Simes Test

$H_{0,i}$, $i = 1, \dots, n$

p_i is p-value for $H_{0,i}$

$H_0 = \bigcap_i H_{0,i}$

Step 1 Let us order p-values:

$$p_{(1)} \le p_{(2)} \le \dots \le p_{(n)}$$

Step 2 $T_n = \min\{p_{(i)} \frac{n}{i}\}$

$$/ \; T_n = n \cdot \min\left(p_{(1)}, \frac{p_{(2)}}{2}, \dots, \frac{p_{(n)}}{n}\right) \; /$$

Step 3 Reject H_0 if $T_n \le \alpha$.

Step 3 follows from

Theorem 1 Under H_0 and independence of the p_i:

$$T_n \sim \mathcal{U}[0,1]$$

Equivalent Formulation:

$$T_n \le \alpha \iff \min\{p_{(i)} \frac{n}{i}\} \le \alpha \iff \exists i: p_{(i)} \le \alpha \frac{i}{n}$$

vs

Bonferroni test: $p_{(1)} \le \frac{\alpha}{n}$

Remark Simes test is less conservative than Bonferroni test (easily reject H_0)

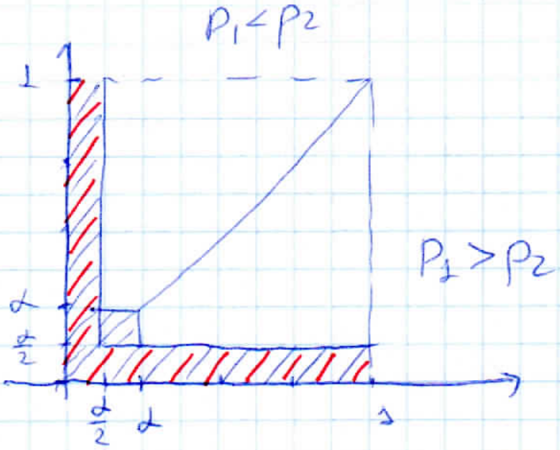
Example

$n=2$ $\quad H_{0,1}, H_{0,2}, \quad H_0 = H_{01} \cap H_{02}$

$p_1 \quad p_2$

$(p_1, p_2) \longmapsto (p_{(1)}, p_{(2)}) \Rightarrow$ reject H_0 if

		if $p_1 < p_2$
Simes $\Rightarrow$	$p_{(1)} \le \frac{\alpha}{2}$ or $p_{(2)} \le \alpha$	$p_1 \le \frac{\alpha}{2}$ or $p_2 \le \alpha$
Bonf $\Rightarrow$	$p_{(1)} \le \frac{\alpha}{2}$	$p_1 \le \frac{\alpha}{2}$



[hatched box] - Sims

[red hatched box] - Bonferroni

$(p_1, p_2) \sim \mathcal{U}([0,1]^2)$

$$\Rightarrow \underset{H_0}{P}(H_0 \text{ is rejected}) = \begin{cases} \alpha - \frac{\alpha^2}{4}, & \text{Bonf} \\ \alpha, & \text{Sims} \end{cases}$$

<u>Proof of the Theorem 1.</u>

<u>Lemma 1</u> $X_1, \dots, X_n$ iid $\sim \mathcal{U}[0,1]$

$$M = \max\{X_1 \dots X_n\} \Rightarrow f_M(t) = \begin{cases} nt^{n-1}, & t \in [0,1] \\ 0, & \text{—//—} \end{cases}$$

$$\ulcorner \; F_{X_i}(t) = P(X_i \le t) = \begin{cases} 0, & t \le 0 \\ t, & t \in (0,1] \\ 1, & t > 1 \end{cases}$$

$$F_M(t) = P(M \le t) = P(X_1 \le t, \dots, X_n \le t) \overset{\text{indep.}}{=} \prod_{i=1}^{n} P(X_i \le t) =$$

$$= \begin{cases} 0, & t \le 0 \\ t^n, & t \in [0,1] \\ 1, & t > 1 \end{cases} \qquad \lrcorner$$

<u>Lemma 2</u> (see Theorem 5.2.1, Hogg ...)

