# Outline

1) Multiple Testing
2) Familywise error rate control
3) Stepdown multiple testing
4) Simultaneous intervals / deduced inference
5) False Discovery Rate control
6) Benjamini - Hochberg Procedure

# Multiple Testing

In many testing problems, we want to test many hypotheses at a time, e.g.

- Test $H_{0j}: \beta_j = 0$ for $j = 1, \ldots, d$ in linear regression

- Test whether each of 2M single nucleotide polymorphisms (SNPs) is associated with a given phenotype (e.g., diabetes / schizophrenia)

- Test whether each of 2000 web site tweaks affects user engagement

Setup: $X \sim P_\theta \in \mathcal{P}$ $\quad H_{0i}: \theta \in \Theta_{0i}$, $i = 1, \ldots, m$

(Commonly, $H_{0i}: \theta_i = 0$)

Goal: Return accept/reject decision for each $i$.

Let $\mathcal{R}(X) = \{i : H_{0i} \text{ rejected}\} \subseteq \{1, \ldots, m\}$

$\mathcal{H}_0(\theta) = \{i : H_{0i} \text{ true}\}$

$R(X) = |\mathcal{R}(X)|$, $\quad m_0 = |\mathcal{H}_0|$

# Familywise Error Rate

**Problem**: Even if all $H_{0i}$ true, might have

$$\mathbb{P}(\text{any } H_{0i} \text{ rejected}) \gg \alpha$$

**Ex** $X_i \overset{\text{ind.}}{\sim} N(\theta_i, 1)$ $\quad i = 1, \ldots, m.$ $\quad H_{0i} : \theta_i = 0$

$$\mathbb{P}_0(\text{any } H_{0i} \text{ rejected}) = 1 - (1-\alpha)^m \to 1$$

Is this a problem? Yes, if all attention will be focused on the (false) rejections and none on the (correct) non-rejections.

Classical solution is to control the familywise error rate (FWER):

$$\text{FWER}_\theta = \mathbb{P}_\theta(\text{any false rejections})$$
$$= \mathbb{P}_\theta(\mathcal{R} \cap \mathcal{H}_0 \neq \emptyset)$$

Want $\displaystyle \sup_{\theta \in \Theta} \text{FWER}_\theta \leq \alpha$

Typically achieved by "correcting" marginal p-values $p_1(X), \ldots, p_m(X)$ $\quad \left(p_i \overset{H_{0i}}{\geq} U[0,1]\right)$

e.g., $p_i(X) = 2(1 - \Phi(|X_i|))$ for Gaussian

# Šidák's Correction

Assume $p_1, \dots, p_m$ are mutually independent, with $p_i \geq U[0,1]$ under $H_{0i}$

What if we reject $H_{0i}$ when $p_i \leq \tilde{\alpha}_m$?

$$\mathbb{P}_\theta(\text{no false rejections})$$

$$= \mathbb{P}_\theta(p_i \geq \tilde{\alpha}_m, \ \forall i \in \mathcal{H}_0)$$

$$\geq (1-\tilde{\alpha}_m)^{m_0} \quad (= \text{ if } p_i \overset{H_{0i}}{\sim} U[0,1])$$

$$\geq (1-\tilde{\alpha}_m)^m$$

Set $= 1-\alpha$: $\quad \tilde{\alpha}_m = 1-(1-\alpha)^{1/m}$

$$\approx \alpha/m \quad \text{for small } \alpha$$

This gives tight control of FWER whenever $p_i \overset{H_{0i}}{\sim} U[0,1]$, $p_i$ independent

Called <u>Šidák's Correction</u>

# Bonferroni Correction

For general dependence, can guarantee control by rejecting $H_{0i}$ iff $p_i \le \alpha/m$:

$$\mathbb{P}_\theta(\text{any false rejections})$$
$$= \mathbb{P}_\theta\Big(\bigcup_{i \in \mathcal{H}_0} \{H_{0i} \text{ rejected}\}\Big)$$
$$\le \sum_{i \in \mathcal{H}_0} \mathbb{P}_\theta(H_{0i} \text{ rejected})$$
$$\le m_0 \cdot \alpha/m \le \alpha$$

No assumptions on dependence, not much worse than Šidák.

