Sequential Anytime-Valid Inference (SAVI) using E-Processes

(Part II)

Xiaoqi (Shirley) Liu

28 October 2025

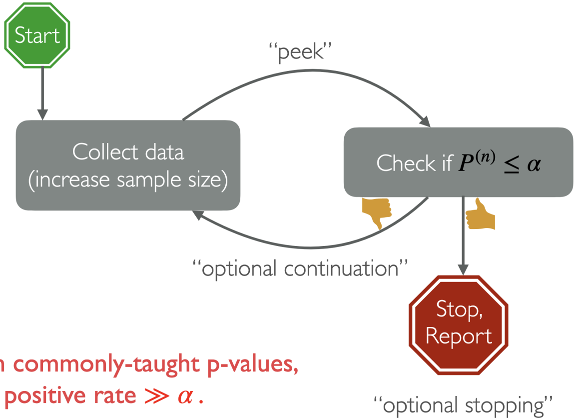
Overview

Outline

Outline

Motivation for SAVI (recap)

What is the problem with continuous monitoring?



With commonly-taught p-values, false positive rate $\gg \alpha$.

Setup (recap)

Setup (recap)

Natural filtration: for $t \in \{0, 1, \ldots\}$, $\mathcal{F}_t = \sigma(X_1, \ldots, X_t)$, $\mathcal{F}_0 = \{\varnothing, \Omega\}$, $\mathcal{F} := (\mathcal{F}_t)_{t \geq 0}$

Process: sequence of r.v.s $Y = (Y_t)_{t \geq 0}$ adapted to $\mathcal{F}$, i.e., every Y_t is $\mathcal{F}_t$-measurable.

Y_t is predictable if Y_t is $\mathcal{F}_{t-1}$-measurable $(t \geq 1)$

Stopping time (or rule) τ: nonnegative integer-valued r.v. s.t.$\{\tau \leq t\} \in \mathcal{F}_t$ for each $t \geq 0$.

Level–α sequential test for $\mathcal{P}$: a binary process $\phi = (\phi_t)_{t \geq 1}$ with

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\big(\exists t \geq 1 : \phi_t = 1\big) \leq \alpha$$

Test supermartingales & e-processes (recap)

Test supermartingales & e-processes (recap)

Test supermartingale for $\mathcal{P}$. For every $\mathbb{P} \in \mathcal{P}$, $M = (M_t)_{t \geq 0}$ satisfies

- $M_t \geq 0$ $\mathbb{P}$-a.s.,
- $\mathbb{E}^{\mathbb{P}}[M_t \mid \mathcal{F}_{t-1}] \leq M_{t-1}$, and
- $\mathbb{E}^{\mathbb{P}}[M_0] \leq 1$.

Test supermartingales & e-processes (recap)

Test supermartingale for $\mathcal{P}$. For every $\mathbb{P} \in \mathcal{P}$, $M = (M_t)_{t \geq 0}$ satisfies

- $M_t \geq 0$ $\mathbb{P}$-a.s.,
- $\mathbb{E}^{\mathbb{P}}[M_t \mid \mathcal{F}_{t-1}] \leq M_{t-1}$, and
- $\mathbb{E}^{\mathbb{P}}[M_0] \leq 1$.

E-process for $\mathcal{P}$. A sequence of e-values $E = (E_t)_{t \geq 0}$ adapted to $\mathcal{F}$ and either (equivalently):

- $\mathbb{E}^{\mathbb{P}}[E_\tau] \leq 1$ for every stopping time τ and $\mathbb{P} \in \mathcal{P}$; or
- $\exists$ a test supermartingale family $(M^{\mathbb{P}})_{\mathbb{P} \in \mathcal{P}}$ with $E_t \leq M_t^{\mathbb{P}}$, $\mathbb{P}$-a.s. for every $t \geq 0$ and $\mathbb{P} \in \mathcal{P}$.

Outline

Optional stopping & Ville's inequality (recap)

Optional stopping. If M is a test supermartingale for $\mathbb{P}$, then for any stopping time τ and $\sigma \leq \tau$:

$$\mathbb{E}^{\mathbb{P}}[M_\tau \mid \mathcal{F}_\sigma] \leq M_\sigma \implies \mathbb{E}^{\mathbb{P}}[M_\tau] \leq \mathbb{E}^{\mathbb{P}}[M_0] \leq 1.$$

Optional stopping & Ville's inequality (recap)

Optional stopping. If M is a test supermartingale for $\mathbb{P}$, then for any stopping time τ and $\sigma \leq \tau$:

$$\mathbb{E}^{\mathbb{P}}[M_\tau \mid \mathcal{F}_\sigma] \leq M_\sigma \implies \mathbb{E}^{\mathbb{P}}[M_\tau] \leq \mathbb{E}^{\mathbb{P}}[M_0] \leq 1.$$

Ville's inequality. If M is a non-negative supermartingale for $\mathbb{P}$, then for any $x > 0$,

$$\mathbb{P}(\exists t \in \mathbb{N} : M_t \geq x) \leq \frac{\mathbb{E}^{\mathbb{P}}[M_0]}{x}$$

Optional stopping & Ville's inequality (recap)

Optional stopping. If M is a test supermartingale for $\mathbb{P}$, then for any stopping time τ and $\sigma \leq \tau$:

$$\mathbb{E}^{\mathbb{P}}[M_\tau \mid \mathcal{F}_\sigma] \leq M_\sigma \implies \mathbb{E}^{\mathbb{P}}[M_\tau] \leq \mathbb{E}^{\mathbb{P}}[M_0] \leq 1.$$

Ville's inequality. If M is an e-process for $\mathcal{P}$, then for any $\alpha \in (0, 1]$,

$$\mathbb{P}\left(\exists t \in \mathbb{N} : M_t \geq \frac{1}{\alpha}\right) \leq \alpha$$

Optional stopping & Ville's inequality (recap)

Optional stopping. If M is a test supermartingale for $\mathbb{P}$, then for any stopping time τ and $\sigma \leq \tau$:

$$\mathbb{E}^{\mathbb{P}}[M_\tau \mid \mathcal{F}_\sigma] \leq M_\sigma \implies \mathbb{E}^{\mathbb{P}}[M_\tau] \leq \mathbb{E}^{\mathbb{P}}[M_0] \leq 1.$$

Ville's inequality. If M is an e-process for $\mathcal{P}$, then for any $\alpha \in (0, 1]$,

$$\mathbb{P}\left(\exists t \in \mathbb{N} : M_t \geq \frac{1}{\alpha}\right) \leq \alpha$$

Proposition

Every level-α sequential test for $\mathcal{P}$ is a e-process for $\mathcal{P}$ thresholded at $1/\alpha$.

