The Bayesian False Discovery Rate

$H_1 \dots H_n$ - hypotheses

$p_1 \dots p_n$ - p-values $\Rightarrow$ assume that iid.

$$\gamma_i = \mathbb{1}_{\{H_{0i} \text{ is false}\}} = \begin{cases} 1, & H_{0i} \text{ is false} \\ 0, & H_{0i} \text{ is true} \end{cases}$$

Cdf of p_i:

$$F_p(t) = P(p_i \le t) \overset{t \in [0,1]}{=} \begin{cases} t, & \text{if } \gamma_i = 0 \\ F_A(t), & \text{if } \gamma_i = 1 \end{cases}$$

Law of Total Prob. $\Rightarrow$ $F_p(t) = F_p(t/\gamma_i = 0) P(\gamma_i = 0) + F_p(t/\gamma_i = 1) P(\gamma_i = 1) =$

$= t \cdot P(\gamma_i = 0) + F_A(t) P(\gamma_i = 1)$

$\Rightarrow$ cdf of p_i can be considered as a mixture of two distributions: $P(\gamma_i = 0)$ and $P(\gamma_i = 1)$

We can think that we generate γ_i randomly from the binomial distribution:

$$P(\gamma_i = 1) = \varepsilon = 1 - P(\gamma_i = 0)$$

In the situation when n is large we can often analyze the data assuming this model (see the Sparse mixture model)

$\Rightarrow$ $F_p(t) = t(1-\varepsilon) + F_A(t)\varepsilon$

Let t_0 be a critical value:

H_{0i} is rejected if $p_i \le t_0$

BFDR $= P(H_{0i} \text{ is true} / H_{0i} \text{ is rejected})$

Let us compare BFDR with FDR:

$$BFDR = \frac{P(H_{0i} \text{ is true}, H_{0i} \text{ is rejected})}{P(H_{0i} \text{ is rejected})}$$

$$FDR = E\frac{V}{R \vee 1} = \left| \begin{array}{l} V = \#\{H_{0i} \text{ is rejected, } H_{0i} \text{ is true}\} \\ R = \#\{H_{0i} \text{ is rejected}\} \end{array} \right| =$$

$$\overset{\text{for large } n + LLN}{=} E\frac{V/n}{(R \vee 1)/n} \approx \frac{P(H_{0i} \text{ is rejected, } H_{0i} \text{ is true})}{P(H_{0i} \text{ is rejected})} = BFDR$$

$$BFDR = \frac{P(H_{0i} \text{ is rejected} / H_{0i} \text{ is true})\, P(H_{0i})}{P(H_{0i} \text{ is rejected})} =$$

$$= \frac{P(p_i \le t_0 / H_{0i})(1-\varepsilon)}{P(p_i \le t_0)} = \frac{t_0(1-\varepsilon)}{F_p(t_0)}$$

$$\tau_{BFDR} = \max\{t_0 : BFDR \le q\} = \max\left\{t : \frac{t(1-\varepsilon)}{F_p(t)} \le q\right\}$$

– we want to control BFDR on level q

let us compare τ_{BFDR} with τ_{BH}

$$\tau_{BFDR} = \max\left\{t : \frac{t(1-\varepsilon)}{F_p(t)} \le q\right\} \qquad 1-\varepsilon \to 1$$

$$\tau_{BH} = \max\left\{t : \frac{t}{\hat{F}_n(t) \vee \frac{1}{n}} \le q\right\} \qquad F_p(t) \to \hat{F}_n(t)$$

if we could calculate/estimate ε, $F_p(t) \Rightarrow$

$\Rightarrow$ we'll get a little bit better procedure

Let us assume than n is large enough s.t.:

$$F_p(t) \approx \hat{F}_n(t)$$

$$\Rightarrow BFDR \approx (1-\varepsilon) FDR_{BH} = (1-\varepsilon) q \approx \frac{n_0}{n} q.$$

