Universal log-optimality of sequential hypothesis tests

Ian Waudby-Smith

Miller Institute & Department of Statistics
University of California, Berkeley

NUS IMS Young Mathematical Scientists Forum, 2025

Ricardo Sandoval

UC Berkeley

Michael I. Jordan

UC Berkeley & INRIA

Outline

1. What is sequential hypothesis testing?

2. How are sequential hypothesis tests derived?

3. ★ Defining and deriving **optimal** sequential tests.

Outline

1. **What is sequential hypothesis testing?**

2. How are sequential hypothesis tests derived?

3. ★ Defining and deriving **optimal** sequential tests.

A motivating example to keep in mind: **experiments**.

Question: Is the new treatment effective?

Question: Is the new treatment effective?

Step 1:

$$H_0 : \text{trt effect} = 0$$
$$H_1 : \text{trt effect} \neq 0$$

$$\alpha := 0.01$$

Step 2: Recruit n patients and randomize (**trt** or **ctrl**).

Question: Is the new treatment effective?

Step 1:

$$H_0 : \text{trt effect} = 0$$
$$H_1 : \text{trt effect} \neq 0$$

$$\alpha := 0.01$$

Step 2:

Recruit n patients and randomize (**trt** or **ctrl**).

Step 3:

Compute a p-value: $p_n = 0.023$

$\implies$ **Cannot reject** H_0 **at** $\alpha = 0.01$

Question: Is the new treatment effective?

Step 1:

$$H_0 : \text{trt effect} = 0$$
$$H_1 : \text{trt effect} \neq 0$$

$$\alpha := 0.01$$

Step 2:

Recruit n patients and randomize (**trt** or **ctrl**).

Step 3:

Compute a p-value: $p_n = 0.023$

$\implies$ **Cannot reject** H_0 **at** $\alpha = 0.01$

Step 4:

Recruit just a few more patients?

Question: Is the new treatment effective?

Step 1:

$$H_0 : \text{trt effect} = 0$$
$$H_1 : \text{trt effect} \neq 0$$

$$\alpha := 0.01$$

Step 2:

Recruit n patients and randomize (**trt** or **ctrl**).

Step 3:

Compute a p-value: $p_n = 0.023$

$\implies$ **Cannot reject** H_0 **at** $\alpha = 0.01$

Step 4:

Recruit just a few more patients?

No! (This is "p-hacking".)

Indeed, n could be insufficiently large to reject H_0.

Indeed, n could be insufficiently large to reject H_0.

On the flip side, even if n was large enough to reject H_0, it is possible that $n' \ll n$ could have sufficed.

$\rightarrow$ e.g. $p_n < 0.00001 \ll \alpha = 0.01$.

Indeed, n could be insufficiently large to reject H_0.

On the flip side, even if n was large enough to reject H_0, it is possible that $n' \ll n$ could have sufficed.

$\rightarrow$ e.g. $p_n < 0.00001 \ll \alpha = 0.01$.

Sequential testing ameliorates these unsettling possibilities.

What is sequential hypothesis testing?

What is sequential hypothesis testing?

(Foundations laid by Wald in the '40s, Robbins & co in the '60s-'70s)

What is sequential hypothesis testing?

(Foundations laid by Wald in the '40s, Robbins & co in the '60s-'70s)

(Modern breakthroughs by Ramdas, Grünwald, and others in 2010's onwards)

There is a composite null $\mathcal{P}$ and a composite alternative $\mathcal{Q}$.

$$\textit{(e.g. } \mathcal{P} = \{P : \textit{trt effect} = 0\} \textit{ versus } \mathcal{Q} = \{P : \textit{trt effect} > 0\}\textit{)}$$

We are tasked with finding a test $\phi_n^{(\alpha)} \equiv \phi^{(\alpha)}(X_1, \ldots, X_n)$ that outputs 1 (rejects $\mathcal{P}$ in favour of $\mathcal{Q}$) with small probability under $\mathcal{P}$.

There is a composite null $\mathcal{P}$ and a composite alternative $\mathcal{Q}$.

$$\textit{(e.g. } \mathcal{P} = \{P : \textit{trt effect} = 0\} \textit{ versus } \mathcal{Q} = \{P : \textit{trt effect} > 0\}\textit{)}$$

We are tasked with finding a test $\phi_n^{(\alpha)} \equiv \phi^{(\alpha)}(X_1, \ldots, X_n)$ that outputs 1 (rejects $\mathcal{P}$ in favour of $\mathcal{Q}$) with small probability under $\mathcal{P}$.