# Holm's Procedure

We can directly improve on Bonferroni by using a <u>step-down procedure</u>

First, order p-values $p_{(1)} \le p_{(2)} \le \cdots \le p_{(m)}$

Order hyp. to match: $H_{(1)}, H_{(2)}, \ldots, H_{(m)}$

Holm's step-down procedure

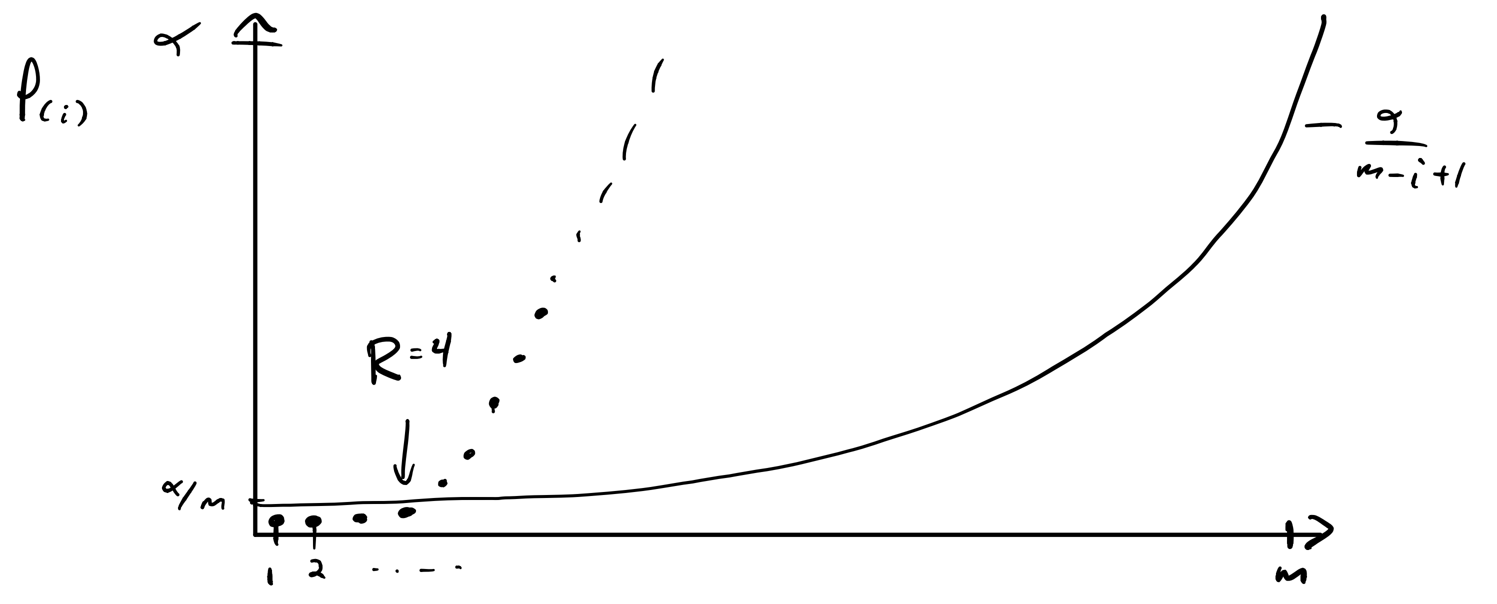
1. If $p_{(1)} \le \alpha/m$, reject $H_{(1)}$ and continue.
   Else, accept $H_{(1)}, \ldots, H_{(m)}$ and halt.
2. If $p_{(2)} \le \alpha/(m-1)$, reject $H_{(2)}$ and continue.
   Else, accept $H_{(2)}, \ldots, H_{(m)}$ and halt.

$\vdots$

m. If $p_{(m)} \le \alpha$, reject $H_{(m)}$. Else accept $H_{(m)}$

More compactly: $R = \max\{r : p_{(i)} \le \frac{\alpha}{m-i+1}, \forall i \le r\}$

Reject $p_{(1)}, \ldots, p_{(R)}$



**Prop** Holm's procedure controls FWER at level $\alpha$

**Proof** Let $p_0^* = \min\{p_i : i \in \mathcal{H}_{0i}\}$

$\mathbb{P}(p_0^* \le \frac{\alpha}{m_0}) \le \alpha$ by union bound.

Suppose $p_0^* > \alpha/m_0$, want to show no false rejections

Let $k = \#\{i : p_i \le p_0^*\} \le m - m_0 + 1$

Note $p_{(k)} = p_0^* > \frac{\alpha}{m_0} \ge \frac{\alpha}{m-k+1}$

So $R < k \Rightarrow p_0^* > p_{(R)} \Rightarrow$ No false rej-s

Holm's procedure strictly dominates Bonferroni.

