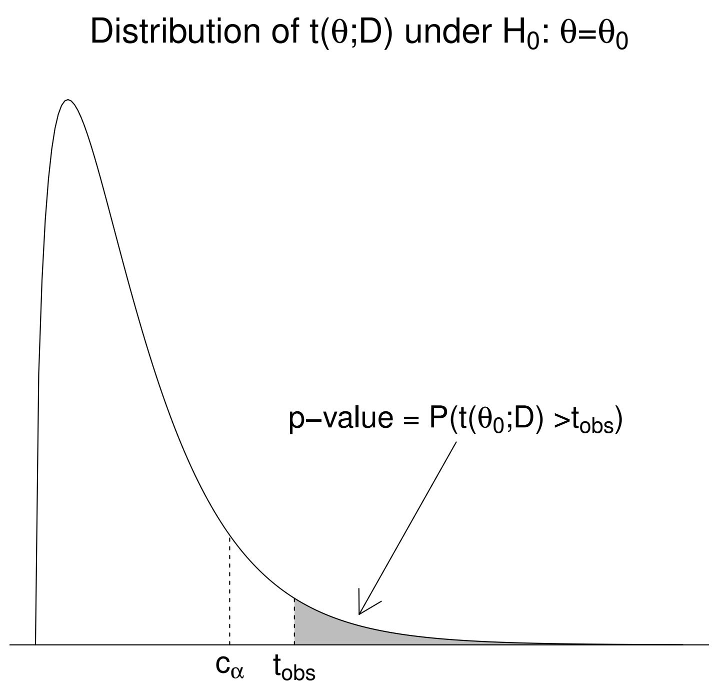
The p-value

Definition The p-value is defined as

$$\text{p-value} \equiv \mathbb{P}[t(\theta_0; \boldsymbol{D}) \geq t_{obs}],$$

where $t_{obs} \equiv t(\theta_0; \boldsymbol{D}_{obs})$ denotes the numeric value of the test statistic $t(\theta; \boldsymbol{D})$ evaluated at the null hypothesis H_0 and the observed data.



$$t_{obs} < c_\alpha \quad \Rightarrow \quad \text{p-value} > \alpha$$

$$t_{obs} = c_\alpha \quad \Rightarrow \quad \text{p-value} = \alpha$$

$$t_{obs} > c_\alpha \quad \Rightarrow \quad \text{p-value} < \alpha$$

A universal scale (whatever test statistic). Ubiquitous jargon.

The p-value: Also a Random Variable

$$T = \underbrace{\text{test statistic}}_{t(\theta;\boldsymbol{D})} \qquad \text{and} \qquad F_0(t) = \text{CDF of } \underbrace{T \text{ under } H_0}_{t(\theta_0;\boldsymbol{D})}$$

$$\Downarrow$$

$$\underset{\text{p-value}}{\overset{\uparrow}{u_{obs}}} \equiv \underbrace{1 - F_0(t_{obs})}_{\mathbb{P}[t(\theta_0;\boldsymbol{D}) \geq t_{obs}]}$$

$$\Downarrow$$

The corresponding RV,

$$U = 1 - F_0(T),$$

describes the stochastic behavior of the p-value under H_0.

Exercise 8.3 (p. 139)

Fact Under H_0, $U = 1 - F_0(T) \sim \text{Uniform}(0, 1)$

$$
\begin{aligned}
F_U(u) &= \mathbb{P}(U \le u) \\
&= \mathbb{P}(1 - F_0(T) \le u) \\
&= \mathbb{P}(F_0(T) \ge 1 - u) \\
&\stackrel{?}{=} \mathbb{P}(T \ge F_0^{-1}(1 - u)) \\
&= 1 - F_0(F_0^{-1}(1 - u)) \\
&= 1 - (1 - u) \\
&= u
\end{aligned}
$$

$$f_U(u) = \frac{d}{du} F_U(u) = 1,$$

clearly, $u \in (0, 1)$

Multiple Testing Problem

Example N independent statistical tests, each with significance level α.

$$\mathbb{P}(\underline{\text{at least one}} \text{ false rejection})$$
$$= 1 - \mathbb{P}(\underline{\text{no}} \text{ false rejection}) = 1 - (1-\alpha)^N$$
$$\underset{\alpha=0.05}{\overset{\uparrow}{=}} \begin{cases} 0.226, & N = 5 \\ 0.401, & N = 10 \\ 0.923, & N = 50 \end{cases}$$

Remark Explains much of what's behind the current replication crisis in some scientific areas. Big problem for the age of Big Data.

