample sizes $n_T$ and $n_C$, analogous results (asymptotic in general and exact for normal data) are obtained by using

$$Z_i = \begin{cases} \frac{1}{n_T} & \text{w.p. } \frac{n_T}{n_T + n_C} \\ \frac{-1}{n_C} & \text{w.p. } \frac{n_C}{n_T + n_C} \end{cases}$$
Relevance to adaptive designs

- One-sample test with an interim (normally distributed data):
  - A trial with $n_1$ patients recruited in stage 1, and then some more.
  - After $n_1$ patients, do an **interim analysis** and change the sample size, based on the total variance $\tilde{\sigma}^2 = \sum_{i=1}^{n_1} D_i^2 / n_1$
  - For example, the new total sample size may be based on the usual power formula $n(\tilde{\sigma}^2) = \frac{\tilde{\sigma}^2 (\Phi^{-1}(1-\alpha) + \Phi^{-1}(1-\beta))}{\delta^2}$; $\alpha$ type I error, $\beta$ type II error, $\delta$ assumed difference treatment-control, $\Phi^{-1}(.)$ standard normal quantile.
  - Let $t_i$ be the usual one-sample $t$-test statistic, calculated from stage-$i$ data only.
  - Under $H_0$: No treatment effect, $\tilde{T}$ and $\tilde{\sigma}^2$ are stoch. independent $\Rightarrow t_1$ and $n(\tilde{\sigma}^2)$ stoch. independent $\Rightarrow t_1$ and $t_2$ stoch. independent
p-value combinations

- Let \( n_2 = n(\bar{s}^2) - n_1 \) and \( p_i \) be the \( p \)-value corresponding to \( t_i \).
- We can combine the \( p \)-values to test \( H_0 \). Possible rules:
  - \( f(p_1, p_2) = -\log(p_1 p_2) \) (Fisher's combination rule)
  - \( g(p_1, p_2) = \sqrt{\frac{n_1}{n_1+n_2}} \Phi^{-1}(1 - p_1) + \sqrt{\frac{n_2}{n_1+n_2}} \Phi^{-1}(1 - p_2) \) (conditional inverse normal rule)

Note that \( g(p_1, p_2) \) is NOT the usual inverse normal \( p \)-value combination rule (Lehmacher and Wassmer, 1999)

\[
g^*(p_1, p_2) = w_1 \Phi^{-1}(1 - p_1) + w_2 \Phi^{-1}(1 - p_2),
\]

where \( w_1, w_2 \) must be fixed before sample size reassessment based on an "anticipated" stage-2-sample size.
Adaptive tests

- Rejection rules: Reject $H_0$ (no treatment benefit) if
  - $f(p_1, p_2) \geq \chi^2_{4,1-\alpha}$
  - $g(p_1, p_2) \geq \Phi^{-1}(1 - \alpha)$
  - $g^*(p_1, p_2) \geq \Phi^{-1}(1 - \alpha)$
  - "naive" $t$-test: $t \geq t_{n_1+n_2-1,1-\alpha}$, where $t$ is the usual $t$-test statistic, calculated as if no sample size calculation after stage 1.

- Can be shown that the naive $t$-test does not keep the type I error in general
  - however, this is hardly detectable in simulations.
Power simulations

![Graph showing power of various adaptive tests when true difference = 0.1, n1=30 and sample size calculated from total variance to retain 80% power](image)
Discussion of power simulations

- $g(p_1, p_2)$ performs a good deal better than both $g^*(p_1, p_2)$ and $f(p_1, p_2)$.

- $g(p_1, p_2)$ performs almost identical as the "naive" t-test (treating $n$ as if it had been fixed from the trial start).

- $g(p_1, p_2)$ is only valid because the sample size re-estimation rule is based on the total variance $\bar{\sigma}^2$ only.

- $g^*(p_1, p_2)$ and $f(p_1, p_2)$ are more generally applicable.
Conclusions

- There is a **close relation** between resampling and *t*-tests.
- Since permutation tests condition on all data except the treatment indicator, we can use them **after having looked at the data** in an interim analysis (with only the treatment indicator hidden).
- Sample size modification based on the total variance alone allows the use of **more powerful p-value-combination methods**.
- All of this is about testing a **strong null hypothesis** of no treatment benefit (in particular: equal variance in the two groups)
References