A different stepdown procedure dominates Šidák when p-values are indep.

1. If $p_{(1)} \le \tilde{\alpha}_m$, reject $H_{(1)}$ & continue
2. If $p_{(2)} \le \tilde{\alpha}_{m-1}$, reject $H_{(2)}$ & continue

$\vdots$

m. If $p_{(m)} \le \alpha$, reject $H_{(m)}$

There is a general framework for making such improvements, called the <u>closure principle</u>

# Closure Principle

Assume we can construct a (marginal) level $\alpha$ test for every intersection null hypothesis: for $S \subseteq \{1, \dots, m\}$

$$H_{0S} : \theta \in \bigcap_{i \in S} \Theta_{0i}$$

e.g., reject $H_{0S}$ if $\min_{i \in S} p_i \leq \alpha / |S|$

**Step 1.** Provisionally reject $H_{0S}$ if the marginal test rejects

**Step 2.** Reject $H_{0i}$ if $H_{0S}$ rejected for all $S \ni i$

**Prop** This two-step procedure controls FWER

**Proof** $\mathbb{P}$(any false rejections)

$$\leq \mathbb{P}(H_{0\mathcal{H}_0} \text{ rejected in Step 1})$$

$$\leq \alpha$$

# Testing with dependence

Bonferroni isn't much worse than Šidák,

e.g. $\alpha = 5\%$ $m = 20$ : $.0025$ vs $.00256$

But when tests are highly dependendent, can often do <u>much</u> better.

<u>Ex.</u> Scheffé's S-method

$$X \sim N_d(\theta, I_d) \qquad \theta \in \mathbb{R}^d$$

$$H_{0,\lambda} : \theta'\lambda = 0 \quad \text{for } \lambda \in S^{d-1} \qquad (\text{"}m\text{"} = \infty)$$

$(\alpha = 5\%)$

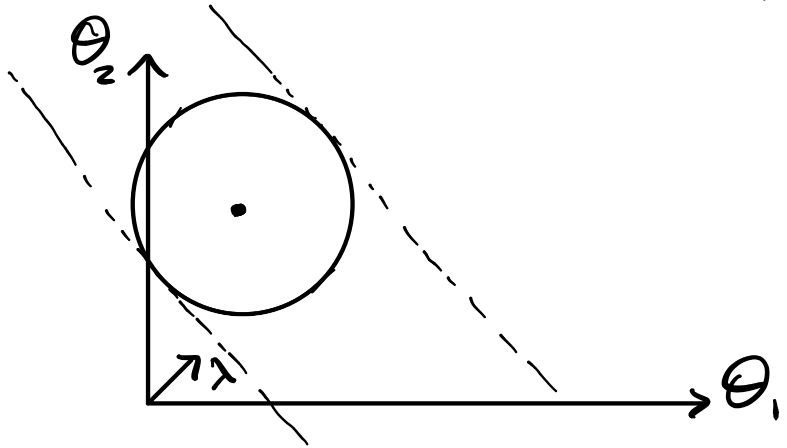
Reject $H_{0,\lambda}$ if $\|X'\lambda\|^2 \geq \chi^2_d(\alpha) \approx d + 3\sqrt{d}$

Controls FWER:

$$\sup_{\lambda : \theta'\lambda = 0} \|X'\lambda\|^2 \leq \sup_{\lambda} \|(X-\theta)'\lambda\|^2 \sim \chi^2_d(\alpha)$$

Can view as <u>deduction</u> from confidence region

$$C(X) = \{\theta : \|\theta - X\|^2 \leq \chi_d^2(\alpha)\}$$



## Deduced inference

Given any joint confidence region $C(X)$ for $\theta \in \Theta$, we may freely assume $\theta \in C(X)$ and "deduce" any and all implied conclusions without any FWER inflation

$$\mathbb{P}_\theta(\text{any deduced inference is wrong})$$
$$\leq \mathbb{P}_\theta(\theta \notin C(X)) \leq \alpha$$

Deduction is often a good paradigm for deriving simultaneous intervals

We say $C_1(X), \ldots, C_m(X)$ are <u>simultaneous $1-\alpha$ confidence intervals</u> for $g_1(\theta), \ldots, g_m(\theta)$ if

$$\mathbb{P}_\theta(g_i(\theta) \in C_i(X), \ \forall i = 1, \ldots, m) \geq 1-\alpha$$