Fixed-n test: $\forall n \in \mathbb{N}, \ \sup_{P \in \mathcal{P}} P\left(\phi_n^{(\alpha)} \text{ rejects}\right) \leqslant \alpha.$

Sequential test: $\sup_{P \in \mathcal{P}} P\left(\exists n \in \mathbb{N} : \phi_n^{(\alpha)} \text{ rejects}\right) \leqslant \alpha.$

There is a composite null $\mathcal{P}$ and a composite alternative $\mathcal{Q}$.

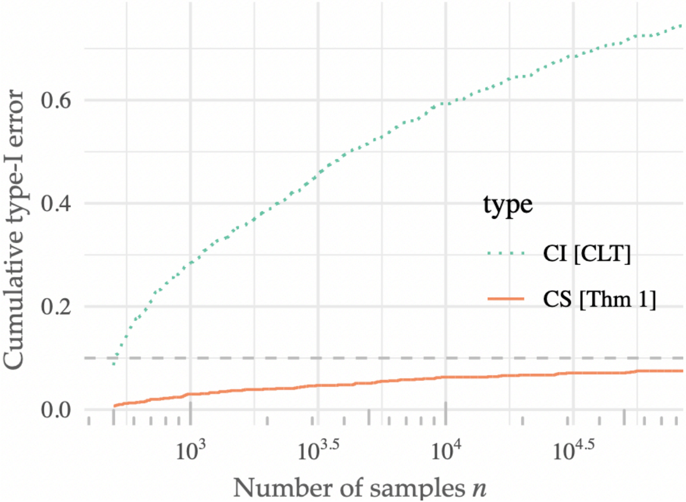
$$\textit{(e.g. } \mathcal{P} = \{P : \textit{trt effect} = 0\} \textit{ versus } \mathcal{Q} = \{P : \textit{trt effect} > 0\}\textit{)}$$

We are tasked with finding a test $\phi_n^{(\alpha)} \equiv \phi^{(\alpha)}(X_1, \ldots, X_n)$ that outputs 1 (rejects $\mathcal{P}$ in favour of $\mathcal{Q}$) with small probability under $\mathcal{P}$.

Fixed-n test: $\forall n \in \mathbb{N}, \ \sup_{P \in \mathcal{P}} P\left(\phi_n^{(\alpha)} \text{ rejects}\right) \leqslant \alpha.$

Sequential test: $\sup_{P \in \mathcal{P}} P\left(\exists n \in \mathbb{N} : \phi_n^{(\alpha)} \text{ rejects}\right) \leqslant \alpha.$

$$\iff \sup_{P \in \mathcal{P}} P\left(\phi_\tau^{(\alpha)} \text{ rejects}\right) \leqslant \alpha \quad \forall \tau.$$



$$\forall n, \ P(\phi_n^{(\alpha)} = 1) \leqslant \alpha$$

$$P(\exists n : \phi_n^{(\alpha)} = 1) \leqslant \alpha$$

Outline

1. What is sequential hypothesis testing?

2. **How are sequential hypothesis tests derived?**

3. ★ Defining and deriving **optimal** sequential tests.

Suppose data come in an online stream: $X_1, X_2, \cdots \sim P$, $P \in \mathcal{P} \cup \mathcal{Q}$.

Suppose data come in an online stream: $X_1, X_2, \dots \sim P$, $P \in \mathcal{P} \cup \mathcal{Q}$.

Sequential tests result from the following two-step procedure:

1. Derive a statistic $W_n \equiv W(X_1, \dots, X_n)$ that forms a nonnegative P-supermartingale with mean $\mathbb{E}_P[W_1] \leqslant 1$ for every $P \in \mathcal{P}$.

   We often call these **test supermartingales** or **e-processes**.
   (We saw these in YJ Choe's talk yesterday)

Suppose data come in an online stream: $X_1, X_2, \cdots \sim P$, $P \in \mathcal{P} \cup \mathcal{Q}$.

Sequential tests result from the following two-step procedure:

1. Derive a statistic $W_n \equiv W(X_1, \ldots, X_n)$ that forms a nonnegative P-supermartingale with mean $\mathbb{E}_P[W_1] \leqslant 1$ for every $P \in \mathcal{P}$.

   We often call these **test supermartingales** or **e-processes**.
   (We saw these in YJ Choe's talk yesterday)

$$\left(\textit{Example: likelihood ratios. } W_n = \prod_{i=1}^{n} \frac{\mathrm{d}Q(X_i)}{\mathrm{d}P(X_i)} \right)$$

Suppose data come in an online stream: $X_1, X_2, \cdots \sim P$, $P \in \mathcal{P} \cup \mathcal{Q}$.

Sequential tests result from the following two-step procedure:

1. Derive a statistic $W_n \equiv W(X_1, \ldots, X_n)$ that forms a nonnegative P-supermartingale with mean $\mathbb{E}_P[W_1] \leqslant 1$ for every $P \in \mathcal{P}$.

