STAT 689: Large-scale and high-dimensional statistical inference **Fall 2024**

Lecture 8

1 Empirical process viewpoint

Recall that the BH procedure rejects all H_i with $p_i \leq T$, where T is defined as

$$T = \sup\left\{0 < t \leq 1 \colon \frac{nt}{1 \vee R(t)} \leq \alpha\right\}$$

with $R(t) = \sum_{i=1}^n \mathbf{1}\{p_i \leq t\}$ being the number of rejections given the threshold t. As argued in Lecture 6, nt can be viewed as a conservative estimate for the number of false discoveries given the threshold t.

This formulation has a simple interpretation. Let $t \in (0,1)$ be fixed and consider rejecting H_i if and only if $p_i \leq t$. We can construct the rejection/acceptance table for the hypotheses whose values depend on t, where $U(t), V(t), T(t), S(t)$ and $R(t)$ are all stochastic processes indexed by $t \in [0,1]$.

	H_0 accepted	H_0 rejected	Total
H_0 true	$U(t)$	$V(t)$	n_0
H_0 false	$T(t)$	$S(t)$	$n - n_0$
	$n - R(t)$	$R(t)$	n

Table 1: Potential outcomes for testing multiple hypotheses based on the rejection rule of the form $p_i \leq t$ for $t \in [0,1]$.

Define

$$\mathrm{FDP}(t) = \frac{V(t)}{1 \vee R(t)}, \quad \mathrm{FDR}(t) = \mathbb{E}[\mathrm{FDP}(t)].$$

In general, we let $\widehat{\mathrm{FDR}}(t)$ be an estimate for $\mathrm{FDR}(t)$. Then, an FDR-controlling procedure can be described as finding

$$T = \sup\{0 \leq t \leq 1 : \widehat{\mathrm{FDR}}(t) \leq \alpha\}$$

and rejecting all H_i with $p_i \leq T$.

Recall that

- BH procedure: $\widehat{\mathrm{FDR}}(t) = nt/\{1 \vee R(t)\}$.
- BC procedure: $\widehat{\mathrm{FDR}}(t) = \sum_{i=1}^n \mathbf{1}\{1 - p_i \leq t\}/\{1 \vee R(t)\}$.

1.1 FDR control based on martingale theory

We now focus on the BH procedure, where $\widehat{\mathrm{FDP}}(t) = nt/\{1 \vee R(t)\}$. We give an alternate proof of the FDR control result for the BH procedure using martingale theory

Define the filtration $\mathcal{F}_t = \sigma(V(s), R(s) : t \leq s \leq 1)$. This is a backward filtration as $\mathcal{F}_s \subset \mathcal{F}_t$ for $t < s$. We define the reverse time martingale

$$\frac{V(t)}{t}, \quad 0 \leq t \leq 1$$

adaptive to $\mathcal{F}_t$. Note that for $s \le t$, conditional on $\mathcal{F}_t$, $V(s) = \#\{p_i : p_i \le s, H_i \text{ is true}\}$ follows $\text{Bin}(V(t), s/t)$. Therefore,

$$\mathbb{E}\left[\frac{V(s)}{s}\middle|\mathcal{F}_t\right] = \frac{1}{s}\mathbb{E}[V(s)|\mathcal{F}_t] = \frac{1}{s}\frac{s}{t}V(t) = \frac{V(t)}{t}.$$

This shows that $\{V(t)/t : 0 \le t \le 1\}$ is a reverse time martingale.