Optional stopping & Ville's inequality (recap)

Optional stopping. If M is a test supermartingale for $\mathbb{P}$, then for any stopping time τ and $\sigma \leq \tau$:

$$\mathbb{E}^{\mathbb{P}}[M_\tau \mid \mathcal{F}_\sigma] \leq M_\sigma \implies \mathbb{E}^{\mathbb{P}}[M_\tau] \leq \mathbb{E}^{\mathbb{P}}[M_0] \leq 1.$$

Ville's inequality. If M is an e-process for $\mathcal{P}$, then for any $\alpha \in (0, 1]$,

$$\mathbb{P}\left(\exists t \in \mathbb{N} : M_t \geq \frac{1}{\alpha}\right) \leq \alpha$$

Proposition

Every level-α sequential test for $\mathcal{P}$ is a e-process for $\mathcal{P}$ thresholded at $1/\alpha$.

Optional continuation. If M^A, M^B are e-processes for $\mathcal{P}$ and $M^B_t = 1$ on $\{t \leq \tau\}$, then $M = (M_t)$ with $M_t := M^A_{\tau \wedge t} M^B_t$ is an e-process for $\mathcal{P}$.

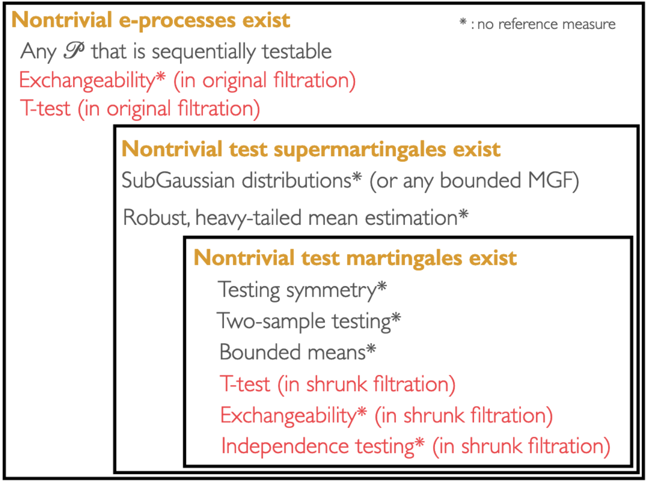


Figure: Hierarchy of tools for constructing sequential tests: nontrivial e-processes (outermost), test supermartingales (middle), and test martingales (innermost), with examples of what each can test.

Outline

Log-optimality (testing $\mathbb{P}$ vs $\mathbb{Q}$) (recap)

The LR process M^* given by $M_0^* = 1$ and $M_t^* = \frac{\mathrm{d}\mathbb{Q}|_{\mathcal{F}_t}}{\mathrm{d}\mathbb{P}|_{\mathcal{F}_t}}(X_1, \ldots, X_t)$ is a test martingale for $\mathbb{P}$.

Theorem (Log-optimality)

For any stopping time τ that is finite $\mathbb{Q}$-a.s. and any e-process M for $\mathbb{P}$:

$$\mathbb{E}^{\mathbb{Q}}[\log M_\tau^*] \;\geq\; \mathbb{E}^{\mathbb{Q}}[\log M_\tau].$$

$\implies$ LR processes are useful benchmark for testing against composite alternatives.

Asymptotic log-optimality (testing $\mathbb{P}$ vs $\mathcal{Q}$) (recap)

Definition

An e-process M is asymptotically log-optimal for $\mathbb{P}$ against $\mathcal{Q}$ if for every $\mathbb{Q} \in \mathcal{Q}$,

$$\lim_{t\to\infty} \frac{1}{t}\left(\log M_t - \log M_t^{\mathbb{Q}}\right) \geq 0 \quad \text{in } L^1\text{-convergence under } \mathbb{Q}$$

where $M^{\mathbb{Q}}$ is the LR process of $\mathbb{Q}$ to $\mathbb{P}$.

$\Longrightarrow$ covers any e-process M that grows an $e^{o(t)}$ factor slower than $M^{\mathbb{Q}}$.

Asymptotic log-optimality (testing $\mathbb{P}$ vs $\mathcal{Q}$) (recap)

Definition

An e-process M is asymptotically log-optimal for $\mathbb{P}$ against $\mathcal{Q}$ if for every $\mathbb{Q} \in \mathcal{Q}$,

$$\lim_{t\to\infty} \frac{1}{t}\left(\log M_t - \log M_t^{\mathbb{Q}}\right) \geq 0 \quad \text{in } L^1\text{-convergence under } \mathbb{Q}$$

where $M^{\mathbb{Q}}$ is the LR process of $\mathbb{Q}$ to $\mathbb{P}$.

$\implies$ covers any e-process M that grows an $e^{o(t)}$ factor slower than $M^{\mathbb{Q}}$.

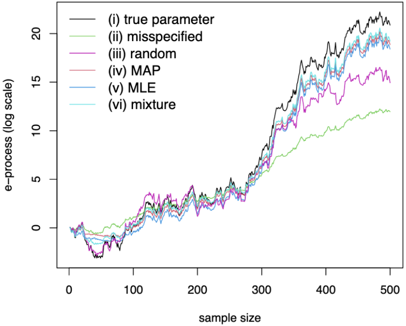
Testing $\mathbb{P}$ against $\mathcal{Q} = \{\mathbb{Q}_\theta : \theta \in \Theta_1\}$ with iid data: $M_0 = 1$, and with predictable $\hat{\theta}_{i-1}$, use

- **Plug-in LR:** $M_t = \prod_{i=1}^t \frac{q_{\hat{\theta}_{i-1}}(X_i)}{p(X_i)}$, or
- **Mixture LR:** $M_t = \int_{\Theta_1} \prod_{i=1}^t \frac{q_\theta(X_i)}{p(X_i)} \nu(\mathrm{d}\theta)$, $\quad \nu$ a prior on Θ_1.

Plug-in is asymptotically log-optimal when $\theta_i \to \theta$ under $\mathbb{Q}_\theta$ in a suitable sense, given log-LR is concave, score function has bounded variance.