   We often call these **test supermartingales** or **e-processes**. *(We saw these in YJ Choe's talk yesterday)*

   $$\left(\textit{Example: likelihood ratios. } W_n = \prod_{i=1}^{n} \frac{\mathrm{d}Q(X_i)}{\mathrm{d}P(X_i)} \right)$$

2. Set the test as $\phi_n^{(\alpha)} := \mathbb{1}\{W_n \geqslant 1/\alpha\}$.

Claim: If $(W_n)_{n\in\mathbb{N}}$ is an e-process, then $\phi_n^{(\alpha)} := \mathbb{1}\{W_n \geqslant 1/\alpha\}$ yields a sequential test for the null $\mathcal{P}$.

Claim: If $(W_n)_{n\in\mathbb{N}}$ is an e-process, then $\phi_n^{(\alpha)} := \mathbb{1}\{W_n \geqslant 1/\alpha\}$ yields a sequential test for the null $\mathcal{P}$.

Proof.

$$\sup_{P\in\mathcal{P}} P\left(\exists n \in \mathbb{N} : \phi_n^{(\alpha)} \text{ rejects}\right) = \sup_{P\in\mathcal{P}} P\left(\exists n \in \mathbb{N} : W_n \geqslant 1/\alpha\right) \leqslant \alpha.$$

□

Claim: If $(W_n)_{n \in \mathbb{N}}$ is an e-process, then $\phi_n^{(\alpha)} := \mathbb{1}\{W_n \geqslant 1/\alpha\}$ yields a sequential test for the null $\mathcal{P}$.

Proof.

$$\sup_{P \in \mathcal{P}} P\left(\exists n \in \mathbb{N} : \phi_n^{(\alpha)} \text{ rejects}\right) = \sup_{P \in \mathcal{P}} P\left(\exists n \in \mathbb{N} : W_n \geqslant 1/\alpha\right) \leqslant \alpha.$$

□

The final inequality follows from Ville [1939] which states that for a nonnegative $\mathcal{P}$-supermartingale $(M_n)_{n \in \mathbb{N}}$,

$$\forall x > 0, \quad P\left(\sup_{n \in \mathbb{N}} M_n \geqslant x\right) \leqslant \frac{\mathbb{E}_P[M_1]}{x}.$$

So, does this mean it is "easy" to come up with sequential tests? **No**.

It is often challenging to devise nontrivial non-parametric e-processes.

So, does this mean it is “easy” to come up with sequential tests? **No**.

It is often challenging to devise nontrivial non-parametric e-processes.

However, lots of progress has been made in recent years.

Example: Testing the mean of a bounded random variables.

(Vovk & Shafer (2001), Hendriks (2018), ***W-S*** *& Ramdas (2023), Orabona & Jun (2023), others)*

Example: Testing the mean of a bounded random variables.

*(Vovk & Shafer (2001), Hendriks (2018), **W-S** & Ramdas (2023), Orabona & Jun (2023), others)*

Let $X_1, X_2, \dots \in [0,1]$. The null is $\mathcal{P} := \{P : \mathbb{E}_P[X_1] = 1/2\}$.

Example: Testing the mean of a bounded random variables.

*(Vovk & Shafer (2001), Hendriks (2018), **W-S** & Ramdas (2023), Orabona & Jun (2023), others)*

Let $X_1, X_2, \dots \in [0,1]$. The null is $\mathcal{P} := \{P : \mathbb{E}_P[X_1] = 1/2\}$. Then,

$$W_n := \prod_{i=1}^{n} (1 + \lambda_i \cdot (X_i - 1/2))$$

forms a test martingale for any $[-2,2]$-valued *predictable* $(\lambda_n)_{n\in\mathbb{N}}$.

(*Quick definition of "predictable"*: $\lambda_i \in \sigma(X_1, \dots, X_{i-1})$).

Example: Testing the mean of a bounded random variables.

(Vovk & Shafer (2001), Hendriks (2018), ***W-S*** *& Ramdas (2023), Orabona & Jun (2023), others)*

Let $X_1, X_2, \dots \in [0,1]$. The null is $\mathcal{P} := \{P : \mathbb{E}_P[X_1] = 1/2\}$. Then,

$$W_n := \prod_{i=1}^{n} (1 + \lambda_i \cdot (X_i - 1/2))$$

forms a test martingale for any $[-2,2]$-valued *predictable* $(\lambda_n)_{n \in \mathbb{N}}$.

(*Quick definition of "predictable"*: $\lambda_i \in \sigma(X_1, \dots, X_{i-1})$).

Therefore, $\phi_n^{(\alpha)} := \mathbb{1}\{W_n \geqslant 1/\alpha\}$ yields a sequential test for $\mathcal{P}$.

