# MT 4113: Computing in Statistics

Computer intensive statistics

Lecture 6: Permutation, randomization, and Monte-Carlo tests

Len Thomas

Oct 10 2018

# 1 Introduction

## 1.1 Revision – Traditional parametric and nonparametric methods

**Problem statement – inference from one sample**

- We have a sample of data $x = x_1, \dots, x_n$ that are realizations of iid random variables $X = X_1, \dots, X_n$ with pdf $f()$.
- We wish to make inferences about some population characteristic – say the mean $\mu$ of $f()$.

**Parametric methods**

- Assume $f()$ is a known function with parameters $\theta$ – e.g., $f(\theta) = \mathrm{N}(\mu, \sigma^2)$.
- For a (two-sided) hypothesis test
  - Specify the null hypothesis, $H_0$ – e.g., $\mu = \mu_0$
  - Construct a test statistic $T(X)$ that has a known distribution under $H_0$
  - e.g.,
$$T(x) = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$
    which has a $t$-distribution with mean 0, variance 1 and d.f. $= n - 1$, if $H_0$ is true
  - Calculate $P(|T(X)| \geq |T(x)|\,; \mu_0)$ – the $p$-value.
  - If $p$-value $\leq \alpha$ the result is statistically significant.
- For (two-sided) confidence intervals on $\mu$
  - For $((1 - \alpha) \times 100)\%$ limits, find the scalar values $L(X)$ and $U(X)$ such that
$$P\left(L(X) < \mu \text{ and } U(X) > \mu\right) = 1 - \alpha$$
  - Many methods for finding the limits
  - E.g., inverting the test statistic
    * limits are the smallest and largest values of $\mu_0$ that are not rejected under $H_0 : \mu = \mu_0$
    * E.g., $t$-based intervals $\bar{x} \pm t_{\alpha/2,n-1} s/\sqrt{n}$
- So, for parametric methods, you need to specify two distributions:
  - The pdf generating the data: $f(\theta)$
  - The distribution of the test statistic: $T(X)$

### Traditional nonparametric methods

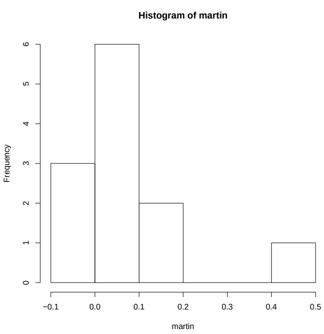
- Avoid fully specifying the distribution of $f()$.
- E.g., Wilcoxon signed ranks test
  - $f()$ is symmetric about some median m
  - $H_0 : m = m_0$
  - Test statistic
  $$T(x) = \sum_{i:x_i \neq m_0} \text{rank}(|x_i - m_0|)\text{sign}(x_i - m_0)$$
  where rank$(x)$ denotes the rank order of $x$ (smallest has value 1, next smallest 2, etc.) and sign$(x - y)$ is a sign function, having value 1 if $x > y$ and -1 if $x < y$.
  - Under $H_0$, T(X) has (approximately) a normal distribution
  - For confidence limits on $m$, can invert the test statistic

### Computer-intensive statistical methods

- With computer intensive methods, you can avoid the need to fully specify either the distribution of $f()$ and/or $T(X)$
- Examples:
  - Permutation-based approach to Wilcoxon test (avoids the need to specify distribution for $T(X)$; Wilcoxon already avoids $f()$)
  - Monte-Carlo-based version of the $t$-test (avoids specification of $T(X)$, but not $f(\theta)$)
  - Parametric and non-parametric bootstrap confidence intervals (again avoids $T(X)$; either specify or not $f(\theta)$)

## 1.2 Motivating example

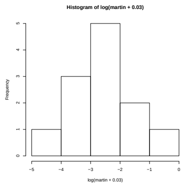
### Pine martin habitat preference data




- 0.13, -0.01, -0.01, 0.42, -0.02, 0.01, 0.09, 0.03, 0.04, 0.06, 0.12, 0.03
- Shapiro-Wilk test for normality W = 0.70, p<0.001