$X_1 \dots X_n$ iid; $Y_1 < Y_2 < \dots < Y_n$ - n order statistics based on $X_1 \dots X_n$ from a continuous distribution with pdf f and support (a,b) $\Rightarrow$ the joint pdf of $Y_1 \dots Y_n$ is given by

$$f(t_1 \dots t_n) = \begin{cases} n!\, f(t_1) \dots f(t_n), & a < t_1 < \dots < t_n < b \\ 0, & \text{—//—} \end{cases}$$

<u>Corollary 1</u> the joint distribution of $p_{(1)} \dots p_{(n)}$:

$$f(t_1 \dots t_n) = \begin{cases} n! & 0 < t_1 < \dots < t_n < 1. \\ 0 & \text{—//—} \end{cases}$$

Example

$n = 2$: X_1, X_2 iid $\sim U[0,1]$, Y_1, Y_2 - order statistics $\iff$ $Y_1 = \min(X_1, X_2)$; $Y_2 = \max(X_1, X_2)$

$$\Rightarrow f(t_1, t_2) = \begin{cases} 2, & 0 < t_1 < t_2 < 1 \\ 0, & \text{—}\,{}''\,\text{—} \end{cases}$$

$$f_{Y_1/Y_2}(t_1/t_2) = \frac{f(t_1,t_2)}{f(t_2)} = \left| \begin{array}{l} f(t_2) = \int_0^{t_2} f(t_1, t_2)\,dt_1 = \\ = 2t_2 \end{array} \right| =$$

$$= \begin{cases} \frac{1}{t_2}, & 0 < t_1 < t_2 \\ 0 & \text{—}\,{}''\,\text{—} \end{cases} \Rightarrow Y_1/Y_2 = t \sim U[0,t]$$

Corollary 2 : Let $p_1 \dots p_{n-1} \overset{iid}{\sim} U[0,t]$ = the joint distribution of $p_{(1)} \dots p_{(n-1)}$:

$$f(t_1 \dots t_{n-1}) = \frac{n-1}{t^{n-1}}, \quad 0 < t_1 \dots < t_{n-1} < t$$

On the other hand: let us calculate pdf of $p_{(1)} \dots p_{(n-1)} / p_{(n)} = t$:

$$f_{1\dots(n-1)/n}(t_1 \dots t_{n-1}) = \frac{n!}{n \cdot t^{n-1}} =$$

$\leftarrow$ pdf of $p_{(1)} \dots p_{(n)}$

$\leftarrow$ pdf of $p_{(n)}$

$$= \frac{(n-1)!}{t^{n-1}}, \quad 0 < t_1 < \dots < t_{n-1} < \underbrace{t_n}_{t} < 1$$

$\Rightarrow$ conditional on $p_{(n)} = t$ the other p values are independently uniform on $[0,t]$

Proof of the Theorem 1. By induction.

$n = 1$ is true

Assume that it is true for $n-1$: $T_{n-1} \sim U[0,1]$

We have $p_{(1)} \le \dots \le p_{(n)}$

$f(t) = nt^{n-1}$, $t \in [0,1]$ - pdf of $p_{(n)}$

Then

$$P(T_n \le \alpha) = |\text{Law of Total Probability}| =$$

$$= \int_0^1 P(T_n \le \alpha / p_{(n)} = t)\, f(t)\, dt =$$

$$= \int_0^\alpha \underbrace{P(T_n \le \alpha / p_{(n)} = t)}_{T_n = \min(p_{(i)} \frac{n}{i})} f(t)\, dt + \int_\alpha^1 P(T_n \le \alpha / p_{(n)} = t) f(t) dt$$

$$\geq P\left(p_{(n)} \frac{n}{n} \le \alpha / p_{(n)} = t\right) =$$

$$= P(p_{(n)} \le \alpha / p_{(n)} = t) = 1, \quad t \le \alpha.$$

$$= \underbrace{\int_0^\alpha f(t)\, dt}_{\alpha^n} + \int_\alpha^1 \underbrace{P(T_n \le \alpha / p_{(n)} = t)}\, f(t)\, dt \;=$$

$$P(T_n \le \alpha / p_{(n)} = t) = P\left(\min_{1 \le i \le n} \left(p_{(i)} \frac{n}{i}\right) \le \alpha / p_{(n)} = t\right) =$$

$$\overset{t > \alpha \Rightarrow p_{(n)} \frac{n}{n} > \alpha}{=} \quad P\left(\min_{1 \le i \le n-1} \left(p_{(i)} \frac{n}{i}\right) \le \alpha / p_{(n)} = t\right) =$$

$$= P\Bigg(\underbrace{\min_{1 \le i \le n-1} \bigg(\underbrace{\frac{p_{(i)}}{t}}_{\sim U[0,1]} \frac{n-1}{i}\bigg)}_{T_{n-1} \sim U[0,1] \text{ by assumption}} \le \frac{\alpha}{t} \cdot \frac{n-1}{n} \Big/ p_{(n)} = t\Bigg) =$$

$$= \frac{\alpha}{t} \cdot \frac{n-1}{n}$$

$$= \alpha^n + \int_\alpha^1 \frac{\alpha}{t} \frac{n-1}{n}\, n\, t^{n-1}\, dt = \alpha^n + \alpha(n-1) \int_\alpha^1 t^{n-2}\, dt =$$

