STA 104 Applied Nonparametric Statistics

Chapter 2: One-Sample Methods for Location Problem

Xiner Zhou

Department of Statistics, University of California, Davis

Table of contents

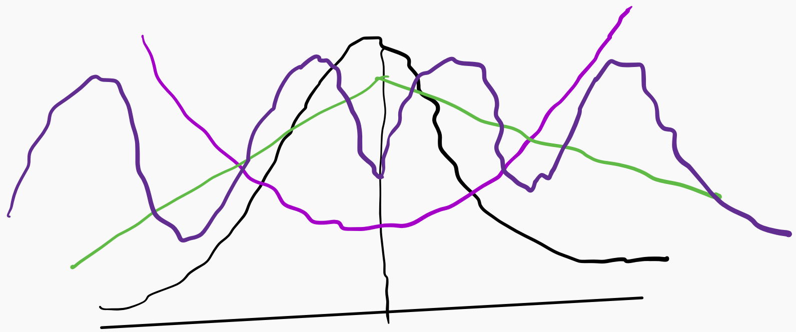
Wilcoxon Signed Rank Test

Setting

Suppose we have a random sample $x_1 \dots x_n$, i.e. data

Assumptions

- The x's are mutually independent.
- they are from a population that is
    - continuous
    - symmetric about the median θ



Hypothesis

Two-Sided Test:

$$H_0 : \theta = \theta_0 \text{ versus } H_a : \theta \neq \theta_0$$

One-Sided Upper-Tail Test:

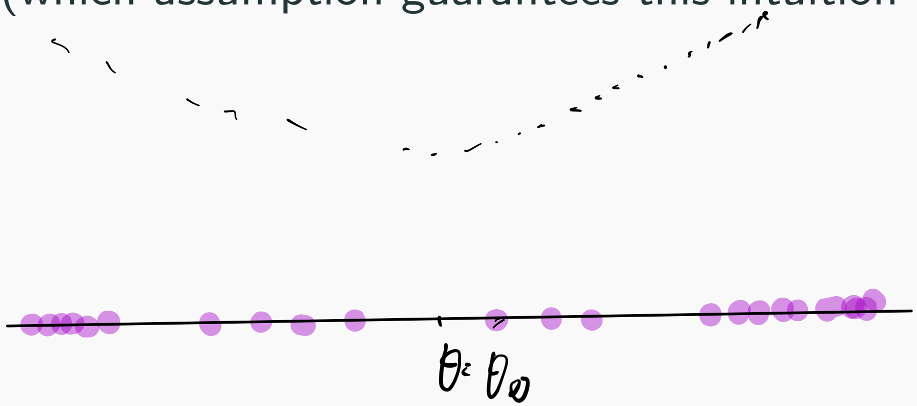
$$H_0 : \theta = \theta_0 \text{ versus } H_a : \theta > \theta_0$$

One-Sided Lower-Tail Test:

$$H_0 : \theta = \theta_0 \text{ versus } H_a : \theta < \theta_0$$

Motivation

Intuition: If θ_0 was the true median of the population, then the magnitude in terms of absolute value of the centered data is nothing to do with the sign of the centered data (which assumption guarantees this intuition valid?)



Assume : Symmetric dist.

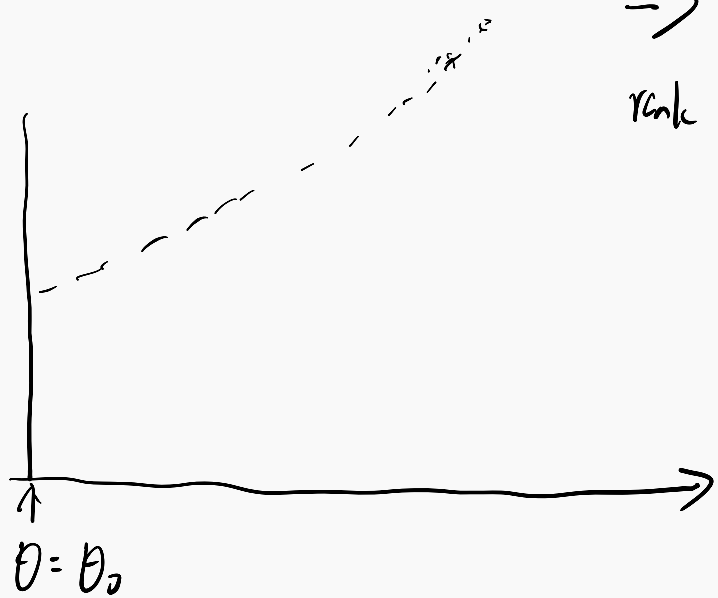
Motivation

- **Centering:** subtract θ_0 from each observation $x_1, \ldots, x_n$ to form a modified sample $\boxed{x_1' = x_1 - \theta_0, \ldots, x_n' = x_n - \theta_0}$ centered data
- **Flip:** form absolute values $|x_1'| \ldots |x_n'|$
- **Rank:** Order them from least to greatest, let R_i denote the rank of ith observation $|x_i'|$

1 3 4 2

$\Rightarrow$ 1 2 3 4

rank : 1 2 3 4



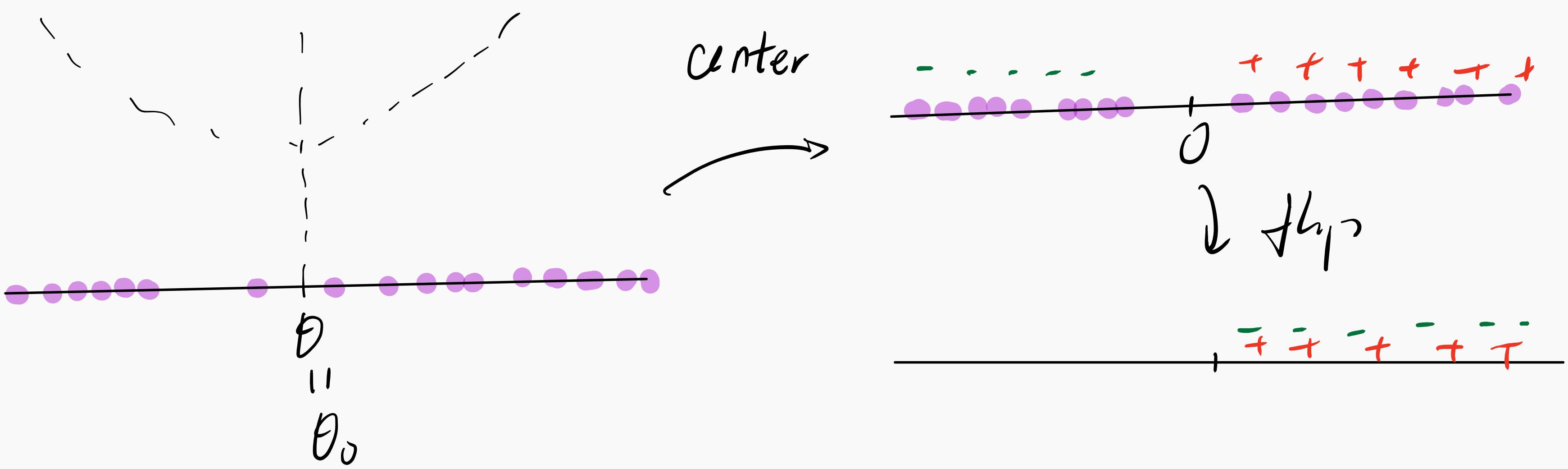
- **Sign:**

Define indicator variable for signs $\psi_i = \begin{cases} 1, & \text{if } \overset{x_i'}{\cancel{Z_i}} > 0 \\ 0, & \text{if } \underset{x_i'}{\cancel{Z_i}} < 0 \end{cases}$

Define positive signed rank of ith observation $\psi_i R_i = \begin{cases} R_i, & \text{if } \overset{x_i'}{\cancel{Z_i}} > 0 \\ 0, & \text{if } \underset{x_i'}{\cancel{Z_i}} < 0 \end{cases}$

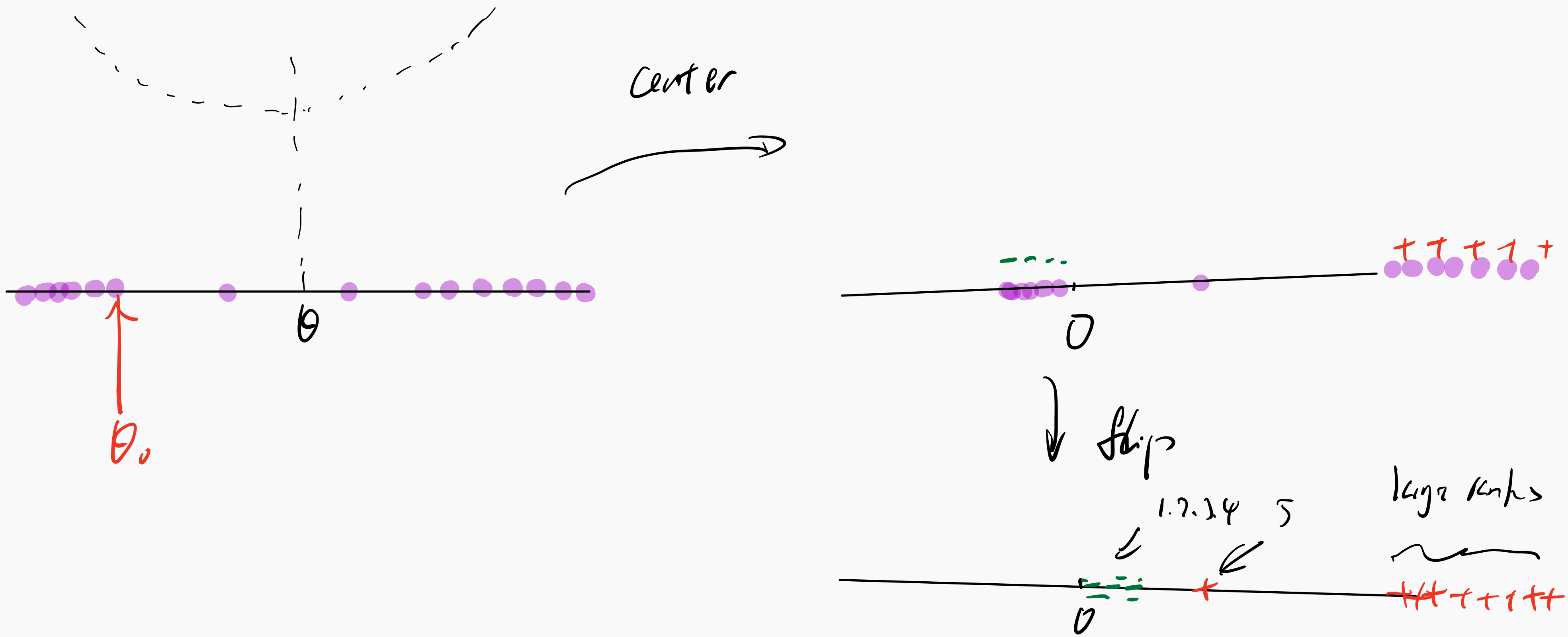
Motivation

- When the true unknown median θ is equal to the hypothesized value θ_0:
  - the centered data will tend to be spread symmetrically around 0, and roughly half of the observations have positive signs, due to symmetric underlying distribution,
  - the ranks associated with negative and positive observations are roughly equal,
  - so the total ranks of those positive signed observation roughly is half of the total ranks of all observations.