**Pine martin example: parametric test**

- Assume data come from a normal distribution
  - $H_0 : \mu = 0$
  - $t$-test: $t = 2.147$, d.f.=11, $p = 0.055$
  - $\bar{x} = 0.074$, 95% $t$-based CI: (-0.002, 0.150)
- Could try transforming – e.g., $\log(x_i + 0.03)$:




  - $H_0 : \log(\mu + 0.03) = \log(0.03)$
  - $t$-test: $t = 2.484$, d.f.=11, $p = 0.030$
  - 95% $t$-based CI (back-transformed): (0.003, 0.095)

**Pine martin example: parametric test (contd.)**

- Problems:
  - Result not invariant to transformation: what transformation to use?
  - If we use log, then what about bias correction?

**Pine martin example: traditional nonparametric test**

- Wilcoxon signed-ranks test:
  - $H_0 : m = 0$
  - $p = 0.012$
  - 95% CI on m: (0.010, 0.120)
- Disadvantages:
  - Requires us to formulate $H_0$ in terms of the median
  - Less powerful than a $t$-test when parametric assumptions met

# 2 Permutation methods

## 2.1 Permutation test

**Pine martin example**

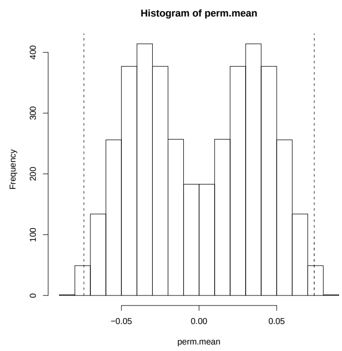
- Under $H_0 : \mu = \mu_0$, and assuming the distribution of $x_i$ is symmetric about $\mu_0$, then
- $+(x_i - \mu_0)$ is just as likely to occur in the data as $-(x_i - \mu_0)$
- For $\mu_0 = 0$ then $+x_i$ is just as likely as $-x_i$
- So, can produce alternative equally likely realizations of the data by swapping any $x_i$ for $-x_i$

**Permutations**

- Given 12 data points, there are $2^{12} = 4096$ ways to do this:
  - 0.13, 0.01, 0.01, 0.42, 0.02, 0.01, 0.09, 0.03, 0.04, 0.06, 0.12, 0.03
  - -0.13, 0.01, 0.01, 0.42, 0.02, 0.01, 0.09, 0.03, 0.04, 0.06, 0.12, 0.03
  - 0.13, -0.01, 0.01, 0.42, 0.02, 0.01, 0.09, 0.03, 0.04, 0.06, 0.12, 0.03
  - 0.13, 0.01, -0.01, 0.42, 0.02, 0.01, 0.09, 0.03, 0.04, 0.06, 0.12, 0.03
  - $\vdots$
  - -0.13, -0.01, -0.01, -0.42, -0.02, -0.01, -0.09, -0.03, -0.04, -0.06, -0.12, -0.03

**Permutation test**

- Find what proportion of the permutations have $|\bar{X}|$ greater than the observed $|\bar{x}|$.




- 48 out of 4096 are as extreme - so $p = 48/4096 = 0.012$

**Permuatation tests - the basic idea**

- Pick a null hypothesis $H_0$ and test statistic $T(X)$
- Under $H_0$, all possible appropriate data permutations (re-orderings) are equally likely
- So, calculate $T(X)$ for each permutation and see how extreme $T(x)$ is compared with these
  - If $T(x)$ is extreme, reject $H_0$
  - If $T(x)$ is not extreme, don't reject $H_0$

No need to specify distributions for:

- $f()$ - no parametric definition needed
- $T(X)$ - obtain this via the randomization procedure

**Choice of test statistic**

- With permutation-based methods (and many other computer-intensive methods) you are free to choose whatever test statistic you like
- For example, we could use the median, trimmed mean, $t$-statistic, etc., etc.
- The method readily extends to other situations – e.g., two sample tests, ANOVA, etc., etc.

