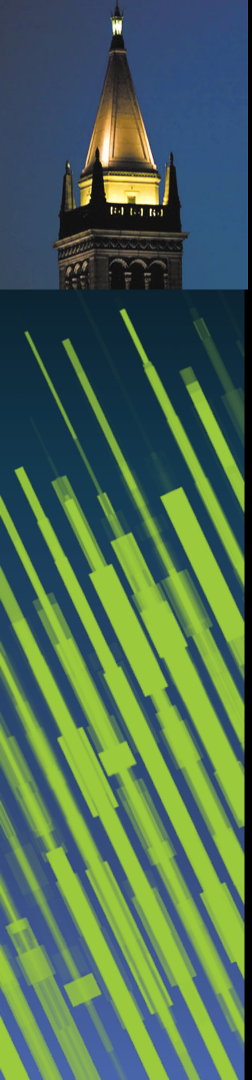
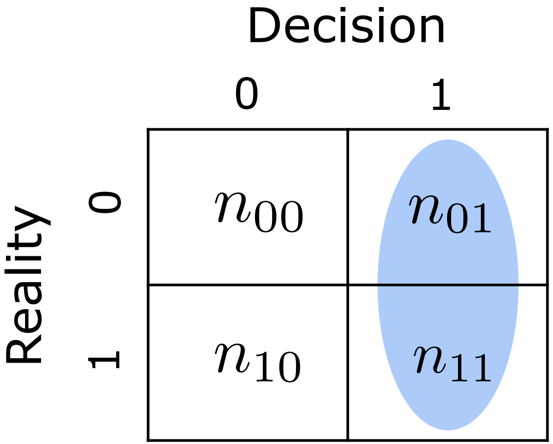


DS 102: Data, Inference, and Decisions

Lecture 4

Michael Jordan
University of California, Berkeley

Some Column-Wise Rates



Decision

		0	1
Reality	0	n_{00}	n_{01}
	1	n_{10}	n_{11}

false discovery proportion $= \dfrac{n_{01}}{n_{01}+n_{11}}$

The Goal: Control Errors A Priori

- The row-focused Neyman-Pearson paradigm, with its Type I and Type II errors, provides a priori control
  - meaning that if my assumptions about the null and alternative distributions are correct, then I can guarantee that these errors will be small (in an average, frequentist sense---over multiple draws of data)

The Goal: Control Errors A Priori

- The row-focused Neyman-Pearson paradigm, with its Type I and Type II errors, provides a priori control
  - meaning that if my assumptions about the null and alternative distributions are correct, then I can guarantee that these errors will be small (in an average, frequentist sense---over multiple draws of data)
- If I’m only testing one hypothesis, that’s satisfying

The Goal: Control Errors A Priori

- The row-focused Neyman-Pearson paradigm, with its Type I and Type II errors, provides a priori control
  - meaning that if my assumptions about the null and alternative distributions are correct, then I can guarantee that these errors will be small (in an average, frequentist sense---over multiple draws of data)
- If I’m only testing one hypothesis, that’s satisfying
- The problem that arose with our A/B testing example arose because we were doing many tests

The Goal: Control Errors A Priori

- The row-focused Neyman-Pearson paradigm, with its Type I and Type II errors, provides a priori control
  - meaning that if my assumptions about the null and alternative distributions are correct, then I can guarantee that these errors will be small (in an average, frequentist sense---over multiple draws of data)
- If I'm only testing one hypothesis, that's satisfying
- The problem that arose with our A/B testing example arose because we were doing many tests
- Can we find a way to obtain a priori control when there are many tests?