Example: Given iid data from $N(\theta^\dagger, 1)$, goal is to test $H_0 : \theta^\dagger = 0$ vs $H_1 : \theta^\dagger > 0$. For illustration, take $\theta^\dagger = 0.3$.

Example: Given iid data from $N(\theta^\dagger, 1)$, goal is to test $H_0 : \theta^\dagger = 0$ vs $H_1 : \theta^\dagger > 0$. For illustration, take $\theta^\dagger = 0.3$.



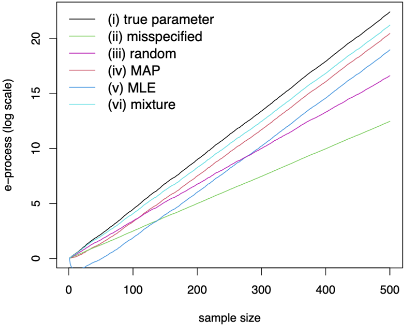


Figure: few ways of constructing e-values from LR processes. Left: one run. Right: average of 1000 runs.

(i) true parameter: choose $\theta_i = \theta^\dagger = 0.3$
(ii) misspecified: choose $\theta_i = 0.1$
(iii) random: take iid θ_i from $U[0, 0.5]$
(iv) MAP: choose θ_i by the MAP estimator with prior $\theta \sim N(0.1, 0.2^2)$
(v) MLE: choose θ_i with $\theta_1 := 0.1$ and θ_i the MLE of θ based on $X_1, \ldots, X_{i-1}$
(vi) mixture: compute a mixture of M each with a fixed $\theta_i \in [0, 0.6]$, uniformly weighted via a discrete grid

Outline

Outline

Universal Inference (UI) e-variable & e-process

Goal: build an e-variable (or an e-process) to test a composite null $\mathcal{P}$ against composite alternative $\mathcal{Q}$ under no regularity conditions.
Recipe: a reduction from hypothesis testing to MLE under $\mathcal{P}$.

Universal Inference (UI) e-variable & e-process

Goal: build an e-variable (or an e-process) to test a composite null $\mathcal{P}$ against composite alternative $\mathcal{Q}$ under no regularity conditions.

Recipe: a reduction from hypothesis testing to MLE under $\mathcal{P}$.

Theorem (UI/Split LR e-variable)

Split data $X_{1:n}$ into $D_0 \perp D_1$, define

$$E = \prod_{i \in D_0} \frac{\hat{q}_1(X_i)}{\hat{p}_0(X_i)}, \qquad \hat{q}_1 \in \mathcal{Q} \text{ learned from } D_1, \quad \hat{p}_0 \in \mathcal{P} \text{ is the MLE on } D_0,$$

then E is an e-variable for $\mathcal{P}$.

Universal Inference (UI) e-variable & e-process

Goal: build an e-variable (or an e-process) to test a composite null $\mathcal{P}$ against composite alternative $\mathcal{Q}$ under no regularity conditions.

Recipe: a reduction from hypothesis testing to MLE under $\mathcal{P}$.

Theorem (UI/Split LR e-variable)

Split data $X_{1:n}$ into $D_0 \perp D_1$, define

$$E = \prod_{i \in D_0} \frac{\hat{q}_1(X_i)}{\hat{p}_0(X_i)}, \qquad \hat{q}_1 \in \mathcal{Q} \text{ learned from } D_1, \quad \hat{p}_0 \in \mathcal{P} \text{ is the MLE on } D_0,$$

then E is an e-variable for $\mathcal{P}$.

Proposition (UI e-process)

Define $E_0 = 1$ and

$$E_t = \prod_{i=1}^{t} \frac{\hat{q}_{i-1}(X_i)}{\hat{p}_t(X_i)} \text{ for } t \geq 1, \quad \hat{q}_{i-1} \in \mathcal{Q} \text{ learned from } X_{1:i-1}, \quad \hat{p}_t \in \mathcal{P} \text{ is the MLE on } X_{1:t},$$

then $E = (E_t)$ is an e-process for $\mathcal{P}$.

Universal Inference (UI) e-variable & e-process

Proposition (UI e-process)

Define $E_0 = 1$ and

$$E_t = \prod_{i=1}^{t} \frac{\hat{q}_{i-1}(X_i)}{\hat{p}_t(X_i)} \text{ for } t \geq 1, \quad \hat{q}_{i-1} \in \mathcal{Q} \text{ learned from } X_{1:i-1}, \quad \hat{p}_t \in \mathcal{P} \text{ is the MLE on } X_{1:t},$$

then $E = (E_t)$ is an e-process for $\mathcal{P}$.

Universal Inference (UI) e-variable & e-process

Proposition (UI e-process)

Define $E_0 = 1$ and

$$E_t = \prod_{i=1}^{t} \frac{\hat{q}_{i-1}(X_i)}{\hat{p}_t(X_i)} \text{ for } t \geq 1, \quad \hat{q}_{i-1} \in \mathcal{Q} \text{ learned from } X_{1:i-1}, \quad \hat{p}_t \in \mathcal{P} \text{ is the MLE on } X_{1:t},$$

then $E = (E_t)$ is an e-process for $\mathcal{P}$.

Proof. For every $\mathbb{P} \in \mathcal{P}$, let $M_0^{\mathbb{P}} = E_0 = 1$, and

$$M_t^{\mathbb{P}} := \prod_{i=1}^{t} \frac{\hat{q}_{i-1}(X_i)}{p(X_i)} \geq \prod_{i=1}^{t} \frac{\hat{q}_{i-1}(X_i)}{\hat{p}_t(X_i)} = E_t \text{ for } t \geq 1.$$

Then $M_t^{\mathbb{P}}$ is a test martingale for $\mathbb{P}$.

$\Rightarrow$ optional stopping: $\mathbb{E}^{\mathbb{P}}[M_\tau^{\mathbb{P}}] \leq \mathbb{E}^{\mathbb{P}}[M_0^{\mathbb{P}}] = 1 \Rightarrow \mathbb{E}^{\mathbb{P}}[E_\tau] \leq \mathbb{E}^{\mathbb{P}}[M_\tau^{\mathbb{P}}] \leq 1$ for every stopping time τ.