## Ex Simultaneous intervals for multivar. Gaussian

Assume $X \sim N_d(\theta, \Sigma)$, $\Sigma$ known, $\Sigma_{ii} = 1$

Let $t_\alpha$ = upper-$\alpha$ quantile of $\|X - \theta\|_\infty$

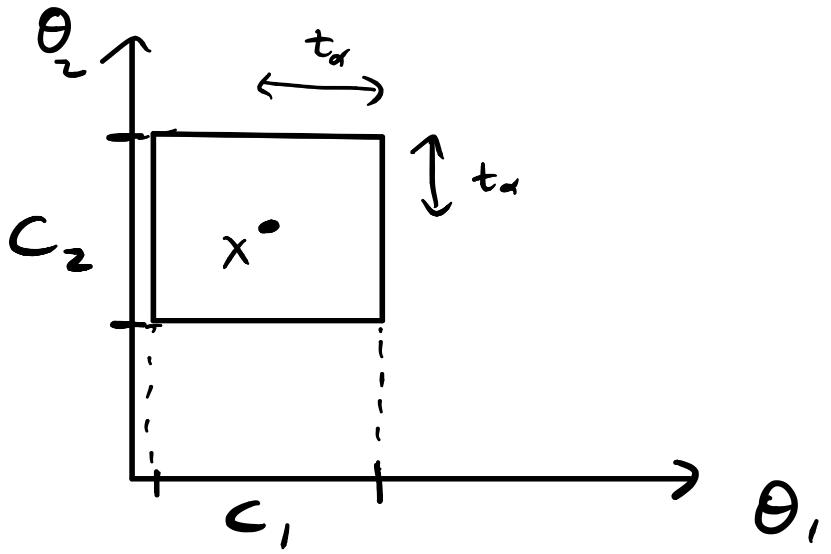
$$C(X) = \{\theta : |\theta_i - X_i| \le c_\alpha, \ \forall i\}$$

$$= (X_1 \pm t_\alpha) \times (X_2 \pm t_\alpha) \times \cdots \times (X_d \pm t_\alpha)$$

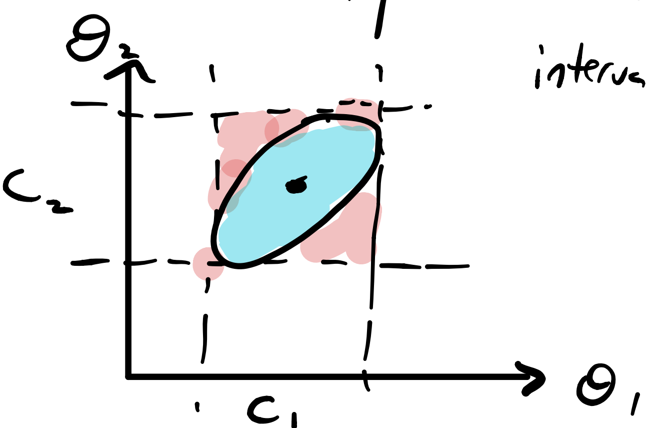
$$= C_1(X_1) \times \cdots \times C_d(X_d)$$

$$\mathbb{P}_\theta(C_i(x) \not\ni \theta_i, \text{ any } i) = \mathbb{P}_\theta(\theta \notin C(X)) = \alpha$$

$(d = 2)$




Note we could have instead constructed an elliptical conf. region, but then the intervals would be conservative.




$\mathbb{P}($ blue $) = 1 - \alpha$

$\mathbb{P}(\theta_1 \in C_1, \theta_2 \in C_2) =$

$\mathbb{P}($ blue $+$ pink $) > 1 - \alpha$

Ex Linear regression $\quad$ $n$ obs, $d$ variables

$X \in \mathbb{R}^{n \times d}$ design $\rightsquigarrow$ $\hat{\beta} \sim N_d(\beta, \sigma^2 (X'X)^{-1})$

Estimate $\hat{\sigma}^2 = RSS/(n-d) \perp\!\!\!\perp \hat{\beta}$

Then $$\frac{\hat{\beta} - \beta}{\hat{\sigma}} = \frac{Z}{\sqrt{V/(n-d)}}$$

where $Z = (\hat{\beta} - \beta)/\sigma \sim N_d(0, (X'X)^{-1})$

$V = RSS/\sigma^2 \sim \chi^2_{n-d}$

$Z \perp\!\!\!\perp V \Rightarrow$ Distr. of $\dfrac{\hat{\beta} - \beta}{\hat{\sigma}}$ fully known.

