同济大學
TONGJI UNIVERSITY
地址：中国上海市四平路1239号　邮编：200092
1239 SIPING ROAD SHANGHAI CHINA 200092
电话（TEL）：+86 21-　传真（FAX）：+86 21-
网址（WEB）：www.tongji.edu.cn

最优FDR控制：基于Compound-Decision Framework（Adaptive z-value procedure）

决策框架下考虑混合决策时，p值方法是不可容许的

1、改进前提 ①基于p值的方法不能做有效的检验，i.e. min $E(FNR)$ s.t. $FDR\le\alpha$。本文基于决策理论的框架提出最优的FDR控制方法（基于z值，min mFNR s.t. $mFDR\le\alpha$）②将该决策问题等价为一分类问题，后者可解，进而转换为决策问题，为 $\delta(Lfdr(\cdot),c)=\{I[Lfdr(x_i)<c]: i=1,\cdots,m\}$，因为 $Lfdr(\cdot)=\frac{(1-p)f_0(\cdot)}{f(\cdot)}$，这和p值方法不同。Lfdr不仅考虑了全局特征p，和局部特征 $\frac{f_0}{f}$，且由Lfdr意味着对"相对重要性"进行排序根据似然比。而对p值排序是要Alternative关于null对称，而决策理论可用在非对称情形（有这种可能：一个距离null很远的观测可能显著性水平很低，在用Lfdr准则下，即该检验可以接受更多极端的假设，拒绝更多不极端的假设，即拒绝域不是对称的）

由此知该方法是最优的

③利用 $mFDR=FDR+O(m^{-\frac{1}{2}})$，现要决定cutoff。因为 $mFNR\downarrow$，$mFDR\uparrow$，故当 $mFDR=\alpha$ 时mFNR取min。又因为mFDR，（$Q(\lambda)$）是阈值λ的增函数，故 $\lambda_{OR}=\sup\{t\in(0,1): Q(t)\le\alpha\}=Q^{-1}_{OR}(\alpha)$

④但是 p, f_0, f 需要估计（Jin, Cai, 2007），由于Efron（2007）指出，FDR为fdr在 $Z\le z$ 的条件期望（积分一阶矩即为方，或简单理解为条件样本均值），估计 $\hat{Q}_{OR}(\hat{Lfdr}_{(k)})=\frac{1}{k}\sum_{i=1}^{k}\hat{Lfdr}_{(i)}$ 但 $\hat{\lambda}_{OR}$ 不好估计，故跳过。找 $k=\max\{i: \hat{Q}_{OR}(\hat{Lfdr}_{(i)})\le\alpha\}$，拒绝 $H_{(1)},\cdots,H_{(k)}$。

2. 假设前提 ① $f>0$ on real line，f连续 ② $T_{OR}(z_i)=\frac{(1-p)f_0(z_i)}{f(z_i)}$ 的边缘pdf为 $g=(1-p)g_0+pg_1$
③ $T_{OR}\in\mathcal{T}$ 满足SMLR假设（plug-in方法要求 G_1 为凹函数，此为该假设的一般化）
④ $\hat{p}\xrightarrow{P}p$，$E\|\hat{f}-f\|^2\to0$，$E\|\hat{f_0}-f\|^2\to0$

NOTE：Plug-in方法不能对z值方法使用，因为不知道 T_{OR} 的分布（需要估计），而z值分布一般为混合正态模型。z值方法（本文）有效的原因是从single rule拓展到compound rule中去考虑。