Comments on the Column-Wise Rates

- They can be thought of as estimates of conditional probabilities
- They are dependent on the prevalence (i.e., the probabilities of the two states of Reality in the population), via Bayes' Theorem
  - as such, they are more Bayesian
  - this is arguably a good thing
- Notation: let H denote Reality, and let D denote the decision

A Bayesian Calculation

$$P(H = 0 \mid D = 1) = \frac{P(H = 0, D = 1)}{P(D = 1)}$$

A Bayesian Calculation

$$P(H=0 \,|\, D=1) = \frac{P(H=0, D=1)}{P(D=1)}$$

$$= \frac{P(D=1 \,|\, H=0)P(H=0)}{P(D=1)}$$

$$= \frac{P(\text{Type I error}) \cdot \pi_0}{P(D=1)}$$

- We could upper bound π_0 with 1, and so the numerator can be controlled; what about the denominator?

A Bayesian Calculation

- Using the law of total probability, we have:

$$P(D = 1) = P(D = 1 \mid H = 0)P(H = 0) + P(D = 1 \mid H = 1)P(H = 1)$$

A Bayesian Calculation

- Using the law of total probability, we have:

$$\begin{aligned} P(D=1) &= P(D=1 \mid H=0)P(H=0) + P(D=1 \mid H=1)P(H=1) \\ &= \pi_0 P(D=1 \mid H=0) + (1-\pi_0)P(D=1 \mid H=1) \end{aligned}$$

- So we see that $P(D=1)$ depends on the prior π_0

A Bayesian Calculation

- Using the law of total probability, we have:

$$
\begin{aligned}
P(D=1) &= P(D=1 \,|\, H=0)P(H=0) + P(D=1 \,|\, H=1)P(H=1) \\
&= \pi_0 P(D=1 \,|\, H=0) + (1-\pi_0)P(D=1 \,|\, H=1)
\end{aligned}
$$

- So we see that $P(D=1)$ depends on the prior π_0
- Is this a problem?
  - i.e., do we have to either decide to be Bayesian and supply the prior, or decide to be frequentist and abandon this approach?

- No! Note that it's easy to estimate $P(D=1)$ directly from the data!

Controlling the FDR

- Benjamini & Hochberg (1995) proposed an algorithm that does it

Controlling the FDR

- Benjamini & Hochberg (1995) proposed an algorithm that does it
- Given m tests, obtain p-values P_i , and sort them from smallest to largest, denoting the sorted p-values as $P_{(k)}$

Controlling the FDR

- Benjamini & Hochberg (1995) proposed an algorithm that does it
- Given m tests, obtain p-values P_i , and sort them from smallest to largest, denoting the sorted p-values as $P_{(k)}$
  - the small ones are the safest to reject

Controlling the FDR

- Benjamini & Hochberg (1995) proposed an algorithm that does it
- Given m tests, obtain p-values P_i , and sort them from smallest to largest, denoting the sorted p-values as $P_{(k)}$
    - the small ones are the safest to reject
- Now, find the largest k such that:

$$P_{(k)} \leq \frac{k}{m}\alpha$$

Controlling the FDR

- Benjamini & Hochberg (1995) proposed an algorithm that does it
- Given m tests, obtain p-values P_i , and sort them from smallest to largest, denoting the sorted p-values as $P_{(k)}$
  - the small ones are the safest to reject
- Now, find the largest k such that:

$$P_{(k)} \leq \frac{k}{m}\alpha$$

- Reject the null hypothesis (i.e., declare discoveries) for all hypotheses H_i such that $i \leq k$

Controlling the FDR

- Benjamini & Hochberg (1995) proposed an algorithm that does it
- Given m tests, obtain p-values P_i , and sort them from smallest to largest, denoting the sorted p-values as $P_{(k)}$
  - the small ones are the safest to reject
- Now, find the largest k such that:

$$P_{(k)} \leq \frac{k}{m}\alpha$$

- Reject the null hypothesis (i.e., declare discoveries) for all hypotheses H_i such that $i \leq k$
- This controls the FDR!

