Lec 4/5

Wednesday, April 5, 2017 14:56

What if lots of observations equal to 0.

Test is not robust to symmetric assumption.

Can still use sign test for median if no assump of symmetry.

Can't use SR test though.

$$\mathbb{P}(T^+ = 0) = 2^{-5} = \frac{1}{32}$$

$$\mathbb{P}(T^+ = 1) = 5 \cdot 2^{-5} \cdot \frac{1}{5} = \frac{1}{32}$$

5 perms

$$\mathbb{P}(T^+ = 2) = 5 \cdot 2^{-5} \cdot \frac{1}{5} = \frac{1}{32}$$

$$\mathbb{P}(T^+ = 3) = 10 \cdot 2^{-5} \cdot \frac{1}{25} + 5 \cdot 2^{-5} \cdot \frac{1}{3} = \frac{1}{2 \cdot 5 \cdot 32} + \frac{1}{32}$$

Let i_k be the corresponding index of rank k i.e.

$$|x_{i_1} - \mu_0| \le |x_{i_2} - \mu_0| \le \cdots \le |x_{i_n} - \mu_0|$$

Let $T^+ = 1 \cdot \xi_1 + 2 \cdot \xi_2 + 3 \cdot \xi_3 + 4 \cdot \xi_4 + 5\, \xi_5$

Where $\xi_j = \begin{cases} 1 & \text{if } x_{i_j} - \mu_0 > 0 \\ 0 & \text{if } x_{i_j} - \mu_0 < 0 \end{cases}$

$\xi_i \sim \text{Bernoulli}(1/2)$ under H_0.

$$\mathbb{P}(T^+ = 0) = \mathbb{P}(\xi_i = 0 \ \forall i) = \frac{1}{32}$$

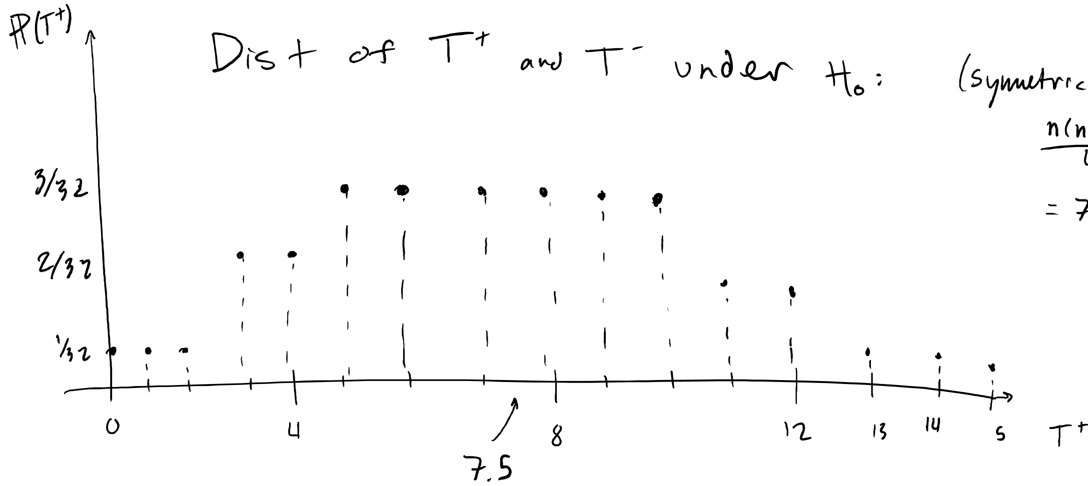
$$\mathbb{P}(T^+ = 1) = \mathbb{P}(\xi_1 = 1, \xi_{i > 1} = 0) = \frac{1}{32}$$

$$\mathbb{P}(T^+ = 2) = \mathbb{P}(\xi_2 = 1, \xi_{i \ne 2} = 0) = \frac{1}{32}$$