Universal Inference (UI) e-variable & e-process

Proposition (UI e-process)

Define $E_0 = 1$ and

$$E_t = \prod_{i=1}^{t} \frac{\hat{q}_{i-1}(X_i)}{\hat{p}_t(X_i)} \text{ for } t \geq 1, \quad \hat{q}_{i-1} \in \mathcal{Q} \text{ learned from } X_{1:i-1}, \quad \hat{p}_t \in \mathcal{P} \text{ is the MLE on } X_{1:t},$$

then $E = (E_t)$ is an e-process for $\mathcal{P}$.

Remarks.

- Learn alternative out-of-sample; fit null in-sample.
- The derived sequential test $\mathbb{1}\{E_t \geq \frac{1}{\alpha}\}$ is computationally expensive: while numerator can be updated in an online fashion, all terms in denominator need to be recalculated after observing each new data point costing $O(t)$ at step t.
- Can use mixture instead of plug-in: $E_t = \int_{\mathcal{Q}} \prod_{i=1}^{t} \frac{q(X_i)}{\hat{p}_t(X_i)} \nu(\mathrm{d}q)$ for any distribution ν over $\mathcal{Q}$.

Outline

Testing by betting

Initialize wealth $M_0 = 1$.
For $t = 1, 2, \ldots$:

- Declare a bet $E_t : \mathcal{X} \to [0, \infty)$ with $\mathbb{E}^{\mathbb{P}}[E_t(X_t) \mid \mathcal{F}_{t-1}] \leq 1 \ \forall \mathbb{P} \in \mathcal{P}$.
- Observe data X_t.
- Update wealth: $M_t = M_{t-1} \cdot E_t(X_t) = \prod_{s=1}^{t} E_s(X_s)$.

Testing by betting

Initialize wealth $M_0 = 1$.
For $t = 1, 2, \ldots$:

- Declare a bet $E_t : \mathcal{X} \to [0, \infty)$ with $\mathbb{E}^{\mathbb{P}}[E_t(X_t) \mid \mathcal{F}_{t-1}] \leq 1 \;\; \forall \mathbb{P} \in \mathcal{P}$.
- Observe data X_t.
- Update wealth: $M_t = M_{t-1} \cdot E_t(X_t) = \prod_{s=1}^{t} E_s(X_s)$.

Definition (Sequential e-variables)

The e-variables E_t with $t \geq 1$ for $\mathcal{P}$, adapted to the filtration $\mathcal{F}$, are *sequential* if

$$\mathbb{E}^{\mathbb{P}}[E_t \mid \underbrace{E_1, \ldots, E_{t-1}}_{\mathcal{F}_{t-1}}] \leq 1, \quad \forall \, \mathbb{P} \in \mathcal{P}, \; \forall t \geq 1.$$

Testing by betting

Initialize wealth $M_0 = 1$.
For $t = 1, 2, \ldots$:

- Declare a bet $E_t : \mathcal{X} \to [0, \infty)$ with $\mathbb{E}^{\mathbb{P}}[E_t(X_t) \mid \mathcal{F}_{t-1}] \leq 1 \ \forall \mathbb{P} \in \mathcal{P}$.
- Observe data X_t.
- Update wealth: $M_t = M_{t-1} \cdot E_t(X_t) = \prod_{s=1}^{t} E_s(X_s)$.

Proposition

If $(E_t)_{t \geq 1}$ are sequential e-variables for $\mathcal{P}$,

$$M_t = \prod_{s=1}^{t} E_s \quad \text{for } t \geq 1, \qquad M_0 = 1,$$

is a test supermartingale (hence e-process) for $\mathcal{P}$.

Proof. $\mathbb{E}^{\mathbb{P}}[M_t \mid \mathcal{F}_{t-1}] = M_{t-1} \, \mathbb{E}^{\mathbb{P}}[E_t \mid \mathcal{F}_{t-1}] \leq M_{t-1}$ for every $\mathbb{P} \in \mathcal{P}$.

Testing by betting

Initialize wealth $M_0 = 1$.
For $t = 1, 2, \ldots$:

- Declare a bet $E_t : \mathcal{X} \to [0, \infty)$ with $\mathbb{E}^{\mathbb{P}}[E_t(X_t) \mid \mathcal{F}_{t-1}] \leq 1 \;\; \forall \mathbb{P} \in \mathcal{P}$.
- Observe data X_t.
- Update wealth: $M_t = M_{t-1} \cdot E_t(X_t) = \prod_{s=1}^{t} E_s(X_s)$.

Question: What are the optimal bets?

Testing by betting

Initialize wealth $M_0 = 1$.
For $t = 1, 2, \ldots$:

- Declare a bet $E_t : \mathcal{X} \to [0, \infty)$ with $\mathbb{E}^{\mathbb{P}}[E_t(X_t) \mid \mathcal{F}_{t-1}] \leq 1 \;\; \forall \mathbb{P} \in \mathcal{P}$.
- Observe data X_t.
- Update wealth: $M_t = M_{t-1} \cdot E_t(X_t) = \prod_{s=1}^{t} E_s(X_s)$.

Question: What are the optimal bets?

- For simple $\mathcal{P} = \{\mathbb{P}\}$ and $\mathcal{Q} = \{\mathbb{Q}\}$, $E_t(X_t) = \frac{q(X_t|\mathcal{F}_{t-1})}{p(X_t|\mathcal{F}_{t-1})}$ ensures (M_t) is log-optimal.

Testing by betting

Initialize wealth $M_0 = 1$.
For $t = 1, 2, \ldots$:

- Declare a bet $E_t : \mathcal{X} \to [0, \infty)$ with $\mathbb{E}^{\mathbb{P}}[E_t(X_t) \mid \mathcal{F}_{t-1}] \leq 1 \ \forall \mathbb{P} \in \mathcal{P}$.
- Observe data X_t.
- Update wealth: $M_t = M_{t-1} \cdot E_t(X_t) = \prod_{s=1}^{t} E_s(X_s)$.

Question: What are the optimal bets?

- For simple $\mathcal{P} = \{\mathbb{P}\}$ and $\mathcal{Q} = \{\mathbb{Q}\}$, $E_t(X_t) = \frac{q(X_t|\mathcal{F}_{t-1})}{p(X_t|\mathcal{F}_{t-1})}$ ensures (M_t) is log-optimal.
- For composite $\mathcal{P}$,
  (i) No known analogue of the LR increments that makes (M_t) log-optimal;
  (ii) While numeraires are log-optimal, they are designed for fixed sample size, $\implies E_t = E^{(t)*}/E^{(t-1)*}$ need not satisfy $\mathbb{E}^{\mathbb{P}}[E_t \mid \mathcal{F}_{t-1}] \leq 1$ and $(E^{(t)*})_{t\geq 1}$ is valid for the "least-favourable" $\mathbb{P}^* \in \mathcal{P}^{\circ\circ}$ via RIPr but not for all $\mathbb{P} \in \mathcal{P}$.