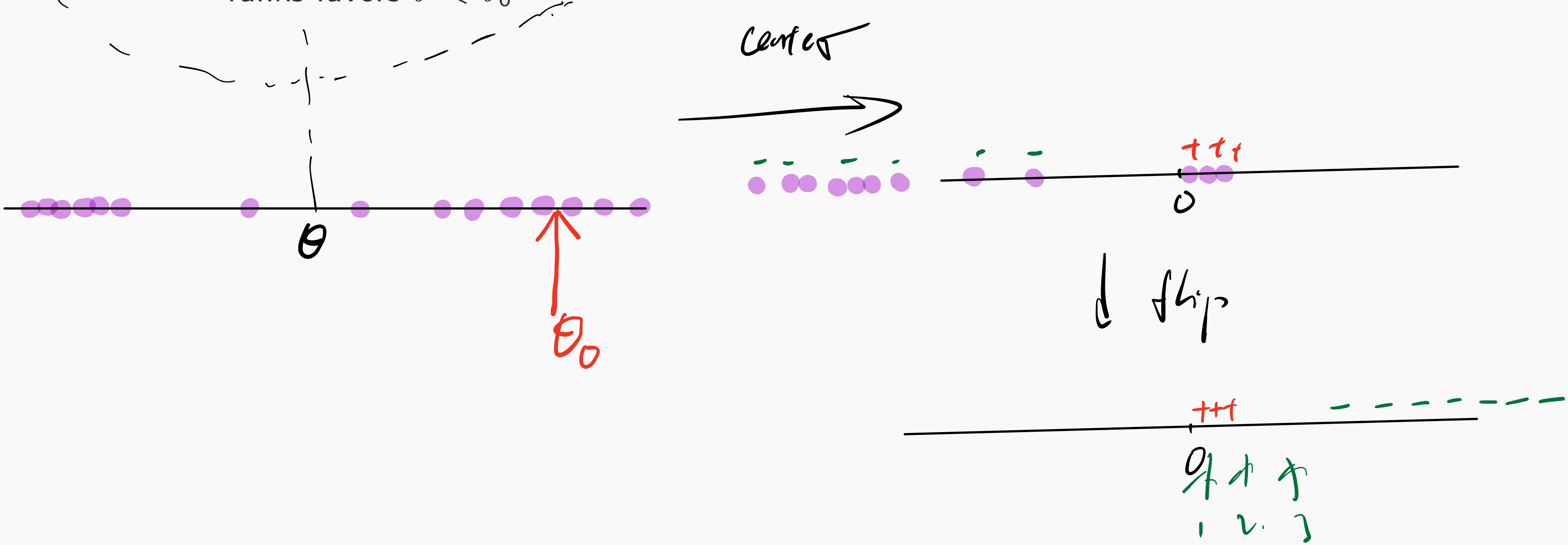
Motivation

- When the true unknown median $\theta > \theta_0$:
  - there will tend to be a larger portion of observations have positive signs and larger ranks associated with them,
  - so the total ranks of those positive signed observation is larger, so large total signed ranks favors $\theta > \theta_0$



Motivation

- When the true unknown median $\theta < \theta_0$:
    - there will tend to be a smaller portion of observations have positive signs and smaller ranks associated with them,
    - so the total ranks of those positive signed observation is smaller, so small total signed ranks favors $\theta < \theta_0$



Motivation

Define Wilcoxon signed rank statistics

$$T^+ = \underbrace{\sum_{i=1}^{n} \psi_i R_i}_{\text{sum of positive signed ranks}}$$

Derivation of (exact) null distribution

When $H_0 : \theta = \theta_0$ is true :

- Think the process as randomly split ranks $1, 2, 3 \dots n$ into two groups
- T^+ is sum of one of groups

Because ranks and signs are independent under H_0

$\Rightarrow$ each rank is equally likely to be + or − : $P(\psi_i = 1) = \frac{1}{2}$

$\Rightarrow$ each configuration of positive signed ranks $\{1\psi_1, 2\psi_2 \dots n\psi_n\}$ occurs with $\left(\frac{1}{2}\right)^n$

Example

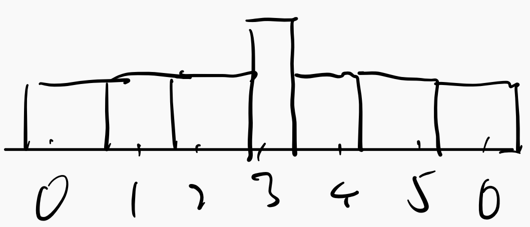
$n=3$

$(1\psi_1, 2\psi_2, 3\psi_3)$	Prob. under H_0	T^+
$(0, 0, 0)$	$(\frac{1}{2})^3 = \frac{1}{8}$	0
$(1, 0, 0)$	$\frac{1}{8}$	1
$(0, 2, 0)$	$\frac{1}{8}$	2
$(0, 0, 3)$	$\frac{1}{8}$	3
$(1, 2, 0)$	$\frac{1}{8}$	3
$(1, 0, 3)$	$\frac{1}{8}$	4
$(0, 2, 3)$	$\frac{1}{8}$	5
$(1, 2, 3)$	$\frac{1}{8}$	6

null dist. of T^+

$\Rightarrow$

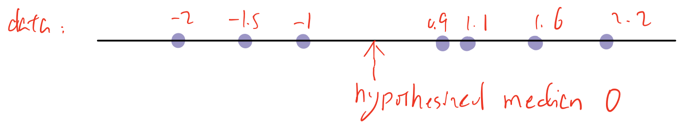
T^+	prob
0	1/8
1	1/8
2	1/8
3	2/8
4	1/8
5	1/8
6	1/8



We have derived the null distribution of T^+ without specifying the forms of the underlying populations beyond the point of requiring that they be continuous and symmetric about zero. This is why the test procedures based on T^+ are called distribution-free procedures.

From the null distribution of T^+we can determine the critical value t_α and control the probability α of falsely rejecting H_0 when H_0 is true.

■ Wilcoxon signed rank test statistic

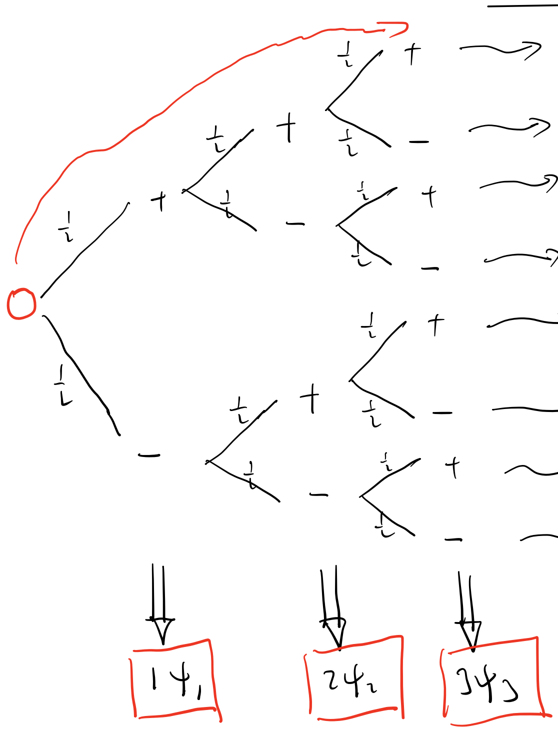


rank : 6 4 2 1 3 5 7

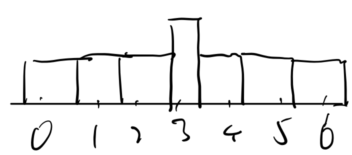
$\Rightarrow T^+ = 1+3+5+7 = 16$

■ (exact) null distribution of T^+

key : each rank 1, 2, 3 ... n has equal probability to have + or − sign



$(1\psi_1, 2\psi_2, 3\psi_3)$	T^+	Prob
(1, 2, 3)	6	$\frac{1}{8}$
(1, 2, 0)	3	$\frac{1}{8}$
(1, 0, 3)	4	$\frac{1}{8}$
(1, 0, 0)	1	$\frac{1}{8}$
(0, 2, 3)	5	$\frac{1}{8}$
(0, 2, 0)	2	$\frac{1}{8}$
(0, 0, 3)	3	$\frac{1}{8}$
(0, 0, 0)	0	$\frac{1}{8}$



null dist. of T^+

distribution-free property

= nonparametric

Large sample approximation of null distribution:

Reorder the data according to their absolute ranks, let $V_i = \cancel{iS_i}$ $i\psi_i$ be the ith observation.

$$
\begin{aligned}
T^+ &= \sum_{i=1}^{n} \psi_i R_i \\
&= \sum_{i=1}^{n} V_i
\end{aligned}
$$

V_i are mutually indepdendent dichotomous random variables with $P(V_i = i) = P(V_i = 0) = \frac{1}{2}$. T^+ is sum of independent random variables, follows from standard theory for sums of mutually independent, but not identically distributed, random variables, such as the Liapounov central limit theorem (cf. Randles and Wolfe (1979, p. 423)), it has an asymptotic normality distribution.

