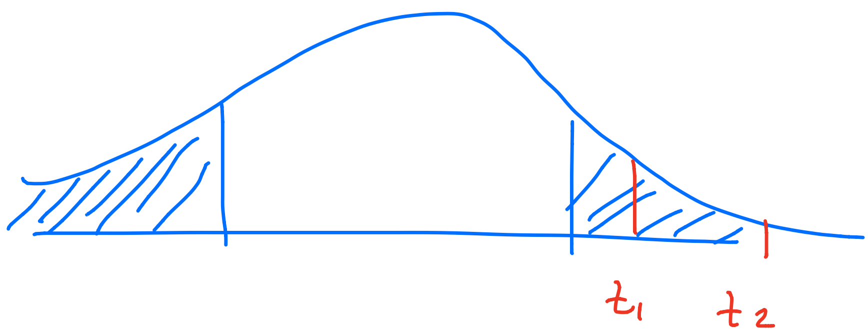
Chp 5: Local False Discovery Rates

FDR_D associated with a decision rule based on the rejection region R_D



Here t_2 stronger than t_1 but the FDR remains the same.

Goal: Define a local FDR

$$fdr(z) = \mathbb{P}(\text{null} \mid z) = \frac{\pi_0 f_0(z)}{f(z)}$$

$$f(z) = \pi_0 f_0 + \pi_1 f_1$$

We then estimate

$$\widehat{\mathrm{fdr}}(z) = \frac{\hat{\pi}_0 f_0(z)}{\hat{f}(z)}$$

$$\widehat{\mathrm{Fdr}}(z) = \frac{\pi_0 F_0(z)}{\hat{F}(z)}$$ ← smoothed estimate

Q: How do we estimate $\hat{f}$?

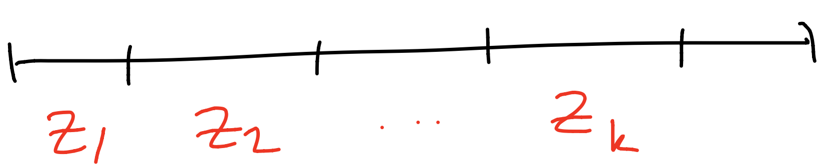
Poisson Regression Model

Suppose our data falls into the exp. family.

$$f(z) = \exp\left\{ \sum_{j=1}^{J} \beta_j z^j \right\}$$

How to estimate $\{\beta_j : 1 \le j \le J\}$

Partion the test statistics



$y_k = \# \, z_i \in Z_k$

x_k = center of Z_k

$r_k = N d \times f(x_k)$ $\quad d$ width of bin

So we want to model the rate in each bin $Y_k \sim \text{Pois}(r_k)$

$$\log(r_k) = \sum_{j=0} \beta_j \cdot x_k^j$$

Issue: independent trials for the Poisson? Probably not.

General objection: what happens if the true model

$$f_i(z_i) = \pi_{0i} f_{0i}(z_i) + \pi_{1i} f_{1i}(z_i)$$

So if were to aggregate

$$\bar{f}(z) = \frac{1}{N}\sum \pi_{0i} f_{0i}(z) + \frac{1}{N}\sum \pi_{1i} f_{1i}(z)$$

through a simple renaming we can recover the original model!

<u>Rmk:</u> $\lim_{N \to \infty} \text{FWER} = \text{bad}$

$$\lim_{N \to \infty} \widehat{fdr} = fdr$$

Combining Fdr and fdr

$$\widehat{Fdr} = \frac{\int_{-\infty}^{z} \widehat{fdr}(x)\bar{f}(x)\,dx}{\int_{-\infty}^{z} \hat{\bar{f}}(x)\,dx}$$

Let

$$\alpha_i = \int_{c_i}^{\infty} f_{i0}(z_i)\,dz_i \quad \text{(size)}$$

$$\beta_i = \int_{c_i}^{\infty} f_{i1}(z_i)\,dz_i \quad \text{(power)}$$

- Expected # of false positives

$$\sum_{i=1}^{N} w_i \pi_{i0} \alpha_i$$

- Expected # of true positives

$$\sum_{i=1}^{N} w_i \pi_{i1} \beta_i$$

What is c_i s.t.

