Conditional Calibration For False Discovery Rate Control Under Dependence[1]

William Fithian and Lihua Lei

Shubha Sankar Banerjee

Department of Statistics
University of Pittsburgh

December 19, 2023

[1]Main Reference: Fithian and Lei (2022)

Contents

The Problem Setup

- In a multiple testing problem, we observe a data set $X \sim P$ and reject a subset of hypotheses $H_1, \ldots, H_m$.
- Assuming $P \in \mathscr{P}$, each hypothesis $H_i \subsetneq \mathscr{P}$ represents a submodel; WLOG, the i^{th} alternative hypothesis is $\mathscr{P} \backslash H_i$.
- We assume that the computed p-value $p_i(X)$ to test each H_i is marginally super-uniform under H_i.
- Let $\mathscr{H}_0(P) = \{i : P \in \mathscr{H}_i\}$ denote the set of true null hypotheses and define $m_0 = |\mathscr{H}_0|$.
- *Multiple testing procedure* is a decision $\mathscr{R}(X) \subseteq [m]$ designating the set of rejected hypotheses.
- An analyst who rejects H_i for each $i \in \mathscr{R}(X)$ makes $V = |\mathscr{R} \cap \mathscr{H}_0|$ false rejections (discoveries).

FDP and FDR

- Benjamini and Hochberg (1995) define *False Discovery Proportion (FDP)* as

$$FDP(\mathscr{R}(X); P) = \frac{V}{R \vee 1}$$

- The *False Discovery Rate (FDR)* is defined as expected FDP:

$$FDR_P(\mathscr{R}) = \mathbb{E}_P[FDP(\mathscr{R}; X)]$$

- A standard goal in multiple testing is to maximize the expected number of rejections while controlling the FDR at a preset significance level α.

Benjamini-Hochberg (BH) procedure

- The most widely used method for FDR control is BH procedure, which is an example of more general class *step-up procedure*.
- For $p_{(1)} \leq \cdots \leq p_{(m)}$, the step-up procedure for increasing sequence of thresholds $0 \leq \Delta(1) \leq \cdots \leq \Delta(m) \leq 1$ finds the largest index r for which $p_{(r)} \leq \Delta(r)$ and rejects all $H_i \ni 1 \leq i \leq r$.
- Basically, we reject the hypotheses with the smallest $R(X)$ p-values where $R(X) = \max\{r : p_{(r)}(X) \leq \Delta(r)\}$.
- The BH procedure takes $\Delta_\alpha(r) = \alpha r/m$.
- For general family of thresholds $\Delta_\alpha(r)$ that are non-decreasing in α and r, we denote the generic step-up procedure as $SU_\Delta(\alpha)$.
- We denote the corresponding testing procedures as $\mathscr{R}^{BH(\alpha)}$ and $\mathscr{R}^{SU_\Delta(\alpha)}$.

FDR control issues I

- Note:

$$FDR = \mathbb{E}\left[\frac{V}{R \vee 1}\right] = \sum_{i \in \mathscr{H}_0} \mathbb{E}\left[\frac{V_i}{R \vee 1}\right]$$

  where $V_i = 1\{H_i \text{ is rejected}\}$.
- Benjamini and Hochberg (1995) proved that $BH(\alpha)$ procedure controls FDR at exactly $\alpha m_0/m$ if the p-values are independent and each term of the above sum is controlled at α/m.
- Benjamini and Yekutieli (2001) showed that $BH(\alpha)$ procedure controls FDR conservatively at $\alpha m_0/m$ provided $p_{-1} = (p_1, \ldots, p_{i-1}, p_{i+1}, \ldots, p_m)$ is *Positive Regression Dependent (PRD)* on p_i, for every $i \in \mathscr{H}_0$: this condition being called *Positive Regression Dependence on a Subset (PRDS)*.