Next, we note that $T = \sup\{0 \le t \le 1 : nt/\{1 \vee R(t)\} \le \alpha\}$ is a stopping time as $\{T \le t\} \in \mathcal{F}_t$. Applying the optional stopping theorem, we get

$$\begin{aligned}
\text{FDR} =& \mathbb{E}\left[\frac{V(T)}{1 \vee R(T)}\right] \\
=& \frac{1}{n}\mathbb{E}\left[\frac{V(T)}{T}\frac{nT}{1 \vee R(T)}\right] \\
\le& \frac{\alpha}{n}\mathbb{E}\left[\frac{V(T)}{T}\right] \\
=& \frac{\alpha}{n}\mathbb{E}\left[V(1)\right] \\
=& \frac{\alpha n_0}{n}.
\end{aligned}$$

Background: A forward filtration is an increasing sequence $\{\mathcal{F}_t : 0 \le t \le 1\}$ of σ-algebras:

$$\mathcal{F}_s \subseteq \mathcal{F}_t$$

for $s \le t$. A forward martingale with respect to $\{\mathcal{F}_t : 0 \le t \le 1\}$ is a stochastic process $X(t)$ such that

1. $X(t)$ is measurable with respect to $\mathcal{F}_t$;
2. For $s \le t$,
$$\mathbb{E}[X(t)|\mathcal{F}_s] = X(s).$$

A random variable T is called a stopping time if

$$\{T \le t\} \in \mathcal{F}_t$$

for all t.

Optional stopping theorem. For a forward martingale, we have

$$\mathbb{E}[X(T)] = \mathbb{E}[X(0)].$$

A reverse filtration is a decreasing sequence $\{\mathcal{F}_t : 0 \le t \le 1\}$ of σ-algebras:

$$\mathcal{F}_s \supseteq \mathcal{F}_t$$

for $s \le t$. A reverse martingale with respect to $\{\mathcal{F}_t : 0 \le t \le 1\}$ is a stochastic process $X(t)$ such that

1. $X(t)$ is measurable with respect to $\mathcal{F}_t$;
2. For $s \le t$,
$$\mathbb{E}[X(s)|\mathcal{F}_t] = X(t).$$

If T is a stopping time, we have

$$\mathbb{E}[X(T)] = \mathbb{E}[X(1)].$$

2 Storey's procedure

Storey's procedure improves the BH procedure by using the p-values to estimate the null proportion $\pi_0 := n_0/n$. Specifically, we define

$$\pi_0^\lambda = \frac{1 + n - R(\lambda)}{(1-\lambda)n},$$

where $\lambda \in [0, 1)$ is fixed.

We briefly explain the intuition behind the construction of π_0^λ. Note that for large n_0,

$$1 + n - R(\lambda) = 1 + \sum_{i=1}^{n} \mathbf{1}\{p_i > \lambda\} \geq 1 + \sum_{i \in \mathcal{N}_0} \mathbf{1}\{p_i > \lambda\} \approx n_0(1-\lambda)$$

and thus

$$\pi_0^\lambda \geq \frac{1 + \sum_{i \in \mathcal{N}_0} \mathbf{1}\{p_i > \lambda\}}{(1-\lambda)n} \approx \frac{(1-\lambda)n_0}{(1-\lambda)n} = \pi_0.$$

Storey's procedure rejects all H_i with $p_i \leq T$, where T is defined as

$$T = \sup\left\{0 < t \leq \lambda \colon \frac{n\pi_0^\lambda t}{1 \vee R(t)} \leq \alpha\right\}.$$

When $\pi_0^\lambda < 1$, Storey's procedure makes more rejections than the BH procedure because Storey's procedure has a less conservative estimate for the FDR.

2.1 FDR control theory

We now show that Storey's procedure also provides FDR control. As before, we can show that $V(t)/t = \sum_{i \in \mathcal{N}_0} \mathbf{1}\{p_i \leq t\}/t$ for $0 < t \leq \lambda$ is a martingale with time running backward with respect to the filtration $\mathcal{F}_t$ and T is a stopping time with respect to $\mathcal{F}_t$. Thus

$$\begin{aligned} \text{FDR} =& \mathbb{E}\left[\frac{V(T)}{1 \vee R(T)}\right] \\ =& \frac{1}{n}\mathbb{E}\left[\frac{V(T)}{\pi_0^\lambda T}\frac{n\pi_0^\lambda T}{1 \vee R(T)}\right] \\ \leq& \frac{\alpha}{n}\mathbb{E}\left[\frac{V(T)}{\pi_0^\lambda T}\right]. \end{aligned}$$

