Bayesian FDR

I) FDR and pFDR: Frequentist definitions

Definition 1:

$(\Omega, \mathcal{F}, (P_\vartheta)_{\vartheta \in \Theta}, \mathcal{H}_m)$ multiple test problem,

$\mathcal{H}_m = (H_i : 1 \leq i \leq m)$; $\varphi = (\varphi_i : 1 \leq i \leq m)$ multiple test

(a) $I_0 \equiv I_0(\vartheta) = \{1 \leq i \leq m : \vartheta \in H_i\}$, $|I_0| = m_0$

$I_1 \equiv I_1(\vartheta) = \{1, \ldots, m\} \setminus I_0$

$= \{1 \leq i \leq m : \vartheta \in K_i = \Theta \setminus H_i\}$,

$|I_1| = m_1$.

(b) $V_m \equiv V_m(\vartheta) = |\{i \in I_0(\vartheta) : \varphi_i = 1\}|$,

$R_m \equiv R_m(\vartheta) = |\{1 \leq i \leq m : \varphi_i = 1\}|$.

(c) $FWER_\vartheta(\varphi) = P_\vartheta(V_m > 0)$,

$FDP_\vartheta(\varphi) = \dfrac{V_m}{\max(R_m, 1)}$, (false discovery proportion)

$$\mathrm{FDR}_\vartheta(\varphi) = \mathbb{E}_\vartheta[\mathrm{FDP}_\vartheta(\varphi)] = \mathbb{E}_\vartheta\left[\frac{V_m}{\max(R_m, 1)}\right],$$

$$\mathrm{pFDR}_\vartheta(\varphi) = \mathbb{E}_\vartheta\left[\frac{V_m}{R_m} \mid R_m > 0\right].$$

<u>(positive) false discovery rate</u>

<u>Theorem 1:</u>

(a) $\mathrm{FDR}_\vartheta(\varphi) = \mathrm{pFDR}_\vartheta(\varphi) \; \mathbb{P}_\vartheta(R_m > 0)$.

(b) If $m_0(\vartheta) = m$, then $\mathrm{FDR}_\vartheta(\varphi) = \mathrm{FWER}_\vartheta(\varphi)$.

(c) $\forall \vartheta \in \Theta$: $\mathrm{FDR}_\vartheta(\varphi) \leq \mathrm{FWER}_\vartheta(\varphi)$.

<u>Proof:</u> $\mathrm{FDR}_\vartheta(\varphi) = \mathbb{E}_\vartheta\left[\frac{V_m}{R_m \vee 1}\right]$

$$= \mathbb{E}_\vartheta\left[\frac{V_m}{R_m \vee 1} \mid R_m > 0\right] \mathbb{P}_\vartheta(R_m > 0)$$

$$+ \mathbb{E}_\vartheta\left[\frac{V_m}{R_m \vee 1} \mid R_m = 0\right] \mathbb{P}_\vartheta(R_m = 0)$$

$$= \mathrm{pFDR}_\vartheta(\varphi) \; \mathbb{P}_\vartheta(R_m > 0) + 0,$$ hence (a).

If $m_0(\vartheta) = m \Rightarrow V_m = R_m$

$$\Rightarrow \mathrm{pFDR}_\vartheta(\varphi) \equiv 1 \Rightarrow$$

$\mathrm{FDR}_\vartheta(\varphi) = \mathbb{P}_\vartheta(R_m > 0) = \mathbb{P}_\vartheta(V_m > 0) = \mathrm{FWER}_\vartheta(\varphi)$, which is (b).

In general, $\mathrm{FDP}_\vartheta(\varphi) \leq \mathbb{1}_{\{V_m > 0\}}$

$$\Rightarrow \mathbb{E}_\vartheta[\mathrm{FDP}_\vartheta(\varphi)] \leq \mathbb{E}_\vartheta[\mathbb{1}_{\{V_m > 0\}}],$$

hence (c). $\blacksquare$

<u>Remark:</u> As one can see from the proof of part (b), pFDR cannot be controlled from a frequentist point of view.

II) A Bayesian mixture model

<u>Definition 2 (Two-class mixture model):</u>

$(T_1, H_1), \ldots, (T_m, H_m)$ iid sequence of test statistics and indicator variables

$H_i = 0 \;:\Leftrightarrow$ i-th null hypothesis true

$H_i = 1 \;:\Leftrightarrow$ i-th null hypothesis false

$T_j \mid H_j \sim (1-H_j) F_0 + H_j F_1$,

F_0, F_1 : cdfs of absolutely continuous distributions with densities f_0, f_1 respectively.

$$\pi_0 := \mathbb{P}(H_1 = 0), \quad \pi_1 := 1 - \pi_0.$$

$\Rightarrow (H_i : 1 \le i \le m)$ are iid. from Bernoulli (π_1).