FDR control issues II

- Benjamini and Yekutieli (2001) also showed that under arbitrary dependence structure of p-values, a much more conservative $BH(\alpha/L_m)$ controls FDR at level α, where $L_m = \sum_{i=1}^{m} i^{-1} = \log m + \mathscr{O}(1)$. This method is called the Benjamini-Yekutieli (BY) procedure.
- It has also been shown that a general *SU* procedure with $\Delta_\alpha(r) = \alpha\beta(r)/m$ where $\beta(r) = \sum_{i=1}^{r} i\nu(i)$ with ν being a probability measure on $\{1, \ldots, m\}$, controls FDR conservatively under arbitrary dependence between p-values.
- These methods control FDR under worse-case dependence assumptions, but their generality comes at a price of substantial conservatism and diminished power compared to BH procedure.

Conditional Calibration I

- This paper introduces a method to adaptively calibrate separate rejection threshold for each p-value to control each term of FDR sum, which we call *FDR contribution of H_i*.
- Let $\tau_i(c; X)$ be rejection threshold for p_i, with *calibration paramater $c \geq 0$*.
- We will aim to calibrate the threshold for p_i, choosing $\hat{c}_i$ to directly control the i^{th} term of the sum

$$\mathbb{E}_{H_i}\left[\frac{V_i}{R \vee 1}\right] = \sup_{P \in H_i} \mathbb{E}_P\left[\frac{1\{p_i \leq \tau_i(\hat{c}_i)\}}{R \vee 1}\right] \leq \frac{\alpha}{m}$$

- We try to control a more tractable conditional expectation to free the calculation from nuisance parameters, given some *conditioning statistic S_i*.
- Only requirement of S_i:

$$\sup_{P \in H_i} \mathbb{P}_P(p_i \leq \alpha | S_i) \leq \alpha \ a.s. \ \forall \alpha \in [0, 1]$$

Conditional Calibration II

- Under independence the super-uniformity condition holds for $S_i = p_{-i}$.

Lemma

Let $p^{(i \leftarrow 0)} = (p_1, \ldots, p_{i-1}, 0, p_{i+1}, \ldots, p_m)$. If $\mathscr{R}$ is a step-up procedure with threshold sequence $\Delta(1), \ldots, \Delta(m)$, then the following are equivalent:

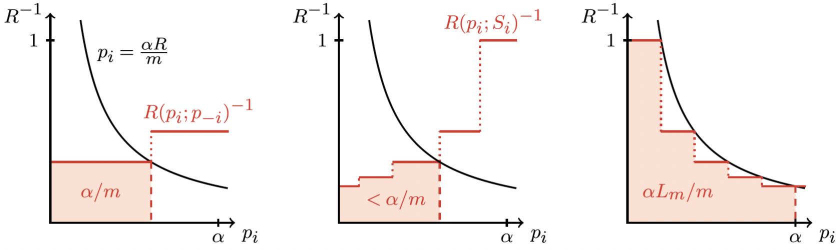
1. $p_i \leq \Delta(R(p^{(i \leftarrow 0)}))$,
2. $i \in \mathscr{R}(p)$,
3. $\mathscr{R}(p) = \mathscr{R}(p^{(i \leftarrow 0)})$.

- Let $R^0 = R(p^{(i \leftarrow 0)})$, which only depends on p_{-i}. Then for standard *BH* procedure under independence,

$$\mathbb{E}_{H_i}\left[\frac{V_i}{R \vee 1} | p_{-i}\right] = \mathbb{E}_{H_i}\left[\frac{1\{p_i \leq \alpha R^0 / m\}}{R^0} | p_{-i}\right] \leq \frac{1}{R^0} \frac{\alpha R^0}{m} = \frac{\alpha}{m}$$

Marginalizing over p_{-i} and summing over $i \in \mathscr{H}_0$ yields $FDR \leq \alpha m_0 / m$.

Conditional Calibration III



(a) Independent case: R is constant on the event where p_i is rejected, $\{p_i \leq \alpha R/m\}$, as a result of Lemma 1.