Testing by betting

Initialize wealth $M_0 = 1$.
For $t = 1, 2, \ldots$:

- Declare a bet $E_t : \mathcal{X} \to [0, \infty)$ with $\mathbb{E}^{\mathbb{P}}[E_t(X_t) \mid \mathcal{F}_{t-1}] \leq 1 \ \forall \mathbb{P} \in \mathcal{P}$.
- Observe data X_t.
- Update wealth: $M_t = M_{t-1} \cdot E_t(X_t) = \prod_{s=1}^{t} E_s(X_s)$.

Question: What are the optimal bets?

- For simple $\mathcal{P} = \{\mathbb{P}\}$ and $\mathcal{Q} = \{\mathbb{Q}\}$, $E_t(X_t) = \frac{q(X_t|\mathcal{F}_{t-1})}{p(X_t|\mathcal{F}_{t-1})}$ ensures (M_t) is log-optimal.
- For composite $\mathcal{P}$,
  (i) No known analogue of the LR increments that makes (M_t) log-optimal;
  (ii) While numeraires are log-optimal, they are designed for fixed sample size, $\implies E_t = E^{(t)*}/E^{(t-1)*}$ need not satisfy $\mathbb{E}^{\mathbb{P}}[E_t \mid \mathcal{F}_{t-1}] \leq 1$ and $(E^{(t)*})_{t\geq 1}$ is valid for the "least-favourable" $\mathbb{P}^* \in \mathcal{P}^{\circ\circ}$ via RIPr but not for all $\mathbb{P} \in \mathcal{P}$.
  (iii) Possible solution: Avoid all-in; pick $\lambda_t \in [0, 1]$ to hedge misspecification.

Testing by betting (generalised)

Initialize wealth $M_0 = 1$.
For $t = 1, 2, \ldots$:

- Declare a bet $E_t : \mathcal{X} \to [0, \infty)$ with $\mathbb{E}^{\mathbb{P}}[E_t(X_t) \mid \mathcal{F}_{t-1}] \leq 1 \ \forall \mathbb{P} \in \mathcal{P}$.
- Choose stake $\lambda_t \in [0, 1]$.
- Observe data X_t.
- Update wealth: $M_t = \underbrace{(1 - \lambda_t) M_{t-1} \cdot 1}_{\text{guaranteed wealth}} + \underbrace{\lambda_t M_{t-1} \cdot E_t}_{\text{risky payoff}} = \prod_{s=1}^{t} \big((1 - \lambda_s) + \lambda_s E_s\big)$

Testing by betting (generalised)

Initialize wealth $M_0 = 1$.
For $t = 1, 2, \ldots$:

- Declare a bet $E_t : \mathcal{X} \to [0, \infty)$ with $\mathbb{E}^{\mathbb{P}}[E_t(X_t) \mid \mathcal{F}_{t-1}] \leq 1 \ \forall \mathbb{P} \in \mathcal{P}$.
- Choose stake $\lambda_t \in [0, 1]$.
- Observe data X_t.
- Update wealth: $M_t = \underbrace{(1 - \lambda_t) M_{t-1} \cdot 1}_{\text{guaranteed wealth}} + \underbrace{\lambda_t M_{t-1} \cdot E_t}_{\text{risky payoff}} = \prod_{s=1}^{t} \big((1 - \lambda_s) + \lambda_s E_s\big)$

Proposition

$(M_t)_{t \geq 0}$ is a test supermartingale (hence e-process) for $\mathcal{P}$.

Proof. $\mathbb{E}^{\mathbb{P}}[(1 - \lambda_t) + \lambda_t E_t \mid \mathcal{F}_{t-1}] \leq (1 - \lambda_t) + \lambda_t \cdot 1 = 1$ for every $\mathbb{P} \in \mathcal{P}$.

Testing by betting (generalised)

Initialize wealth $M_0 = 1$.
For $t = 1, 2, \ldots$:

- Declare a bet $E_t : \mathcal{X} \to [0, \infty)$ with $\mathbb{E}^{\mathbb{P}}[E_t(X_t) \mid \mathcal{F}_{t-1}] \leq 1 \ \forall \mathbb{P} \in \mathcal{P}$.
- Choose stake $\lambda_t \in [0, 1]$.
- Observe data X_t.
- Update wealth: $M_t = \underbrace{(1 - \lambda_t) M_{t-1} \cdot 1}_{\text{guaranteed wealth}} + \underbrace{\lambda_t M_{t-1} \cdot E_t}_{\text{risky payoff}} = \prod_{s=1}^{t} \big((1 - \lambda_s) + \lambda_s E_s\big)$

Proposition

$(M_t)_{t \geq 0}$ is a test supermartingale (hence e-process) for $\mathcal{P}$.

Proof. $\mathbb{E}^{\mathbb{P}}[(1 - \lambda_t) + \lambda_t E_t \mid \mathcal{F}_{t-1}] \leq (1 - \lambda_t) + \lambda_t \cdot 1 = 1$ for every $\mathbb{P} \in \mathcal{P}$.

Definition

$(M_t)_{t \geq 0}$ is called an e-process built on $(E_t)_{t \geq 1}$.

Optimising predictable stakes $(\lambda_t)_{t\geq 1}$

Definitions

(i) For an alternative measure $\mathbb{Q}$, the $\mathbb{Q}$-log-optimal e-process built on (E_t) is (M_t) with

$$\lambda_t \in \arg\max_{\lambda\in[0,1]} \mathbb{E}^{\mathbb{Q}}\big[\log\big((1-\lambda)+\lambda E_t\big) \mid \mathcal{F}_{t-1}\big].$$

(ii) For $\gamma \in (0,1]$, the empirically adaptive e-process is (M_t) with

$$\lambda_t \in \arg\max_{\lambda\in[0,\gamma]} \frac{1}{t-1}\sum_{s=1}^{t-1} \log\big((1-\lambda)+\lambda E_s\big), \quad \lambda_1 = 0.$$

