11/8/2023

## False Discovery Rate (FDR)

**Setting:** $X \sim P_\theta \quad \theta \in \Theta$

$H_{0,i}: \theta \in \Theta_{0,i} \quad i = 1, \dots, n$

p-values $P_i \sim \text{unif}[0,1]$ for $i \in \mathcal{N}$ ← set of true nulls

| | $H_{0,i}$ accepted | $H_{0,i}$ rejected | |
|---|---|---|---|
| $H_{0,i}$ true | $U$ | $V$ | $n_0 = \lvert\mathcal{N}\rvert$ |
| $H_{0,i}$ false | $T$ | $S$ | $n - n_0$ |
| | $n - R$ | $R$ | |

FWER: $P(V \geq 1)$ Too strict when $n$ is large.

**FDR**

$$FDP = \frac{V}{\max(R,1)}$$ fraction false positives

$$FDR = \mathbb{E}[FDP]$$ average of FDP

We will discuss a procedure that guarantees FDR ≤

**note 1** $\mathbb{1}_{V \geq 1} \geq \frac{V}{\max(R,1)}$ so $FDR \leq \overbrace{P(V \geq 1)}^{\text{FWER}}$

**note 2** under the global null, $\frac{V}{\max(R,1)} = \begin{cases} 1 & \text{if } V \neq \emptyset \\ 0 & \text{o/w} \end{cases}$

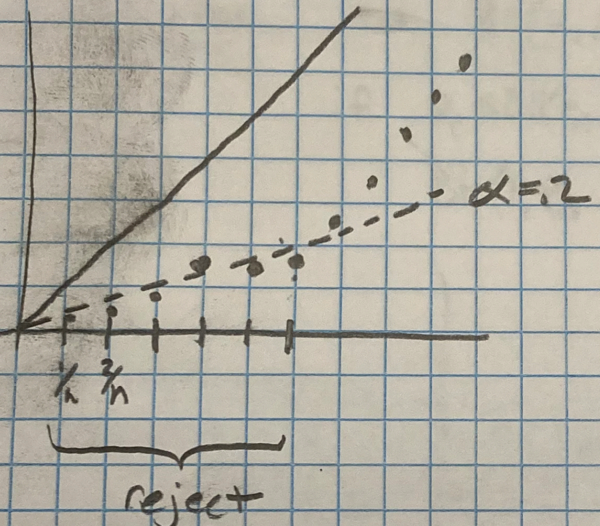
so $FDR = P(V \geq 1)$

## Benjamin-Hochberg Algorithm (1995)

Let $P_{(1)} \le \dots \le P_{(n)}$ be sorted p-values

$i_0 = \max\left\{ i : \boxed{P_{(i)} \le \frac{i}{n}\alpha} \right\}$ $\qquad \alpha \in [0,1]$ FDR level

reject $H_{0,(i)}$ for $(i) \le i_0$




Suppose $P_i$ are independent

**theorem** The BH procedure gives

$$FDR \le \frac{n_0}{n} \cdot \alpha .$$

Equality holds when $P_i \sim unif[0,1]$ for nulls.

Pf (BH controls FDR)

$$FDR = \mathbb{E}\frac{V}{\max(R,1)} = \mathbb{E}\left[\sum_{i\in\mathcal{N}} \frac{V_i}{\max(R,1)}\right]$$

$V_i = \mathbb{1}\{i\in R\}$ for $i \in \mathcal{N}$

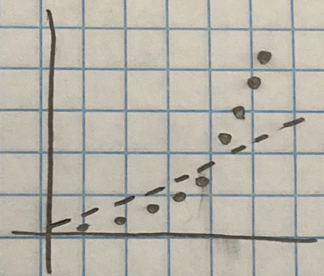
$$\frac{V_i}{\max(R,1)} = \sum_{k=1}^{n} \frac{V_i \cdot \mathbb{1}_{R=k}}{k}$$

Clever observation

Let $BH(P_1, P_2, \dots, P_n) \subseteq \{1, \dots, n\}$ be rejections

if $P_i \in BH(P_1, \dots, P_n)$ then

$$BH(P_1, P_2, \dots, P_{i-1}, 0, P_{i+1}, \dots, P_n) = BH(P_1, P_2, \dots, P_{i-1}, P_i, P_{i+1}, \dots)$$



Let $R(P_i \to 0)$ be size of rejection set if we set $P_i$ to zero.
This only depends on $P_{-i}$ (p-values except $P_i$)

$$FDR = \sum_{i\in\mathcal{N}} \mathbb{E}\left[\sum_{k=1}^{n} \frac{V_i \mathbb{1}_{R(P_i\to 0)=k}}{k}\right]$$

$$= \sum_{i\in\mathcal{N}} \mathbb{E}\left[\sum_{k=1}^{n} \mathbb{E}\left[\frac{V_i \mathbb{1}_{R(P_i\to 0)=k}}{k} \,\middle|\, P_{-i}\right]\right]$$