Example: Replication Crisis

S. S. Young, A. Karr (2011), "Deming, data and observational studies: A process out of control and needing fixing", *Significance* **8**, pp. 116–120.

- 12 clinical trials between 1990 and 2010
- tested 52 scientific claims about the health benefits (or hazards) of vitamin E, vitamin D, calcium, selenium, hormone replacement therapy, folic acid, beta-carotne, ...
- unable to replicate <u>ANY</u> of the 52 claims

Bonferroni

hypothesis:	$H_{0,1}$	$H_{0,2}$	$\ldots$	$H_{0,N}$
p-value:	u_1	u_2	$\ldots$	u_N

Reject all $H_{0,i}$ for which

$$u_i \leq \alpha/N;$$

accept the rest.

Remark This guarantees

$$\begin{aligned} \mathbb{P}(\underline{\text{at least one}} \text{ false rejection}) &\leq \sum_{i=1}^{N} \mathbb{P}(\text{false rejection of } H_{0,i}) \\ &= N \times (\alpha/N) = \alpha \end{aligned}$$

is controlled at the level of α, but very severe correction!

Benjamini-Hockberg

sort :	$H_{0,(1)}$	$\succ$	$H_{0,(2)}$	$\succ$	$\ldots$	$\succ$	$H_{0,(N)}$
	$u_{(1)}$	$<$	$u_{(2)}$	$<$	$\ldots$	$<$	$u_{(N)}$

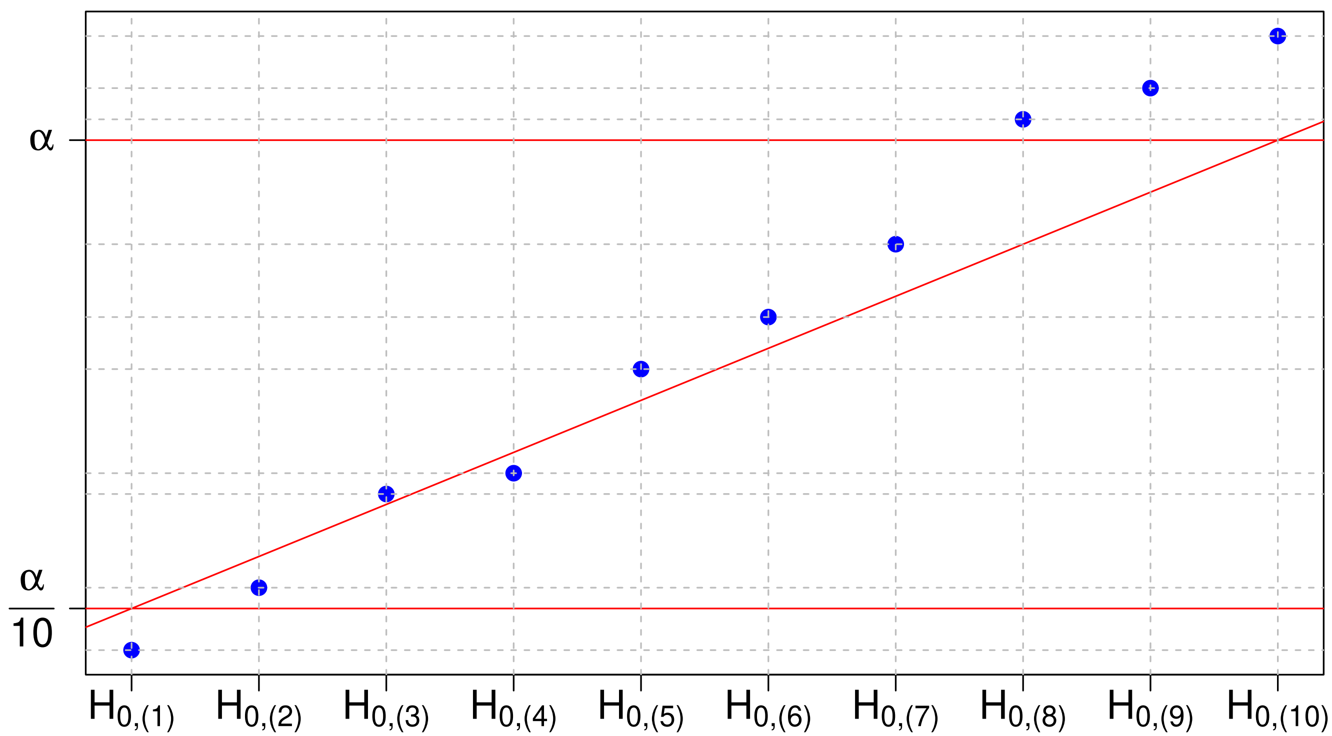
Reject $H_{0,(i)}$ for all $i \leq i_{max}$, where

$$i_{max} \equiv \max_{1,\ldots,N}\{i : \; u_{(i)} \leq i\alpha/N\}$$

is the largest index for which the inequality above holds; accept the rest.

