Lec 8 MATH 341/641

m: # of hypothesis tests

m_0: # of tests s.t. H_0 is true, $m_0 \le m$

r: # of rejections, $r \le m$, R is its r.v.

v: # of Type I errors, $v \le r$, V is its rv

If $m = m_0 = 1$, $P(R=1) = P(V=1) \ge \alpha$, the level which is set by you

If $m = m_0$

$V = R = \text{Binom}(m, \alpha)$

$P(V \ge 1) = 1 - P(V=0) = 1 - (1-\alpha)^m$

for $\alpha = 5\%$, $m = 30$, $P(V \ge 1) = 76\%$

for any $\alpha > 0$, $m \to \infty$, $P(V \ge 1) \to 1$

Family-wise error (FWER) control

$FWER := P(V \ge 1)$

A family of tests is a logical collection of tests where its meaningful to talk about a notion of error on the collection.

Not well-defined!

The analogue of α is now denoted $FWER_0$, the value you control e.g. $FWER_0 = 5\%$

If $FWER \le FWER_0$ for any $m_0 \le m$, this is called "strong control of FWER".

If $FWER \le FWER_0$ for all $m_0 = m$, this is called "weak control of FWER".

We will focus on weak control of FWER. The r.v.'s of interest are ($m = m_0$)

$$R_1 := \begin{cases} 1 & \text{if first test's } H_0 \text{ is rejected} \\ 0 & \text{if first test's } H_0 \text{ is retained} \end{cases} \sim \text{Bern}(\alpha)$$

$$R_2 := \begin{cases} 1 & \text{if second test's } H_0 \text{ is rejected} \\ 0 & \text{if second test's } H_0 \text{ is retained} \end{cases} \sim \text{Bern}(\alpha)$$

$\vdots$

$$R_m = \{ \qquad \sim \text{Bern}(\alpha)$$

no knowledge of the dependence structure of these Bernoullis

$R := R_1 + R_2 + \ldots + R_m$, the total # of rejections

Assume level α for each of the m tests

Bode's Inequality

$$FWER := P(R \ge 1) = P(R_1 = 1 \cup R_2 = 1 \cup \ldots \cup R_m = 1) \le \sum_{l=1}^{m} P(R_l = 1) = m\alpha$$

set $FWER_0 = m\alpha$ and solve for α

$$\Rightarrow FWER \le m\alpha \Rightarrow \text{if } \alpha = \frac{FWER_0}{m} \Rightarrow FWER \le FWER_0$$

e.g. $FWER_0 = 5\%$, $m = 100 \Rightarrow \alpha = 0.0005$.

This called the Bonferroni Correction (1936).

It always works but it's extremely conservative i.e. it's very difficult to reject H_0 if H_0 is false i.e. Power is very low!

Let's assume $R_1, R_2, \ldots, R_m \overset{iid}{\sim} \text{Bern}(\alpha)$

$$FWER = P(R \geq 1) = 1 - P(R = 0) = 1 - (1-\alpha)^m$$

set $FWER_0 = 1 - (1-\alpha)^m$ and ~~solve~~ solve for α

$$1 - FWER_0 = (1-\alpha)^m \Rightarrow 1 - \alpha = (1 - FWER_0)^{1/m} \Rightarrow \alpha = 1 - (1 - FWER_0)^{\frac{1}{m}}$$

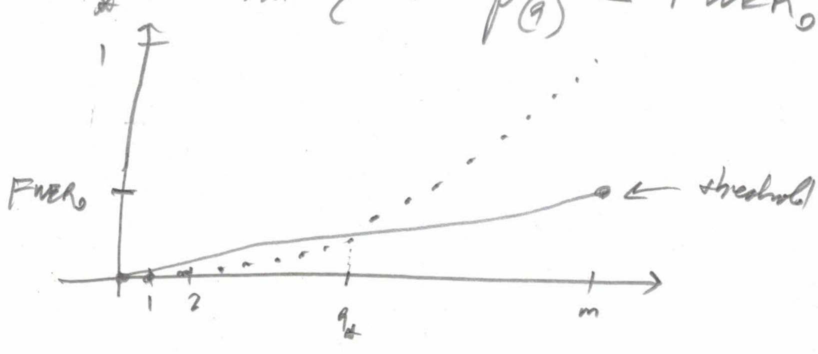
This is called the Dunn-Sidak correction (1967).

if $FWER_0 = 5\%$, $m = 100$ $\Rightarrow \alpha = .000513 >$ Bonferroni α but not by much.

There needs to be a better way! Note that Neither Bonferroni nor Dunn examined the p-vals themselves, which gauge "strength of rejection". A pval of .0001 is much stronger than a pval of .01. Simes in 1986 decided to use this idea. For all m independent tests, there are m pval's:

$p_1, p_2, \ldots, p_m$. Sort them from smallest to largest $\underset{\text{min}}{p_{(1)}}, p_{(2)}, \ldots, \underset{\text{max}}{p_{(m)}}$.

Let $a_* := \max \left\{ a : p_{(a)} \leq FWER_0 \frac{a}{m} \right\}$ "linear step-up" or ∞ if max doesn't exist



Then reject all hypotheses corresponding to p_{i}'s $\leq FWER_0 \frac{a_*}{m}$

Proof comes after we do order statistics in 340.