3. 理论结果

命题一（该命题说明了SMLR假设保证了多重检验以及分类问题的良好性质）
设 $\theta_i\overset{iid}{\sim}B(1,p)$，$i=1,\cdots,m$ 设 ~~x_i为~~ $x_i|\theta_i$，$i=1,\cdots,m$ 为模型 $[X_i|\theta_i\sim(1-\theta_i)F_0+\theta_iF_1]$ 的独立观测。设 $T\in\mathcal{T}$（下降的MLR），则 $\delta(T,c)=\{I[T(X_i)<c], i=1,\cdots,m\}$（无论是多重检验还是分类问题），有如下性质：
(a) $P(\theta_i=1|T(X_i)\le c)$ 为阈值c的单调降函数
(b) mFDR为阈值c和拒绝总数期望r的增函数
(c) mFNR为c和r的减函数
(d) mFNR随着mFDR的上升而下降。
(e) 在分类问题中，c和r随着分类权重λ的上升而下降。

证明：(a) ∵ T 有单调减的MLR. ∴ $\int_{-\infty}^{c} g_0(t)dt \,/ \int_{-\infty}^{c} g_1(t)dt = \int_{-\infty}^{c} g_0(t)dt \,/ \int_{-\infty}^{c} \frac{g_1(t)}{g_0(t)} g_0(t)dt$

$< \int_{-\infty}^{c} g_0(t)dt \,/ \int_{-\infty}^{c} \frac{g_1(c)}{g_0(c)} g_0(t)dt = \frac{g_0(c)}{g_1(c)}$

对$G(x)$对应$T(X)$的cdf. $F(x)$对应X的cdf

再由分布函数定义. 有 $g_1(c)G_0(c) < g_0(c)G_1(c)$

对 $P(\theta_i=1 \mid T(X_i)\le c) = \frac{pG_1(c)}{G(c)}$ 中 c 求导, $\frac{(1-p)p[g_1(c)G_0(c)-g_0(c)G_1(c)]}{G(c)^2} < 0$

故 $P(\theta_i=1 \mid T(X_i)\le c)$ 是 c 的单调减函数

(b)、(c) 类似可证，其中, $mFDR = \frac{p_0 G_0(c)}{G(c)}$, $mFNR = \frac{p_1(1-G_1(c))}{1-G(c)}$

由(b), (c) 即证 (d)

(e). 易得分类风险 $R_\lambda = E(L_\lambda(\theta,\delta)) = \lambda(1-p)G_0(c) + p[1-G_1(c)]$，易求得使风险最小的阈值满足 $\lambda = \frac{pg_1(c^*)}{(1-p)g_0(c^*)}$，由MLR性质，$\lambda$是$c$的减函数. $\gamma = G(c^*)$为c^*的增函数. ∴ λ是γ的减函数.

定理一（多重检验问题与分类问题等价）. $\theta_i, X_i|\theta_i$同前. 令 $\Lambda \in \mathcal{T}$. 设分类损失函数 $L_\lambda(\theta,\delta) = \frac{1}{m}\sum_i \{\lambda I(\theta_i=0)\delta_i + I(\theta_i=1)(1-\delta_i)\}$ 最小化时对应 $\delta^\lambda[\Lambda, c(\lambda)] = \{\delta_1^\lambda, \cdots, \delta_m^\lambda\}$, $\delta_i^\lambda = I\{\Lambda(X_i) < c(\lambda)\}$ 对每任意水平为α的mFDR多重检验，存在唯一$\lambda(\alpha)$，其对应分类问题的最优解 $\delta^{\lambda(\alpha)}[\Lambda, c(\lambda(\alpha))]$ 也是多重检验问题的最优解. 即 $\min\ mFNR \quad s.t.\ mFDR \le \alpha$.

