36-700: Probability and Mathematical Statistics I **Fall 2016**

Lecture 26: November 2

Lecturer: Siva Balakrishnan

26.1 Review and Outline

In the last class we discussed hypothesis testing:

1. Permutation based two-sample tests
2. Multiple testing: FWER and FWER control

Today we will continue our discussion of multiple testing, and then discuss confidence intervals.

26.2 The False Discovery Rate

The Bonferroni and Sidak corrections we discussed in the last lecture can be very conservative. They ensure that we do not make even a single false rejection. In scientific experiments it often makes sense to control what is called the **False Discovery Rate** (FDR). The FDR is the expected number of false rejections divided by the number of rejections.

Denote the number of false rejections as V, and the total number of rejections as R. Then the false discovery *proportion* is:

$$\text{FDP} = \begin{cases} \frac{V}{R} & \text{if} \quad R > 0 \\ 0 & \text{if} \quad R = 0. \end{cases}$$

The FDR is then defined as:

$$\text{FDR} = \mathbb{E}[\text{FDP}].$$

In this notation we can see that the FWER is:

$$\text{FWER} = \mathbb{P}(V \geq 1).$$

We will next consider how one can control the FDR. We will describe a procedure known as the Benjamini-Hochberg (BH) procedure. We will not prove its correctness (its not difficult but it is a bit involved).

26.2.1 The BH procedure

The BH procedure is one that controls the FDR under independence (i.e. the p-values are independent). There is a much weaker form of this procedure that works under dependence (see the Wasserman book). It turns out to be very challenging to tightly control FDR under strong dependence.

The procedure is:

1. Suppose we do T tests. Let us take the p-values $p_1, \ldots, p_T$, and sort them, i.e. we create the list: $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(T)}$.

2. Define the thresholds:

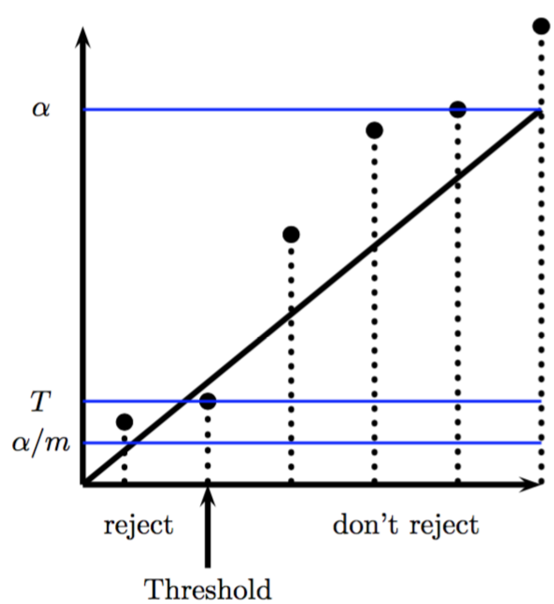
$$t_i = \frac{i\alpha}{T}.$$

3. Find the largest $i_{\max}$ such that

$$i_{\max} = \arg\max_i \{i : p_{(i)} < t_i\}.$$

4. Reject all nulls upto and including $i_{\max}$.

This might seem a bit confusing but here is a simple picture:



26.2.2 Properties of FDR

We have now see a procedure that controls the FDR under some assumptions. One question of interest is how does FDR control compare to FWER control? Another is just: how do we interpret FDR control?

Interpreting FDR control: The way to think about FDR control is: if we repeat our experiment many times, on average we control the FDP. This is not a statement about the individual experiment we did conduct, and really it does not say much about how likely it is that on a given experiment we have an FDP that is larger than a threshold (think about using Markov's inequality).

FWER on the other hand, does control the error rate for a single experiment. That is, with FWER control, we have managed our false discoveries unless we are very unlucky; with FDR control, on average our test will control FDP, but in our particular experiment we may not have done a very good job. We will see in a second controlling FWER does control the FDR. The way to interpret all of this is that: FDR control is a very weak notion of error control.

Connection to FWER:

1. The first connection is that under the global null (when all the null hypotheses are true) FDR control is equivalent to FWER control.

   Proof: Under the global null, any rejection is a false rejection. There are two possibilities: either we do not reject anything: in this case the FDP $= 0$. If we do reject any null hypothesis then our FDP is 1 (since $V = R$). So we have that:

$$\text{FDR} = \mathbb{E}[\text{FDP}] = \mathbb{P}(V > 0) * 1 + \mathbb{P}(V = 0) * 0 = \mathbb{P}(V > 0) = \text{FWER}.$$

2. The second connection is that the FWER $\geq FDR$ always. This implies that controlling the FWER implies FDR control.

   Proof: We can see that the following is a simple upper bound on the FDP:

$$\text{FDP} \leq \mathbb{I}(V \geq 1),$$

   since if $V = 0$, FDP $= 0$, and if $V > 0$ then $V/R \leq 1$. Taking expectations of this expression gives:

$$\text{FDR} \leq \mathbb{P}(V \geq 1) = \text{FWER}.$$

   The flip-side of this is that FDR control is less stringent so if this is the correct measure for you then you will have *more* power by controlling FDR (rather than controlling FWER).

26.3 Confidence Intervals

We have already discussed the construction of confidence intervals and asymptotic confidence intervals at various points in the course. Today, we will discuss a way to construct a confidence interval by *inverting a hypothesis test.*

Before we do this lets review a simple relation between tests and intervals.

From intervals to tests: Suppose that:

$$P_\theta(\theta \in C_n(x_1, \dots, x_n)) = 1 - \alpha,$$

then the test: reject $H_0 : \theta = \theta_0$ if $\theta_0 \notin C_n(x_1, \dots, x_n)$ has size at most α. To see this we just note that under the null, the probability of rejecting the null is at most α.

From tests to intervals: A natural question is whether we can somehow construct a confidence interval from a procedure that performs hypothesis tests for us. This is called inverting a test. We suppose that for every parameter θ_0, we have a hypothesis tester with level α for the hypothesis test:

$$\begin{aligned} H_0 : &\quad \theta = \theta_0 \\ H_1 : &\quad \theta \neq \theta_0. \end{aligned}$$

The procedure is then simple: we consider every parameter θ and put it in the confidence set if our tester fails to reject the null hypothesis, i.e. in some sense we are not "sure" that the parameter is not θ. This results in a confidence set $C_n(x_1, \dots, x_n)$ with coverage at least $1 - \alpha$.

This is easy to verify:

$$P_{\theta_0}(\theta_0 \notin C_n(x_1, \dots, x_n)) \leq P(\text{falsely rejecting null } H_0 : \theta = \theta_0) \leq \alpha.$$

Example: Suppose we saw samples $X_1, \dots, X_n \sim N(\theta, 1)$, and our hypothesis test of choice was the Wald test. We would reject the null hypothesis: $\theta = \theta_0$ at level α if:

$$T_n = \sqrt{n} \left| \frac{1}{n} \sum_{i=1}^{n} X_i - \theta_0 \right| \geq z_{\alpha/2}.$$

So the confidence interval we would construct is:

$$C_n(X_1, \dots, X_n) = \left\{ \theta : \sqrt{n} \left| \frac{1}{n} \sum_{i=1}^{n} X_i - \theta \right| \leq z_{\alpha/2} \right\}.$$

This is just the usual $1 - \alpha$ confidence interval.