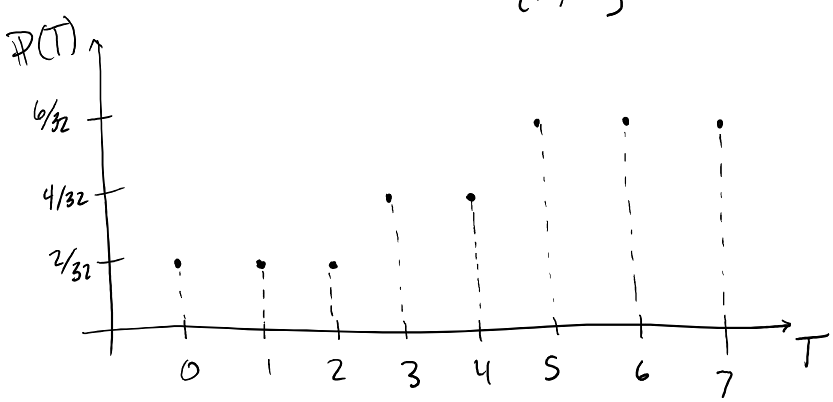
$$\mathbb{P}(T^+ = 3) = \mathbb{P}(\xi_3 = 1, \xi_{i \ne 3} = 0) + \mathbb{P}(\xi_1 = \xi_2 = 1, \xi_{1 \ne i \ne 2} = 0) = \frac{2}{32}$$

$\ldots + \mathbb{P}(\xi_1 = \xi_2 = 1, \xi_{1 \neq i \neq 2} = 0) = \frac{2}{32}$

Dist of T^+ and T^- under H_0: (symmetric about $\frac{n(n+1)}{4} = 7.5$)



Dist of $T = \min\{T^+, T^-\}$



<u>Remark</u>: Since T^+ and T^- are discrete, we may not get type I error rate exactly equal to α.

So, choose T_α = largest integer s.t. $\mathbb{P}(T \leq T_\alpha) \leq \alpha$ under H_0.

Ex: $\mu_0 = 10$. $T^+ = 53$, $T^- = 13$, $T = 13$.

$H_1: \mu \neq \mu_0 \Rightarrow$ reject H_0 if $T < T_\alpha = 11$ fail.

$\mu > \mu_0 \Rightarrow$ reject H_0 if $T^- < T_{2\alpha} = 14$ reject.

$\mu < \mu_0 \Rightarrow$ reject H_0 if $T^+ < T_{2\alpha} = 14$ fail.

<u>Theorem 16.1</u> under assumptions req'd by sign rank test,
T^+ is a RV with mean $\frac{n(n+1)}{4}$ and variance $\frac{n(n+1)(2n+1)}{24}$

<u>Proof</u>: Let $T^+ = 1 \cdot \xi_1 + 2 \cdot \xi_2 + \cdots + n \cdot \xi_n$ where $\xi_i \sim \text{Bernoulli}(1/2)$

$$\mathbb{E}[T^+] = 1 \cdot \mathbb{E}[\xi_1] + 2 \cdot \mathbb{E}[\xi_2] + \cdots + n\,\mathbb{E}[\xi_n]$$

$$= \frac{1}{2} \sum_{i=1}^{n} i = \frac{n(n+1)}{4}$$

$$\text{Var}(T^+) = 1^2\,\text{Var}(\xi_1) + 2^2\,\text{Var}(\xi_2) + \cdots + n^2\,\text{Var}(\xi_n)$$

$$= \frac{1}{4} \sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{24}$$

(Note this is the same for T^-).

* for large n, use normal approximation $T^+ \overset{\text{Approx}}{\approx} N\left(\frac{n(n+1)}{4}, \frac{n(n+1)(2n+1)}{24}\right)$

or $$Z = \frac{T^+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \sim N(0,1) \quad \text{as } n \to \infty$$

** when dealing w/ paired data, we can also use signed-rank test to test $H_0: \mu_1 = \mu_2$ vs $H_1: \mu_1 \gtrless \mu_2$

<u>Ex</u>