<u>Bayesian classifier</u>

X - data ; $X \in \Omega$

$H_0 : X \sim P_0$ vs $H_A : X \sim P_A$

P_0, P_A are defined (see, Neyman-Pearson lemma)

let $\Omega = \Gamma_0 \cup \Gamma_A$, where

Γ_0 is acceptance region : $X \in \Gamma_0 \Leftrightarrow H_0$ is accepted

Γ_A is rejection region : $X \in \Gamma_A \Leftrightarrow H_0$ is rejected

Let $\delta(x) = \begin{cases} 1, & \text{if } X \in \Gamma_A \\ 0, & \text{if } X \in \Gamma_0 \end{cases}$ be a test function.

Cost function

C:

	H_0 is accepted	H_0 is rejected
H_0 is true	0	C_0
H_A is true	C_A	0

"we have no loss when we make correct decision and we pay some cost in the case of wrong decision (C_0 for Type I Error and C_A for Type II Error).

$$E(C/H_0) = 0 \cdot P(X \in \Gamma_0/H_0) + C_0 P(X \in \Gamma_A/H_0) = C_0 \underbrace{P(X \in \Gamma_A/H_0)}_{\text{Type I Er}}$$

$$E(C/H_A) = C_A P(X \in \Gamma_0/H_A) + 0 \cdot P(X \in \Gamma_A/H_A) = C_A \underbrace{P(X \in \Gamma_0/H_A)}_{\text{Type II Er}}$$

$$\Rightarrow E(C) = E(C/H_0) \cdot P(H_0) + E(C/H_A) \cdot P(H_A) =$$
$$= C_0 P(X \in \Gamma_A/H_0) P(H_0) + C_A P(X \in \Gamma_0/H_A) \cdot P(H_A) \ominus$$

let us assume that $P(X \in \cdot/H_0)$ and $P(X \in \cdot/H_A)$ have densities $f(X/H_0)$ and $f(X/H_A)$ respectively. ($\Leftrightarrow P_0, P_A$ have densities).

$$\ominus\ C_0 P(H_0) \int_{\Gamma_A} f(X/H_0)\,dx + C_A P(H_A) \int_{\Gamma_0} f(x/H_A)\,dx =$$

$$= /\Gamma_0 = \Omega \setminus \Gamma_A/ =$$

$$= C_0 P(H_0) \int_{\Gamma_A} f(x/H_0)\,dx + C_A P(H_A)\left(1 - \int_{\Gamma_A} f(x/H_A)\,dx\right) =$$

$$= C_A P(H_A) + \int_{\Gamma_A} \left[C_0 P(H_0) f(X/H_0) - C_A P(H_A) f(X/H_A)\right] dx$$

we would like to minimize $E(C) \Leftrightarrow$

we would like to pick Γ_A so as $E(C)$ is minimal.

$$E(C) = c_A P(H_A) + \int_{\Gamma_A} g(x)\,dx \to \min$$

$$\Longleftrightarrow \Gamma_A := \{x : g(x) \le 0\}$$

$$\Rightarrow \Gamma_A = \{x : c_0 P(H_0) \cdot f(x/H_0) - c_A P(H_A) f(x/H_A) \le 0\}$$

$$\Rightarrow \Gamma_A = \{x : \frac{c_0 P(H_0) f(X/H_0)}{c_A P(H_A) f(X/H_A)} \le 1\}$$

$$\Rightarrow \Gamma_A = \{x : \frac{f(X/H_A)}{f(X/H_0)} \ge \frac{c_0 P(H_0)}{c_A P(H_A)}\}$$

Let us compare with Neyman-Pearson rule:

Γ $\frac{f(X/H_A)}{f(X/H_0)}$ is likelihood ratio but we specify the probability of Type I Error

Γ_A : $P(\Gamma_A/H_0) = \alpha$ where α is given

for Bayesian classifier: critical value depends on $P(H_0), P(H_1)$ and on the cost functions

If c_0 or $P(H_0)$ are large $\Rightarrow$ likelihood ratio should be large.