定理二（分类问题的最优解）$\theta_i, X_i|\theta_i$同前. 设 p, f_0, f_1已知. 对所有$\mathcal{D}_s$中的决策(SMLR决策)分类最优解为

$\delta_i = I\{\Lambda(X_i) = \frac{(1-p)f_0(X_i)}{pf_1(X_i)} < \frac{1}{\lambda}\}$. $i=1,\cdots,m$. [显然. i.e. $\lambda(1-p)f_0 < pf_1$, 维损失小, 做维的决定]

检验方法：令 $k = \max\{i, \frac{1}{i}\sum_{j=1}^{i} \widehat{Lfdr}_{(j)} \le \alpha\}$, 拒绝$H^{(i)}$, $i=1,\cdots,k$.
mFDR的估计

定理三（检验方法的渐近有效性）$\theta_i, X_i|\theta_i$同前. pdf为$f=(1-p)f_0+pf_1$. 设 $T_{OR}(z_i) = \frac{(1-p)f_0(z_i)}{f(z_i)}$ 的mpdf为 $g=(1-p)g_0+pg_1$, $T_{OR}\in\mathcal{T}$满足SMLR假设. 设$f\in C^1$. $f>0$. 设$\hat{p}, \hat{f}_0, \hat{f}$分别为$p, f_0, f$的估计，满足$\hat{p} \xrightarrow{P} p$. $E\|\hat{f}-f\| \to 0$. $E\|\hat{f}_0-f_0\| \to 0$. 定义检验统计量 $\hat{T}_{OR}(z_i) = \frac{(1-\hat{p})\hat{f}_0(z_i)}{\hat{f}(z_i)}$, 设 $\widehat{Lfdr}_{(1)}, \cdots, \widehat{Lfdr}_{(m)}$ 为$\hat{T}_{OR}(z_i)$的排序结果，则上述检验方法渐近将mFDR控制在水平α下.

定理四（检验方法的最优性）. 符号同定理三，则检验方法的mFNR水平渐近为 $\tilde{Q}_{OR}(\lambda_{OR})+o(1)$, $\tilde{Q}_{OR}(\lambda_{OR}) = \frac{p\tilde{G}_{OR}^{1}(\lambda_{OR})}{\tilde{G}_{OR}(\lambda_{OR})}$ 为Oracle情形下Z值方法的mFNR水平.

故为证明Thm 3 和Thm 4，需先证明下面几个引理.

同济大学
TONGJI UNIVERSITY
地址：中国上海市四平路1239号 邮编：200092
1239 SIPING ROAD SHANGHAI CHINA 200092
电话（TEL）：+86 21- 传真（FAX）：+86 21-
网址（WEB）：www.tongji.edu.cn

引理一：设 $\hat p, \hat f, \hat f_0$ 满足 $\hat p \xrightarrow{P} p$, $E\|\hat f_0 - f_0\| \to 0$, $E\|\hat f - f\| \to 0$, 则有 $E\|\hat T_{OR} - T_{OR}\|^2 \to 0$.

Pf: $f \in C(-\infty, +\infty)$ 且 $f>0$. $\exists \underbrace{K_1}_{=[-M,M]}$ s.t. $Pr(z \in K_1^c) \to 0$. ~~M~~ $M \to +\infty$. 设 $\inf_{z\in K_1} f(z) = l_0$

$A_\varepsilon^f = \{z : |\hat f - f| \ge \frac{l_0}{2}\}$ $\Rightarrow$ $E\|\hat f - f\|^2 \ge (\frac{l_0}{2})^2 P(A_\varepsilon^f)$, 由题意, $P(A_\varepsilon^f) \to 0$. 设 $\sup_{z\in K_1} f(z) = l_1$

$B_\varepsilon^f = \{z : |\hat f - f| \ge \frac{l_1}{2}\}$ $\Rightarrow$ $P(B_\varepsilon^f) \to 0$. 故对充分大的 M, f, $\hat f$ 有上下界，除了一个测度很小的集合. f_0, $\hat f_0$ 同理. 故 $\exists\, 0 < l_a < l_b < \infty$ 对充分大的 M 和 A_ε (任相联字集), 有.