So, for $X_1, X_2, \dots \in [0,1]$ and $\mathcal{P} := \{P : \mathbb{E}_P[X_1] = 1/2\}$,

$$W_n := \prod_{i=1}^{n} (1 + \lambda_i \cdot (X_i - 1/2))$$

forms a test martingale (e-process) under $\mathcal{P}$.

So, for $X_1, X_2, \dots \in [0,1]$ and $\mathcal{P} := \{P : \mathbb{E}_P[X_1] = 1/2\}$,

$$W_n := \prod_{i=1}^{n} (1 + \lambda_i \cdot (X_i - 1/2))$$

forms a test martingale (e-process) under $\mathcal{P}$.

Many interesting nonparametric problems have a similar form.

★ We re-cast several testing problems *(bounded means, two-sample, independence, equality of bounded tuples, **testing randomness online**, etc.)* from the literature with the following unified test supermartingale:

$$W_n := \prod_{i=1}^{n} \left((1 - \lambda_i) E_i^{(1)} + \lambda_i E_i^{(2)} \right), \quad \textit{(write on board.)}$$

for some iid e-values $(E_n^{(1)})_{n \in \mathbb{N}}$ and $(E_n^{(2)})_{n \in \mathbb{N}}$ where $(\lambda_n)_{n \in \mathbb{N}}$ is any $[0, 1]$-valued predictable sequence.

★ We re-cast several testing problems *(bounded means, two-sample, independence, equality of bounded tuples,* ***testing randomness online****, etc.)* from the literature with the following unified test supermartingale:

$$W_n := \prod_{i=1}^{n} \left((1 - \lambda_i) E_i^{(1)} + \lambda_i E_i^{(2)} \right), \quad \textit{(write on board.)}$$

for some iid e-values $(E_n^{(1)})_{n \in \mathbb{N}}$ and $(E_n^{(2)})_{n \in \mathbb{N}}$ where $(\lambda_n)_{n \in \mathbb{N}}$ is any $[0, 1]$-valued predictable sequence.

Definition: e-value. (See Shafer & Vovk, Grünwald et al., Vovk & Wang)

A nonnegative random variable E is said to be an e-value under P if

$$\mathbb{E}_P[E] \leqslant 1.$$

★ Some special cases found in the literature

One-sided bounded mean testing: Set $E_i^{(1)} = 1$ and $E_i^{(2)} = X_i/\mu_0$.
Two-sided bounded mean testing: Set $E_i^{(1)} = (1-X_i)/(1-\mu_0)$ and $E_i^{(2)} = X_i/\mu_0$.
Two-sample testing: Set $E_i^{(1)} = 1$ and $E_i^{(2)} = g^\star(X_i) - g^\star(Y_i)$ for a witness f'n $g^\star$.
Testing randomness: Set $E_i^{(1)} = 2(1 - s_i)$ and $E_i^{(2)} = 2s_i$ for a conformal score s_i.

★ Some special cases found in the literature

One-sided bounded mean testing: Set $E_i^{(1)} = 1$ and $E_i^{(2)} = X_i/\mu_0$.
Two-sided bounded mean testing: Set $E_i^{(1)} = (1-X_i)/(1-\mu_0)$ and $E_i^{(2)} = X_i/\mu_0$.
Two-sample testing: Set $E_i^{(1)} = 1$ and $E_i^{(2)} = g^\star(X_i) - g^\star(Y_i)$ for a witness f'n $g^\star$.
Testing randomness: Set $E_i^{(1)} = 2(1 - s_i)$ and $E_i^{(2)} = 2s_i$ for a conformal score s_i.

There has been one lingering question this entire discussion:

How should one choose $(\lambda_n)_{n\in\mathbb{N}}$?

Outline

1. What is sequential hypothesis testing?

2. How are sequential hypothesis tests derived?

3. **⋆ Defining and deriving optimal sequential tests.**

"Power" in the traditional sense is not the right concept for sequential tests.

"Power" in the traditional sense is not the right concept for sequential tests.

There are two common power desiderata for sequential tests:

(i) **Growth-rate optimality**

"Power" in the traditional sense is not the right concept for sequential tests.

There are two common power desiderata for sequential tests:

(i) **Growth-rate optimality**

(ii) **Small expected rejection times**

"Power" in the traditional sense is not the right concept for sequential tests.