If $c_0 = c_A = 1 \Longleftrightarrow$ c = Probability of making wrong decision

this rule minimizes misclassification error (choosing H_0 when H_1 is true and otherwise)

(this is close to naive Bayessian classifier).

Example Sparse Mixture Model

$H_0: X_i \sim N(0,1)$ vs $H: X_i \sim (1-\varepsilon) N(0,1) + \varepsilon N(\mu, 1)$

cost:

	H_0 is accepted	H_0 is rejected
H_0 is true	0	c_0
H_A is true	c_A	0

$P(H_0) = 1-\varepsilon$; $\quad P(H_A) = \varepsilon$

$t_{1i} = P(H_{0i}$ is rejected $/ H_{0i}$ is true$) = P(\text{Type I Er})$ in H_i

$t_{2i} = P(H_{0i}$ is accepted $/ H_{Ai}$ is true$) = P(\text{Type II Er})$ in H_i

$$\Rightarrow E(C_i) = c_0 (1-\varepsilon) t_{1i} + c_A \varepsilon t_{2i} \to \min$$

if $\dfrac{f(X_i / H_A)}{f(X_i / H_0)} \geq \dfrac{c_0}{c_A} \dfrac{(1-\varepsilon)}{\varepsilon}$

$$\frac{(1-\varepsilon) e^{-\frac{X_i^2}{2}} + \varepsilon e^{-\frac{(X_i-\mu)^2}{2}}}{e^{-X_i^2/2}} = (1-\varepsilon) + \varepsilon e^{\mu X_i - \frac{\mu^2}{2}} \geq \frac{c_0}{c_A} \frac{1-\varepsilon}{\varepsilon}$$

$$e^{\mu X_i} \geq e^{\frac{\mu^2}{2}} \left(\frac{c_0}{c_A} \frac{1-\varepsilon}{\varepsilon} - (1-\varepsilon) \right) = e^{\frac{\mu^2}{2}} (1-\varepsilon) \left(\frac{c_0}{c_A} \cdot \frac{1}{\varepsilon} - 1 \right)$$

$$X_i \geq \underbrace{\frac{1}{\mu} \left(\frac{\mu^2}{2} + \log(1-\varepsilon) + \log\left(\frac{c_0}{c_A} \frac{1}{\varepsilon} - 1 \right) \right)}_{\tau_{Bcl}}$$

$\Rightarrow$ we reject all H_{0i} : $X_i \geq \tau_{Bcl}$.

$$BFDR = P(H_{0i} \text{ is true} / H_{0i} \text{ is rejected}) = \frac{(1-\varepsilon) t_1}{(1-\varepsilon) t_1 + \varepsilon (1-t_2)}$$

(Storey, 2003) in the case when individual test statistics are generated by the two-component mixture model and the multiple testing procedure uses the same fixed threshold for each of the tests, BFDR coincides with the positive FDR (pFDR)

$$pFDR = E\left(\frac{V}{R} / R > 0 \right) = \frac{FDR}{P(R>0)}$$

threshold for BFDR: $\frac{(1-\varepsilon)(1-F_0(c_{BFDR}))}{1-F(c_{BFDR})} = q$,

where F_0 is the cdf of $N(0,1)$,
F is the cdf of $(1-\varepsilon)N(0,1)+\varepsilon N(\mu,1)$.

<u>Remark</u> $BFDR = \frac{(1-\varepsilon)t_1}{(1-\varepsilon)t_1+\varepsilon(1-t_2)} \leq q$

$$\underbrace{(1-q)(1-\varepsilon)t_1 + q\varepsilon t_2}_{\text{Bayes Risk with } c_0 = 1-q;\ c_A = q} \leq q\varepsilon.$$

BFDR controls Bayes Risk with $c_0 = 1-q$; $c_A = q$ on the level $q\varepsilon$.