Remark Interesting, but what's going on?

p–values $u_{(1)} < u_{(2)} < ... < u_{(10)}$ from testing 10 different null hypotheses



Exercise Which of these null hypotheses will be rejected, respectively under "ordinary" circumstances, by Bonferroni, and by Benjamini-Hockberg?

Think Bayesian!

$$\mathbb{P}(H_0|U \le u) = \frac{\overbrace{\mathbb{P}(U \le u|H_0)}^{\stackrel{?}{=}u}\,\overbrace{\mathbb{P}(H_0)}^{\le 1}}{\mathbb{P}(U \le u)}$$

$$\underset{u=u_{(i)}}{\overset{\uparrow}{\Rightarrow}} \quad \widehat{\mathbb{P}}(H_0|U \le u_{(i)}) \le \frac{u_{(i)}}{i/N} \quad \overset{u_{(i)} \le i\alpha/N}{\overset{\downarrow}{\le}} \quad \alpha$$

Denominator

$$\mathbb{P}(U \le u) = \underbrace{\mathbb{P}(U \le u|H_0)}_{u}\,\underbrace{\mathbb{P}(H_0)}_{?} + \underbrace{\mathbb{P}(U \le u|H_0^c)}_{?}\,\underbrace{\mathbb{P}(H_0^c)}_{?}$$

but, with multiple $H_{0,1}, ..., H_{0,N}$ being tested (some true and some not), can be estimated from $u_1, ..., u_N$ (i.e., empirical Bayes!!) by

$$\widehat{\mathbb{P}}(U \le u) = \frac{1}{N}\sum_{j=1}^{N} I(u_j \le u) \underset{u=u_{(i)}}{\overset{}{\underset{\uparrow}{=}}} \frac{i}{N}.$$

Empirical Bayes:
Nothing Bayesian?

Example 6.2 (p.96–97): Remember this earlier example?

$$
\begin{aligned}
X_i|\lambda_i &\overset{ind}{\sim} \text{Poisson}(\lambda_i) \\
\lambda_i &\overset{iid}{\sim} \text{Exponential}(\theta)
\end{aligned}
$$

Can simply regard each λ_i as latent variable, and use EM to estimate θ (details next slide).

$$\Downarrow$$

Critics (aka "true" Bayesians ☺):
"There is nothing Bayesian about empirical Bayes."

E-Step

$$\mathbb{E}(\lambda_i | X_i; \widehat{\theta}^{(t-1)}) = \frac{X_i + 1}{1 + \widehat{\theta}^{(t-1)}}.$$

M-Step Given $\lambda_1, ..., \lambda_n$, easy to derive[a] MLE of θ is simply $n/(\lambda_1 + ... + \lambda_n)$, but λ_i is latent, so replace[b] with "current guess" $\mathbb{E}(\lambda_i | X_i; \widehat{\theta}^{(t-1)})$, and get

$$\widehat{\theta}^{(t)} = \frac{n(1 + \widehat{\theta}^{(t-1)})}{\sum_{i=1}^{n} (X_i + 1)} \qquad (\dagger)$$

Convergence Must have $\widehat{\theta}^{(t-1)} = \widehat{\theta}^{(t)} = \widehat{\theta}^{(\infty)}$, so solve equation ($\dagger$) above[c] to get convergent solution $\widehat{\theta}^{(\infty)} = 1/\bar{X}$.

Exercise Do [a] + [c]. For [b], verify that a "direct replacement" of each λ_i is indeed sufficient for this EM.

Benjamini-Hockberg:

Fundamentally Bayesian!

Classical

☞ Choose the "right" threshold u to control $\mathbb{P}(U \leq u | H_0)$.†
[†In fact, here we are even borrowing/abusing Bayesian notation!]

Benjamini-Hockberg

☞ Choose the "right" threshold u to control $\mathbb{P}(H_0 | U \leq u)$.‡
☞ Only possible with multiple tests.
☞ Key is to borrow information from other tests!
[‡More precisely, $\widehat{\mathbb{P}}(H_0 | U \leq u)$, an estimated version of it.]

$\Downarrow$

Response to critics:
"Benjamini-Hockberg is fundamentally Bayesian,
and it is empirical Bayes."