Define standardized Wilcoxon signed rank statistics

$$
\Rightarrow \text{ standardized form } T^* = \frac{T^+ - E_0\left(T^+\right)}{\left\{\mathrm{var}_0\left(T^+\right)\right\}^{1/2}} = \frac{T^+ - \frac{n(n+1)}{4}}{\left\{\frac{n(n+1)(2n+1)}{24}\right\}^{1/2}}
$$

$$
\sim N(0,1)
$$

if n is large

n: sample size

Optimal :

①
$$E_0\left(T^+\right) = E\left[\sum_{i=1}^{n} V_i\right] = \sum_{i=1}^{n} E\left[V_i\right]$$

$$E_0\left(V_i\right) = i\left(\frac{1}{2}\right) + 0\left(\frac{1}{2}\right) = \frac{i}{2}$$

$$\Rightarrow E_0\left(T^+\right) = \frac{1}{2}\sum_{i=1}^{n} i = \frac{1}{2}\left[\frac{n(n+1)}{2}\right] = \boxed{\frac{n(n+1)}{4}}$$

②
$$\mathrm{var}_0\left(T^+\right) = \mathrm{var}\left(\sum_{i=1}^{n} V_i\right) = \sum_{i=1}^{n} \mathrm{var}\left(V_i\right)$$

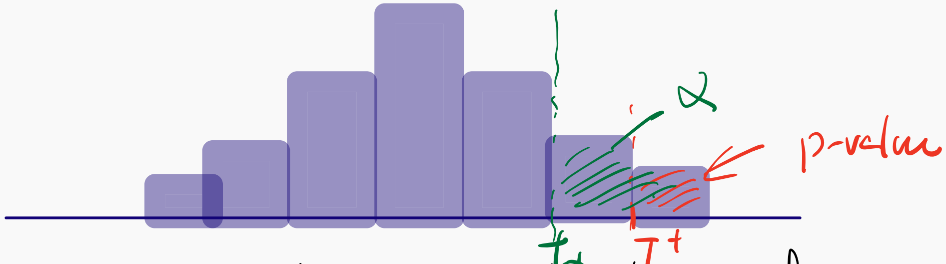
$$\mathrm{var}_0\left(V_i\right) = E_0\left(V_i^2\right) - \left[E_0\left(V_i\right)\right]^2$$

$$= \left[i^2\left(\frac{1}{2}\right) + 0^2\left(\frac{1}{2}\right)\right] - \left[\frac{i}{2}\right]^2$$

$$= \frac{i^2}{2} - \frac{i^2}{4} = \frac{i^2}{4}$$

$$\Rightarrow \mathrm{var}_0\left(T^+\right) = \frac{1}{4}\sum_{i=1}^{n} i^2 = \frac{1}{4}\left[\frac{n(n+1)(2n+1)}{6}\right] = \boxed{\frac{n(n+1)(2n+1)}{24}}$$

Procedure



(exact) null dist. of T^+

a. One-Sided Upper-Tail Test.

To test

$$H_0 : \theta = \cancel{0}\ \theta_0$$

versus

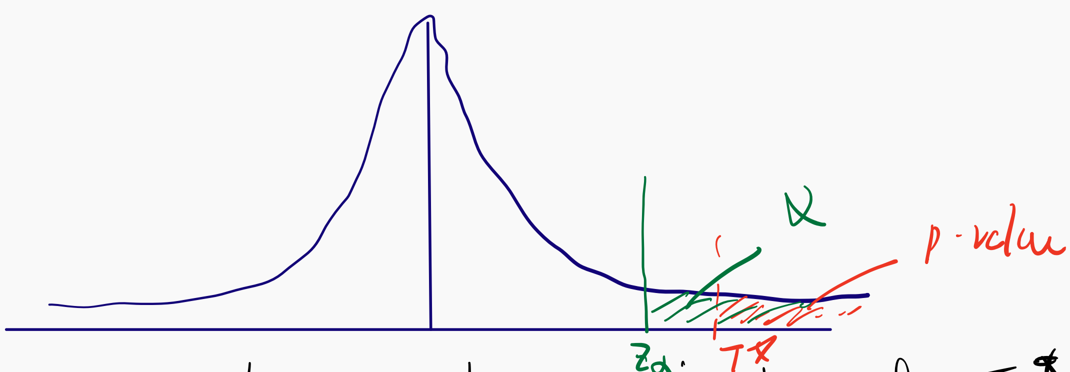
$$H_a : \theta > \cancel{0}\ \theta_0$$

Large T^+ / T^* favor H_a

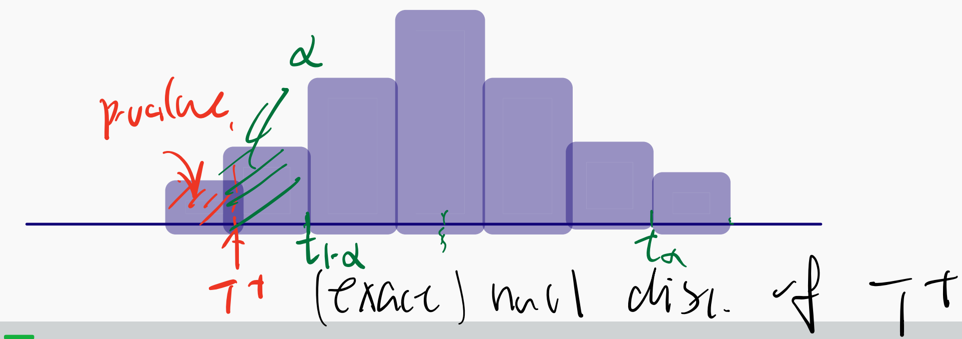
at the α level of significance, Reject H_0 if $T^+ \geq t_\alpha$; otherwise do not reject.

The normal approximation:

Reject H_0 if $T^* \geq z_\alpha$; otherwise do not reject.



large-sample approximation of T^*



b. One-Sided Lower-Tail Test.

To test

$$H_0 : \theta = \cancel{0}\ \theta_0$$

versus

$$H_a : \theta < \cancel{0}\ \theta_0$$

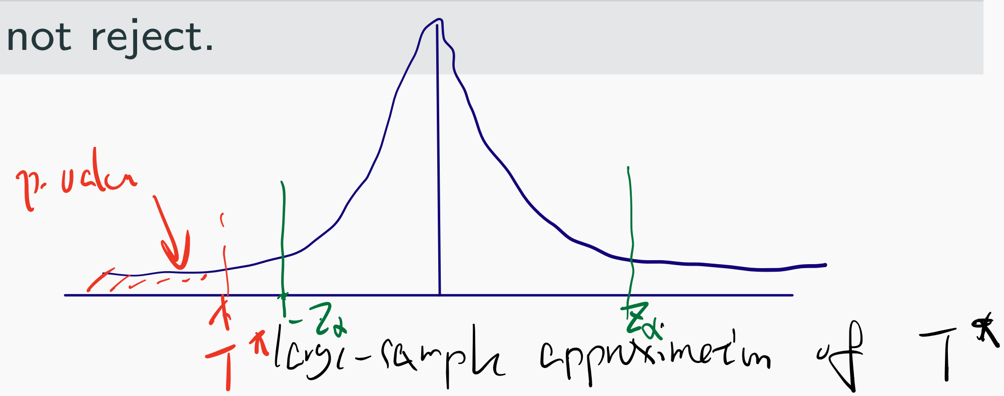
Small T^+ / T^* favor H_a

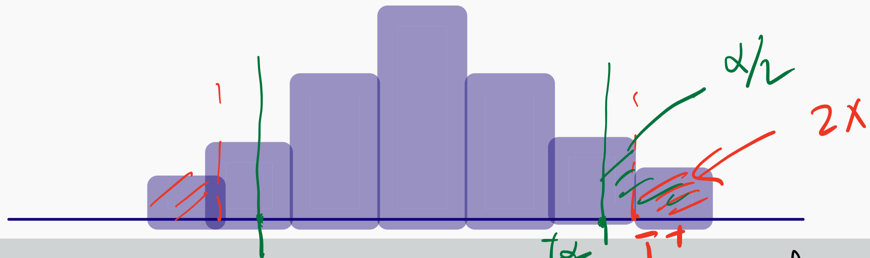
at the α level of significance, Reject H_0 if $T^+ \le \boxed{t_{1-\alpha}} = \frac{n(n+1)}{2} - \boxed{t_\alpha}$; otherwise do not reject.

The normal approximation:

Reject H_0 if $T^* \le -z_\alpha$; otherwise do not reject.

$= z_{1-\alpha}$





$t_{1-\frac{\alpha}{2}}$ (exact) null dist. of T^+

c. Two-Sided Test.

To test

$$H_0 : \theta = \cancel{0}\ \theta_0$$

versus

$$H_a : \theta \neq \cancel{0}\ \theta_0$$

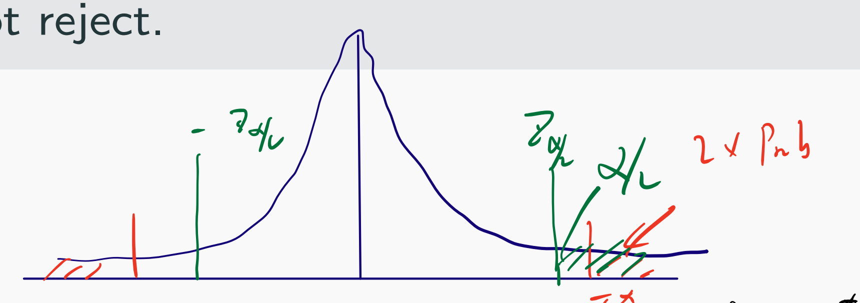
Both small and large T^+/T^* favor Ha

at the α level of significance, Reject H_0 if $T^+ \geq t_{\alpha/2}$ or $T^+ \leq t_{1-\alpha/2} = \frac{n(n+1)}{2} - t_{\alpha/2}$; otherwise do not reject.

This two-sided procedure is the two-sided symmetric test with $\alpha/2$ probability in each tail of the null distribution of T^+.

The normal approximation:

Reject H_0 if $|T^*| \geq z_{\alpha/2}$; otherwise do not reject.