There are two common power desiderata for sequential tests:

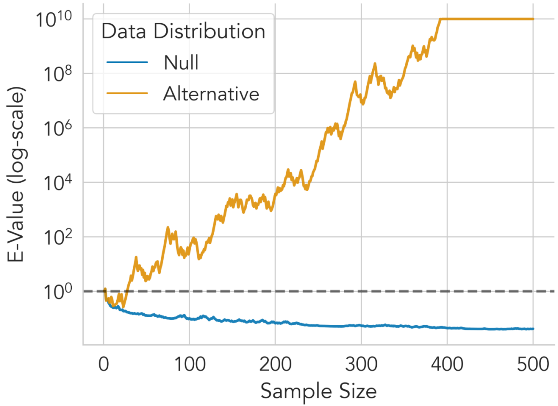
(i) **Growth-rate optimality**

(ii) **Small expected rejection times**

★ We show that these are optimized via the same criterion and derive matching lower and upper bounds for both.

(i) Growth-rate-optimality.

(Kelly ['56], Long Jr. ['90], Grünwald et al. [2024], Larsson et al. [2024])



An e-process is expected to be small under the ***null****;*
we want it to grow large under the ***alternative****.*

Image credit: YJ Choe.

Since we *reject* the null $\mathcal{P}$ when $W_n \geqslant 1/\alpha$, we aim to choose $(\lambda_n)_{n \in \mathbb{N}}$ so that W_n diverges "quickly".

Since we *reject* the null $\mathcal{P}$ when $W_n \geqslant 1/\alpha$, we aim to choose $(\lambda_n)_{n\in\mathbb{N}}$ so that W_n diverges "quickly".

Observe by the strong law of large numbers:

$$
\begin{aligned}
W_n &= \exp\left\{n \cdot \frac{1}{n}\sum_{i=1}^{n}\log((1-\lambda_i)E_i^{(1)} + \lambda_i E_i^{(2)})\right\} \\
&\approx \exp\left\{n \cdot \mathbb{E}_Q[\log((1-\lambda)E^{(1)} + \lambda E^{(2)})]\right\} \\
&= \exp\left\{n \cdot \ell_Q(\lambda)\right\},
\end{aligned}
$$

where $\ell_Q(\lambda) := \mathbb{E}_Q\left[\log\left((1-\lambda)E^{(1)} + \lambda E^{(2)}\right)\right]$. *(write on board).*

Since we *reject* the null $\mathcal{P}$ when $W_n \geqslant 1/\alpha$, we aim to choose $(\lambda_n)_{n\in\mathbb{N}}$ so that W_n diverges "quickly".

Observe by the strong law of large numbers:

$$
\begin{aligned}
W_n &= \exp\left\{n \cdot \frac{1}{n}\sum_{i=1}^{n} \log((1-\lambda_i)E_i^{(1)} + \lambda_i E_i^{(2)})\right\} \\
&\approx \exp\left\{n \cdot \mathbb{E}_Q[\log((1-\lambda)E^{(1)} + \lambda E^{(2)})]\right\} \\
&= \exp\left\{n \cdot \ell_Q(\lambda)\right\},
\end{aligned}
$$

where $\ell_Q(\lambda) := \mathbb{E}_Q\left[\log\left((1-\lambda)E^{(1)} + \lambda E^{(2)}\right)\right]$. *(write on board).*

So, should we just maximize $\ell_Q(\lambda)$ over $\lambda \in [0,1]$?

This is the famous "Kelly criterion" from gambling / info. theory.

A New Interpretation of Information Rate

By J. L. KELLY, JR.

(Manuscript received March 21, 1956)

$$\text{“Kelly bet”:} \quad \lambda_Q^\star := \operatorname*{argmax}_{\lambda\in[0,1]} \ell_Q(\lambda).$$

Another justification of Kelly betting *(Long Jr. '90)*:

Let W'_n be *any* process built from predictable $(\lambda_n)_{n\in\mathbb{N}}$. Then for all n sufficiently large,

$$W_n(\lambda^\star_Q) \geqslant W'_n \quad Q\text{-almost surely.}$$

Another justification of Kelly betting *(Long Jr. '90)*:

Let W'_n be *any* process built from predictable $(\lambda_n)_{n\in\mathbb{N}}$. Then for all n sufficiently large,

$$W_n(\lambda^\star_Q) \geqslant W'_n \quad Q\text{-almost surely.}$$

However, Q is unknown, so $\lambda^\star_Q$ is unknown.

Another justification of Kelly betting *(Long Jr. '90)*:

Let W'_n be *any* process built from predictable $(\lambda_n)_{n\in\mathbb{N}}$. Then for all n sufficiently large,

$$W_n(\lambda^\star_Q) \geqslant W'_n \quad Q\text{-almost surely.}$$

However, Q is unknown, so $\lambda^\star_Q$ is unknown.

Question: Can we choose $(\lambda_n)_{n\in\mathbb{N}}$ so that W_n adaptively behaves like $W_n(\lambda^\star_Q)$ *regardless of* $Q \in \mathcal{Q}$?

Answer: Yes. We call this $\mathcal{Q}$-universal log-optimality.