$\mathbb{P}$ and $\mathbb{E}$ refer to this model throughout the remainder!

Remark: In Baysian statistics, uncertainty about parameter values before data are ascertained is expressed by probability distributions. Here, $H_1, \dots, H_m$ are the parameters.

Theorem 2:

Under the model from Definition 2, let $\varphi = (\varphi_i, 1 \le i \le m)$ be such that

$$\varphi_i(x) = 1 \iff T_i(x) \in \Gamma \equiv \Gamma_\alpha, \quad x \in \Omega.$$

Then it holds

$$pFDR(\Gamma) := \mathbb{E}\left[\frac{V(\Gamma)}{R(\Gamma)} \,\middle|\, R(\Gamma) > 0\right]$$

$$= \mathbb{P}(H=0 \mid T \in \Gamma),$$

where $T \overset{d}{=} T_1$, $H \overset{d}{=} H_1$,

$$V(\Gamma) = \{1 \le i \le m : T_i \in \Gamma, H_i = 0\},$$

$$R(\Gamma) = \{1 \le i \le m : T_i \in \Gamma\}.$$

Proof:

$$pFDR(\Gamma) = \sum_{k=1}^{m} \mathbb{E}\left[\frac{V(\Gamma)}{R(\Gamma)} \,\middle|\, R(\Gamma) = k\right] \cdot \mathbb{P}(R(\Gamma) = k \mid R(\Gamma) > 0)$$

$$= \sum_{k=1}^{m} \mathbb{E}\left[\frac{V(\Gamma)}{k} \,\middle|\, R(\Gamma) = k\right] \mathbb{P}(R(\Gamma) = k \mid R(\Gamma) > 0)$$

$$= (*)$$

Observe that $\mathcal{L}(V(\Gamma) \mid R(\Gamma) = k) = \mathrm{Bin}(k, p)$,

where $p = \mathbb{P}(H=0 \mid T \in \Gamma)$.

$$\Rightarrow (*) = \sum_{k=1}^{m} \mathbb{P}(H=0 \mid T \in \Gamma)\, \mathbb{P}(R(\Gamma) = k \mid R(\Gamma) > 0)$$

$$= \mathbb{P}(H=0 \mid T \in \Gamma) \cdot 1$$

$\blacksquare$

III) Local false discovery rate

Definition 3:

Under the assumption of theorem 2, consider

- $\Gamma = (-\infty, t]$. (small values are in favor of alternatives!)

Let $\mathrm{Fdr}(t) := p\mathrm{FDR}(\Gamma) = \pi_0 \dfrac{F_0(t)}{F(t)}$

$= \mathbb{P}(H=0 \mid T \leq t)$, where

$F(t) = \pi_0 F_0(t) + \pi_1 F_1(t)$ is the mixture cdf.

$$\mathrm{fdr}(t) = \pi_0 \frac{f_0(t)}{f(t)} = \mathbb{P}(H=0 \mid T=t)$$

is called local false discovery rate

$$\mathrm{Fdr}(\Gamma) = \frac{\pi_0 \, \mathrm{Prob}_{f_0}(T \in \Gamma)}{\mathrm{Prob}_f(T \in \Gamma)}$$ is called <u>Bayesian FDR</u>

of the (general) rejection region Γ.

<u>Remark:</u>

(a) $f(t) = \frac{d}{dt} F(t) = \pi_0 f_0(t) + \pi_1 f_1(t)$.

(b) $\mathrm{fdr}(t) = \mathbb{P}(H = 0 \mid T = t)$

$$= \pi_0 f_0(t) \Big/ \sum_{i=0}^{1} \pi_i f_i(t) = \frac{\pi_0 f_0(t)}{f(t)}.$$

is also Bayesian posterior probability.

(c) $$\mathrm{Fdr}(\Gamma) = \frac{\pi_0 \, \mathbb{P}(T \in \Gamma \mid H = 0)}{\mathbb{P}(T \in \Gamma)}$$

$$= \frac{\mathbb{P}(H = 0, T \in \Gamma)}{\mathbb{P}(T \in \Gamma)} = \mathbb{P}(H = 0 \mid T \in \Gamma)$$

is a Bayesian posterior probability, too.

(d) Nonparametric estimates for $\mathrm{Fdr}(t)$ and $\mathrm{fdr}(t_i)$ are given by

$$\widehat{\mathrm{Fdr}}(t) = \frac{\pi_0 F_0(t)}{\hat{F}_m(t)}, \quad \hat{F}_m: \text{ empirical cdf of } T_1, \dots, T_m;$$

$$\widehat{\mathrm{fdr}}(t_i) = \frac{\pi_0 f_0(t_i)}{\hat{f}(t_i)}, \quad t_i = T_i(x),$$

$\hat{f}$: some density estimator.