$0 < l_a < \inf_{z \in A_\varepsilon^c} \min\{f_0, \hat f_0, f, \hat f\} < \sup_{z \in A_\varepsilon^c} \max\{f_0, \hat f_0, f, \hat f\} < l_b < +\infty$

$$\because \hat T_{OR} - T_{OR} = \frac{(1-\hat p)\hat f_0}{\hat f} - \frac{(1-p) f_0}{f} = \frac{\hat f_0 f (p - \hat p) + (1-p) f (\hat f_0 - f) + (1-p) f_0 (f - \hat f)}{f \hat f}$$

在 A_ε^c 上(所有元素有界), 有 $(\hat T_{OR} - T_{OR})^2 \le c_1 (p - \hat p)^2 + c_2 (f_0 - \hat f_0)^2 + c_3 (\hat f - f)^2$

易知 $\hat T_{OR}$, $T_{OR} \in (0,1)$, 有界, 设为 L

$\therefore E\|\hat T_{OR} - T_{OR}\|^2 \le L P(A_\varepsilon) + c_1 E(p - \hat p)^2 + c_2 E\|\hat f - f\|^2 + c_3 E\|\hat f_0 - f\|^2$

$\because \hat p \in (0,1)$ 有界. 易得 $\hat p$ 一致可积. 又 $\because \hat p \xrightarrow{P} p$. 故 $\hat p \xrightarrow{L^2} p$. 代入条件 $\hat f_0 \xrightarrow{L^2} f_0$, $\hat f \xrightarrow{L^2} f$

有 $E\|\hat T_{OR} - T_{OR}\|^2 \to 0$.

NOTE. 已证 $\hat T_{OR} \xrightarrow{L_2} T_{OR}$. 由Marcov不等式有 $\hat T_{OR} \xrightarrow{P} T_{OR}$, 进而 $\hat T_{OR} \xrightarrow{D} T_{OR}$. 由弱收敛定义, 对 $\forall$ 有界连续函数 h, 有 $E h(\hat T_{OR}) \to E h(T_{OR})$, 取 $h(x) = I\{x<t\} x$ (可数间断点函数)
得到 $E\{I(\hat T_{OR}(z_i) < t) \hat T_{OR}(z_i)\} \to E\{I(T_{OR}(z_i) < t) T_{OR}(z_i)\}$

引理二. 设 $\hat G_{OR}(t) = \frac{1}{m} \sum_{i=1}^m I(\hat T_{OR}(z_i) \le t)$, $\hat G_{OR}^0 = \frac{1}{m}\sum_{i=1}^m I(\hat T_{OR}(z_i) \le t) \hat T_{OR}(z_i)$, 定义 $\hat Q_{OR}(t) = \frac{\hat G_{OR}^0(t)}{\hat G_{OR}(t)}$
即 mFDR 估计, 对于 $\alpha < t < 1$, 有 $\hat Q_{OR}(t) \xrightarrow{P} Q_{OR}(t)$

NOTE: 用 $\hat G_{OR}^0(t)$ 估计 $(1-p) G_{OR}^0(t)$ 的原因: $(1-p) G_{OR}^0(t) = (1-p) \int_{T_{OR}(z) < t} f_0(z) dz = \int I(T_{OR}(z) < t) T_{OR}(z) f(z) dz$

由此得到 $\hat Q_{OR}(t) = [\sum_{i=1}^m I(\hat T_{OR}(z_i) \le t) \hat T_{OR}(z_i)] / [\sum_{i=1}^m I(\hat T_{OR}(z_i) \le t)]$. $\hat Q_{OR}(Lfdr(k)) = \frac{1}{k}\sum_{i=1}^k Lfdr(i)$