(b) Positive-dependent case: R is decreasing in p_i, since the other p-values are increasing in p_i.

(c) Worst case: the other p-values are adversarially configured so that R is always just large enough to reject p_i.

- The number of rejections R also depend on all $\hat{c}_1, \ldots, \hat{c}_m$. We substitute an estimatior $\hat{R}_i \geq 1$ for eventual value of $|\mathscr{R}(X) \cup \{i\}|$.
- Note that $\hat{R}_i$ should be lower bound for R, so that $V_i/\hat{R}_i$ is a good upper bound for $V_i/(R \vee 1)$.

Step 1: Calibration

- We use $\hat{R}_i$ to estimate $\mathbb{E}_{H_i}[V_i/(R \vee 1)|S_i]$ as a function of the calibration parameter c, which we call to be *valid* if:

$$g_i^*(c; S_i) = \sup_{P \in H_i} \mathbb{E}_P \left[\frac{1\{p_i \leq \tau_i(c)\}}{\hat{R}_i} | S_i \right] \leq \frac{\alpha}{m}$$

- We choose $c_i^*(S_i) = \sup\{c \geq 0; g_i^*(c; S_i) \leq \alpha/m\}$.
- $g_i^*(c; S_i)$ is non-decreasing and any $c < c_i^*$ is valid. However, if g_i^* is discontinuous, c_i^* itself may not be valid.
- We consider a sequence $\{\hat{c}_{i,t}^*\}_t \ni \hat{c}_{i,t} \uparrow \hat{c}_i^*$ and say the $\hat{c}_i^*$ is *maximal* if $\cup_t[0, \hat{c}_{i,t}]$ includes every valid c, almost surely.

Step 2: Initial rejection

- We now initialize the rejection set via:

$$\mathscr{R}_+ = \{i : p_i \leq \tau_i(\hat{c}_i)\}$$

- In practice instead of calculating $\hat{c}_i$, we use $q - values$ $q_i(X) = \min\{c : p_i \leq \tau_i(c)\}$ since they are easy to calculate for SU procedures.
- For the maximal $\hat{c}_i$, $i \in \mathscr{R}_+$ iff the observed q_i is a valid calibration parameter, so we alternatively write,

$$\mathscr{R}_+ = \{i : g_i^*(q_i; S_i) \leq \alpha/m\}$$

- Let $R_+ = |\mathscr{R}_+|$. If $R_+ \geq \hat{R}_i$, $\forall\, i \in \mathscr{R}_+$, we set $\mathscr{R} = \mathscr{R}_+$. Otherwise, we prune the rejection set further.

Step 3: Randomized Pruning

- If there is some $i \in \mathscr{R}_+$ for which $\hat{R}_i > R_+$, then we must prune the rejection set via a secondary BH procedure.
- For use generated random variables $u_1, \ldots, u_m \overset{i.i.d.}{\sim} Unif(0,1)$, let,

$$R(X;u) = \max\{r : |\{i \in \mathscr{R}_+ : u_i \leq r/\hat{R}_i\}| \geq r\}$$

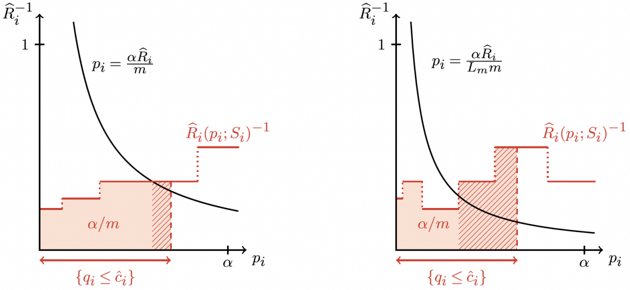
  and reject H_i for the R indices with $i \in \mathscr{R}_+$ and $u_i \leq R/\hat{R}_i$.
- This procedure is equivalent to $BH(1)$ procedure on “p-value” $\tilde{p}_i = u_i \hat{R}_i / R_+$ for $i \in \mathscr{R}_+$.
- While this pruning step introduces extra randomness, the rejection set includes $\{i : \hat{R}_i \leq R(X;\mathbf{1}_n)\}$ almost surely.
- We call a calibrated procedure *safe* is pruning is never necessary.