By the optional stopping theorem, we have

$$\mathbb{E}\left[\frac{V(T)}{\pi_0^\lambda T}\right] = \mathbb{E}\left[\frac{n(1-\lambda)}{1+n-R(\lambda)}\frac{V(\lambda)}{\lambda}\right].$$

Since

$$1 + n - R(\lambda) = 1 + (n_0 - V(\lambda)) + \{(n - n_0) - (R(\lambda) - V(\lambda))\} \geq 1 + n_0 - V(\lambda),$$

we have

$$\mathbb{E}\left[\frac{V(T)}{\pi_0^\lambda T}\right] \leq \frac{n(1-\lambda)}{\lambda}\mathbb{E}\left[\frac{V(\lambda)}{1 + n_0 - V(\lambda)}\right].$$

Because the p-values follow the uniform distribution on $[0, 1]$ under the null, we have $V(\lambda) \sim \text{Bin}(n_0, \lambda)$, which implies

$$
\begin{aligned}
\mathbb{E}\left[\frac{V(\lambda)}{1+n_0-V(\lambda)}\right] &= \sum_{i=1}^{n_0} P\big(V(\lambda)=i\big)\frac{i}{1+n_0-i} \\
&= \sum_{i=1}^{n_0} \binom{n_0}{i}\lambda^i(1-\lambda)^{n_0-i}\frac{i}{1+n_0-i} \\
&= \sum_{i=1}^{n_0} \lambda^i(1-\lambda)^{n_0-i}\frac{n_0!\times i}{(1+n_0-i)\times(n_0-i)!\times i!} \\
&= \sum_{i=1}^{n_0} \lambda^i(1-\lambda)^{n_0-i}\frac{n_0!}{(1+n_0-i)!(i-1)!} \\
&= \sum_{j=0}^{n_0-1} \lambda^{j+1}(1-\lambda)^{n_0-j-1}\frac{n_0!}{(n_0-j)!j!} \\
&= \frac{\lambda}{1-\lambda}\sum_{j=0}^{n_0-1} \lambda^{j}(1-\lambda)^{n_0-j}\frac{n_0!}{(n_0-j)!j!} \\
&= \frac{\lambda}{1-\lambda}\big((\lambda+1-\lambda)^{n_0}-\lambda^{n_0}\big) \\
&= \frac{\lambda(1-\lambda^{n_0})}{1-\lambda}.
\end{aligned}
$$

Hence,

$$
\mathrm{FDR} \le \frac{\alpha}{n}\mathbb{E}\left[\frac{V(T)}{\pi_0^\lambda T}\right] \le \alpha(1-\lambda^{n_0}) \le \alpha.
$$

3 Bayesian viewpoint

The Bayesian perspective provides an insightful way to approach the multiple-testing problem. It allows us to derive the optimal testing procedure from the Bayesian viewpoint.

3.1 Two-group mixture models

For the ith hypothesis, we let θ_i be its underlying truth. More specifically, $\theta_i = 1$ if H_i is non-null/alternative and $\theta_i = 0$ if H_i is null. From the Bayesian perspective, we can view θ_i as a sequence of i.i.d random variables generated from $\mathrm{Bern}(1-\pi_0)$, where π_0 is the probability of a randomly selected hypothesis being null. When $\theta_i = 0$, it is common to assume that p_i follows the uniform distribution over $[0,1]$. Under the alternative (i.e., $\theta_i = 1$), we will assume that p_i is drawn from a distribution concentrated around zero. Therefore, conditional on θ_i,

$$
p_i|\theta_i \sim (1-\theta_i)f_0 + \theta_i f_1
$$

and marginally, the p-values follow the mixture model:

$$
p_i \sim \pi_0 f_0 + (1-\pi_0)f_1,
$$

where f_0 and f_1 denote the p-value densities under the null and alternative, respectively.