Optimising predictable stakes $(\lambda_t)_{t\geq 1}$

Definitions

(i) For an alternative measure $\mathbb{Q}$, the $\mathbb{Q}$-log-optimal e-process built on (E_t) is (M_t) with

$$\lambda_t \in \arg\max_{\lambda\in[0,1]} \mathbb{E}^{\mathbb{Q}}\left[\log\left((1-\lambda)+\lambda E_t\right) \mid \mathcal{F}_{t-1}\right].$$

(ii) For $\gamma \in (0,1]$, the empirically adaptive e-process is (M_t) with

$$\lambda_t \in \arg\max_{\lambda\in[0,\gamma]} \frac{1}{t-1}\sum_{s=1}^{t-1} \log\left((1-\lambda)+\lambda E_s\right), \quad \lambda_1 = 0.$$

Remarks:

Optimising predictable stakes $(\lambda_t)_{t\geq 1}$

Definitions

(i) For an alternative measure $\mathbb{Q}$, the $\mathbb{Q}$-log-optimal e-process built on (E_t) is (M_t) with

$$\lambda_t \in \arg\max_{\lambda\in[0,1]} \mathbb{E}^{\mathbb{Q}}\left[\log\big((1-\lambda)+\lambda E_t\big) \mid \mathcal{F}_{t-1}\right].$$

(ii) For $\gamma \in (0,1]$, the empirically adaptive e-process is (M_t) with

$$\lambda_t \in \arg\max_{\lambda\in[0,\gamma]} \frac{1}{t-1}\sum_{s=1}^{t-1} \log\big((1-\lambda)+\lambda E_s\big), \quad \lambda_1 = 0.$$

Remarks:

- Choose λ_t to maximise the (empirical) e-power of $(1-\lambda_t)+\lambda_t E_t$ given $\mathcal{F}_{t-1}$.

Optimising predictable stakes $(\lambda_t)_{t\geq 1}$

Definitions

(i) For an alternative measure $\mathbb{Q}$, the $\mathbb{Q}$-log-optimal e-process built on (E_t) is (M_t) with

$$\lambda_t \in \arg\max_{\lambda\in[0,1]} \mathbb{E}^{\mathbb{Q}}\left[\log\left((1-\lambda)+\lambda E_t\right) \mid \mathcal{F}_{t-1}\right].$$

(ii) For $\gamma \in (0,1]$, the empirically adaptive e-process is (M_t) with

$$\lambda_t \in \arg\max_{\lambda\in[0,\gamma]} \frac{1}{t-1}\sum_{s=1}^{t-1} \log\left((1-\lambda)+\lambda E_s\right), \quad \lambda_1 = 0.$$

Remarks:

- Choose λ_t to maximise the (empirical) e-power of $(1-\lambda_t)+\lambda_t E_t$ given $\mathcal{F}_{t-1}$.
    - If $\mathcal{P} = \{\mathbb{P}\}$, $\mathcal{Q} = \{\mathbb{Q}\}$ and E_t = conditional LR, then $\lambda_t = 1$.

Optimising predictable stakes $(\lambda_t)_{t\geq 1}$

Definitions

(i) For an alternative measure $\mathbb{Q}$, the $\mathbb{Q}$-log-optimal e-process built on (E_t) is (M_t) with

$$\lambda_t \in \arg\max_{\lambda\in[0,1]} \mathbb{E}^{\mathbb{Q}}\big[\log\big((1-\lambda)+\lambda E_t\big) \mid \mathcal{F}_{t-1}\big].$$

(ii) For $\gamma \in (0,1]$, the empirically adaptive e-process is (M_t) with

$$\lambda_t \in \arg\max_{\lambda\in[0,\gamma]} \frac{1}{t-1}\sum_{s=1}^{t-1} \log\big((1-\lambda)+\lambda E_s\big), \quad \lambda_1 = 0.$$

Remarks:

- Choose λ_t to maximise the (empirical) e-power of $(1-\lambda_t)+\lambda_t E_t$ given $\mathcal{F}_{t-1}$.
  - If $\mathcal{P} = \{\mathbb{P}\}$, $\mathcal{Q} = \{\mathbb{Q}\}$ and E_t = conditional LR, then $\lambda_t = 1$.
  - Theorem 3.14: for $E \geq 0$, $\mathbb{E}^{\mathbb{Q}}[E] > 1 \Leftrightarrow \exists\lambda \in [0,1]$ s.t. $\mathbb{E}^{\mathbb{Q}}[\log((1-\lambda)+\lambda E)] > 0$ $\implies$ If (empirical) mean of $E_1,\ldots,E_{t-1}$ is ≤ 1, choose $\lambda_t = 0$.

Optimising predictable stakes $(\lambda_t)_{t\geq 1}$

Definitions

(i) For an alternative measure $\mathbb{Q}$, the $\mathbb{Q}$-log-optimal e-process built on (E_t) is (M_t) with

$$\lambda_t \in \arg\max_{\lambda\in[0,1]} \mathbb{E}^{\mathbb{Q}}\left[\log\left((1-\lambda)+\lambda E_t\right) \mid \mathcal{F}_{t-1}\right].$$

(ii) For $\gamma \in (0,1]$, the empirically adaptive e-process is (M_t) with

$$\lambda_t \in \arg\max_{\lambda\in[0,\gamma]} \frac{1}{t-1}\sum_{s=1}^{t-1} \log\left((1-\lambda)+\lambda E_s\right), \quad \lambda_1 = 0.$$

Remarks:

- Choose λ_t to maximise the (empirical) e-power of $(1-\lambda_t)+\lambda_t E_t$ given $\mathcal{F}_{t-1}$.
  - If $\mathcal{P} = \{\mathbb{P}\}$, $\mathcal{Q} = \{\mathbb{Q}\}$ and $E_t =$ conditional LR, then $\lambda_t = 1$.
  - Theorem 3.14: for $E \geq 0$, $\mathbb{E}^{\mathbb{Q}}[E] > 1 \Leftrightarrow \exists\lambda \in [0,1]$ s.t. $\mathbb{E}^{\mathbb{Q}}[\log((1-\lambda)+\lambda E)] > 0$ $\implies$ If (empirical) mean of $E_1,\ldots,E_{t-1}$ is ≤ 1, choose $\lambda_t = 0$.
- (M_t) generated from method (i) is log-optimal among e-processes built on (E_t)