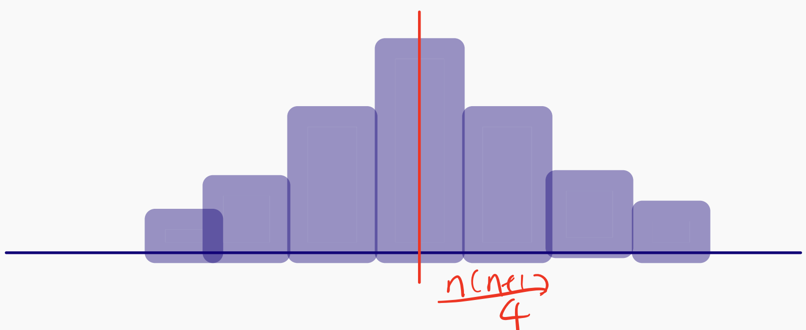
large-sample approximation of T^*

An estimator associated with Wilcoxon's signed rank statistics (Hodges-Lehmann)

The null distribution of the statistic T^+ is symmetric about its mean, $n(n+1)/4$. A natural estimator of θ is the amount that should be subtracted from each Z_i so that the value of T^+, when applied to the shifted sample $X_1 - \hat{\theta}, \ldots, X_n - \hat{\theta}$, is as close to $n(n+1)/4$ as possible.

Roughly speaking, we estimate θ by the amount ($\hat{\theta}$) that the Z sample should be shifted in order that $Z_1 - \hat{\theta}, \ldots, Z_n - \hat{\theta}$ appears (when "viewed" by the signed rank statistic T^+) as a sample from a population with median 0.

If we have guessed correct median :



(exact) null dist. of T^+

Optional:

$$T^+ = \sum_{i=1}^n S_i R_i$$

$$\Rightarrow \frac{1}{\frac{n(n+1)}{2}} \sum \mathbb{I}\left(\frac{X_i+X_j}{2} \geq \theta_0\right) \approx \frac{1}{2}$$

$$\psi_i = I(X_i > \theta_0)$$

$$R_i = \sum_{j=1}^n I\left(\left|X_j - \theta_0\right| \leq \left|X_i - \theta_0\right|\right)$$

	X_1	X_2	X_3
X_1	$\frac{X_1+X_1}{2}$	$\frac{X_1+X_2}{2}$	$\frac{X_1+X_3}{2}$
X_2		$\frac{X_2+X_2}{2}$	$\frac{X_2+X_3}{2}$
X_3			$\frac{X_3+X_3}{2}$

re-arrange

$$= \sum_{i=1}^n \sum_{j=1}^n I\left(\left|X_j - \theta_0\right| \leq \left|X_i - \theta_0\right|, X_i > \theta_0\right)$$

$$= \sum_{i=1}^n \sum_{j=1}^n I\left(\left|X_j - \theta_0\right| \leq X_i - \theta_0, X_i > \theta_0\right)$$

$$= \sum_{i=1}^n \sum_{j=1}^n I\left(\theta_0 - X_i \leq X_j - \theta_0 \leq X_i - \theta_0\right)$$

$$= \sum_{i=1}^n \sum_{j=1}^n I\left(2\theta_0 \leq X_j + X_i \leq 2X_i\right)$$

$$= \sum_{i=1}^n \sum_{j=1}^n I\left(\frac{X_i + X_j}{2} \geq \theta_0, X_j \leq X_i\right)$$

$$= \boxed{\sum_{1 \leq i \leq j \leq n} \mathbb{1}\left(\frac{X_i + X_j}{2} \geq \theta_0\right)} \approx \frac{n(n+1)}{4}$$

$\Rightarrow$ The median θ_0 should be the value that splits the pairwise averages into 50% / 50%.

Procedure

The Walsh Averages.

Each of the $n(n+1)/2$ averages $\left(X_i + X_j\right)/2$, $i \leq j = 1, \ldots, n$, is called a Walsh average.

Estimate

To estimate the median θ, form the $M = n(n+1)/2$ averages $\left(X_i + X_j\right)/2$, for $i \leq j = 1, \ldots, n$. The estimator of θ associated with the Wilcoxon signed rank statistic T^{+} is

$$\hat{\theta} = \text{median}\left\{\frac{\cancel{X}_i + \cancel{X}_j}{2}, i \leq j = 1, \ldots, n\right\}.$$

Let $W^{(1)} \leq \cdots \leq W^{(M)}$ denote the ordered values of $\left(X_i + X_j\right)/2$, where $M = \frac{n(n+1)}{2}$

- Then if M is odd, say $M = 2k + 1$, we have $k = (M-1)/2$ and

  $$\hat{\theta} = W^{(k+1)},$$

  the value that occupies position $k+1$ in the list of the ordered $\left(X_i + X_j\right)/2$ averages.
- If M is even, say $M = 2k$, then $k = M/2$ and

  $$\hat{\theta} = \frac{W^{(k)} + W^{(k+1)}}{2}$$

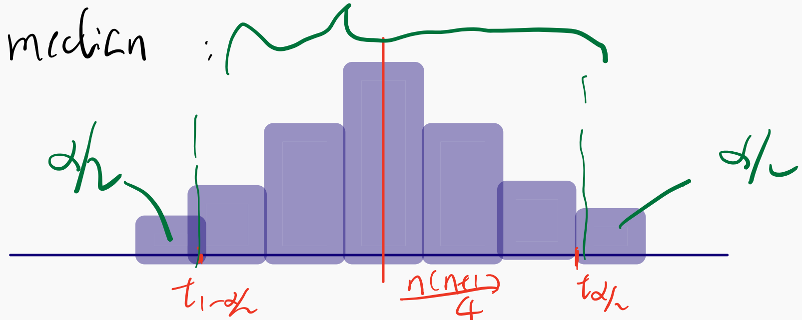
  That is, when M is even, $\hat{\theta}$ is the average of the two $\left(X_i + X_j\right)/2$ values that occupy positions k and $k+1$ in the ordered list of the M $\left(X_i + X_j\right)/2$ averages. The $\left(X_i + X_j\right)/2$.

Confidence interval based on Wilcoxon's signed rank test

The true population median θ_0 is the value such that the number of Walsh averages above it is the Wilcoxon's signed rank statisitcs T^+ which should be centered at $\frac{n(n+1)}{4}$ with some natural variation. So we use the natural variation of T^+ reverse engineer the most probable region of θ_0 .

$$T^+ = \underbrace{\sum_{1\le i\le j\le n} I\left(\frac{X_i+X_j}{2} \ge \theta_0\right)}_{\text{total number of Walsh averages above true median } \theta_0}$$

If we have guessed correct median :



(exact) null dist. of T^+

$$\underbrace{P\left(\boxed{t_{1-\alpha/2}} \le T^+ \le \boxed{t_{\alpha/2}}\right) = \boxed{1-\alpha}}_{\text{with probability } (1-\alpha)100\%} \approx 95\%$$

$\Rightarrow$ total number of Walsh averages above true median should be between $t_{1-\alpha/2}$ and $t_{\alpha/2}$

$W^{(1)} \quad \cdots \quad W^{(\frac{n(n+1)}{2})}$

$\uparrow \quad W^{(\frac{n(n+1)}{2}+1-t_{\alpha/2})} \quad \uparrow \quad W^{(t_{\alpha/2})}$

$(1-\alpha)100\%$ Confidence Interval

For a symmetric two-sided confidence interval for median θ, with confidence coefficient $1-\alpha$, set

$$\cancel{t_\alpha = \frac{n(n+1)}{2} + 1 - t_{\alpha/2},}$$

where $t_{\alpha/2}$ is the upper $(\alpha/2)$ th percentile point of the null distribution of T^+.

The $100(1-\alpha)\%$ confidence interval (θ_L, θ_U) for θ that is associated with the Wilcoxon signed rank statistics is

$$\theta_L = \cancel{W^{(t_\alpha)}}\ W^{(\frac{n(n+1)}{2}+1-t_{\alpha/2})}, \quad \theta_U = \cancel{W^{(M+1-C_\alpha)}} = W^{(t_{\alpha/2})}$$

~~where $M = n(n+1)/2$ and $W^{(1)} \leq \cdots \leq W^{(M)}$ are the ordered values of the $(X_i + X_j)/2$ Walsh averages. θ_L is the Walsh average that occupies position t_α in the list. The upper end point θ_U is the Walsh average that occupies the position $M+1-t_\alpha = t_{\alpha/2}$ in this ordered list.~~

$W^{(1)} \quad W^{(2)} \quad W^{(3)} \quad W^{(4)} \quad W^{(5)} \quad W^{(6)} \quad W^{(7)} \quad W^{(8)}$

$t_{\alpha/2} = 7$

$8+1-7 = 2$

$\Rightarrow$ C.I.: $(\theta_L = W^{(2)}, \ \theta_U = W^{(7)})$

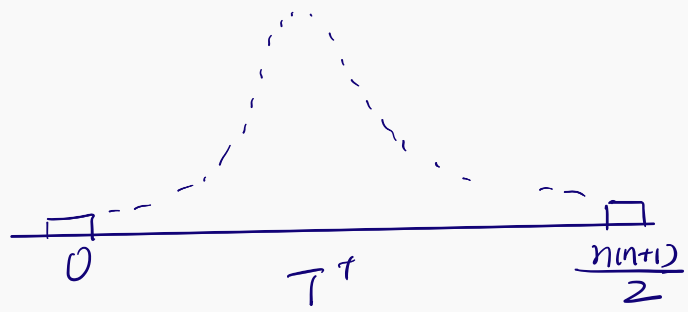
Example: The Mariner and the Pioneer Spacecraft Data

The data were reported by Anderson, Efron, and Wong (1970). The seven observations represent average measurements of, the ratio of the mass of the Earth to that of the moon, obtained from seven different spacecraft.

On the basis of the previous (2-3 years earlier) Ranger spacecraft findings, scientists had considered the value of the ratio of the mass of the Earth to that of the moon to be approximately 81.3035. Thus, we are interested in testing $H_0 : \theta = 81.3035$ versus the alternative $\theta \neq 81.3035$.