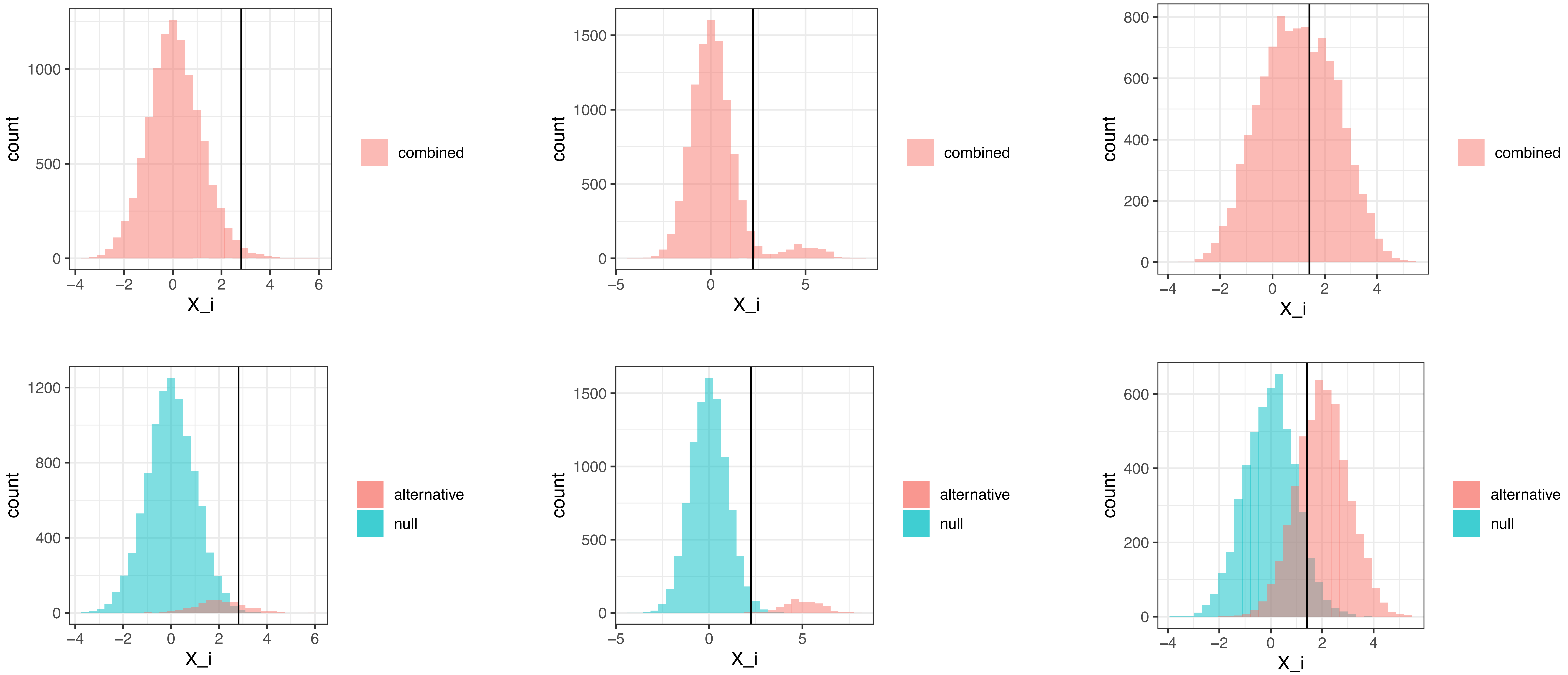
$$= \sum_{i\in\mathcal{N}} \mathbb{E}\left[\sum_{k=1}^{n} \frac{\mathbb{1}_{R(P_i\to 0)=k}}{k} \cdot P\left(P_i \le \frac{k\cdot\alpha}{n}\right)\right]$$

$$\le \sum_{i\in\mathcal{N}} \mathbb{E}\left[\sum_{k=1}^{n} \frac{\alpha}{n} \mathbb{1}_{R(P_i\to 0)=k}\right] = \frac{n_0}{n}\cdot\alpha$$