Optimising predictable stakes $(\lambda_t)_{t\geq 1}$

Definitions

(i) For an alternative measure $\mathbb{Q}$, the $\mathbb{Q}$-log-optimal e-process built on (E_t) is (M_t) with

$$\lambda_t \in \arg\max_{\lambda\in[0,1]} \mathbb{E}^{\mathbb{Q}}\left[\log\left((1-\lambda)+\lambda E_t\right) \mid \mathcal{F}_{t-1}\right].$$

(ii) For $\gamma \in (0,1]$, the empirically adaptive e-process is (M_t) with

$$\lambda_t \in \arg\max_{\lambda\in[0,\gamma]} \frac{1}{t-1}\sum_{s=1}^{t-1} \log\left((1-\lambda)+\lambda E_s\right), \quad \lambda_1 = 0.$$

Remarks:

- Choose λ_t to maximise the (empirical) e-power of $(1-\lambda_t)+\lambda_t E_t$ given $\mathcal{F}_{t-1}$.
  - If $\mathcal{P} = \{\mathbb{P}\}$, $\mathcal{Q} = \{\mathbb{Q}\}$ and $E_t =$ conditional LR, then $\lambda_t = 1$.
  - Theorem 3.14: for $E \geq 0$, $\mathbb{E}^{\mathbb{Q}}[E] > 1 \Leftrightarrow \exists \lambda \in [0,1]$ s.t. $\mathbb{E}^{\mathbb{Q}}[\log((1-\lambda)+\lambda E)] > 0$
    $\implies$ If (empirical) mean of $E_1, \ldots, E_{t-1}$ is ≤ 1, choose $\lambda_t = 0$.
- (M_t) generated from method (i) is log-optimal among e-processes built on (E_t)
- (M_t) generated from method (ii), GRAPA, has good e-power if (E_t) are roughly iid under $\mathbb{Q}$.

=Growth Rate Adaptive to the Particular Alternative

Empirically adaptive e-processes

Theorem

Let $(E_t)_{t\geq 1}$ be iid under the alternative distribution $\mathbb{Q}$ such that $\mathbb{E}^{\mathbb{Q}}[\log E_1]$ is finite. The empirically adaptive e-process $(M_t)_{t\geq 0}$ with $\gamma = 1$ satisfies the following:

(i) Asymptotic log-optimality in the sense that

$$\lim_{t\to\infty} \frac{1}{t}\left(\log M_t - \log M_t^{\mathbb{Q}}\right) \geq 0 \quad \text{in } L^1\text{-convergence under } \mathbb{Q}$$

for the $\mathbb{Q}$-log-optimal e-process $(M_t^{\mathbb{Q}})_{t\geq 0}$ built on $(E_t)_{t\geq 1}$.

(ii) Consistency, i.e., if $\mathbb{E}^{\mathbb{Q}}[E_1] > 1$, then $M_t \to \infty$ $\mathbb{Q}$-a.s. as $t \to \infty$.

Empirically adaptive e-processes

Theorem

Let $(E_t)_{t\geq 1}$ be iid under the alternative distribution $\mathbb{Q}$ such that $\mathbb{E}^{\mathbb{Q}}[\log E_1]$ is finite. The empirically adaptive e-process $(M_t)_{t\geq 0}$ with $\gamma = 1$ satisfies the following:

(i) Asymptotic log-optimality in the sense that

$$\lim_{t\to\infty} \frac{1}{t}\left(\log M_t - \log M_t^{\mathbb{Q}}\right) \geq 0 \quad \text{in } L^1\text{-convergence under } \mathbb{Q}$$

for the $\mathbb{Q}$-log-optimal e-process $(M_t^{\mathbb{Q}})_{t\geq 0}$ built on $(E_t)_{t\geq 1}$.

(ii) Consistency, i.e., if $\mathbb{E}^{\mathbb{Q}}[E_1] > 1$, then $M_t \to \infty$ $\mathbb{Q}$-a.s. as $t \to \infty$.

Proof.

(i) follows from LLN.

(ii) applies Theorem 3.14: for $E \geq 0$, $\mathbb{E}^{\mathbb{Q}}[E] > 1 \Leftrightarrow \exists \lambda \in [0,1]$ s.t. $\mathbb{E}^{\mathbb{Q}}[\log((1-\lambda) + \lambda E)] > 0$.

Empirically adaptive e-processes

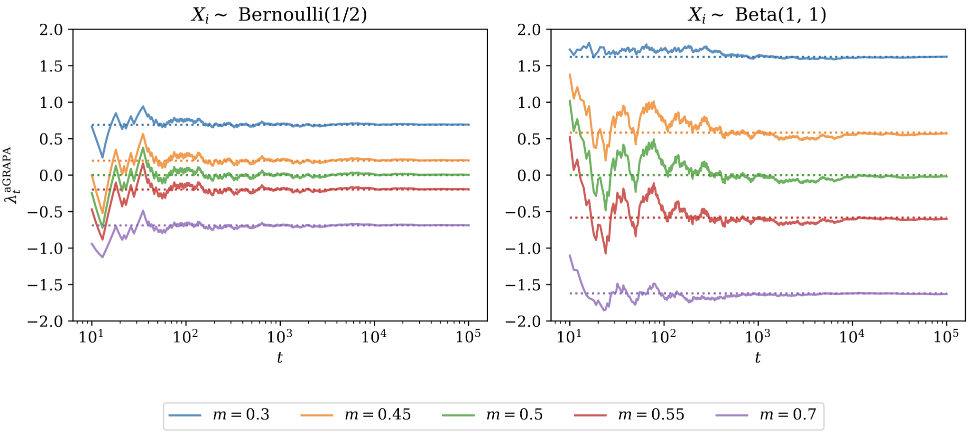


Figure: λ_t chosen using approximated GRAPA under 2 data distributions; dotted lines show oracle stakes.

Outline

Mixing fixed-sample-size e-variables into e-process

For all $t \geq 1$, let $E^{(t)}$ be e-variable for $\mathcal{P}$ based on $X_{1:t}$. With any pmf w on $\mathbb{N}$, define

$$M_t = \sum_{j=1}^{t} w(j)\, E^{(j)}, \quad M_0 = 1.$$

Proposition

$(M_t)_{t\geq 1}$ is an e-process for $\mathcal{P}$.

Mixing fixed-sample-size e-variables into e-process

For all $t \geq 1$, let $E^{(t)}$ be e-variable for $\mathcal{P}$ based on $X_{1:t}$. With any pmf w on $\mathbb{N}$, define

$$M_t = \sum_{j=1}^{t} w(j)\, E^{(j)}, \quad M_0 = 1.$$

Proposition

$(M_t)_{t \geq 1}$ is an e-process for $\mathcal{P}$.