P-Values

- Consider a point-null hypothesis, $\theta = 0$, and $\mathbb{P}$ denote that null
- Consider a statistic, $T(X)$, which has a continuous distribution under the null, and let $F(t)$ denote its tail cdf:

$$F(t) = \mathbb{P}(T > t)$$

- Define the P-value as $P = F(T)$
- The P-value has a uniform distribution under the null:

$$\mathbb{P}(P < p) = \mathbb{P}(F(T) < p) = \mathbb{P}(T > F^{-1}(p)) = F(F^{-1}(p)) = p$$

A Generic Decision Rule

- Reject H_i if the random variable T_i is equal to 1:

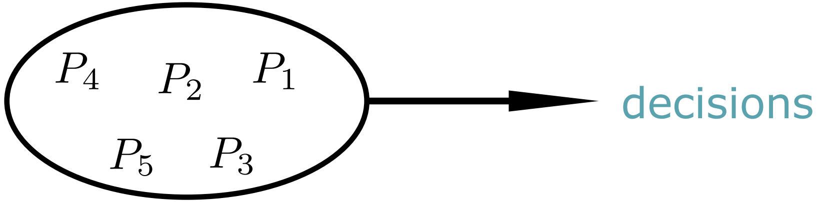
$$T_i = \begin{cases} 1, & \text{if } P_i \leq \alpha_i \\ 0, & \text{otherwise} \end{cases}$$

The Online Problem

- Classical statistics, and also the Benjamini & Hochberg algorithm focused on a batch setting in which all data has already been collected
- E.g., for Benjamini & Hochberg, you need all of the p-values before you can get started
- Is is possible to consider methods that make sequences of decisions, and provide FDR control at any moment in time
- Is it conceivable that one can achieve lifetime FDR control?

Online vs Offline FDR Control

- Classical FDR procedures (such as BH) which make all decisions simultaneously are called “offline”

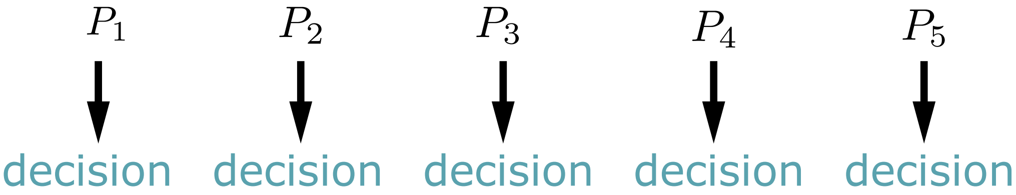


P_4 P_2 P_1

P_5 P_3

decisions

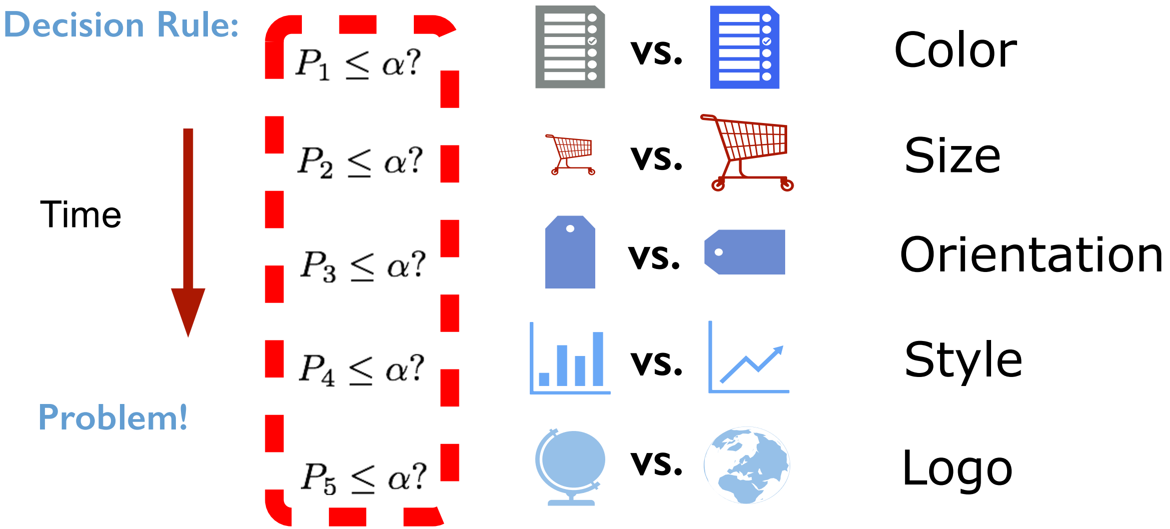
- “Online” FDR procedures make decisions one at a time



