Outline

1. Motivation for SAVI
   1.1 Problem with peeking at p-values
   1.2 Wald's Sequential Probability Ratio Test
2. Validity: e-processes under $\mathcal{P}$
   2.1 Setup & definitions
   2.2 Martingales, test (super)martingales & e-processes
   2.3 Optional stopping & Ville's inequality
3. **Efficiency: e-processes under $\mathcal{Q}$**
   3.1 Simple $\mathbb{P}$ vs. simple $\mathbb{Q}$
   3.2 Simple $\mathbb{P}$ vs. composite $\mathcal{Q}$
   3.3 Composite $\mathcal{P}$ vs. composite $\mathcal{Q}$: Testing by betting
4. Further discussions
5. Summary

From validity to efficiency

Consider an e-process $M = (M_t)_{t \geq 1}$.
Validity: $\mathbb{E}^{\mathbb{P}}[M_t] \leq 1 \; \forall t \geq 1$
Efficiency: maximise ~~$\mathbb{E}^{\mathbb{Q}}[M_t]$~~ $\mathbb{E}^{\mathbb{Q}}[\log M_t]$

Example:

- $(X_t)_{t \geq 1}$ iid with $\mathbb{P} = \text{Bern}(0.5)$ vs. $\mathbb{Q} = \text{Bern}(0.6)$
- For a parameter $\kappa \in [0, 1]$, let

$$E_t = \begin{cases} 1 + \kappa, & \text{if } X_t = 1 \\ 1 - \kappa, & \text{if } X_t = 0 \end{cases} \qquad \text{for } t \geq 1$$

$\implies \mathbb{E}^{\mathbb{P}}[E_t] = 0.5(1 + \kappa) + 0.5(1 - \kappa) = 1$, so E_t is an e-variable for $\mathbb{P}$
$\implies M_t = \prod_{i=1}^{t} E_i$ for $t \geq 1$ is an e-process for $\mathbb{P}$

- $\mathbb{E}^{\mathbb{Q}}[M_t] = t\mathbb{E}^{\mathbb{Q}}[E_1] = t(1 + 0.2\kappa)$ is maximised at $\kappa = 1$

Maximise $\mathbb{E}^{\mathbb{Q}}[M_t]$ for efficiency?

Example: $(X_t)_{t\geq 1}$ iid with $\mathbb{P} = \text{Bern}(0.5)$ vs. $\mathbb{Q} = \text{Bern}(0.6)$. $M_t = \prod_{i=1}^t E_i$ with $E_t = \begin{cases} 1+\kappa & X_t = 1 \\ 1-\kappa & X_t = 0 \end{cases}$

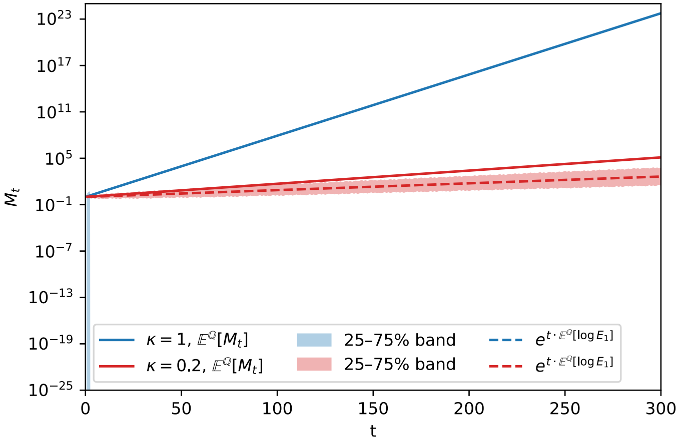


Figure: $M_t = \prod_{i=1}^t E_i$ under $\mathbb{Q}$, with 6000 runs.

Mean path: In $\kappa = 1$ case, only the all-ones path $X_1, \ldots, X_t = 1$ contributes to $\mathbb{E}^{\mathbb{Q}}[M_t]$ with tiny probability 0.6^t

Typical path?

$$\frac{1}{t}\log M_t \xrightarrow[t\to\infty]{\mathbb{Q}\text{-a.s.}} \mathbb{E}^{\mathbb{Q}}[\log E_1]$$

$$\Rightarrow \text{median}(M_t) \sim e^{t\mathbb{E}^{\mathbb{Q}}[\log E_1]}$$

Maximise $\mathbb{E}^{\mathbb{Q}}[\log M_t]$ for efficiency

- Evidence multiplies ~~(sums)~~
  $\implies$ **Typical path** governed by geometric (~~arithmetic~~) mean
  $\implies$ Look at logs
- $\mathbb{E}^{\mathbb{Q}}[\log M_t]$: **e-power/expected log-growth**