Assume wlog $\left((X'X)^{-1}\right)_{jj} = 1 \quad \forall j$

Let $t_\alpha$ denote upper-$\alpha$ quantile of $\left\|\frac{\hat{\beta}-\beta}{\hat{\sigma}}\right\|_\infty$.

Then $C_j = \hat{\beta}_j \pm \hat{\sigma} t_\alpha$ are simultaneous CIs for $\hat{\beta}_j$, $j = 1, \ldots, d$. (Compute $t_\alpha$ by simulation)

$$\mathbb{P}(\beta_j \in C_j, \forall j) = \mathbb{P}(|\hat{\beta}_j - \beta_j| \leq \hat{\sigma} t_\alpha, \forall j) = 1 - \alpha.$$

# False Discovery Rate (FDR)

**Motivation**: Suppose we test 10,000 hypotheses with independent test statistics, all at level $\alpha = 0.001$. We expect 10 rejections just by chance. What if we get 50? Probably only $\approx 20\%$ of them are fake rejections.

Can we accept 10 false rejections as long as most rejections are valid?

Benjamini & Hochberg (95) proposed a more liberal error control criterion, called FDR

$$R(X) = \#\mathcal{R}(X) \qquad \text{\# rejections/ "discoveries"}$$
$$V(X;\theta) = \#\left(\mathcal{R}(X) \cap \mathcal{H}_0(\theta)\right) \qquad \text{\# false discoveries}$$

The false discovery proportion (FDP) is

$$FDP = V/R \quad \text{where} \quad 0/0 \triangleq 0 \quad \left(\frac{V}{R \vee 1}\right)$$

The FDR is $\mathbb{E}_\theta[FDP] = \mathbb{E}_\theta\left[\frac{V}{R \vee 1}\right]$