Theorem

Assuming conditional superuniformity of p-values and that $\hat{c}_i$ is chosen to guarantee term-wise FDR control $\forall i$, the three step procedure controls FDR at or below level $\alpha m_0/m$.

The dependence-adjusted BH and BY procedures

- If we use the effective threshold $\tau_i = \tau^{BH}$ and the estimator $\hat{R}_i = |\mathscr{R}^{BH(\gamma\alpha)} \cup \{i\}|$, we call the method *dependence-adjusted BH procedure* and denote it using $dBH(\gamma\alpha)$.
- If $\gamma = 1/L_m$, we denote it by $dBY(\alpha)$



(a) The $\mathrm{dBH}_1(\alpha)$ procedure in the positive-dependent case.

(b) The $\mathrm{dBY}(\alpha)$ procedure, which is often much more powerful than $\mathrm{BY}(\alpha)$.

$dBH_1(\alpha)$ and $dBY(\alpha)$ procedure

- **Definition:** For given conditioning statistic S_i, we say p_{-i} is *conditionally positive regression dependent (CRPD)* if $\mathbb{P}(p_{-i} \in A | p_i, S_i)$ is a.s. non-decreasing in p_i for any increasing set A. If p_{-i} is CPRD on p_i $\forall i \in \mathscr{H}_0$, we say p-values are CPRD on subset (CPRDS).

Theorem

Assume $\hat{c}_1, \ldots, \hat{c}_m$ are maximal, then:

1. *If the p-values are independent with p_i uniform under H_i, then $dBH_1(\alpha)$ procedure with $S_i = p_{-i}$ is identical to $BH(\alpha)$ procedure.*
2. *If the p-values are CPRDS $\forall\ P \in \mathscr{P}$, then $BH_1(\alpha)$ procedure is safe and uniformly more powerful than the $BH(\alpha)$ procedure.*
3. *For arbitrary dependence, the $dBY(\alpha)$ procedure is safe, and uniformly more powerful than the $BY(\alpha)$ procedure.*
4. *Recall the thresholds Δ_α for generic SU procedure, then for arbitrary dependence, the $dSU_\Delta(\alpha)$ procedure is safe, and uniformly more powerful than the $SU_\Delta(\alpha)$ procedure.*

Identifying conditioning statistic S_i

- To facilitate calibration, S_i should eliminate or mitigate the influence of nuisance parameters on the conditional distribution of X.
- Calibration is conceptually simplified if S_i is a sufficient statistic for the null submodel H_i, so that conditional distribution of X is known under H_i, in which case we call H_i to be *conditionally simple*. Otherwise, we call it *conditionally composite* and P_i^* is *least favorable* for calibrating $\hat{c}_i$ if it almost surely attains the supremum $g_i^*(c_i; S_i)$.
- **Example:** For full ranked exponential family,

$$X \sim f_\theta(x) = e^{\theta' T(x) - A(\theta)} f_0(x), \ \theta \in \Theta \subset \mathbb{R}^d$$

  For $i = 1, \dots, m \leq d$, H_i takes form $H_i : \theta_i = 0$ or $H_i : \theta \leq 0$. The UMPU test rejects H_i when $T_i(X)$ is extreme, conditional on the value of $S_i = T_{-i}$.
- For one sided testing, $H_i : \theta_i \leq 0$ is conditionally composite and it turns our that under some mild condition, $\theta_i = 0$ is least favorable.