3.2 Local false discovery rate

The goal of multiple testing is to separate the alternative cases ($\theta_i = 1$) from the null cases ($\theta_i = 0$). To motivate an optimal procedure, we consider the so-called marginal FDR (mFDR):

$$
\mathrm{mFDR} = \frac{\mathbb{E}[V]}{1 \vee \mathbb{E}[R]}.
$$

Suppose that p_i follows the mixture model and we intend to reject the ith null hypothesis if $p_i \le t_i$. The mFDR is defined as

$$\text{mFDR}(\mathbf{t}) = \frac{\mathbb{E}\left[\sum_{i=1}^n (1-\theta_i)\mathbf{1}\{p_i \le t_i\}\right]}{\mathbb{E}\left[\sum_{i=1}^n \mathbf{1}\{p_i \le t_i\}\right]}, \quad \mathbf{t} = (t_1, \dots, t_n).$$

Under the two-group mixture model, the mFDR can be simplified as

$$\text{mFDR}(\mathbf{t}) = \frac{\sum_{i=1}^n \pi_0 t_i}{\sum_{i=1}^n \{\pi_0 t_i + (1-\pi_0)F_1(t_i)\}},$$

where F_1 is the CDF associated with f_1. Consider the problem

$$\max_{\mathbf{t}} \sum_{i=1}^n (1-\pi_0)F_1(t_i) \quad \text{subject to} \quad \text{mFDR}(\mathbf{t}) \le \alpha,$$

where $\sum_{i=1}^n (1-\pi_0)F_1(t_i) = \mathbb{E}[\sum_{i=1}^n \theta_i \mathbf{1}\{p_i \le t\}]$ is the expected number of true discoveries.

Define the local false discovery rate as

$$\text{LFDR}(t) = \frac{\pi_0 f_0(t)}{\pi_0 f_0(t) + (1-\pi_0)f_1(t)},$$

where $f_0(t) = 1$ is the density of Unif$[0,1]$.

Theorem. The optimal solution $\mathbf{t}^* = (t_1^*, t_2^*, \dots, t_n^*)$ to the above constraint optimization problem satisfies that $\text{LFDR}(t_i^*) = c$ for all $1 \le i \le n$.

Proof. Note that we can rewrite the constraint as

$$\sum_{i=1}^n \pi_0 t_i - \alpha \sum_{i=1}^n \{\pi_0 t_i + (1-\pi_0)F_1(t_i)\} \le 0.$$

The Lagrangian function associated with the optimization problem is given by

$$\mathcal{L}(\mathbf{t}, \lambda) = \sum_{i=1}^n (1-\pi_0)F_1(t_i) - \lambda \sum_{i=1}^n \pi_0 t_i + \lambda\alpha \sum_{i=1}^n \{\pi_0 t_i + (1-\pi_0)F_1(t_i)\}$$

where $\lambda \ge 0$ is called the Lagrange multiplier. Taking the derivative with respect to t_i in $\mathcal{L}$ and setting it to be zero, we have

$$(1-\pi_0)f_1(t_i) - \lambda\pi_0 + \lambda\alpha\{\pi_0 + (1-\pi_0)f_1(t_i)\} = 0,$$

which leads to

$$\text{LFDR}(t_i^*) = \frac{\pi_0}{\pi_0 + (1-\pi_0)f_1(t_i^*)} = \frac{1+\lambda\alpha}{1+\lambda}.$$

Assumption. Assume that $f_1(t)$ is strictly decreasing.

Under the above assumption, $\text{LFDR}(t)$ is strictly increasing. Then $p_i \le t_i^*$ is equivalent to

$$\text{LFDR}(p_i) \le \text{LFDR}(t_i^*) = c.$$

Thus, our goal becomes finding c such that the FDR is controlled while making a large number of true rejections.