P_1 P_2 P_3 P_4 P_5

decision decision decision decision decision

Example: Many Enterprises Run Thousands of So-Called A/B Tests Each Day



Decision Rule:

$P_1 \leq \alpha?$ vs. Color

$P_2 \leq \alpha?$ vs. Size

$P_3 \leq \alpha?$ vs. Orientation

$P_4 \leq \alpha?$ vs. Style

$P_5 \leq \alpha?$ vs. Logo

Time

Problem!

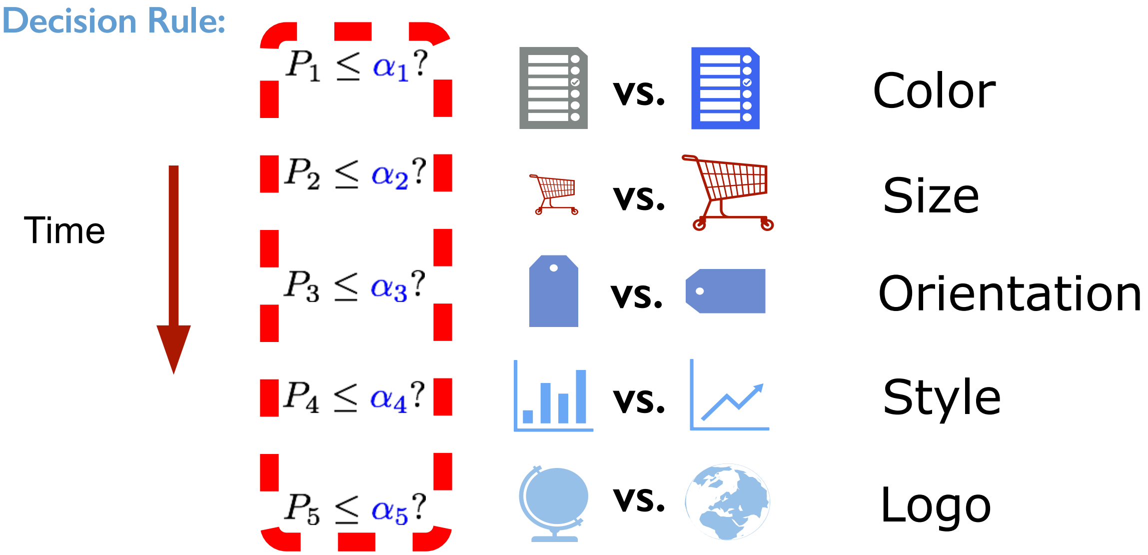
Challenges

- It’s not clear how to do change batch procedures such aws Benjamini-Hochberg procedure to be online

Challenges

- It’s not clear how to do change batch procedures such aws Benjamini-Hochberg procedure to be online
- We might retreat to Bonferroni, which would allow us to set α to $0.05/n$ and thereby have a FWER of 0.05 after n tests
  - but what do we do on the $(n+1)th$ test?
  - we eventually can’t do any more tests
  - we’ve used up our “alpha wealth”

A More General Approach: Time-Varying Alpha



Decision Rule:

Time

$P_1 \le \alpha_1?$ vs. Color

$P_2 \le \alpha_2?$ vs. Size

$P_3 \le \alpha_3?$ vs. Orientation

$P_4 \le \alpha_4?$ vs. Style

$P_5 \le \alpha_5?$ vs. Logo

More Challenges

- We want to keep going for an arbitrary amount of time, so we need $\sum_{t=1}^{\infty} \alpha_t = 1$, and $\sum_{t=1}^{T} \alpha_t < 1$ for any fixed T
- An example: $\alpha_t = 2^{-t}$
- But now we have less and less power to make discoveries over time, and eventually we may as well quit
- Is there any way out of this dilemma?