Pf: 设 $\rho_m = cov(\hat T_{OR}(z_i), \hat T_{OR}(z_j))$, z_i, z_j 相互独立且分布为 F, 由 $\hat T_{OR}(z_i) \xrightarrow{P} T_{OR}(z_i)$ 得.
$E(\hat T_{OR}(z_i) \hat T_{OR}(z_j)) \to E(T_{OR}(z_i) T_{OR}(z_j))$, 即 $\rho_m \to cov(T_{OR}(z_i), T_{OR}(z_j)) = 0$, 设 $\sigma_m^2 = Var(\hat T_{OR}(z_i))$
则由 $\hat T_{OR}(z_i) = \frac{(1-\hat p)\hat f_0}{\hat f} \wedge 1 \le 1$ 得, $\sigma_m^2 \le E(\hat T_{OR}(z_i))^2 \le 1$
设 $u_m = P(\hat T_{OR}(z_i) < t)$, $S_m = \sum_{i=1}^m I(\hat T_{OR}(z_i) \le t)$, 由方差展开式可得
$\frac{Var(S_m)}{m^2} = \frac{1}{m}\sigma_m^2 + \frac{m-1}{m}\rho_m \le \frac{1}{m} + \rho_m \to 0$. 由弱大数定律得 $\frac{1}{m}\sum_{i=1}^m I(\hat T_{OR}(z_i) < t) - u_m \xrightarrow{P} 0$

又 $\because u_m \to P(T_{OR}(z_i) < t) = G_{OR}(t)$. $\therefore \hat G_{OR}(t) \xrightarrow{P} G_{OR}(t)$, 类似证 $\hat G_{OR}^0(t) - v_m \xrightarrow{P} 0$, 其中
$v_m = E(I(\hat T_{OR}(z_i) \le t) \hat T_{OR}(z_i))$, 由上述NOTE有 $E(I(\hat T_{OR}(z_i) \le t) \hat T_{OR}(z_i)) \to E(I(T_{OR}(z_i) \le t) T_{OR}(z_i))$
$\therefore v_m \to E(I(T_{OR}(z_i) < t) T_{OR}(z_i)) = (1-p) \int I(T_{OR}(z_i) < t) f_0 dt = (1-p) G_{OR}^0(t)$.
$\therefore \hat G_{OR}^0(t) \xrightarrow{P} (1-p) G_{OR}^0(t)$. 对于 $t > \alpha$, $G_{OR}(t) \ge \varepsilon > 0$. 故有 $\hat Q_{OR}(t) \xrightarrow{P} \frac{(1-p) G_{OR}^0(t)}{G_{OR}(t)} = Q_{OR}(t)$

依概率收敛的性质

同濟大學
TONGJI UNIVERSITY

地址：中国上海市四平路1239号　邮编：200092
1239 SIPING ROAD SHANGHAI CHINA 200092
电话（TEL）：+86 21-　传真（FAX）：+86 21-
网址（WEB）：www.tongji.edu.cn

引理三．定义 $\hat{\lambda}_{OR}=\sup\{t\in(0,1):\hat{Q}_{OR}(t)\le\alpha\}$．若 $\hat{Q}_{OR}(t)\xrightarrow{P}Q_{OR}(t)$，则 $\hat{\lambda}_{OR}\xrightarrow{P}\lambda_{OR}$

Pf：注意到 $\hat{Q}_{OR}(t)$ 是不连续的，故 $\hat{Q}_{OR}(t)$ 的相合性推不出 $\hat{\lambda}_{OR}$ 的相合性．构造 $\hat{Q}^-_{OR}(t)$，$\hat{Q}^+_{OR}(t)$ 如下，当 $Lfdr_{(k)}<t<Lfdr_{(k+1)}$ 时．

$$\hat{Q}^-_{OR}(t)=\frac{Lfdr_{(k+1)}-t}{Lfdr_{(k+1)}-Lfdr_{(k)}}\hat{Q}_{OR}(Lfdr_{(k-1)})+\frac{t-Lfdr_{(k)}}{Lfdr_{(k+1)}-Lfdr_{(k)}}\hat{Q}_{OR}(Lfdr_{(k)})$$ 滞后＋线性连接

$$\hat{Q}^+_{OR}(t)=\frac{Lfdr_{(k+1)}-t}{Lfdr_{(k+1)}-Lfdr_{(k)}}\hat{Q}_{OR}(Lfdr_{(k)})+\frac{t-Lfdr_{(k)}}{Lfdr_{(k+1)}-Lfdr_{(k)}}\hat{Q}_{OR}(Lfdr_{(k+1)})$$ 线性连接．