Theorem 3 (Averaging Theorem):

$$\mathrm{Fdr}(\Gamma) = E_f[\,\mathrm{fdr}(T) \mid T \in \Gamma\,].$$

Proof: $E_f[\,\mathrm{fdr}(T) \mid T \in \Gamma\,]$

$$= \int_\Gamma \frac{\pi_0 f_0(t)}{f(t)} f(t)\, dt \Big/ \int_\Gamma f(t)\, dt$$

$$= \pi_0 \int_\Gamma f_0(t)\, dt \,/\, \int_\Gamma f(t)\, dt$$

$$= \frac{\pi_0 \,\mathrm{Prob}_{f_0}(T \in \Gamma)}{\mathrm{Prob}_f(T \in \Gamma)} = \mathrm{Fdr}(\Gamma). \quad \blacksquare$$

Corollary: Based on Theorem 3, a Bayesian FDR-controlling multiple test can be constructed as follows ($\alpha \in (0,1)$ given):

(1) Calculate $\widehat{\mathrm{fdr}}(t_i)$ for all $1 \le i \le m$.

(2) Denote by $t_{1:m} \le t_{2:m} \le \ldots \le t_{m:m}$ the ordered values of the t_i.

(3) Define the threshold

$$t^* = \max_{1 \le j \le m} \{ t_{j:m} : j^{-1} \sum_{i=1}^{j} \widehat{fdr}(t_{i:m}) \le \alpha \}.$$

(4) Declare t_j significant if $t_j \le t^*$.

IV) Connection to binary classification

Consider the classification loss function

$$L_\Gamma(H_i, T_i) = \lambda_1 \, \mathbb{1}\{H_i = 0, T_i \in \Gamma\} + \lambda_2 \, \mathbb{1}\{H_i = 1, T_i \notin \Gamma\},$$

$\lambda_1, \lambda_2 > 0$ given cost parameters.

This leads to the classification risk (expected loss):

$$\mathbb{E}[L_\Gamma(H_i, T_i)]$$

$$= \mathbb{E}[\lambda_1 \, \mathbb{1}\{H_i = 0, T_i \in \Gamma\} + \lambda_2 \, \mathbb{1}\{H_i = 1, T_i \notin \Gamma\}]$$

$$= \mathbb{E}_{T_i \sim f}\big[\mathbb{E}_{H_i | T_i}[\lambda_1 \, \mathbb{1}\{T_i \in \Gamma\} \, \mathbb{1}\{H_i = 0\}] + \mathbb{E}_{H_i | T_i}[\lambda_2 \, \mathbb{1}\{T_i \notin \Gamma\} \, \mathbb{1}\{H_i = 1\}] \big]$$

$$= \mathbb{E}_{T_i \sim f} \Big[\lambda_1 \, \mathbb{1}\{T_i \in \Gamma\} \, \mathbb{P}(H_i = 0 \mid T_i)$$

$$+ \lambda_2 \, \mathbb{1}\{T_i \notin \Gamma\} \, \mathbb{P}(H_i = 1 \mid T_i) \Big].$$

It follows that the Bayes-optimal rejection region is such that for every given realization $T_i = t_i$ the expression

$$\lambda_1 \, \mathbb{1}\{t_i \in \Gamma\} \, \pi_0 f_0(t_i) + \lambda_2 \, \mathbb{1}\{t_i \notin \Gamma\} \, \pi_1 f_1(t_i)$$

is minimal

● $\Rightarrow$ Decision in favor of "1" - class if

$$\lambda_1 \pi_0 f_0(t_i) \leq \lambda_2 \pi_1 f_1(t_i)$$

$$\Leftrightarrow \frac{\pi_0 f_0(t_i)}{\pi_1 f_1(t_i)} \leq \frac{\lambda_2}{\lambda_1}$$

$$\Leftrightarrow \frac{\mathrm{fdr}(t_i)}{1 - \mathrm{fdr}(t_i)} \leq \frac{\lambda_2}{\lambda_1}$$

$$\Leftrightarrow \mathrm{fdr}(t_i) \leq \frac{\lambda_2}{\lambda_1 + \lambda_2}.$$

References:

Dickhaus, T. (2014). Simultaneous Statistical Inference with Applications in the Life Sciences. Springer-Verlag. Chapters 1 and 6.

Efron, B. and Tibshirani, R. (2002). Empirical Bayes Methods and False Discovery Rates for Microarrays. Genetic Epidemiology 23, 70-86.

Storey, J. D. (2003). The positive false discovery rate: A Bayesian interpretation and the q-value. Ann. Statist. 31(6): 2013-2035.