A Glimmer of Hope

- Recall that the FDP is a ratio of two counts
- We can make a ratio small in one of two ways:
  - make the numerator small
  - make the denominator big

A Glimmer of Hope

- Recall that the FDP is a ratio of two counts
- We can make a ratio small in one of two ways:
  - make the numerator small
  - make the denominator big
- The numerator has the false-positive rate in it, and so we're back to the same problem of controlling sums of α_i values

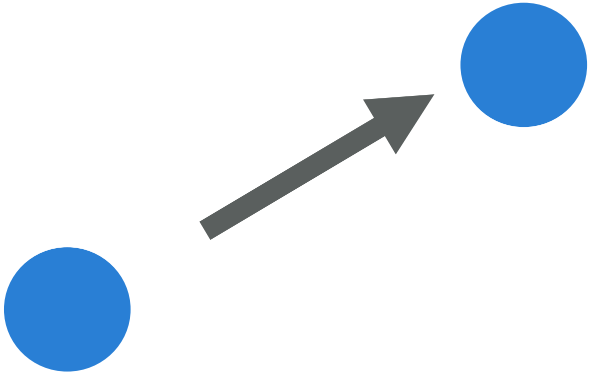
A Glimmer of Hope

- Recall that the FDP is a ratio of two counts
- We can make a ratio small in one of two ways:
  - make the numerator small
  - make the denominator big
- The numerator has the false-positive rate in it, and so we're back to the same problem of controlling sums of α_i values
- The denominator can be made large by making lots of discoveries

A Glimmer of Hope

- Recall that the FDP is a ratio of two counts
- We can make a ratio small in one of two ways:
  - make the numerator small
  - make the denominator big
- The numerator has the false-positive rate in it, and so we're back to the same problem of controlling sums of α_i values
- The denominator can be made large by making lots of discoveries
- Perhaps we can earn a bit of alpha whenever we make a discovery, to be invested and used for false discoveries later

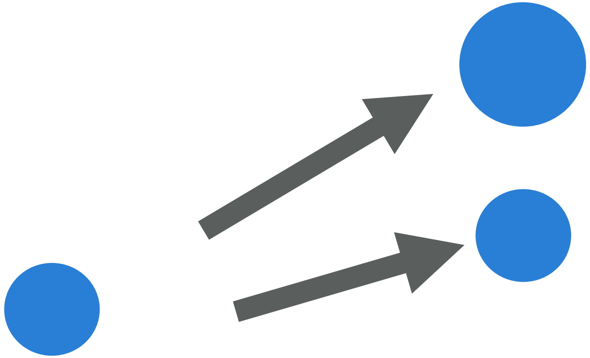
Online FDR Control : High-Level Picture



Error budget
for first test

Remaining error budget
or "alpha-wealth"

Online FDR Control : High-Level Picture

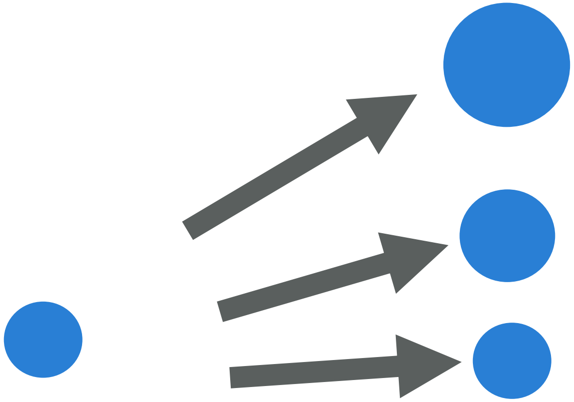


Error budget for first test

Error budget for second test

Remaining error budget or "alpha-wealth"