代入有 $\hat{Q}_{OR}(Lfdr_{(k+1)})-\hat{Q}_{OR}(Lfdr_{(k)})=\frac{1}{k(k+1)}[kLfdr_{(k+1)}-\sum_{i=1}^{k}Lfdr_{(i)}]>0$

显然有 $\hat{Q}^-_{OR}(t)\le\hat{Q}_{OR}(t)\le\hat{Q}^+_{OR}(t)$，且 $\hat{Q}^-_{OR}(t)$ 和 $\hat{Q}^+_{OR}(t)$ 为 t 的严格增函数，定义 $\hat{\lambda}^-_{OR}=\sup\{t\in(0,1):\hat{Q}^-_{OR}(t)\le\alpha\}$．$\hat{\lambda}^+_{OR}=\sup\{t\in(0,1):\hat{Q}^+_{OR}(t)\le\alpha\}$，则有 $\hat{\lambda}^+_{OR}\le\hat{\lambda}_{OR}\le\hat{\lambda}^-_{OR}$

断言 $\hat{\lambda}^-_{OR}\xrightarrow{P}\lambda_{OR}$．否则 $\exists\,\varepsilon_0,y_0$ s.t. $P(|\hat{\lambda}^-_{OR}-\lambda_{OR}|>\varepsilon_0)>2y_0$．$\forall m\ge M$．则有 $P(\hat{\lambda}^-_{OR}>\lambda_{OR}+\varepsilon_0)>2y_0$

$\hat{Q}^-_{OR}\xrightarrow{P}Q_{OR}$ (~~格里纹科定理~~)．($\because\hat{Q}^-_{OR}(t)-\hat{Q}^+_{OR}(t)\xrightarrow{a.s.}0$．SLLN．$\hat{Q}^-_{OR}(t)\le\hat{Q}_{OR}(t)\le\hat{Q}^+_{OR}(t)$ ~~且 $\hat{Q}^-_{OR}(t)-\hat{Q}^+_{OR}(t)\xrightarrow{a.s.}$~~ $\hat{Q}_{OR}(t)\xrightarrow{P}\hat{Q}_{OR}(t)$)

$\therefore\forall m\ge M$．$P(|\hat{Q}^-(\lambda_{OR}+\varepsilon_0)-Q_{OR}(\lambda_{OR}+\varepsilon_0)|<\delta_0)>1-y_0$，$2\delta_0=Q_{OR}(\lambda_{OR}+\varepsilon_0)-\alpha$

设 $K_{1m}=\{w:|\hat{Q}^-_{OR}(\lambda_{OR}+\varepsilon_0)-Q_{OR}(\lambda_{OR}+\varepsilon_0)|<\delta_0$．$K_{2m}=\{w:\hat{\lambda}^-_{OR}>\lambda_{OR}+\varepsilon_0\}$．

$P(K_{1m})>1-y_0$．$P(K_{2m})>2y_0$．$K_m=K_{1m}\cap K_{2m}$．$P(K_m)\ge1-y_0+2y_0-1=y_0>0$．

在 K_m 中．$\alpha\overset{定义}{=}\hat{Q}^-_{OR}(\hat{\lambda}^-_{OR})\overset{\wedge}{\geq}\hat{Q}^-_{OR}(\lambda_{OR}+\varepsilon_0)\overset{K_{1m}}{>}Q_{OR}(\lambda_{OR}+\varepsilon_0)-\delta_0\overset{\delta_0}{=}\delta_0+\alpha$．矛盾．