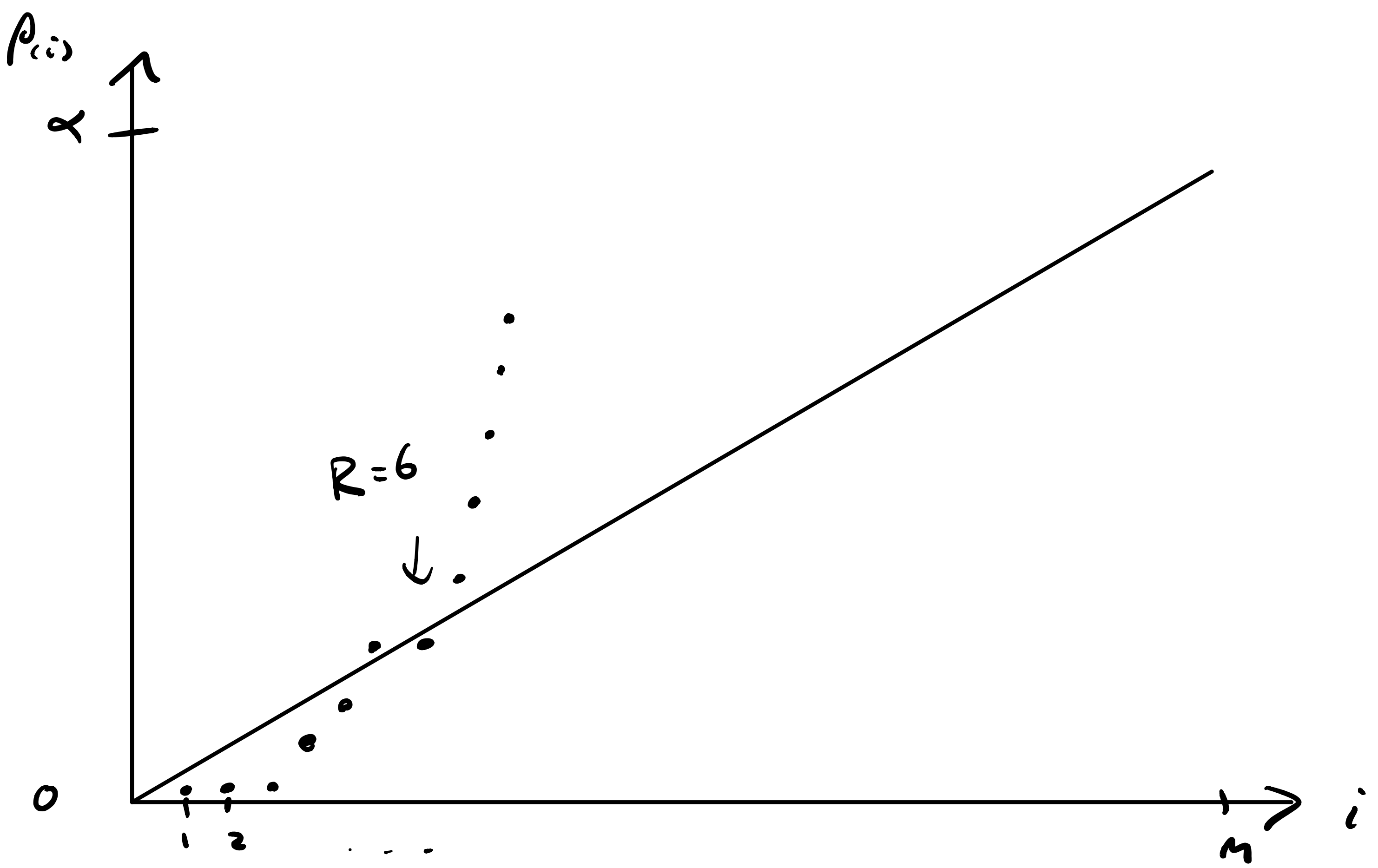
# Benjamini-Hochberg Procedure

B & H also proposed a method to control FDR given ordered p-values $p_{(1)} \le p_{(2)} \le \cdots \le p_{(m)}$:

$$R(X) = \max\left\{r : p_{(r)} \le \frac{\alpha r}{m}\right\}$$

(called step-up procedure)

Reject $H_{(1)}, \ldots, H_{(R)}$




This is much more liberal than Holm's procedure when $1 \ll R \ll m$. BH rejects (at least) $r$ p-values if $p_{(r)} \le \frac{\alpha}{m} \cdot r$

Holm needs $p_{(r)} \le \frac{\alpha}{m-r+1} \approx \frac{\alpha}{m} \cdot \left(1 + \frac{r}{m}\right)$

# BH as empirical Bayes

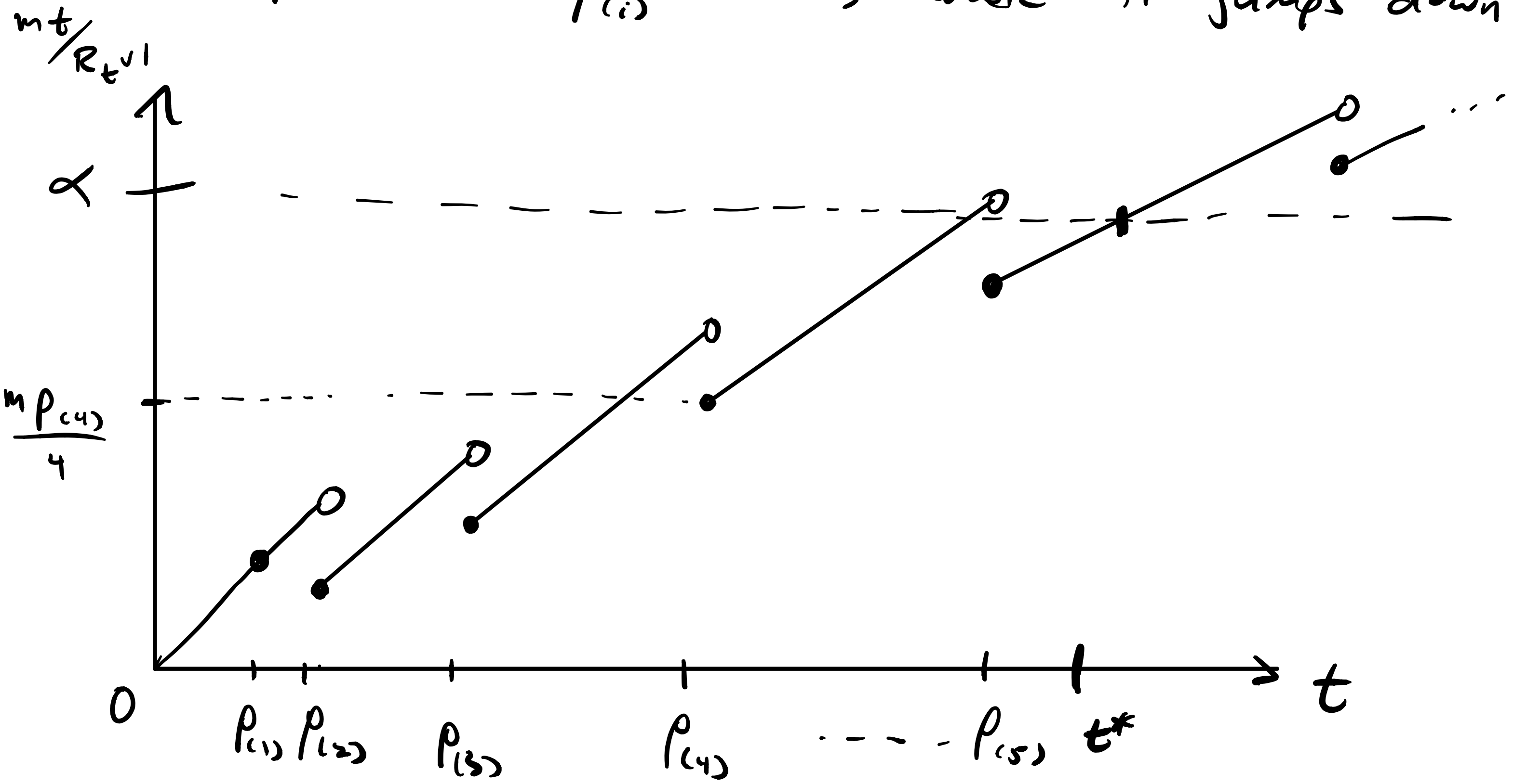
<u>Equivalent formulation:</u> for $R_t = \#\{i : p_i \le t\}$,