Recursive refinement of $\hat{R}_i$

Performing Steps 1 and 2 our method will make us estimate the rejection rejections once for all the m p-values and hence changing $(\hat{R}_1, \dots, \hat{R}_m)$ will affect the entire procedure and change $\mathscr{R}$ in turn. We however use a better procedure called *recursive refinement* of the estimator.

- Denote the original estimator as $\hat{R}_i^{(1)}$, which leads to original calibration parameter, $\hat{c}_i^{(1)}$ and initial rejection set $\mathscr{R}_+^{(1)}$.
- We define the recursively refined estimator as:

$$\hat{R}_i^{(k+1)} = |\mathscr{R}_+^{(k)}(x) \cup \{i\}|,\ k > 1$$

If we use effective *BH* threshold with the $dBH_\gamma(\alpha)$ estimator, we call the resulting procedure $dBH_\gamma^2(\alpha)$ procedure, or $dBY^2(\alpha)$ if $\gamma = 1/L_m$.

Theorem

Assume $\hat{c}_1^{(k)}, \dots, \hat{c}_m^{(k)}$ are maximal $\forall i,\ k$. If $\mathscr{R}^{(1)}$ is safe, then for every $k \geq 1$, $\mathscr{R}^{(k+1)}$ is safe and uniformly more powerful than $\mathscr{R}^{(k)}$.

Multivariate z-statistics

Assume $Z \sim N_d(\mu, \Sigma)$ with $\Sigma \succ 0$ and all $\Sigma_{i,i} = 1$ and we wish to test $H_i : \mu_i = 0$ or $H_i : \mu \leq 0$ for $i = 1, \ldots, m \leq d$. p_i are the standard one- or two-sided p-values based on Z_i. Now,

$$f_\mu(z) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left\{\mu'\Sigma^{-1}z - \frac{1}{2}z'\Sigma^{-1}z - \frac{1}{2}\mu'\Sigma^{-1}\mu\right\}$$

It can be shown that taking $S_i = z_{-i} - \Sigma_{-i,i}\Sigma_{i,i}^{-1}z_i$, eliminates the influence of μ_{-i} from the problem, leaving a one-parameter exponential family model in μ_i with Z_i as sufficient statistic.
To carry out $dBH_\gamma(\alpha)$ procedure, we plug in $c = q_i(Z)$ in

$$\mathbb{E}_0\left[\frac{1\{q_i \leq c\}}{|\mathscr{R}^{BH(\gamma\alpha)} \cup \{i\}|} S_i\right] \leq \frac{\alpha}{m}$$

The expectation is now easy to calculate since given S_i, we simply integrate with respect to $Z_i \sim N(0,1)$.

Multivariate t-statistics

In the previous problem, we now assume that $\Sigma = \sigma^2\Psi$ where $\Psi \succ 0$ is known but $\sigma^2 > 0$ is unknown. WLOG, assume $\Psi_{i,i} = 1$. To estimate σ^2, we observe another independent vector $W \sim N_{n-d}(0, \sigma^2 I_{n-d})$.
As before, we test $H_i : \mu = 0$ or $H_i : \mu_i \leq 0$ for $i = 1, \ldots, m \leq d$, the usual test statistic is

$$T_i = \frac{Z_i}{\hat{\sigma}} \text{ where } (n-d)\hat{\sigma}^2 = ||W||^2 \sim \sigma^2 \chi^2_{n-d}$$

It can be shown that conditioning the resulting $(d+1)$-parameter exponential family on $S_i = \left(z_{-i} - \Psi_{-i,i}\Psi_{i,i}^{-1} z_i, \sqrt{||W||^2 + Z_i^2}\right)$ yields a one-parameter exponential family with parameter μ_i/σ^2.

Therefore, solving the expectation to control the contribution to FDR boils down to a simple integration w.r.t. $T_i \sim t_{n-d}$.

Reference

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300.

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29(4):1165 – 1188.

Fithian, W. and Lei, L. (2022). Conditional calibration for false discovery rate control under dependence. *The Annals of Statistics*, 50(6):3091 – 3118.