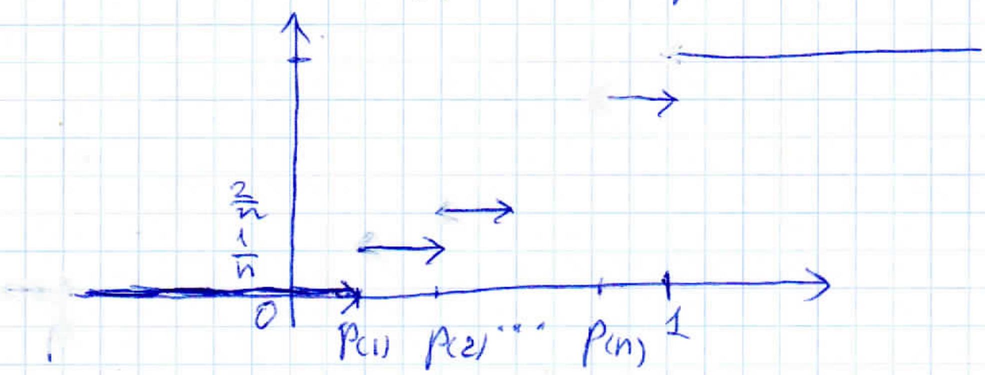
$$= \alpha^n + \alpha(1 - \alpha^{n-1}) = \alpha.$$

Tests based on empirical CDF's

goodness of fit tests – tests that verify that the sample comes from given distribution

Empirical CDF:

$$\hat{F}_n(t) = \frac{1}{n} \#\{i : p_i \le t\}$$



Under H_0, assume that p_i are independent

$$y_i = \begin{cases} 1, & p_i \le t \\ 0, & p_i > t \end{cases} \qquad y_1 \dots y_n \text{ are independent}$$

$P(y_i = 1) = P(p_i \le t) = t$ if $t \in [0,1]$

$\Leftrightarrow$ $y_1 \dots y_n$ iid, have Bernoulli distribution with probability of success $= t$.

$$S_n(t) = \#\{i : p_i \le t\} = \sum_{i=1}^{n} y_i \sim B(n,t)$$

$$\Rightarrow F_n(t) = \frac{1}{n} S_n(t) \Rightarrow E\,F_n(t) = \frac{nt}{n} = t$$

We know:

1) $EF_n(t) = t = F(t)$
2) $\forall t \quad F_n(t) \longrightarrow F(t)$ with $P1$ (LLN)
3) Glivenko – Cantelli theorem
$P(\sup_{t \in \mathbb{R}} |F_n(t) - F(t)| \to 0) = 1 \Leftrightarrow$ convergence is uniform
4) $\text{var}\, F_n(t) = \frac{1}{n^2} \underbrace{\text{var}\, S_n(t)}_{nt(1-t)} = \frac{t(1-t)}{n}$

$$\xrightarrow{CLT} \sqrt{n}\, \frac{F_n(t) - F(t)}{\sqrt{F(t)(1-F(t))}} \to N(0,1)$$

$$\sqrt{n}\,\frac{F_n(t)-F(t)}{t(1-t)} \to N(0,1) \qquad \forall t\in[0,1]$$

$w(t) = \sqrt{n}\,(F_n(t)-F(t))$ — empirical process

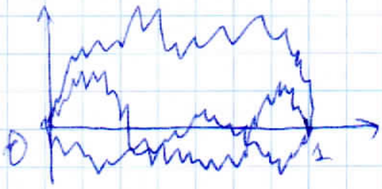
in our case $w(t) = \sqrt{n}\,(F_n(t)-t)$.

$w(t) \to B(t) \leftarrow$ Brownian bridge

/ Brownian bridge - Gaussian process ; $EB(t)=0$

$\mathrm{var}\,B(t)=t(1-t)$

$\mathrm{cov}(B(t), B(s)) = t\wedge s - ts$



<u>Kolmogorov - Smirnov test</u>.

$KS = \sup_{t\in[0,1]} \sqrt{n}\,|\hat{F}_n(t)-t| \to \sup_{[0,1]} B(t)$

↑ Brownian Bridge

are tabulated

<u>Anderson - Darling test</u>

$$A^2 = n\int_0^1 (\hat{F}_n(t)-t)^2\, w(t)\,dt$$

$w(t)\equiv 1 \Rightarrow$ Cramer-von Mises statistic.

$w(t) = \frac{1}{t(1-t)} \Rightarrow$ Anderson - Darling statistics

$$A^2 = -n - \sum_{i=1}^{n} \frac{2i-1}{n}\left[\log p_{(i)} + \log(1-p_{(n+1-i)})\right]$$

/ compare with Fisher's statistic : $T = -\sum_{i=1}^{n} 2\log p_i$.

A.-D. statistic gives more weight to p-values in the center