Online FDR Control : High-Level Picture



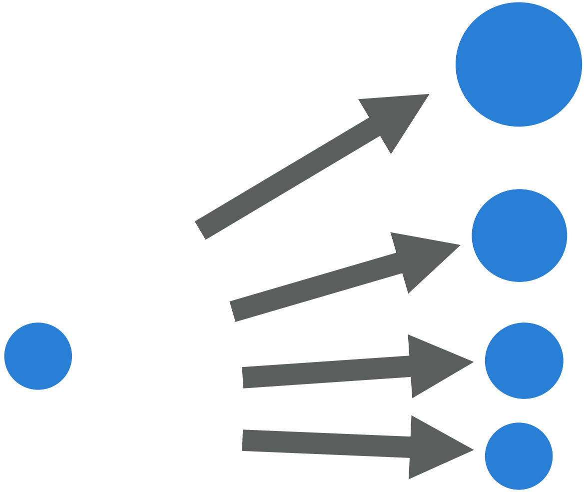
Error budget for first test

Error budget for second test

Tests use wealth

Remaining error budget or “alpha-wealth”

Online FDR Control : High-Level Picture



Error budget for first test

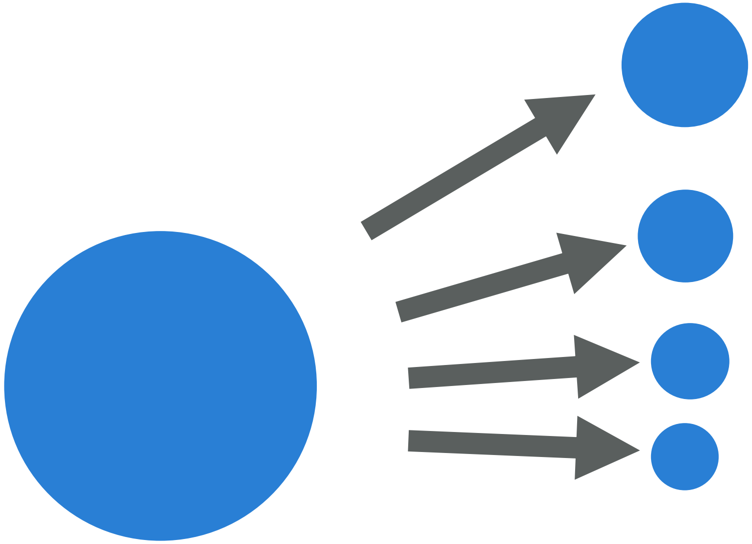
Error budget for second test

Tests use wealth

Discoveries earn wealth

Remaining error budget or "alpha-wealth"