We consider the procedure by finding

$$C = \sup\{0 \le c \le 1 : \widehat{\text{FDR}}(c) \le \alpha\}$$

and reject all H_i with $\text{LFDR}(p_i) \le C$. Given the rejection rule $\text{LFDR}(p_i) \le c$, we can define

$$V(c) = \sum_{i=1}^{n} (1 - \theta_i)\mathbf{1}\{\text{LFDR}(p_i) \le c\}.$$

Exercise 8.1: Show that

$$\mathbb{E}[V(c)] = n\mathbb{E}[\text{LFDR}(p)\mathbf{1}\{\text{LFDR}(p) \le c\}].$$

Therefore, we can estimate the FDR by

$$\widehat{\text{FDR}}(c) = \frac{\sum_{i=1}^{n} \text{LFDR}(p_i)\mathbf{1}\{\text{LFDR}(p_i) \le c\}}{\sum_{i=1}^{n} \mathbf{1}\{\text{LFDR}(p_i) \le c\}}$$

Let $\text{LFDR}_i = \text{LFDR}(p_i)$ and $\text{LFDR}_{(1)} \le \text{LFDR}_{(2)} \le \cdots \le \text{LFDR}_{(n)}$. In this case, show that the procedure is equivalent to the step-up procedure by finding

$$j^* = \max\left\{1 \le j \le n : j^{-1} \sum_{i=1}^{j} \text{LFDR}_{(i)} \le \alpha\right\}$$

and reject $H_{(1)}, \ldots, H_{(j^*)}$, where $H_{(i)}$ is the hypothesis associated with $\text{LFDR}_{(i)}$.

In reality, π_0 and f_1 are unknown and have to be estimated from the data. One approach is to find the estimates by maximizing the likelihood function, i.e.,

$$(\hat{\pi}_0, \hat{f}_1) = \text{argmax}_{\pi_0, f_1} \sum_{i=1}^{n} \log(\pi_0 + (1 - \pi_0)f_1(p_i)).$$

The optimization can be solved by the expectation-maximization algorithm.

4 Positive false discovery rate and q-values

The positive false discovery rate (pFDR) is defined as

$$\text{pFDR} = \mathbb{E}\left[\frac{V}{R}\Big| R > 0\right].$$

Note that

$$\text{FDR} = \text{pFDR} \times P(R > 0).$$

The term "positive" has been added to reflect the fact that we are conditioning on the event that positive findings have occurred. When the FDR is controlled at level α, the pFDR is controlled at level $\alpha/P(R > 0)$.

Let $X_1, \ldots, X_n$ be n statistics for testing $H_1, \ldots, H_n$. Consider the two-group mixture model:

$$X_i|\theta_i \sim (1 - \theta_i)f_0 + \theta_i f_1,$$

and $\theta_i \sim \text{Bern}(1 - \pi_0)$. Denote the critical region (the values of X_i for which H_i is rejected) as Γ. Then both V and R can be viewed as functions of Γ. Given the rejection region Γ, define

$$\text{pFDR}(\Gamma) = \mathbb{E}\left[\frac{V(\Gamma)}{R(\Gamma)}\Big| R(\Gamma) > 0\right].$$

Theorem. We have

$$\text{pFDR}(\Gamma) = \frac{\pi_0 P(X \in \Gamma)}{\pi_0 P(X \in \Gamma|\theta = 0) + (1 - \pi_0)P(X \in \Gamma|\theta = 1)} = P(\theta = 0|X \in \Gamma),$$

where $X|\theta \sim (1-\theta)f_0 + \theta f_1$ and $\theta \sim \text{Bern}(1-\pi_0)$.