Maybe... FWER is not the metric to care about..

FWER $= P(R \geq 1)$ is really conservative. 5% that any rejections happen at all?

How about the following. Let the false discovery proportion $(FDP)_{rv}$ be:

$$FDP := \begin{cases} \frac{V}{R}, & R>0 \\ 0, & R=0 \end{cases}$$, the prop. of all rejections that are Type I errors

Since you need to control a scalar quantity, define the false discovery rate (FDR) to be $FDR := E[FDP]$ which we control to be no more than FDR_0 e.g.

$FDR \leq FDR_0 = 5\%$ we say the FDR is controlled at 5%..

$\Rightarrow$ If you get $r=100$, you expect $v=5$, i.e. 5% and 95 to be true discoveries (i.e. H_a is true).

If $m=m_0$, FWER = FDR. Proof:

Since $m=m_0$, $V=R \Rightarrow FDP = \begin{cases} 1 \text{ if } R\geq 1 \\ 0 \text{ if } R=0 \end{cases} = \text{Bern}(P(R \geq 1)) = \text{Bern}(\text{FWER})$

$$\Rightarrow E[FDP] = FWER$$

So FDR only makes sense if $m_0 < m$, i.e. there are some discoveries to be made. This makes FDR control realistic unlike controlling weak FWER.

Benjamini & Hochberg, 1995

Thm: the Simes procedure controls FDR, i.e.

if you let $a_x := \max \{ a : p_{(a)} \le FDR_0 \frac{a}{m} \}$ or ∞ if no max

and reject all tests where their pval $< FDR_0 \frac{a_x}{m}$. Proof too difficult.

This result is the most cited result in the entire field of statistics.

FDR is less conservative than FWER

$$\frac{V}{R} \le \mathbb{1}_{V \ge 1}$$

if $V=0$, $0 \le 0$

if $V=1$, $1 \le 1$, $\frac{1}{2} \le 1$, $\frac{1}{3} \le 1$, ...

if $V=2$, $1 \le 1$, $\frac{2}{3} \le 1$, $\frac{3}{4} \le 1$, ...

⋮

if $V \ge 1$, $\frac{V}{R} \le 1$

$$\Rightarrow E\left[\frac{V}{R}\right] \le E\left[\mathbb{1}_{V \ge 1}\right]$$

$\Rightarrow$ FDR $\le$ FWER.

to get the same # of false discoveries, you need a higher FWER

to get the same FWER, you have to be okay with much less false discoveries.

Totally different type of test. Called Goodness-of-Fit.

Usually we assumed DGP: $X_1, \ldots, X_n \stackrel{iid}{\sim} f(x; \theta)$

and we wanted to infer θ.

Now, we flip things to reverse. We begin with data

e.g. 1.73, −0.91, 0.93, 2.16, 0.03 and want to show

"is the DGP differ from $X_1, \ldots, X_n \stackrel{iid}{\sim} f(x;\theta)$"?

e.g.

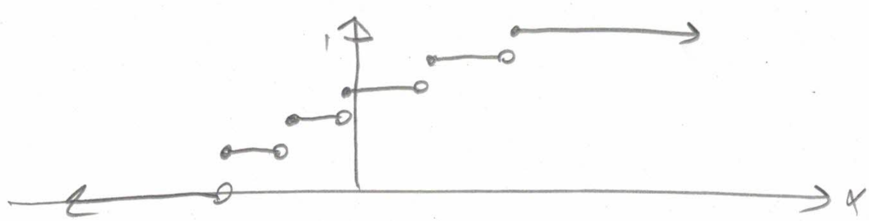
note: needs full specification. There cannot be any unknown params

$H_0: X_1, \ldots, X_n \stackrel{iid}{\sim} N(0,1)$ i.e. $F_X(x) = \Phi(x)$

H_a: DGP is anything else

If the H_0 is a continuous DGP, we can use the Kolmogorov-Smirnov Test. We first compute the estimate of $F(x)$, $\hat{F}_n(x)$ called the empirical CDF (ECDF).

$$\hat{F}_n(x) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{x_i \le x}$$ e.g.



this is a famous estimator for the true CDF.

Under H_0, F_{H_0} is assumed. Now we need to measure difference between F_{H_0} and $\hat{F}_n$. We need a test statistic and we need its distr under H_0.

$$D_n = \overset{\text{scalar}}{\text{diff}}\left(\hat{F}_n, F_{H_0}\right)$$

One possibility is the largest absolute difference between the two functions.

$$D_n := \sup_x \left\{ \left| \hat{F}_n(x) - F_{H_0}(x) \right| \right\}$$ "supremum norm difference"

Glivenko-Cantelli Thm of 1933: $D_n \to 0$ if H_0 is true $\Rightarrow \lim_{n\to\infty} \hat{F}_n(x) = F(x)$

$D_n \to c \neq 0$ if H_0 not true $\Rightarrow$ Power $\to 1$ as $n \to \infty$.

What's the sampling distr of D_n?

Kolmogorov in 1933 proved that

$$\sqrt{n}\, D_n \xrightarrow{d} K$$, the Kolmogorov Distr. (show it on wikipedia)

Amazing result which is completely distr.-free like the CLT.

Tables of critical values are tabulated. Recommended to have $n > 50$.

(show values)