Online FDR Control : High-Level Picture



Error budget
for first test

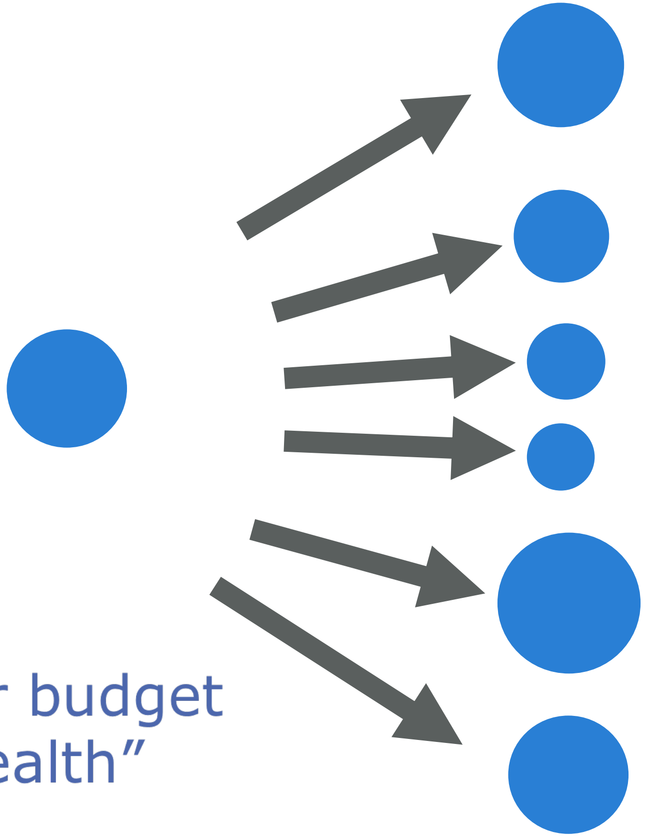
Error budget for
second test

Tests use wealth

**Discoveries
earn wealth**

Remaining error budget
or "alpha-wealth"

Online FDR Control : High-Level Picture



Remaining error budget
or "alpha-wealth"

Error budget
for first test

Error budget for
second test

Tests use wealth

Discoveries
earn wealth

Error budget
is data-dependent

Infinite process

Online FDR Algorithms

- The first online FDR algorithm was known as “alpha investing” and is due to Foster and Stine (2008)
- A more recent (and simpler) online FDR algorithm is due to Javanmard and Montanari, and is called “LORD”
- The basic idea is to assign α_t in a way that ensures

$$\widehat{\mathrm{FDP}}(t) := \frac{\sum_{i=1}^{t} \alpha_i}{\sum_{i=1}^{t} 1\{P_i \le \alpha_i\}} \le \alpha$$

Algorithm 1 The LORD Procedure

```
input: FDR level $\alpha$, non-increasing sequence $\{\gamma_t\}_{t=1}^{\infty}$ such that $\sum_{t=1}^{\infty} \gamma_t = 1$,
       initial wealth $W_0 \leq \alpha$
Set $\alpha_1 = \gamma_1 W_0$

for $t = 1, 2, \ldots$ do
    p-value $P_t$ arrives

    if $P_t \leq \alpha_t$, reject $P_t$

    $\alpha_{t+1} = \gamma_{t+1} W_0 + \gamma_{t+1-\tau_1}(\alpha - W_0)\mathbf{1}\{\tau_1 < t\} + \alpha \sum_{j=1}^{\infty} \gamma_{t+1-\tau_j} \mathbf{1}\{\tau_j < t\}$,

    where $\tau_j$ is time of $j$-th rejection $\tau_j = \min\{k : \sum_{l=1}^{k} \mathbf{1}\{P_l \leq \alpha_l\} = j\}$
end
```

A Stripped-Down Version of LORD

- Only consider the most recent rejection
- This renews the wealth, which further decays
- See description, and proof of mFDR control, on board

A Heuristic Argument for LORD’s Control of FDR

- We make an approximation:

$$\mathrm{FDR} \approx \frac{\mathbb{E}[\sum_{i \le t, i \text{ null}} 1\{P_i \le \alpha_i\}]}{\mathbb{E}[\sum_{i \le t} 1\{P_i \le \alpha_i\}]}$$

and then compute:

$$\mathbb{E}\left[\sum_{i \le t, i \text{ null}} 1\{P_i \le \alpha_i\}\right] = \sum_{i \le t, i \text{ null}} \mathbb{E}[\mathbb{E}[1\{P_i \le \alpha_i\}|\alpha_i]] = \sum_{i \le t, i \text{ null}} \mathbb{E}[\mathbb{P}\{P_i \le \alpha_i|\alpha_i\}]$$

$$= \sum_{i \le t, i \text{ null}} \mathbb{E}[\alpha_i] \le \mathbb{E}[\sum_{i \le t} \alpha_i] \le \alpha \mathbb{E}[\sum_{i \le t} 1\{P_i \le \alpha_i\}]$$

where the last line uses:

$$\widehat{\mathrm{FDP}}(t) := \frac{\sum_{i=1}^{t} \alpha_i}{\sum_{i=1}^{t} 1\{P_i \le \alpha_i\}} \le \alpha$$

- This establishes:

$$\mathrm{FDR} \le \alpha$$