		Step 1	Step 2	Step 3
i	X_i	$X_i' = X_i - 81.3035$	Sign (ψ_i)	Rank (R_i)
1	81.3001	$-.0034$	$-$	6
2	81.3015	$-.0020$	$-$	2
3	81.3006	$-.0029$	$-$	4
4	81.3011	$-.0024$	$-$	3
5	81.2997	$-.0038$	$-$	7
6	81.3005	$-.0030$	$-$	5
7	81.3021	$-.0014$	$-$	1

Exact test:



$$= \frac{7\times 8}{2} = 28$$

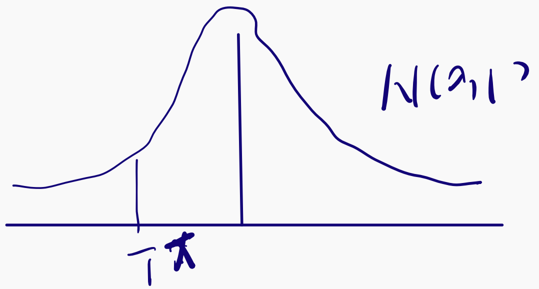
$$T^{+} = 0$$

$$p - value = 1/2^7 = 0.015625$$

$$P(T^+ \le 0 \text{ or } T^+ \ge 28) = 2P(T^+ \le 0)$$
$$= 2P(T^+ = 0)$$
$$= 2 \times \left(\frac{1}{2}\right)^7$$

Confirm with built-in function:

```
> wilcox.test(c(81.3001,81.3015,81.3006,81.3011,81.2997,81.3005,81.3021),mu=81.3035)

        Wilcoxon signed rank exact test

data:  c(81.3001, 81.3015, 81.3006, 81.3011, 81.2997, 81.3005, 81.3021)
V = 0, p-value = 0.01563
alternative hypothesis: true location is not equal to 81.3035
```



Large-sample approximation:

$$T^* = \frac{T^+ - \frac{n(n+1)}{4}}{\left\{\frac{n(n+1)(2n+1)}{24}\right\}^{1/2}} = \frac{0 - [7(8)/4]}{[7(8)(15)/24]^{1/2}} = -2.366$$

$$p - value = 0.01798144 \Leftarrow 2P(\underbrace{Z}_{\sim N(0,1)} \le T^*)$$

> pnorm( ) in R

Both the exact test and the large-sample approximation indicate the existence of strong evidence to reject the findings of the earlier Ranger spacecraft that $\theta = 81.3035$.

Welsh averages	x_1	x_2	x_3	x_4
x_1				
x_2				
x_3				
x_4				

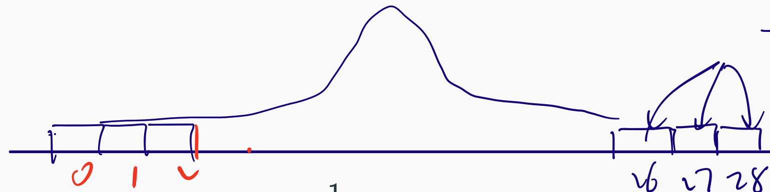
An estimate for median:

```
> library(Rfit)
> sort(walsh(c(81.3001,81.3015,81.3006,81.3011,81.2997,81.3005,81.3021)))
 [1] 81.29970 81.29990 81.30010 81.30010 81.30015 81.30030 81.30035 81.30040
 [9] 81.30050 81.30055 81.30060 81.30060 81.30060 81.30080 81.30080 81.30085
[17] 81.30090 81.30100 81.30105 81.30110 81.30110 81.30130 81.30130 81.30135
[25] 81.30150 81.30160 81.30180 81.30210
```

$$M = 7(8)/2 = 28, \text{ we see that } M = 2k \text{ with } k = 14$$

$$\Rightarrow \hat{\theta} = \frac{W^{(14)} + W^{(15)}}{2} = \frac{81.3008 + 81.3008}{2} = 81.3008$$

95% Confidence interval for median:



$\frac{1}{2^7} = 0.0078125$

With $n = 7$ and $\alpha = .05$, each configuration under null has equal probability of $\frac{1}{2^7} = 0.0078125$, there should be at most 3.2 $= \frac{0.025}{0.0078125}$ configurations to the right of $t_{\alpha/2} = 26$. Thus, $\cancel{t_{1-\alpha/2}} = 28 + 1 - 26 = 3$.

$t_{\alpha/2} = t_{0.025}$

$$\theta_L = W^{(3)} = 81.3001 \text{ and } \theta_U = W^{(26)} = 81.3016$$

so that our 95% confidence interval for θ is

$$(\theta_L, \theta_U) = (81.3001, 81.3016)$$

$W^{(1)}$ $W^{(2)}$ $W^{(3)}$

Confirm with built-in function:

Correction

```
> wilcox.test(c(81.3001,81.3015,81.3006,81.3011,81.2997,81.3005,81.3021),
m=28, μ=81.3035, exact=T, conf.int=T,conf.level=0.95)
```

mu

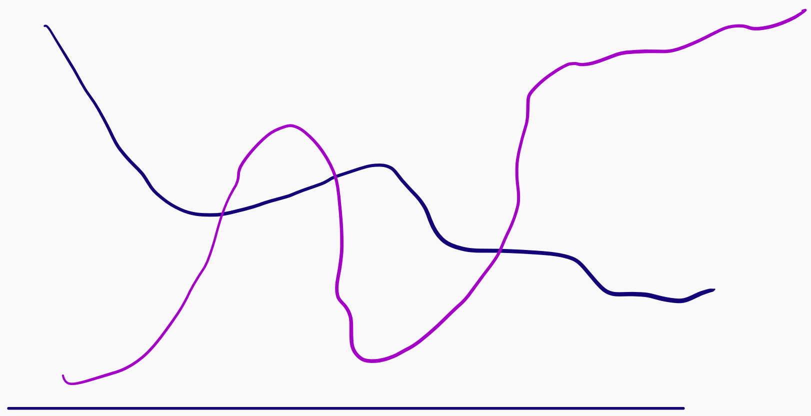
```
        Wilcoxon signed rank exact test

data:  c(81.3001, 81.3015, 81.3006, 81.3011, 81.2997, 81.3005, 81.3021)
V = 28, p-value = 0.01563
alternative hypothesis: true location is not equal to 28
95 percent confidence interval:
 81.3001 81.3016
sample estimates:
(pseudo)median
       81.3008
```

...

$W^{(26)}$ $W^{(27)}$ $W^{(28)}$

Signed Test

Setting



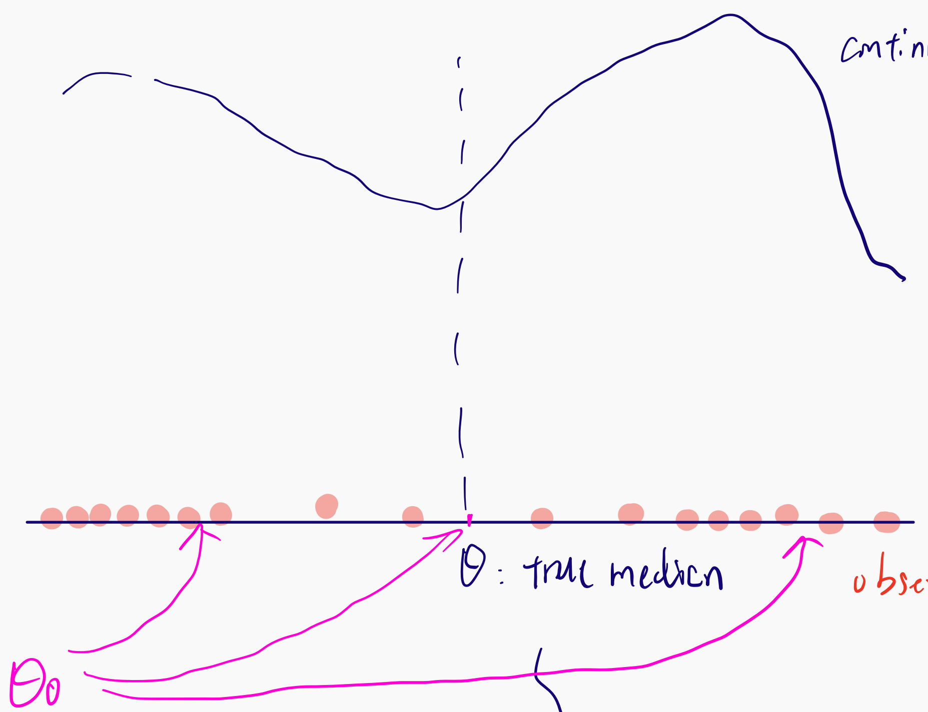
Suppose we have a random sample $x_1 \ldots x_n$

- The x's are mutually independent.
- they are from a population that is continuous with median θ

Hypothesis

$$\begin{aligned}
\text{Two-Sided Test:}& \\
&H_0 : \theta = \theta_0 \text{ versus } H_a : \theta \neq \theta_0 \\
\text{One-Sided Upper-Tail Test:}& \\
&H_0 : \theta = \theta_0 \text{ versus } H_a : \theta > \theta_0 \\
\text{One-Sided Lower-Tail Test:}& \\
&H_0 : \theta = \theta_0 \text{ versus } H_a : \theta < \theta_0
\end{aligned}$$

- Centering: subtract θ_0 from each observation $x_1, \ldots, x_n$ to form a modified sample $x_1' = x_1 - \theta_0, \ldots, x_n' = x_n - \theta_0$
- Define indicator variables $\psi_i, i = 1, \ldots, n$ where $\psi_i = \begin{cases} 1, & \text{if } X_i > \theta_0 \\ 0, & \text{if } X_i < \theta_0 \end{cases}$
- Let $B = \sum_{i=1}^{n} \psi_i$ denote the number of X_i 's out of n that fall above the hypothesized median θ_0.



$\to P(X \geq \theta) = \frac{1}{2}$: probability that each sample is greater than true median is $\frac{1}{2}$.