let $\widehat{FDP}_t = \dfrac{mt}{R_t \vee 1}$ ← "estimate" of $V_t$, # false disc.

BH rejects $H_i$ if $p_i \le t^*(X) = \max\{t : \widehat{FDP}_t \le \alpha\}$

<u>Why?</u> $\widehat{FDP}_t$ is continuously increasing in $t$,

<u>except</u> at $p_{(i)}$ values, where it jumps down




Only values of $t$ that matter for the algorithm

are $t = p_{(i)}$, where $\widehat{FDP}_t = \dfrac{m p_{(i)}}{i}$

$$\frac{m p_{(i)}}{i} \le \alpha \iff p_{(i)} \le \frac{\alpha i}{m}$$

# FDR control

Elegant (but fragile) proof due to Storey, Taylor, & Sigmund (2002)

Assume $p_i$ indep., $p_i \sim U[0,1]$ $i \in \mathcal{H}_0$

Let $V_t = \#\{i \in \mathcal{H}_0 : p_i \leq t\}$

$$\mathrm{FDP}_t = \frac{V_t}{R_t \vee 1}$$

$$= \widehat{\mathrm{FDP}}_t \cdot \underbrace{\frac{V_t}{mt}}_{Q_t}$$

Then $\mathrm{FDR} = \mathbb{E}\left[\mathrm{FDP}_{t^*}\right]$

$$= \mathbb{E}\left[\widehat{\mathrm{FDP}}_{t^*} \cdot \frac{V_{t^*}}{mt^*}\right]$$

$$= \alpha \cdot \mathbb{E}\left[V_{t^*}/mt^*\right] \quad \left(\widehat{\mathrm{FDP}}_{t^*} \overset{a.s.}{=} \alpha\right)$$

Note $Q_t$ is a martingale when $t$ runs <u>backwards</u> from $t=1$ to $t=0$:

$s < t$:

$$\mathbb{E}\left[V_s \mid V_t = v\right]$$

$$= \mathbb{E}\left[\#\{i : p_i \le s\} \mid \#\{i : p_i \le t\} = v\right]$$

$$= v \cdot \frac{s}{t}$$

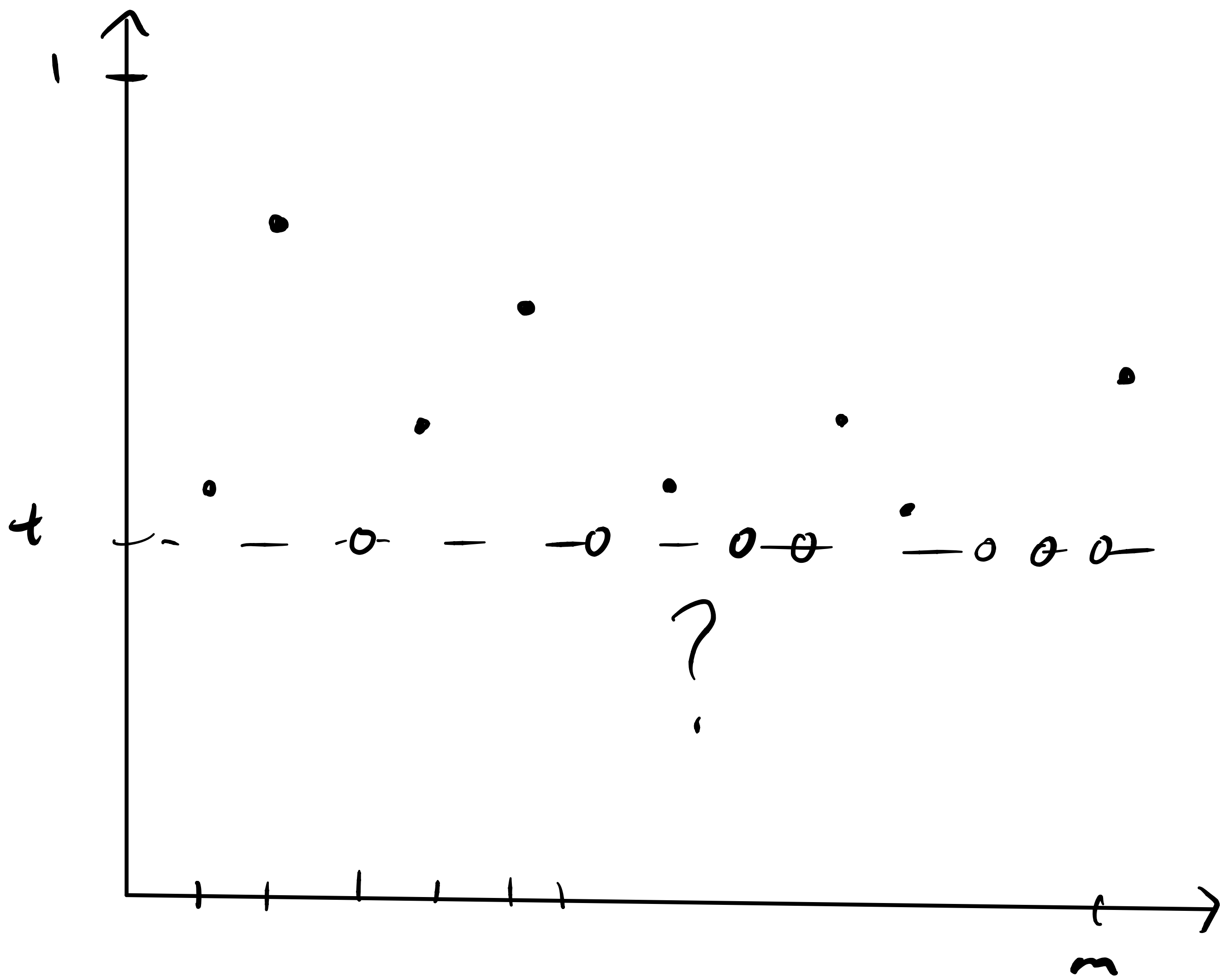
$$\mathbb{E}\left[\frac{V_s}{ms} \mid \frac{V_t}{mt} = q\right] = \frac{1}{ms} \cdot (qmt) \cdot \frac{s}{t} = q$$

And $t^*$ is a stopping time wrt the filtration $\mathcal{F}_t = \sigma(p_1 \vee t, \ldots, p_m \vee t)$

(again, filtration with $t=1 \rightarrow t=0$)

Why? For $s \ge t$, $R_s = \#\{i : p_i \le s\}$
$$= \#\{i : p_i \vee t \le s\}$$

$$\widehat{FDP}_s = \frac{ns}{R_s}$$




$$\text{FDR} = \alpha \, \mathbb{E}\left[ \frac{V_{t^*}}{m t^*} \right]$$

$$= \alpha \, \mathbb{E}\left[ V_1 / m \right]$$

$$= \alpha \, m_0 / m$$

# Remarks

- Proof only works if p-values indep., null ones exactly uniform
- More robust proof shows FDR controlled when null p-values conservative, can be extended to "positive dependence"
- FDR controlled under general dependence if we use corrected level $\alpha / L_m$,

$$L_m = \sum_{i=1}^{m} \frac{1}{i} \approx \log(m)$$