## 2.2 Permutation interval

**Permutation interval**

- Invert the test statistic: Find the set of values of $\mu_0$ such that $H_0 : \mu = \mu_0$ is not rejected
- For each potential value of $\mu_0$, follow the same procedure as before
  - calculating all permutations of $\pm(x_i - \mu_0)$
  - then seeing how extreme the observed $\bar{x}$ is among the permutations
- Can use stochastic search methods[1] to find the confidence limits
- In our example, 95% CI is (0.013, 0.150) (c.f., $t$-based CI of (-0.002, 0.150)).

## 2.3 Summary

**Pros**

- Widely applicable
- Produces exact results
- No specific distribution assumed for the data
- Does not require analytic distribution of test statistic

**Cons**

- Each permutation of the data must be equally likely (or probability of occurrence is known)
- To be generally applicable, requires thought, and ability to program a computer (permutation algorithms can be non-trivial)
- Can be prohibitively computer (and programmer) intensive in complex situations or for large sample sizes

# 3 Randomization methods

## 3.1 Overview

**Introduction**

- Randomization methods are like permutation methods, except that a only random subset of all the possible permutations are generated
- E.g., pine martin example:
  - To generate one randomization, list the set of values $|x_i - \mu_0|$ and randomly assign a sign to each value
  - From one run of 999 randomizations plus the original data value, I obtained $p = 0.015$, very similar to the $p = 0.012$ of the permutation method.
- Confidence intervals can be obtained in a similar way to permutation intervals

[1] e.g., Robbins-Monroe search - see accompanying notes

## 3.2 Summary

**Pros, compared with permutation methods**

- All of the same advantages as permutation, plus
- Much easier to code
- Number of randomizations is fixed, rather than being a function of sample size

**Cons, compared with permutation methods**

- As with permutation tests, each randomization of the data must be equally likely (or probability of occurrence is known)
- Results will vary from run to run
  - Size of this "Monte-Carlo variation" depends on the number of randomizations

# 4 Monte-Carlo tests

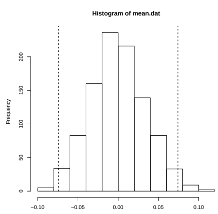
## 4.1 Overview

**Introduction**
A generalization[2] of randomization tests to include parametric distributions for the data.

- Algorithm:
  1. Set up some distribution for the data, $f()$ or $f(\theta)$, and some $H_0$
  2. Repeat the following two steps many times:
     (a) Simulate a data set according to $H_0$
     (b) Calculate $T(X)$ using the simulated data
  3. Add $T(x)$ evaluated from the sample data
  4. Order all of the $T(X)$s
  5. $p$-value is the proportion of the $T(X)$s as extreme or more extreme than the one from the sample data

**Pine martin example - parametric simulation**

- Assume observations are iid N$(\mu, \sigma^2)$, that $H_0 : \mu = 0$, and that $\sigma^2 = s^2$
- Simulate 999 data sets of 12 observations from N$(0, s^2)$.
- Calculate $\bar{x}$ for each of these, and add in the mean from the sample data, to give 1000 values of $\bar{x}$.
- $p$-value is the proportion of these values that are more extreme (e.g., $p = 0.029$)




[2]If the assumed model for $f()$ implies that all data orderings are equally likely then you get back to a randomization test

## 4.2 Summary

**Pros**

- Data can be assumed to follow any distribution - parametric or non-parametric (Randomization tests are a special case)
- The analytic distribution of the test statistic is not required

**Cons, compared with permutation methods**

- For parametric simulations, need to estimate nuisance parameters (e.g., $\sigma^2$ in the pine martin example)
- Results will vary from run to run

**Monte Carlo confidence intervals**
Confidence intervals can be constructed using similar methods to randomization intervals – but there is a better way – see next lecture...