Proof. Note that

$$\begin{aligned}
\text{pFDR}(\Gamma) =& \mathbb{E}\left[\frac{V(\Gamma)}{R(\Gamma)}\Big| R(\Gamma) > 0\right] \\
=& \sum_{k=1}^{n} \mathbb{E}\left[\frac{V(\Gamma)}{R(\Gamma)}\Big| R(\Gamma) = k\right] P(R(\Gamma) = k | R(\Gamma) > 0) \\
=& \sum_{k=1}^{n} \mathbb{E}\left[\frac{V(\Gamma)}{k}\Big| R(\Gamma) = k\right] P(R(\Gamma) = k | R(\Gamma) > 0)
\end{aligned}$$

Because of the i.i.d. assumption, it follows that

$$\begin{aligned}
\mathbb{E}\left[V(\Gamma)\Big| R(\Gamma) = k\right] =& \sum_{i=1}^{n} \mathbb{E}\left[\mathbf{1}\{X_i \in \Gamma, \theta_i = 0\}\Big| R(\Gamma) = k\right] \\
=& \sum_{i=1}^{n} \mathbb{E}\left[\mathbf{1}\{X_i \in \Gamma, \theta_i = 0\}\Big| X_1, \ldots, X_k \in \Gamma, X_{k+1}, \ldots, X_n \notin \Gamma\right] \\
=& \sum_{i=1}^{k} \mathbb{E}\left[\mathbf{1}\{\theta_i = 0\}\Big| X_1, \ldots, X_k \in \Gamma, X_{k+1}, \ldots, X_n \notin \Gamma\right] \\
=& \sum_{i=1}^{k} \mathbb{E}\left[\mathbf{1}\{\theta_i = 0\}\Big| X_i \in \Gamma\right] \\
=& kP(\theta = 0 | X \in \Gamma).
\end{aligned}$$

Thus

$$\begin{aligned}
\text{pFDR}(\Gamma) =& \sum_{k=1}^{n} \mathbb{E}\left[\frac{V(\Gamma)}{k}\Big| R(\Gamma) = k\right] P(R(\Gamma) = k | R(\Gamma) > 0) \\
=& P(\theta | X \in \Gamma) \sum_{k=1}^{n} P(R(\Gamma) = k | R(\Gamma) > 0) = P(\theta | X \in \Gamma).
\end{aligned}$$

Note that $\mathbb{E}[V(\Gamma)] = n\pi_0 P(X \in \Gamma)$ and $\mathbb{E}[R(\Gamma)] = nP(X \in \Gamma)$. As a corollary of the above theorem, we have

$$\text{pFDR}(\Gamma) = \frac{\mathbb{E}[V(\Gamma)]}{\mathbb{E}[R(\Gamma)]}.$$

Now, let us consider a nested set of significance regions without loss of generality by $\{\Gamma_\alpha : 0 < \alpha < 1\}$, where α is such that

$$P(X \in \Gamma_\alpha | \theta = 0) = \alpha.$$

Note that $\Gamma_{\alpha'} \subseteq \Gamma_\alpha$ for $\alpha' \leq \alpha$. Using this notation, the p-value of an observed statistic $X = x$ is defined to be

$$\text{p-value}(x) = \inf_{\Gamma_\alpha : x \in \Gamma_\alpha} P(X \in \Gamma_\alpha | \theta = 0).$$

As a special case, consider $\Gamma_\alpha = \{u : u \geq c_\alpha\}$. Then we have

$$\inf_{\Gamma_\alpha : x \in \Gamma_\alpha} P(X \in \Gamma_\alpha | \theta = 0) = \inf_{c_\alpha : x \geq c_\alpha} P(X \geq c_\alpha | \theta = 0) = P(X \geq x | \theta = 0),$$

which coincides with the usual definition of p-values.

Definition. The q-value of an observed statistic $X = x$ is defined as

$$\text{q-value}(x) = \inf_{\Gamma_\alpha : x \in \Gamma_\alpha} \text{pFDR}(\Gamma_\alpha) = \inf_{\Gamma_\alpha : x \in \Gamma_\alpha} P(\theta = 0 | X \in \Gamma_\alpha).$$

In other words, the above definition says the q-value is a measure of the strength of an observed statistic with respect to the pFDR; it is the minimum pFDR that can occur when rejecting a statistic with value x for the set of nested significance regions.