Outline

1. Motivation for SAVI
   - 1.1 Problem with peeking at p-values
   - 1.2 Wald's Sequential Probability Ratio Test
2. Validity: e-processes under $\mathcal{P}$
   - 2.1 Setup & definitions
   - 2.2 Martingales, test (super)martingales & e-processes
   - 2.3 Optional stopping & Ville's inequality
3. **Efficiency: e-processes under $\mathcal{Q}$**
   - 3.1 Simple $\mathbb{P}$ vs. simple $\mathbb{Q}$
   - 3.2 Simple $\mathbb{P}$ vs. composite $\mathcal{Q}$
   - 3.3 Composite $\mathcal{P}$ vs. composite $\mathcal{Q}$: Testing by betting
4. Further discussions
5. Summary

Simple $\mathbb{P}$ vs. simple $\mathbb{Q}$

The Likelihood Ratio (LR) process M^* given by $M_0^* = 1$ and $M_t^* = \frac{\mathrm{d}\mathbb{Q}|_{\mathcal{F}_t}}{\mathrm{d}\mathbb{P}|_{\mathcal{F}_t}}(X_1, \ldots, X_t)$ for $t \geq 1$ 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 $\mathbb{P}$:

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

Proof. For fixed t, setting $M_t = M_t^*$ maximises $\mathbb{E}^{\mathbb{Q}}[\log M_t]$ (presented by François) + Reduction

Next: Going back to our Bernoulli example...

Simple $\mathbb{P}$ vs. simple $\mathbb{Q}$: log-optimality of LR process

Example: $(X_t)_{t\geq 1}$ iid with $\mathbb{P} = \text{Bern}(0.5)$ vs. $\mathbb{Q} = \text{Bern}(0.6)$. $M_t = \prod_{i=1}^t E_i$ with E_t as a function of $\kappa \in [0,1]$.

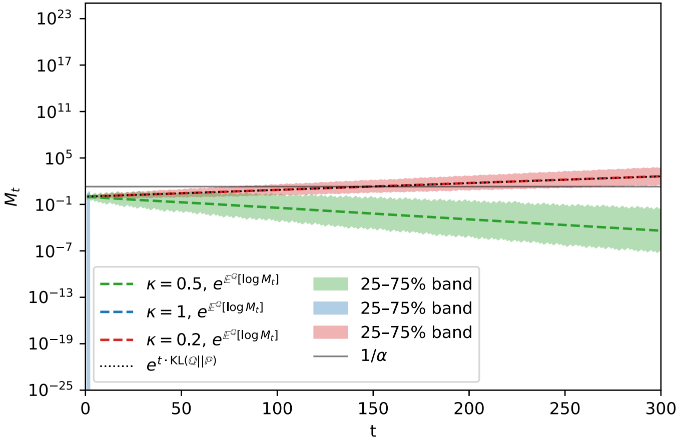


Figure: $M_t = \prod_{i=1}^t E_i$ under $\mathbb{Q}$, with 6000 runs.

LR process: $M_t^* = \prod_{i=1}^t \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}(X_i)$

Remarks:

- For every fixed t,
  $\mathbb{E}^{\mathbb{Q}}[\log M_t^*] \geq \mathbb{E}^{\mathbb{Q}}[\log M_t]$
  $= \sum_{i=1}^t \mathbb{E}^{\mathbb{Q}}[\log \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}(X_i)] = t \cdot \mathrm{KL}(\mathbb{Q}\|\mathbb{P})$
- Asymptotically,
  Power $\mathbb{Q}\left(M_t \geq \frac{1}{\alpha}\right) \xrightarrow[t\to\infty]{} 1$

Outline

1. Motivation for SAVI
   - 1.1 Problem with peeking at p-values
   - 1.2 Wald's Sequential Probability Ratio Test
2. Validity: e-processes under $\mathcal{P}$
   - 2.1 Setup & definitions
   - 2.2 Martingales, test (super)martingales & e-processes
   - 2.3 Optional stopping & Ville's inequality
3. **Efficiency: e-processes under $\mathcal{Q}$**
   - 3.1 Simple $\mathbb{P}$ vs. simple $\mathbb{Q}$
   - 3.2 Simple $\mathbb{P}$ vs. composite $\mathcal{Q}$
   - 3.3 Composite $\mathcal{P}$ vs. composite $\mathcal{Q}$: Testing by betting
4. Further discussions
5. Summary

Simple $\mathbb{P}$ vs. composite $\mathcal{Q}$

Definition (Asymptotic log-optimality)

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 oracle LR process of $\mathbb{Q}$ to $\mathbb{P}$.

- Requires at least the same long-run log-growth rate as $M^{\mathbb{Q}}$
- Covers any e-process M that grows an $e^{o(t)}$ factor slower than $M^{\mathbb{Q}}$

Definition (Consistency)

An e-process M is said to be consistent against $\mathbb{Q}$ if $M_t \to \infty$, $\mathbb{Q}$-a.s. as $t \to \infty$.