10% non null $\mu = 2$

10% non null $\mu = 5$

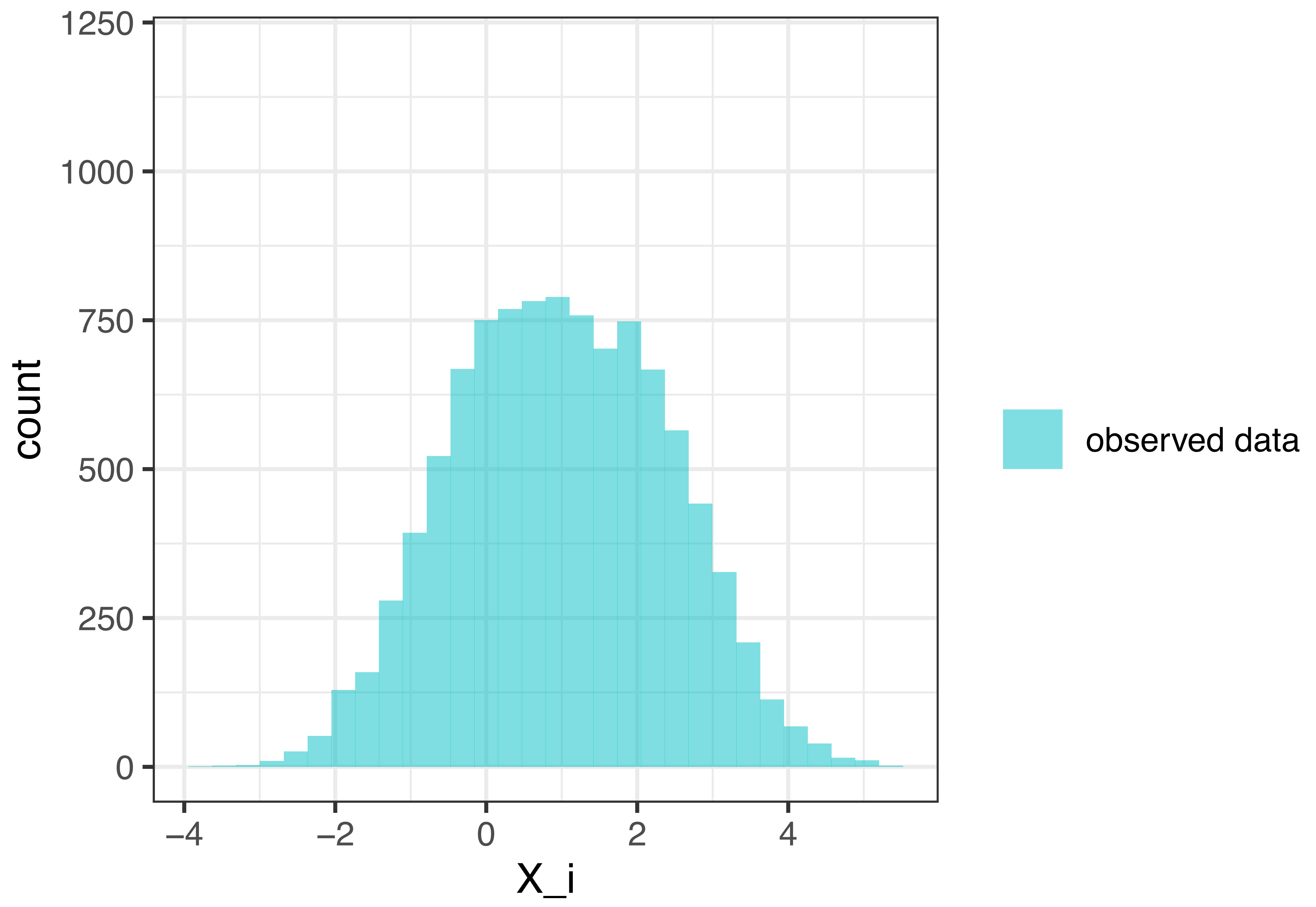
50% non null $\mu = 2$

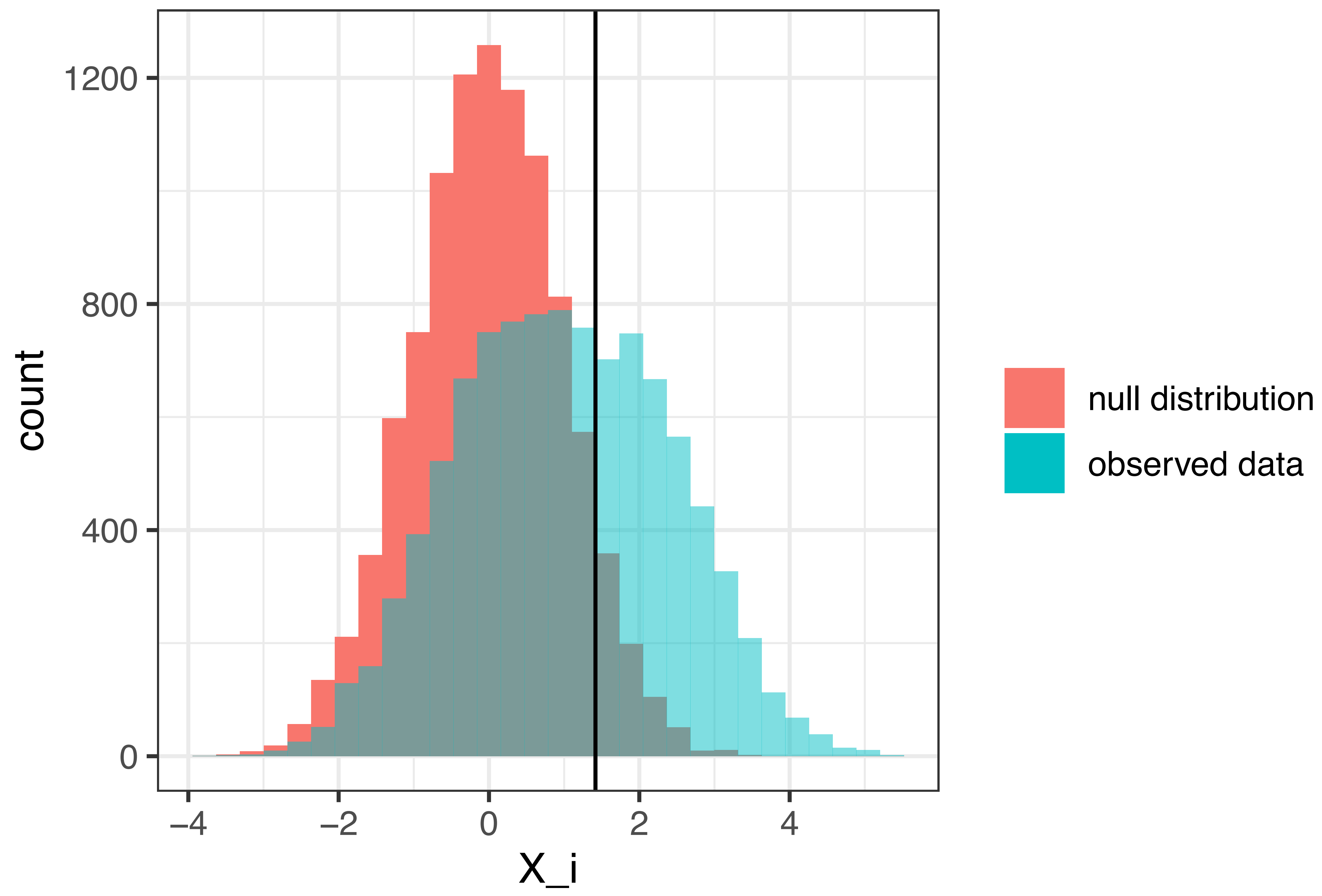



BH cutoff: 2.8

BH cutoff: 2.3

BH cutoff: 1.38




**Benjamini-Hochberg:** find cutoff where blue area to the right is 5x larger than the red area to the right

## A closer look at BH

[Show numerics]

$$\hat{F}_n(t) = \frac{|\{i : P_i \le t\}|}{n} \qquad \text{ECDF}$$

$$p^* = \max\left\{P_{(i)} : P_{(i)} \le \frac{\alpha i}{n}\right\} \qquad \text{BH critical p-value}$$

$$= \max\left\{P_{(i)} : P_{(i)} \le \alpha \hat{F}(P_{(i)})\right\}$$

$$= \max\left\{t \in \{P_1, \dots, P_n\} : t \le \alpha \hat{F}(t)\right\} \qquad 0 \text{ if empty}$$

$$\tau_{BH} = \max\left\{t : \frac{t}{\hat{F}(t)} \le \alpha\right\}$$

BH alg: reject $\{i : P_i \le \tau_{BH}\}$

| | $H_0$ accepted | $H_0$ rejected | |
|---|---|---|---|
| $H_{0,i}$ true | $U(t)$ | $V(t)$ | $n_0$ |
| $H_{0,i}$ false | $T(t)$ | $S(t)$ | $n-n_0$ |
| | | $R(t)$ | |

$$FDP(t) = \frac{V(t)}{\max(R(t), 1)}$$

$$FDR(t) = \mathbb{E}[FDP(t)]$$

now

$$\frac{t}{\hat{F}(t)} = \frac{nt}{n\hat{F}(t)} = \frac{\hat{V}(t)}{R(t)}$$

so $\underbrace{\frac{t}{\hat{F}(t)}}_{\widehat{FDR}(t)}$ can be interpreted as estimate of FDR

BH alg: $\tau_{BH} = \sup\left\{t \le 1 : \widehat{FDP}(t) \le \alpha\right\}$

[Show numerics]