★ **Definition:** Universal, asymptotic, almost-sure log-optimality.

We say that a process $W_n^\star$ is <u>$\mathcal{Q}$-universally log-optimal</u> if for any other W_n' and for any $Q \in \mathcal{Q}$,

$$\liminf_{n\to\infty} \left(\frac{1}{n} \log(W_n^\star) - \frac{1}{n} \log(W_n') \right) \geqslant 0 \quad Q\text{-almost surely.}$$

★ **Definition:** Universal, asymptotic, almost-sure log-optimality.

We say that a process $W_n^\star$ is $\mathcal{Q}$-universally log-optimal if for any other W_n' and for any $Q \in \mathcal{Q}$,

$$\liminf_{n\to\infty} \left(\frac{1}{n}\log(W_n^\star) - \frac{1}{n}\log(W_n') \right) \geqslant 0 \quad Q\text{-almost surely.}$$

What property leads to $\mathcal{Q}$-universal log-optimality?

Sublinear *portfolio* regret.

★ **Definition:** Portfolio regret.

We define the portfolio regret $\mathcal{R}_n$ of an e-process W_n to be

$$\mathcal{R}_n := \max_{\lambda \in [0,1]} \sum_{i=1}^{n} \log\left((1-\lambda)E_i^{(1)} + \lambda E_i^{(2)}\right) - \log W_n.$$

★ **Definition:** Portfolio regret.

We define the portfolio regret $\mathcal{R}_n$ of an e-process W_n to be

$$\mathcal{R}_n := \max_{\lambda \in [0,1]} \sum_{i=1}^{n} \log \left((1-\lambda) E_i^{(1)} + \lambda E_i^{(2)} \right) - \log W_n.$$

This is *precisely* the notion of regret considered by Thomas Cover and Erik Ordentlich in their work on on universal portfolios circa 1990s.

The following theorem: "Portfolio regret $\implies$ $\mathcal{Q}$-universal log-optimality".

★ **Theorem:** Universal log-optimality via sublinear portfolio regret.

Suppose that $(\lambda_n)_{n=1}^{\infty}$ is chosen so that $\mathcal{R}_n$ is (pathwise) sublinear:

$$\mathcal{R}_n \equiv \max_{\lambda \in [0,1]} \sum_{i=1}^{n} \log\left((1-\lambda)E_i^{(1)} + \lambda E_i^{(2)}\right) - \log W_n = o(n)$$

pathwise. Then W_n is $\mathcal{Q}$-universally log-optimal.

★ **Theorem:** Universal log-optimality via sublinear portfolio regret.

Suppose that $(\lambda_n)_{n=1}^{\infty}$ is chosen so that $\mathcal{R}_n$ is (pathwise) sublinear:

$$\mathcal{R}_n \equiv \max_{\lambda \in [0,1]} \sum_{i=1}^{n} \log\left((1-\lambda)E_i^{(1)} + \lambda E_i^{(2)}\right) - \log W_n = o(n)$$

pathwise. Then W_n is $\mathcal{Q}$-universally log-optimal. Moreover, for any $Q \in \mathcal{Q}$,

$$\lim_{n\to\infty} \frac{1}{n} \log W_n = \max_{\lambda \in [0,1]} \ell_Q(\lambda) \quad Q\text{-almost surely.}$$

A natural question: When is sublinear portfolio regret attainable?

A natural question: When is sublinear portfolio regret attainable?

A perhaps surprising answer: **always**. *(Cover & Ordentlich [1996])*.

Define λ_n^{UP} as

$$\lambda_n^{\text{UP}} := \frac{\int_{\lambda\in[0,1]} \lambda W_{n-1}(\lambda) dF(\lambda)}{\int_{\lambda\in[0,1]} W_{n-1}(\lambda) dF(\lambda)}.$$

Define λ_n^{UP} as

$$\lambda_n^{\mathrm{UP}} := \frac{\int_{\lambda \in [0,1]} \lambda W_{n-1}(\lambda) dF(\lambda)}{\int_{\lambda \in [0,1]} W_{n-1}(\lambda) dF(\lambda)}.$$