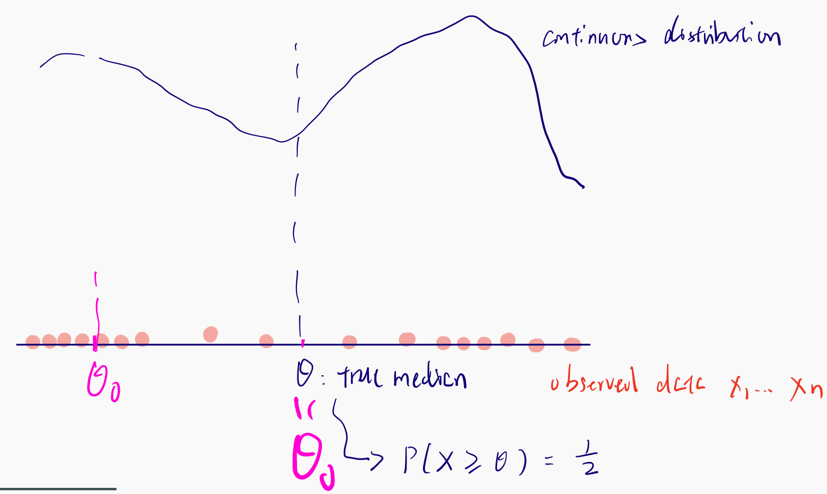
Motivation (upper tail hypothesis $H_0: \theta = \theta_0$ vs $H_a: \theta > \theta_0$)

If H_0 is true, then each X_i has probability .5 of falling above θ_0, so B has a binomial distribution with probability $p = .5$.

If the true median is greater than θ_0, then B has a binomial distribution with probability $p > .5$.

Thus, we can decide between H_0 and H_a based on the value of B; that is, we can test [1]

$$\underbrace{H_0: \theta = \theta_0 \text{ vs } H_a: \theta > \theta_0}_{\text{original one-sample median problem}} \iff \underbrace{H_0 : p = .5, H_a : p > .5}_{\text{Binomial problem}}$$



[1] The test procedures based on the sign statistic B are actually special cases of the general binomial test . The sign test procedures are simply binomial procedures, with "success" corresponding to a positive centered observation, "failure" corresponding to a negative centered observation, and $p = P(\text{"success"}) = P(X_i > \theta_0)$ assuming the value $p_0 = \frac{1}{2}$ when the null hypothesis $H_0 : \theta = \theta_0$ is true.

Procedure

a. One-Sided Upper-Tail Test.

To test

$$H_0 : \theta = \theta_0$$

versus

$$H_1 : \theta > \theta_0$$

at the α level of significance,

Reject H_0 if $B \geq b_\alpha$; otherwise do not reject, where the constant b_α is chosen to make the type I error probability equal to α. The number b_α is the upper α percentile point of the binomial distribution with sample size n and success probability p_0.

The normal approximation:

Reject H_0 if $B^* \geq z_\alpha$; otherwise do not reject.

b. One-Sided Lower-Tail Test.

To test

$$H_0 : \theta = \theta_0$$

versus

$$H_a : \theta < \theta_0$$

at the α level of significance, Reject H_0 if $B \leq b_{1-\alpha}$; otherwise do not reject.

The normal approximation:

Reject H_0 if $B^* \leq -z_\alpha$; otherwise do not reject.

c. Two-Sided Test.

To test

$$H_0 : \theta = \theta_0$$

versus

$$H_{3a} : \theta \neq \theta_0$$

at the α level of significance, Reject H_0 if $B \geq b_{\alpha/2}$ or $B \leq b_{1-\alpha/2}$; otherwise do not reject

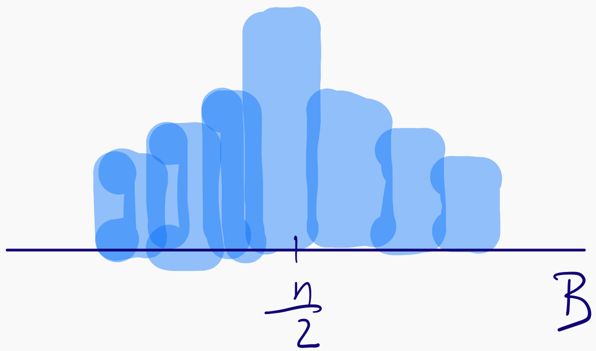
The normal approximation:

Reject H_0 if $|B^*| \geq z_{\alpha/2}$; otherwise do not reject.

An estimator associated with the signed statistics (Hodges-Lehmann)

The null distribution of the statistic $B = \sum_{i=1}^{n} \psi_i \sim Bin(n, \frac{1}{2})$ is around its mean, $n/2$.

A natural estimator of θ is the amount that should be subtracted from each X_i so that the value of B, when applied to the shifted sample $X_1 - \hat{\theta}, \ldots, X_n - \hat{\theta}$, is as close to the center $n/2$ as possible.



$$B = \sum_{i=1}^{n} \psi_i \approx n/2$$

$$\Rightarrow \frac{\sum_{i=1}^{n} \psi_i}{n} \approx 1/2$$

i.e. total observations above the true population median is the sample median.

Intuitively, we estimate θ by the amount that the X sample should be shifted in order that $X_1 - \hat{\theta}, \ldots, X_n - \hat{\theta}$ appears (when "viewed" by the sign statistic B) as a sample from a population with median 0 .

Estimate

The estimator of θ associated with the sign statistic

$$\hat{\theta} = \text{median}\left\{X_i, 1 \leq i \leq n\right\}.$$

Thus,

- if n is odd, say $n = 2k + 1$, we have $k = (n-1)/2$ and

  $$\hat{\theta} = X^{(k+1)},$$

  the value that occupies position $k+1$ in the list of the ordered X_i values.
- If n is even, say $n = 2k$, then $k = n/2$ and

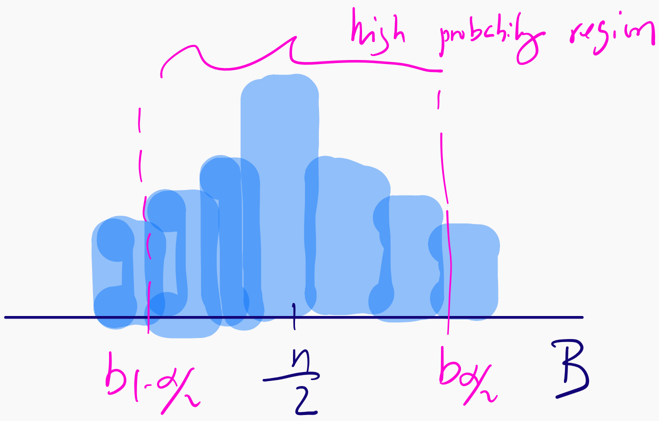
  $$\hat{\theta} = \frac{X^{(k)} + X^{(k+1)}}{2};$$

  that is, when n is even, $\tilde{\theta}$ is the average of the two X_i values that occupy positions k and $k+1$ in the ordered list of the n data values.

Confidence interval based on the signed test

The true population median θ_0 is the value such that the number of observations above it is the signed statisitcs B which should be centered at $\frac{n}{2}$ with some natural variation.

So we use the natural variation of T^+ reverse engineer the most probable region of θ_0 .



$$P(\underbrace{b_{1-\alpha/2} \le B \le b_{\alpha/2}}_{\text{number of observations above the median should be between } b_{1-\alpha/2} \text{ and } b_{\alpha/2}}) = 1-\alpha$$

$\Rightarrow$ median should be between the $b_{1-\alpha/2}$-th and $b_{\alpha/2}$-th ordered observations.

Procedure

$(1-\alpha)100\%$ Confidence Interval

For a symmetric two-sided confidence interval for θ, with confidence coefficient $1-\alpha$, first obtain the upper $(\alpha/2)$ nd percentile point $\boxed{b_{\alpha/2}}$ of the null distribution of $\boxed{B \sim Bin(n, 1/2)}$

$$\cancel{b_{1-\alpha/2}} = \underbrace{n+1-b_{\alpha/2}}.$$ ← lower limit.

The $100(1-\alpha)\%$ confidence interval $(\theta_{\mathrm{L}}, \theta_{\mathrm{U}})$ for θ that is associated with sign test

$$\theta_{\mathrm{L}} = X^{\left(b_{1-\alpha/2}\right)}, \theta_{\mathrm{U}} = X^{\left(b_{\alpha/2}\right)}$$

where $X^{(1)} \leq \cdots \leq X^{(n)}$ are the ordered sample observations; that is, θ_L is the sample observation that occupies position $b_{1-\alpha/2}$ in the list of ordered sample data. The upper end point θ_U is the sample observation that occupies position $b_{\alpha/2}$ in this ordered list.

Then we have

$$P_\theta\left(\theta_{\mathrm{L}} < \theta < \theta_{\mathrm{U}}\right) = 1-\alpha \text{ for all } \theta.$$

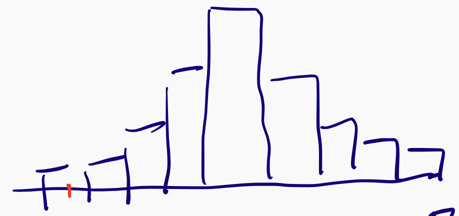
Example: The Mariner and the Pioneer Spacecraft Data

The data were reported by Anderson, Efron, and Wong (1970). The seven observations represent average measurements of, the ratio of the mass of the Earth to that of the moon, obtained from seven different spacecraft.

On the basis of the previous (2-3 years earlier) Ranger spacecraft findings, scientists had considered the value of the ratio of the mass of the Earth to that of the moon to be approximately 81.3035. Thus, we are interested in testing $H_0 : \theta = 81.3035$ versus the alternative $\theta \neq 81.3035$.

i	X_i	$X_i' = X_i - 81.3035$
1	81.3001	$-.0034$
2	81.3015	$-.0020$
3	81.3006	$-.0029$
4	81.3011	$-.0024$
5	81.2997	$-.0038$
6	81.3005	$-.0030$
7	81.3021	$-.0014$

$\alpha = 0.05$



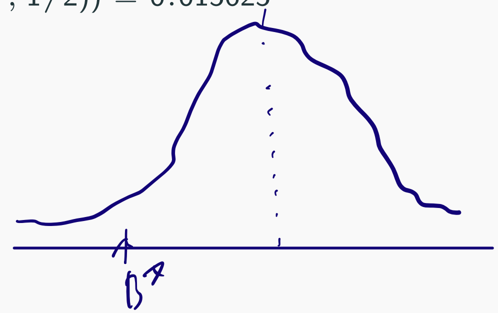
$B \sim Bin(7, \frac{1}{2})$

Exact test:

$$B = 0$$

$$p - value = 2P(B \leq 0 | B \sim Bin(7, 1/2)) = 0.015625$$

Large-sample approximation:



$$B^* = \frac{0 - \left(\frac{7}{2}\right)}{\left(\frac{7}{4}\right)^{1/2}} = -2.645751$$

$Var(B)$

$$p - value = 2P(Z < -2.645751) = 0.008150979$$

Both the exact test and the large-sample approximation indicate the existence of strong evidence to reject the findings of the earlier Ranger spacecraft that $\theta = 81.3035$.

