DS 102 Data, Inference, and Decisions **Fall 2019**

Lecture 5: More on Online FDR; Permutation Testing

Lecturer: Michael I. Jordan

1 LORD Algorithm

We return to the LORD algorithm from the previous lecture. There are several different versions of the LORD algorithm, and for simplicity, in this section we analyze a version slightly different from the one from previous lecture.

Let r_t the time of the last rejection before time t, and let $\{\gamma_t\}_{t=1}^{\infty}$ be a non-negative infinite sequence that sums to 1. The LORD version that we consider assigns significance levels at each time step t as:

$$\alpha_t = \begin{cases} \gamma_t \alpha, & \text{if no rejection has yet been made} \\ \gamma_{t-r_t}\alpha, & \text{otherwise} \end{cases}.$$

First, we show that this update guarantees an upper bound on estimate of the FDP:

$$\widehat{\text{FDP}} := \frac{\sum_{i=1}^t \alpha_i}{\sum_{i=1}^t \mathbf{1}\{P_i \le \alpha_i\}} \le \alpha.$$

Notice that the denominator is equal to the total number of discoveries. And more generally, any version of the LORD algorithm guarantees this inequality (you can check the update from Lecture 4 and prove this yourself!).

We will first show that the LORD update satisfies $\widehat{\text{FDP}} \le \alpha$, and then we will show (approximately) that this is sufficient for FDR control.

Suppose that we have completed t tests. At some of these t time steps, we have made discoveries. Denote by τ_j the time of the j-th discovery. Suppose we have made D total discoveries so far. For every $j \le D$, consider the significance levels in an epoch between two discoveries, i.e. at times $\tau_j, \tau_j + 1, \ldots, \tau_{j+1}$. By definition of the significance level update, these levels are equal to $\alpha\gamma_1, \alpha\gamma_2, \alpha\gamma_3, \ldots$. If we sum up these test levels, we get at most α, because $\sum_{t=1}^{\infty} \gamma_t = 1$. Moreover, for each such epoch we get the sum of significance levels at most α, so:

$$\sum_{i=1}^t \alpha_i \le \alpha \cdot \text{ number of ``epochs''}.$$

However, notice that the number of epochs is exactly equal to the number of rejections, so

$$\sum_{i=1}^t \alpha_i \le \alpha \sum_{i=1}^t \mathbf{1}\{P_i \le \alpha_i\},$$

and after dividing each side by the total number of rejections, we can conclude that $\widehat{\text{FDP}} \leq \alpha$.

Now we want to show that $\widehat{\text{FDP}} \leq \alpha$ implies FDR control. A formal proof of this is slightly more contrived, so we will present a proof stating that a close approximation of the FDR is controlled. The approximation we consider is

$$\text{FDR} \approx \frac{\mathbb{E}[\sum_{i\leq t, i \text{ null}} \mathbf{1}\{P_i \leq \alpha_i\}]}{\mathbb{E}[\sum_{i\leq t} \mathbf{1}\{P_i \leq \alpha_i\}]}.$$

The difference between the approximation and exact FDR is that FDR takes an expectation of the ratio, while here we are taking a ratio of expectations. We will show that approximately FDR $\leq \alpha$ by showing

$$\mathbb{E}\left[\sum_{i\leq t, i \text{ null}} \mathbf{1}\{P_i \leq \alpha_i\}\right] \leq \alpha\, \mathbb{E}\left[\sum_{i\leq t} \mathbf{1}\{P_i \leq \alpha_i\}\right].$$

First, by the tower property, we have:

$$\mathbb{E}\left[\sum_{i\leq t, i \text{ null}} \mathbf{1}\{P_i \leq \alpha_i\}\right] = \mathbb{E}\left[\sum_{i\leq t, i \text{ null}} \mathbb{E}[\mathbf{1}\{P_i \leq \alpha_i\}|\alpha_i]\right].$$

Next, we use the fact that the expectation of an indicator of an event is the probability of that event:

$$\mathbb{E}\left[\sum_{i\leq t, i \text{ null}} \mathbb{E}[\mathbf{1}\{P_i \leq \alpha_i\}|\alpha_i]\right] = \mathbb{E}\left[\sum_{i\leq t, i \text{ null}} \mathbb{P}(P_i \leq \alpha_i|\alpha_i)\right].$$

By uniformity of null p-values, we further have:

$$\mathbb{E}\left[\sum_{i\leq t, i \text{ null}} \mathbb{P}(P_i \leq \alpha_i|\alpha_i)\right] = \sum_{i\leq t, i \text{ null}} \mathbb{E}[\alpha_i].$$

By summing up all the test levels, and not just the null ones (remember $\alpha_i \geq 0$), we get

$$\sum_{i\leq t, i \text{ null}} \mathbb{E}[\alpha_i] \leq \sum_{i\leq t} \mathbb{E}[\alpha_i].$$

Finally, we use $\widehat{\text{FDP}} \leq \alpha$ to conclude the argument:

$$\mathbb{E}\left[\sum_{i\leq t, i \text{ null}} \mathbf{1}\{P_i \leq \alpha_i\}\right] \leq \sum_{i\leq t} \mathbb{E}[\alpha_i] \leq \alpha \mathbb{E}[\sum_{i=1}^{t} \mathbf{1}\{P_i \leq \alpha_i\}].$$

And rearranging gives

$$\frac{\mathbb{E}[\sum_{i\leq t, i \text{ null}} \mathbf{1}\{P_i \leq \alpha_i\}]}{\mathbb{E}[\sum_{i\leq t} \mathbf{1}\{P_i \leq \alpha_i\}]} \leq \alpha\ .$$

2 Permutation Testing

So far, in our hypothesis testing framework we have assumed that we have access to the distribution of the observations under the null. Sometimes, however, we do not have the null distribution in a well-specified closed form. It turns out that when there are far more nulls than non-nulls, we can actually get around this assumption by estimating the null distribution directly from the data.

More specifically, we do so by designing **permutation tests**. An assumption we will need for this construction is *exchangeability*.

An ordered sequence of random variables $(X_1, \ldots, X_n)$ is called *exchangeable* if its permuted sequence $(X_{\pi(1)}, \ldots, X_{\pi(n)})$ has the same distribution as $(X_1, \ldots, X_n)$, for any permutation $\pi(\cdot)$. Notice that this is trivially satisfied for independent and identically distributed samples X_i.

Suppose we have access to a set of samples which is exchangeable under the null hypothesis, and denote by $\widetilde{X}$ the "unordered" set corresponding to this data set (i.e. all of its permutations). Suppose that for all $n!$ permutations we get a test statistic, i.e. we have $T_1, \ldots, T_{n!}$. Then, because all permutations have equal probability, all test statistics have equal probability as well. If the test statistic T^* that we computed on the original, ordered data set is in the top α-quantile of the test statistics $T_1, \ldots, T_{n!}$, we reject the null because such a test statistic belongs to "unlikely" test statistics. This immediately gives us the guarantee that a false discovery happens with probability of at most α, because we reject the upper α quantile in the distribution of computed test statistics.

The discussion so far has assumed that we are conditioning on $\widetilde{X}$, in a Bayesian way. However, the unconditional probability of a false discovery is also controlled under α by the tower property:

$$\mathbb{E}[\mathbf{1}\{\text{false discovery}\}] = \mathbb{E}[\mathbb{E}[\mathbf{1}\{\text{false discovery}\}|\widetilde{X}]] = \mathbb{E}[\mathbb{P}(\text{false discovery}|\widetilde{X})] \leq \mathbb{E}[\alpha] = \alpha.$$