Proof. For any $\mathbb{P} \in \mathcal{P}$ and any random time τ,

$$\mathbb{E}^{\mathbb{P}}[M_\tau] = \sum_{j \geq 1} w(j)\, \mathbb{E}^{\mathbb{P}}\left[E^{(j)} \cdot \mathbb{1}\{\tau \geq j\}\right] \leq \sum_{j \geq 1} w(j) = 1.$$

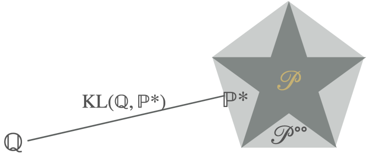
Remark. $(M_t)_{t \geq 1}$ is an increasing process, which is unusual for e-processes.

Mixing fixed-sample-size numeraires into e-process

Mixing fixed-sample-size numeraires into e-process

Recall:

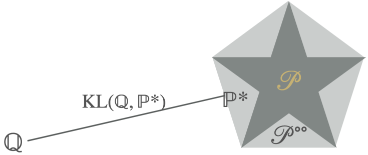
- A numeraire is a $\mathbb{Q}$-a.s. strictly positive e-variable E^* s.t. $\mathbb{E}^{\mathbb{Q}}[E/E^*] \leq 1$ for every e-variable E. $\implies$ $\mathbb{E}^{\mathbb{Q}}[\log(E/E^*)] \leq 0$.
- "Composite LR": E^* is LR between $\mathbb{Q}$ and some element $\mathbb{P}^* \in \mathcal{P}^{\circ\circ}$ (i.e., RIPr of $\mathbb{Q}$ onto $\mathcal{P}$).



Mixing fixed-sample-size numeraires into e-process

Recall:

- A numeraire is a $\mathbb{Q}$-a.s. strictly positive e-variable E^* s.t. $\mathbb{E}^{\mathbb{Q}}[E/E^*] \leq 1$ for every e-variable E. $\implies$ $\mathbb{E}^{\mathbb{Q}}[\log(E/E^*)] \leq 0$.
- "Composite LR": E^* is LR between $\mathbb{Q}$ and some element $\mathbb{P}^* \in \mathcal{P}^{\circ\circ}$ (i.e., RIPr of $\mathbb{Q}$ onto $\mathcal{P}$).



Can mix fixed-sample-size numeraires to obtain an e-process:

Popular choice of mixing pmf: $w(t) = \frac{c}{t(\log t)^2}$ with $\sum_{t \in \mathbb{N}} w(t) = 1$ and constant $c > 0$.

$\implies$ $\log M_t \geq \log\left(w(t) E^{(t)*}\right) = \log E^{(t)*} - \log t - 2 \log\log t + \log c$

$\implies$ $\lim_{t\to\infty} \frac{1}{t}\left(\log M_t - \log E^{(t)*}\right) = 0.$

Outline

Outline

Caveat: pathologies of e-power

Caveat: pathologies of e-power

Heavy tailed $\mathbb{Q}$ can make finite-sample e-power misleading:

Example: let $Y_1, Y_2, \ldots$ be iid Pareto(1) under $\mathbb{Q}$ (i.e., $\mathbb{Q}(Y > x) = 1/x$ for $x \geq 1$), define $E_t = \exp(t^2 - Y_t)$ and set $M_t = \prod_{s=1}^{t} E_s$ for $t \in \mathbb{N}$. Then

$\implies \mathbb{E}^{\mathbb{Q}}[\log E_t] = -\infty$, but $\log E_t$ is $\mathbb{Q}$-a.s. finite

$\implies \mathbb{E}^{\mathbb{Q}}[\log M_t] = -\infty$ for every t, while $M_t \to \infty$ in probability under $\mathbb{Q}$ (so $\mathbb{Q}(M_t \geq 1/\alpha) \to 1$)

Takeaway: Extra caution needed when infinity is involved in the calculation of e-power. Use consistency diagnostics alongside e-power.

Outline

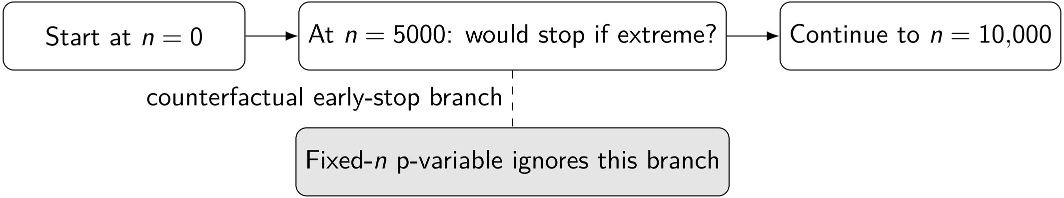
E-processes avoid reasoning about hypothetical worlds

E-processes avoid reasoning about hypothetical worlds

Even if an analyst actually stops at $n = 10{,}000$, the mere willingness to stop earlier if results had been extreme makes the stopping time data-dependent.

Consequence. Reasoning about the validity of p-variables requires reasoning about all possible "hypothetical worlds".

E-process fix. Report E_τ (or threshold at $1/\alpha$); valid for any stopping time τ, including adaptive or hypothetical ones.



Outline

Summary: SAVI with e-processes

Anytime validity: E-process controls type-I error at all stopping times; threshold at $1/\alpha$ via Ville.

Summary: SAVI with e-processes

Anytime validity: E-process controls type-I error at all stopping times; threshold at $1/\alpha$ via Ville.
Simple null:

- LR process is log-optimal
- plug-in or mixture LR for composite alternative can be asymptotically log-optimal.

Summary: SAVI with e-processes

Anytime validity: E-process controls type-I error at all stopping times; threshold at $1/\alpha$ via Ville.
Simple null:

- LR process is log-optimal
- plug-in or mixture LR for composite alternative can be asymptotically log-optimal.

Composite null:

- **UI e-process**: learn alt out-of-sample, fit null in-sample; applicable to irregular problems.
- **Betting via sequential e-variables** (E_t): Tune stakes $\lambda_t \in [0, 1]$ to form empirically adaptive e-process, matching asymptotic growth rate of its oracle counterpart if (E_t) are iid.
- **Mixing fixed-sample-size numeraires:** resulting e-process matches the numeraires' asymptotic growth rate.