Compare with Wilcoxon, no qualitative difference (minor quantitative difference)

An estimate for median:

The ordered Z observations are $Z^{(1)} \leq \cdots \leq Z^{(7)}$: $\cancel{81.299781.300181.300581.300681.301181.301581.3021}$

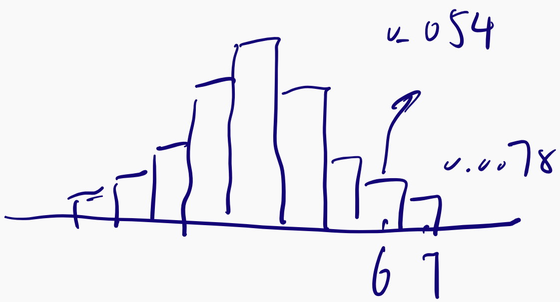
$$\hat{\theta} = Z^{(4)} = 81.3006$$

Confidence interval for median:

With $n = 7$ and $\alpha = .05$, the null distribution of B:

exact null distribution of $B \sim \text{Binomial}(n, p)$
$= \text{Bin}(7, \frac{1}{2})$

```
> dbinom(x=seq(0,7,by=1), size=7, prob=0.5)
[1] 0.0078125 0.0546875 0.1640625 0.2734375 0.2734375 0.1640625 0.0546875 0.0078125
```



not use this notation

$\underbrace{b_{\alpha/2}}_{b_{0.025}} = 7$, $\cancel{b_{1-\alpha/2}} = \underbrace{8}_{n+1 = 7+1} - 7 = 1$

$$\theta_{\rm L} = Z^{(1)} = 81.2997 \text{ and } \theta_{\rm U} = Z^{(7)} = 81.3021$$

so that our 95% confidence interval for θ is

$$(\theta_{\rm L}, \theta_{\rm U}) = (81.2997, 81.3021)$$

corresponds to $\underbrace{\text{Upper Achieved CI}}_{\text{conservative}}$.

conservative : actual confidence level is > 95%

```
> library(BSDA)
> SIGN.test(c(81.3001,81.3015,81.3006,81.3011,81.2997,81.3005,81.3021), md=14)

        One-sample Sign-Test

data:  c(81.3001, 81.3015, 81.3006, 81.3011, 81.2997, 81.3005, 81.3021)
s = 7, p-value = 0.01563
alternative hypothesis: true median is not equal to 14
95 percent confidence interval:
 81.29983 81.30191
sample estimates:
median of x
    81.3006

Achieved and Interpolated Confidence Intervals:

                  Conf.Level  L.E.pt  U.E.pt
Lower Achieved CI     0.8750 81.3001 81.3015
Interpolated CI       0.9500 81.2998 81.3019
Upper Achieved CI     0.9844 81.2997 81.3021
```

← exact p-value

actual confidence > 95%

A Comparison of Statistical Tests

Two statistical issues in choosing between tests

$P(\text{reject } H_0 \mid H_0 \text{ is true}) = P(\text{false positive})$

- Type I error: The probability of a Type I error should be what we claim it to be.
  - t-test:

    from the central limit theorem, valid for large samples, so probabilities of Type I errors found by using the standard normal distribution will be approximately correct.
    Many studies have shown that the approximation is quite good even for moderate sample sizes and a range of population distributions.
  - Wilcox test or Fisher's signed test: stated probability of a Type I error will generally be essentially correct.

Two statistical issues in choosing between tests

$1 - P(\text{Type II error})$

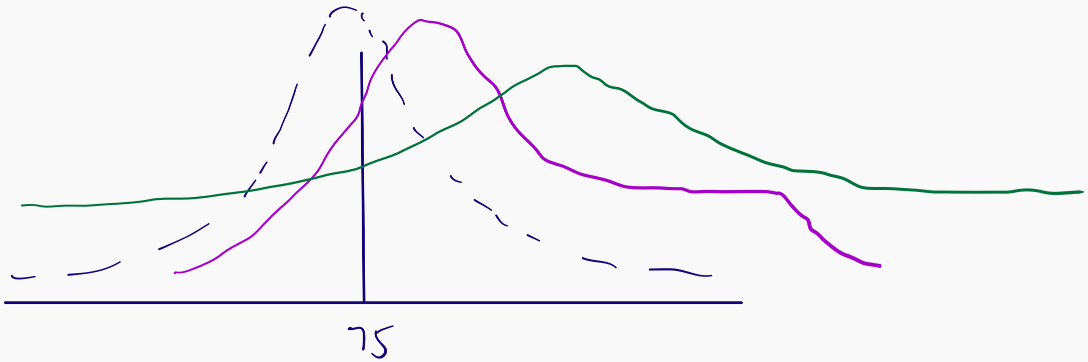
$= P(\text{reject } H_0 \mid H_a \text{ is true}) = P(\text{true positive discovery})$

- Power: probability of rejecting H_0 if it is false.

  Power measures the ability of a test to detect a departure from the null hypothesis, i.e. true positive

  If two tests have the correct probability of a Type I error, then the one with the greater power is the preferred test.

Power is the major concern since it's not controlled by design.



A researcher is testing the sodium contents in packages of food. The desired sodium content of $75\mathrm{mg}$ might not occur because the setting is too high on the machine that puts salt into the product.

For instance, if the amount of sodium per package has a normal distribution with mean $75\mathrm{mg}$ and standard deviation $2.5\mathrm{mg}$ under the correct setting of the salt machine, it may be reasonable to assume that the distribution is normal with mean $\mu > 75\mathrm{mg}$ and standard deviation $2.5\mathrm{mg}$ when the setting is too high.

However, the situation may be even more complicated. There may be a glitch in the machine that causes every tenth package, say, to have twice as much salt as it should.

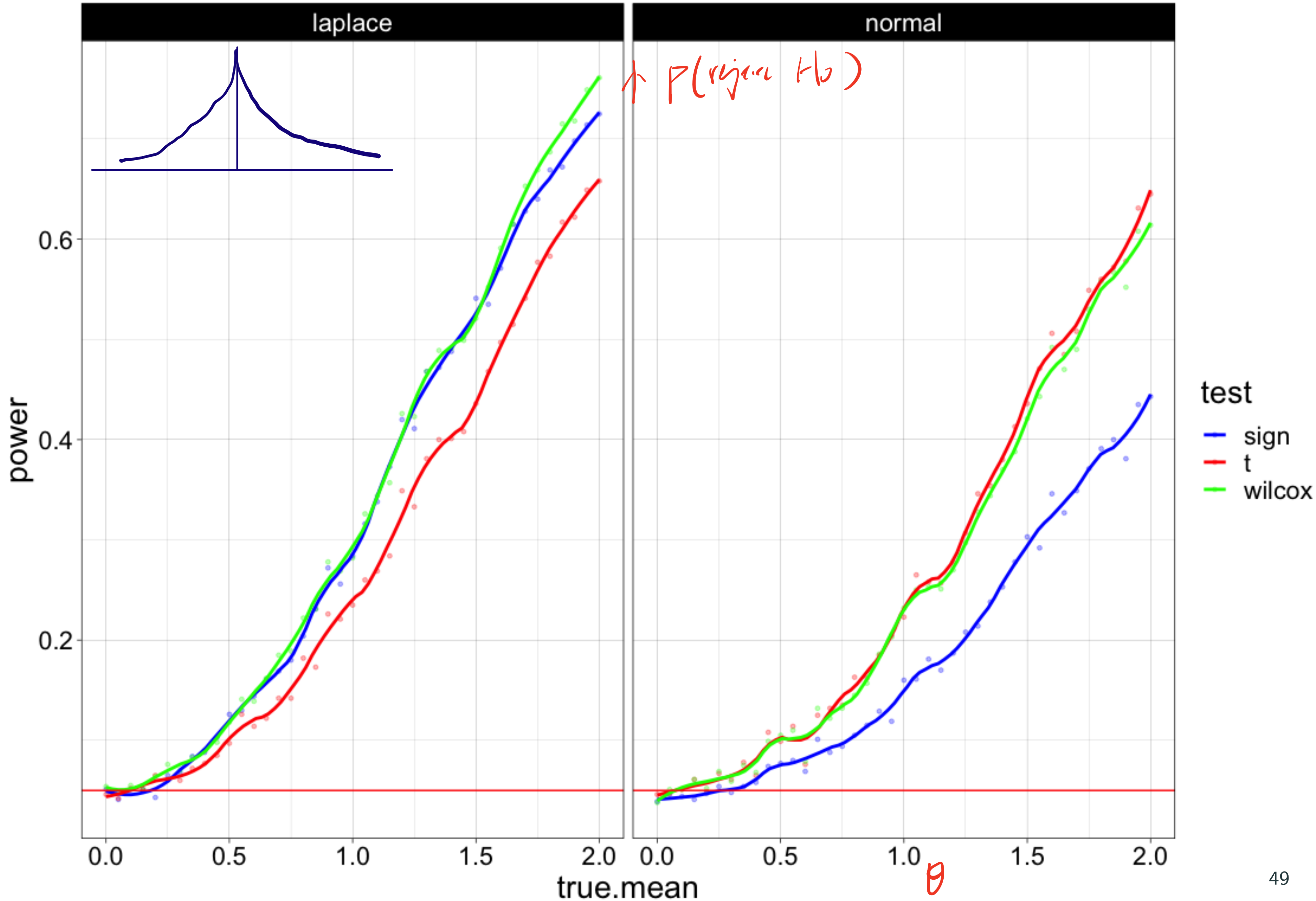
Or an incorrect setting of the machine might cause both the mean and the standard deviation to change. One test may have greater power in one circumstance and the other test may be more powerful in another.

$\Rightarrow$ The strategy is to choose a test that has good power under the alternatives that seem plausible for the problem at hand and/or underlying distribution of the data.