If $F(\lambda)$ is taken to be Beta$(1/2,\ 1/2)$, then $W_n(\lambda_1^{\mathrm{UP}}, \ldots, \lambda_n^{\mathrm{UP}})$ enjoys *logarithmic* portfolio regret:

$$\mathcal{R}_n \leqslant \log(n+1)/2 + \log 2.$$

Define λ_n^{UP} as

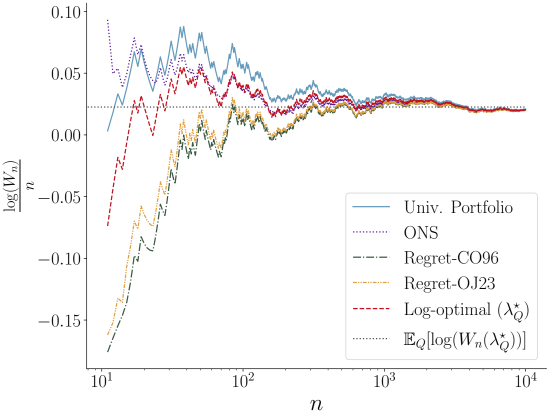
$$\lambda_n^{\mathrm{UP}} := \frac{\int_{\lambda \in [0,1]} \lambda W_{n-1}(\lambda) dF(\lambda)}{\int_{\lambda \in [0,1]} W_{n-1}(\lambda) dF(\lambda)}.$$

If $F(\lambda)$ is taken to be Beta$(1/2, 1/2)$, then $W_n(\lambda_1^{\mathrm{UP}}, \ldots, \lambda_n^{\mathrm{UP}})$ enjoys *logarithmic* portfolio regret:

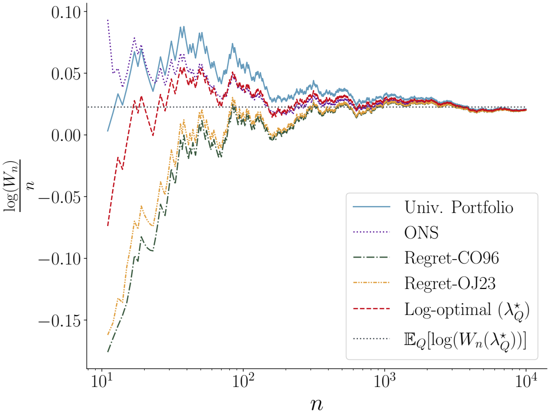
$$\mathcal{R}_n \leqslant \log(n+1)/2 + \log 2.$$

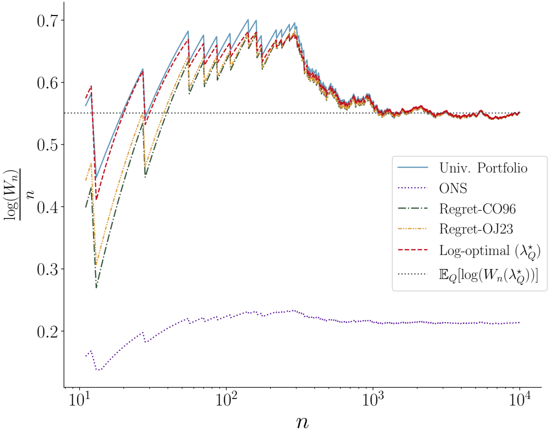
Universal portfolio has been used in some sequential <u>*estimation*</u> *problems by Orabona & Jun [2023], Ryu & Bhatt [2024], Shekhar & Ramdas [2024], and some others, but without proofs of log-optimality.*

Prior state-of-the-art fails to have log-optimal growth rates in general.



Prior state-of-the-art fails to have log-optimal growth rates in general.





Reminder: there were *two* common "power" desiderata in sequential testing:

(i) ~~**Growth-rate optimality**~~ ✓

(ii) **Small expected rejection times**

(ii) Measuring optimality through expected rejection times

(Wald 1945, Breiman 1961, Kaufmann, Agrawal, Koolen, others from the BAI literature)

Recall the "unifying" e-process:

$$W_n := \prod_{i=1}^{n} \left((1 - \lambda_i) E_i^{(1)} + \lambda_i E_i^{(2)} \right).$$

Define the first time at which we can reject the null $\mathcal{P}$ at the level $\alpha \in (0, 1)$:

$$\tau_\alpha := \inf \left\{ n \in \mathbb{N} : W_n \geqslant \frac{1}{\alpha} \right\}.$$

Recall the "unifying" e-process:

$$W_n := \prod_{i=1}^{n} \left((1 - \lambda_i) E_i^{(1)} + \lambda_i E_i^{(2)} \right).$$

Define the first time at which we can reject the null $\mathcal{P}$ at the level $\alpha \in (0, 1)$:

$$\tau_\alpha := \inf \left\{ n \in \mathbb{N} : W_n \geqslant \frac{1}{\alpha} \right\}.$$

Since τ_α is a random variable, let us study its (normalized) Q-expectation

$$\frac{\mathbb{E}_Q[\tau_\alpha]}{\log(1/\alpha)}.$$

★ Theorem: Lower bound on the expected rejection time

For any predictable $(\lambda_n)_{n\in\mathbb{N}}$, any $Q \in \mathcal{Q}$, and any $\alpha \in (0,1)$, it holds that

$$\frac{\mathbb{E}_Q[\tau_\alpha]}{\log(1/\alpha)} \geqslant \frac{1}{\max_{\lambda\in[0,1]} \ell_Q(\lambda)}$$