同理 $\hat{\lambda}^+_{OR}\xrightarrow{P}\lambda_{OR}$．$\because\hat{Q}^-_{OR}(t)-\hat{Q}_{OR}(t)\xrightarrow{a.s.}0$．$\Rightarrow\hat{\lambda}^-_{OR}-\hat{\lambda}^+_{OR}\xrightarrow{a.s.}0\Rightarrow\hat{\lambda}^-_{OR}-\hat{\lambda}_{OR}\xrightarrow{a.s.}0$．

由 Slutsky 定理．$\hat{\lambda}_{OR}\xrightarrow{P}\lambda_{OR}$．

定理三的证明：微分法相同于用 $\hat{T}_{OR}(z_i)$ 为检验统计量．$\hat{\lambda}_{OR}$ 为阈值．则 $\delta(\hat{T}_{OR}(z_i),\hat{\lambda}_{OR})$ 的

$$mFDR=\frac{E(N_{10})}{E(R)}=\frac{(1-p)P_{H_0}(T_{OR}<\hat{\lambda}_{OR})}{P(\hat{T}_{OR}<\hat{\lambda}_{OR})}\ \#\longrightarrow\frac{(1-p)P_{H_0}(T_{OR}<\lambda_{OR})}{P(T_{OR}<\lambda_{OR})}=\frac{(1-p)\tilde{G}_{OR}(\lambda_{OR})}{Q_{OR}(\lambda_{OR})}=Q_{OR}(\lambda_{OR})=\alpha$$

定理四的证明：同上．

$$mFNR=\frac{E(N_{01})}{E(S)}=\frac{pP_{H_1}(\hat{T}_{OR}>\hat{\lambda}_{OR})}{P(\hat{T}_{OR}>\hat{\lambda}_{OR})}\longrightarrow\frac{pP_{H_1}(T_{OR}>\lambda_{OR})}{P(T_{OR}>\lambda_{OR})}=\frac{p\tilde{G}_{OR}(\lambda_{OR})}{\tilde{G}_{OR}(\lambda_{OR})}=\tilde{Q}_{OR}(\lambda_{OR})$$

NOTE：注要说明分母 $\ge\varepsilon>0$．利用依概率收敛的性质．

同濟大學
TONGJI UNIVERSITY

地址：中国上海市四平路1239号 邮编：200092
1239 SIPING ROAD SHANGHAI CHINA 200092
电话（TEL）：+86 21- 传真（FAX）：+86 21-
网址（WEB）：www.tongji.edu.cn

4.方法步骤.

① Oracle Procedure:

p, f_0, f, G^0_{OR}, G_{OR}需要估计

Step 1: 计算 z 值.

Step 2: 计算 local fdr $T_{OR}(z_i) = \frac{(1-p) f_0(z_i)}{f(z_i)}$, $i=1,\cdots,m$

Step 3: 计算 $\lambda_{OR} = \sup\{t \in (0,1): Q_{OR}(t) \le \alpha\}$, $Q_{OR}(t) = \frac{(1-p) G^0_{OR}(t)}{G_{OR}(t)}$

Step 4: 记 $\mathcal{H} = \{i: I(T_{OR}(z_i) < \lambda_{OR}) = 1\}$. 拒绝 H_i. $i \in \mathcal{H}$.

② Data-driven Procedure:

Step 1: 计算 z 值.

Step 2: 用 Jin, Cai(2007) 方法估计 $\hat{p}, \hat{f}_0, \hat{f}$.

Step 3: 计算 local fdr $\cancel{(z_i)}$ $\hat{T}_{OR}(z_i) = \frac{(1-\hat{p}) \hat{f}_0(z_i)}{\hat{f}(z_i)}$, $i=1,\cdots,m$. 并排序得到 $Lfdr_{(1)}, \cdots, Lfdr_{(m)}$.

Step 4: 计算 $k = \max\{i: \hat{Q}_{OR}(Lfdr_{(i)}) = \frac{1}{i}\sum_{j=1}^{i} Lfdr_{(j)} \le \alpha\}$. 拒绝 $H_{(1)}, \cdots, H_{(k)}$