$$\max_{c_1,\dots,c_n} ETP \text{ w.r.t. } EFP \leq \text{threshold}$$

$$L = \sum_{i=1}^{N} w_i \pi_{i0} \int_{c_i}^{\infty} f_{i1}(z_i)\, dz_i$$

$$- \lambda \sum_{i=1}^{N} w_i \pi_{i1} \int_{c_i}^{\infty} f_{i1}(z_i)$$

$$\frac{\partial L}{\partial c_i} = 0 \iff \pi_{i1} f_{i1}(z) = \lambda \pi_{i0} f_{i0}(c_i)$$

$$\text{fdr}(z_i) = \frac{\pi_0 f_{i0}(z_i)}{\pi_0 f_{i0}(z_i) + \pi_1 f_{i1}(z_i)}$$

$$= \frac{1}{1+\lambda}$$

Hence the best decision rule would reject each null at the same threshold of its local fdr.

Large Scale Inference

Type I error $\longrightarrow$ FDR, fdr

Power $\longrightarrow$ True discovery rate.

P_1(Decision Rule)

$$tdr(z) = \mathbb{P}(\text{nonnull} \mid z) = 1 - fdr(z)$$

$$= \frac{\pi_1 f_1(z)}{f(z)}.$$

$$\widehat{tdr}(z) = 1 - \widehat{fdr}$$

$$\hat{y}_{1k} = N d\, \widehat{tdr}_k \cdot \hat{f}_k$$

Non-null cdf of $\widehat{fdr}(z)$

$$\mathbb{P}_1\{\widehat{fdr} \le q\} = \frac{\sum_{\{k:\, \widehat{fdr} \le q\}} \hat{y}_{1k}}{\sum_k \hat{y}_{1k}}$$

$$\mathbb{E}_1 \widehat{fdr}_1 = \frac{\sum \widehat{fdr}\, \hat{y}_{1k}}{\sum \hat{y}_{1k}}$$

lower if we have more power.

Chap 6: Theoretical Null Distribution

Permutation null dist:

permute indices on dataset
to remove all signal to
construct null distribution.

Empirical Null Estimates

$f_0 \sim N(\delta_0, \sigma_0^2)$

Near $z=0$

$\log f_0 \sim \beta_0 + \beta_1 z + \beta_2 z^2$

then use values near $z=0$
in the same poisson regression.

Now assume

$\mu \sim g(\cdot) \quad z|\mu \sim N(\mu, 1)$

$g(\mu) = \pi_0 \, \delta_0(\mu) + \pi_1 g_1(\mu)$

$f(z) = \int_{-\infty}^{\infty} g(\mu) \varphi(z-\mu) \, d\mu$

$\delta_g = \arg\max f(z) \quad \sigma_g = \left(\frac{-\partial^2}{\partial z^2} \log f \right)^{-1/2}_{\delta_g}$

What's the worst case scenario?

$\sigma_{max} = \max \sigma_g$

$f(z) = \pi_0 \varphi(z) + \sum_{j=1}^{J} \pi_j \frac{\left(\varphi(z-\mu_j) + \varphi(z+\mu_j)\right)}{2}$

$\Rightarrow \delta_g = 0$

$\Rightarrow \sigma_g = (1-Q)^{-1/2}$

$Q =$ See notes

Then maximizing we see

$$\theta_{max} = \left(1 - \max_{\mu} \left\{ \frac{\mu^2}{1 + C_0 e^{\mu^2/2}} \right\} \right)$$

$$C_0 = \frac{\pi_0}{1 - \pi_0}$$

<u>MLE</u>

$z = (z_1, \dots, z_n)$ $\quad N_0$: # $z_i \in A_0$ acceptance region

$$I_0 = \{ i : z_i \in A_0 \}$$

$$\vec{z}_0 = \{ z_i : i \in I_0 \}$$

$$f_{\delta_0, \sigma_0, \pi_0}(\vec{z}_0) = \underbrace{\left[\binom{N}{N_0} \theta^{N_0} (1-\theta)^{N-N_0} \right]}_{\text{Binomial probs}} \underbrace{\prod_{I_0} \frac{\varphi_{\delta_0, \sigma_0}(z_i)}{H_0(\delta_0, \sigma_0)}}_{\text{Conditional of being in } A_0}$$