★ Theorem: Lower bound on the expected rejection time

For any predictable $(\lambda_n)_{n\in\mathbb{N}}$, any $Q \in \mathcal{Q}$, and any $\alpha \in (0,1)$, it holds that

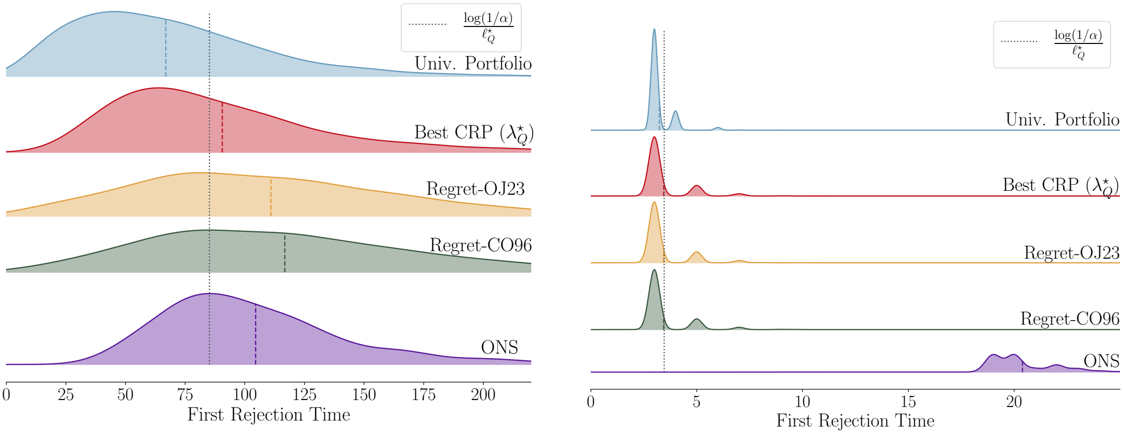
$$\frac{\mathbb{E}_Q[\tau_\alpha]}{\log(1/\alpha)} \geqslant \frac{1}{\max_{\lambda\in[0,1]} \ell_Q(\lambda)}$$

★ Theorem: A matching upper bound for small α

If $(\lambda_n)_{n\in\mathbb{N}}$ is chosen to have sublinear portfolio regret (e.g. UP),

$$\lim_{\alpha\to 0^+} \frac{\mathbb{E}_Q[\tau_\alpha]}{\log(1/\alpha)} \overset{(=)}{\leqslant} \frac{1}{\max_{\lambda\in[0,1]} \ell_Q(\lambda)}$$

Prior state-of-the-art fails to have optimal expected rejection times.



⋆ Summary of results

Given an e-process $(W_n)_{n\in\mathbb{N}}$ for $\mathcal{P}$ of the form

$$W_n := \prod_{i=1}^{n} \left((1 - \lambda_i) E_i^{(1)} + \lambda_i E_i^{(2)} \right),$$

★ Summary of results

Given an e-process $(W_n)_{n\in\mathbb{N}}$ for $\mathcal{P}$ of the form

$$W_n := \prod_{i=1}^{n} \left((1-\lambda_i)E_i^{(1)} + \lambda_i E_i^{(2)} \right),$$

if $(\lambda_n)_{n\in\mathbb{N}}$ are chosen to have sublinear portfolio regret (e.g. via Cover's universal portfolio algorithm), then for any $Q \in \mathcal{Q}$,

$$\lim_{n\to\infty} \frac{1}{n} \log W_n = \max_{\lambda\in[0,1]} \ell_Q(\lambda) \quad Q\text{-almost surely},$$

★ Summary of results

Given an e-process $(W_n)_{n\in\mathbb{N}}$ for $\mathcal{P}$ of the form

$$W_n := \prod_{i=1}^{n} \left((1-\lambda_i) E_i^{(1)} + \lambda_i E_i^{(2)} \right),$$

if $(\lambda_n)_{n\in\mathbb{N}}$ are chosen to have sublinear portfolio regret (e.g. via Cover's universal portfolio algorithm), then for any $Q \in \mathcal{Q}$,

$$\lim_{n\to\infty} \frac{1}{n} \log W_n = \max_{\lambda\in[0,1]} \ell_Q(\lambda) \quad Q\text{-almost surely,}$$

and

$$\lim_{\alpha\to 0^+} \frac{\mathbb{E}_Q[\tau_\alpha]}{\log(1/\alpha)} = \frac{1}{\max_{\lambda\in[0,1]} \ell_Q(\lambda)}.$$

Thank you

ianws.com