A Simulation Study

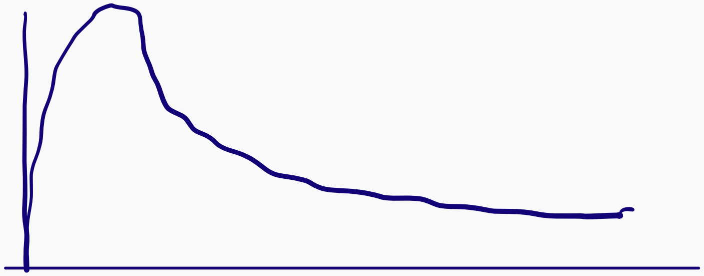
$H_0: \theta = 0$ vs $H_1: \theta \neq 0$

- Type I error ?
- Power
- t. dist df = 1, 2, 3



A Simulation Study

A heavy-tailed probability distribution is a type of probability distribution in which the tail of the distribution (i.e., the portion representing rare events) is "heavy" or thick, meaning that the probability of such events is much higher than in a distribution with a thinner tail, such as the normal distribution.



Examples of heavy-tailed distributions include

- the Pareto distribution
- the Lévy distribution

These types of distributions are often used to model phenomena such as

- income inequality
- popularity of items
- stock market fluctuations

where extreme events (such as large market crashes or a small number of individuals having a large proportion of wealth) occur more frequently than expected based on a normal distribution.

Takeaway

Generally, the nonparametric test will have higher power than the Z/t-test for heavier-tailed population distributions, but the opposite will be true for lighter-tailed distributions.

Statistical theory tells us that Z/t- test has the greatest power among all tests that have the same level of significance, and this advantage holds for all alternatives. The test is said to be uniformly most powerful when the underlying population is truly normal.

Some have mistakenly assumed that the optimal power properties of the (parametric) Z/t-test carry over to the case of sampling from nonnormal populations. This is not the case, not even for large samples.

Paired Comparisons

Setting

We obtain $2n$ observations, two observations on each of n subjects (blocks, patients, etc.).

Subject i	X_i	Y_i
1	X_1	Y_1
2	X_2	Y_2
.	.	.
.	.	.
.	.	.
n	X_n	Y_n

- effect of medical treatment
- effect of new product
- effect of media on public opinion

- The differences $Z_1 = Y_1 - X_1, \ldots, Z_n = Y_n - X_n$ are mutually independent.
- $Z, i = 1, \ldots, n$, comes from a continuous population with a median θ, treatment effect.

Hypothesis

The null hypothesis of interest here is that of zero shift in location due to the treatment, namely,

$$H_0 : \theta = 0.$$

Two-Sided Test:

$$H_0 : \theta = 0 \text{ versus } H_a : \theta \neq 0$$

One-Sided Upper-Tail Test:

$$H_0 : \theta = 0 \text{ versus } H_a : \theta > 0$$

One-Sided Lower-Tail Test:

$$H_0 : \theta = 0 \text{ versus } H_a : \theta < 0$$

Procedure

One-sample tests for location!

Wilcoxon Signed Rank Test, Fisher's Signed Test

Example: Hamilton Depression Scale Factor

The data are a portion of the data obtained by Salsburg (1970). These data, based on nine patients who received tranquilizer, were taken from a double-blind clinical trial involving ~~two~~ one tranquilizers. The measure used was the Hamilton depression scale factor (the "suicidal" factor). The X (pre) value was obtained at the first patient visit after initiation of therapy, whereas the Y (post) value was obtained at the second visit after initiation of therapy. The patients had been diagnosed as having mixed anxiety and depression.

The question of interest is whether the tranquilizer reduce Hamilton depression scale.

$$H_0 : \theta = 0, H_a : \theta < 0$$

Patient i	X_i	Y_i
1	1.83	0.878
2	0.50	0.647
3	1.62	0.598
4	2.48	2.05
5	1.68	1.06
6	1.88	1.29
7	1.55	1.06
8	3.06	3.14
9	1.30	1.29

Wilcoxon's signed rank test

Exact test:

```
> pre<-c(1.83, .50,1.62,2.48,1.68,1.88,1.55,3.06,1.30)
> post<-c(.878, .647, .598,2.05,1.06,1.29,1.06,3.14,1.29)
> z=post-pre
> z
[1] -0.952  0.147 -1.022 -0.430 -0.620 -0.590 -0.490  0.080 -0.010
> sort(abs(z))
[1] 0.010 0.080 0.147 0.430 0.490 0.590 0.620 0.952 1.022
```

$$T^+ = 5$$

$$p - value = 10/2^9 = 0.01953125$$

	# configurations
$T^+ = 0$	(.)
1	(1)
2	(2)
3	(3), (1, 2)
4	(4), (1, 3)
5	(5), (1, 4) (2, 3)

Confirm with built-in function:

```
> wilcox.test(post,pre,paired = T,alternative = "less")

        Wilcoxon signed rank exact test

data:  post and pre
V = 5, p-value = 0.01953
alternative hypothesis: true location shift is less than 0
```

Wilcoxon's signed rank test

Large-sample approximation:

$$T^* = \frac{T^+ - \frac{n(n+1)}{4}}{\left\{\frac{n(n+1)(2n+1)}{24}\right\}^{1/2}} = \frac{5 - (9(10)/4)}{\{9(10)(19)/24\}^{1/2}} = -2.07$$

$$p - value = P(Z < -2.07) = 0.01922617$$

```
> pnorm(-2.07)
[1] 0.01922617
```

Both the exact test and the large-sample approximation indicate that there is strong evidence that tranquilizer does lead to patient improvement, as measured by a reduction in the Hamilton scale factor IV values.

Wilcoxon's signed rank test

An estimate for median:

```
> library(Rfit)
> sort(walsh(z))
 [1] -1.0220 -0.9870 -0.9520 -0.8210 -0.8060 -0.7860 -0.7710 -0.7560 -0.7260
[10] -0.7210 -0.6910 -0.6200 -0.6050 -0.5900 -0.5550 -0.5400 -0.5250 -0.5160
[19] -0.5100 -0.4900 -0.4810 -0.4710 -0.4600 -0.4375 -0.4360 -0.4300 -0.4025
[28] -0.3150 -0.3000 -0.2700 -0.2550 -0.2500 -0.2365 -0.2215 -0.2200 -0.2050
[37] -0.1750 -0.1715 -0.1415 -0.0100  0.0350  0.0685  0.0800  0.1135  0.1470
```

$$M = 45$$

$$\Rightarrow \hat{\theta} = W^{(23)} = -0.46$$

Wilcoxon's signed rank test

Confidence interval for median:

With $n = 9$ and $\alpha = .05$, each configuration under null has equal probability of $\frac{1}{2^9} = 0.001953125$, there should be at most 12.8 configurations to the right of $t_{\alpha/2} = 40$. Thus, $\cancel{t_{1-\alpha/2}} = 45 + 1 - 40 = 6$.

$\theta_{\mathrm{L}} = W^{(6)} = -.786$ and $\theta_{\mathrm{U}} = W^{(40)} = -.010$

so that our 95% confidence interval for θ is

$$(\theta_{\mathrm{L}}, \theta_{\mathrm{U}}) = (-.786, -.010)$$

Wilcoxon's signed rank test

Confirm with built-in function:

```
> wilcox.test(post,pre,paired = T,conf.int = T,conf.level = 0.95)

        Wilcoxon signed rank exact test

data:  post and pre
V = 5, p-value = 0.03906
alternative hypothesis: true location shift is not equal to 0
95 percent confidence interval:
 -0.786 -0.010
sample estimates:
(pseudo)median
         -0.46
```

Fisher's signed test

Exact test:

$$B = 2$$

$$p - value = P(B \leq 2 | B \sim Bin(9, 1/2)) = 0.08984375$$

Large-sample approximation:

$$B^* = \frac{2 - \left(\frac{9}{2}\right)}{\left(\frac{9}{4}\right)^{1/2}} = -1.666667$$

$$p - value = P(Z < -1.666667) = 0.04779032$$

Both the exact test and the large-sample approximation indicate that there is strong evidence that tranquilizer does lead to patient improvement, as measured by a reduction in the Hamilton scale factor IV values.

same qualitative conclusion

Fisher's signed test

Confirm with built-in function:

```
> SIGN.test(z,alt='less',conf.level = 0.95)

        One-sample Sign-Test

data:  z
s = 2, p-value = 0.08984
alternative hypothesis: true median is less than 0
95 percent confidence interval:
  -Inf 0.041
sample estimates:
median of x
      -0.49
```

Fisher's signed test

An estimate for median:

The ordered Z observations are $Z^{(1)} \leq \cdots \leq Z^{(9)}$:

```
> sort(z)
[1] -1.022 -0.952 -0.620 -0.590 -0.490 -0.430 -0.010  0.080  0.147
```

$$\hat{\theta} = Z^{(5)} = -0.49$$

Fisher's signed test

Confidence interval for median: With $n = 9$ and $\alpha = .05$, the null distribution of B:

```
> dbinom(x=seq(0,9,by=1), size=9, prob=0.5)
 [1] 0.001953125 0.017578125 0.070312500 0.164062500 0.246093750 0.246093750
 [7] 0.164062500 0.070312500 0.017578125 0.001953125
```

$b_{\alpha/2} = 8$, $b_{\cancel{1-\alpha/2}} = 9 + 1 - 8 = 2$

$\parallel$

$b_{0.025}$

Correction:

$$\theta_L = Z^{(\cancel{1})\,2} = -0.952 \text{ and } \theta_U = Z^{(\cancel{9})\,8} = 0.080$$

so that our 95% confidence interval for θ is

$$(\theta_L, \theta_U) = (-0.952, 0.080)$$

corresponds to Upper Achieved CI.

actual confidence > 95%

Fisher's signed test

Confirm with built-in function:

```
> library(BSDA)
> SIGN.test(z,alt='two.sided',conf.level = 0.95)

        One-sample Sign-Test

data:  z
s = 2, p-value = 0.1797
alternative hypothesis: true median is not equal to 0
95 percent confidence interval:
 -0.9261778  0.0730000
sample estimates:
median of x
      -0.49

Achieved and Interpolated Confidence Intervals:

                  Conf.Level  L.E.pt U.E.pt
Lower Achieved CI     0.8203 -0.6200 -0.010
Interpolated CI       0.9500 -0.9262  0.073
Upper Achieved CI     0.9609 -0.9520  0.080
```