Simple $\mathbb{P}$ vs. composite $\mathcal{Q}$

Example: Testing $\mathbb{P}$ against $\mathcal{Q} = \{\mathbb{Q}_\theta : \theta \in \Theta_1\}$ with iid data.

Plug-in LR: Set $M_0 = 1$, and for $t \geq 1$ use $M_t = \prod_{i=1}^{t} \frac{q_{\hat{\theta}_{i-1}}(X_i)}{p(X_i)}$ with predictable $\hat{\theta}_{i-1} \in \Theta_1$

Proposition (informal)

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 $\mathcal{H}_0 : \theta^\dagger = 0$ vs $\mathcal{H}_1 : \theta^\dagger > 0$. For illustration, take $\theta^\dagger = 0.3$.

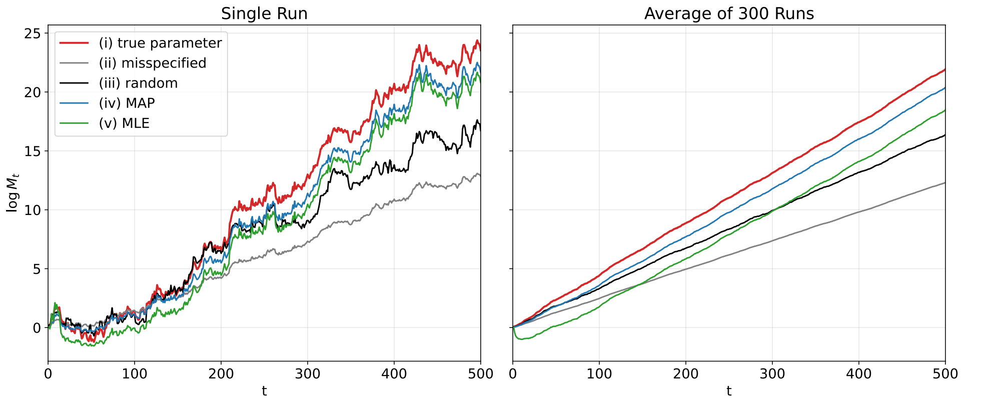


Figure: Few ways of constructing e-processes from LR processes.

(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 $X_1, \ldots, X_{i-1}$

Outline

1. Motivation for SAVI
   - 1.1 Problem with peeking at p-values
   - 1.2 Wald's Sequential Probability Ratio Test

2. Validity: e-processes under $\mathcal{P}$
   - 2.1 Setup & definitions
   - 2.2 Martingales, test (super)martingales & e-processes
   - 2.3 Optional stopping & Ville's inequality

3. **Efficiency: e-processes under $\mathcal{Q}$**
   - 3.1 Simple $\mathbb{P}$ vs. simple $\mathbb{Q}$
   - 3.2 Simple $\mathbb{P}$ vs. composite $\mathcal{Q}$
   - 3.3 Composite $\mathcal{P}$ vs. composite $\mathcal{Q}$: Testing by betting

4. Further discussions

5. Summary

Testing by betting

Key idea: e-process for $\mathcal{P}$ = wealth of a bettor wagering against $\mathcal{P}$

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 $\mathbb{E}^{\mathbb{P}}[E_t \mid \mathcal{F}_{t-1}] \leq 1$ for all $\mathbb{P} \in \mathcal{P}$ and $t \geq 1$, then $M_t = \prod_{s=1}^{t} E_s$ for $t \geq 1$ with $M_0 = 1$ 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

Key idea: e-process for $\mathcal{P}$ = wealth of a bettor wagering against $\mathcal{P}$

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 $(M_t)_{t\geq 0}$ is log-optimal.
- For composite $\mathcal{P}$ and $\mathcal{Q}$,
  (i) No known analogue of the LR increments that makes $(M_t)_{t\geq 0}$ log-optimal;
  (ii) Compromise: Avoid all-in; pick stake $\lambda_t \in [0, 1]$ to hedge misspecification.

Testing by betting (composite $\mathcal{P}$ vs. $\mathcal{Q}$)

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: $M_t = \underbrace{(1 - \lambda_t) M_{t-1} \cdot 1}_{\text{guaranteed wealth}} + \underbrace{\lambda_t M_{t-1} \cdot E_t}_{\text{risky payoff}} = \prod_{s=1}^{t} \left((1 - \lambda_s) + \lambda_s E_s\right)$

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 $\mathbb{P} \in \mathcal{P}$.

Next: How to optimise the stakes $(\lambda_t)_{t \geq 1}$?

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

Definitions

(i) For an alternative measure $\mathbb{Q}$, the oracle e-process built on $(E_t)_{t\geq 1}$ is $(M_t)_{t\geq 0}$ with

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

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

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

Remarks:

- Choose λ_t to maximise the (empirical) e-power of $(1-\lambda_t)+\lambda_t E_t$ given $\mathcal{F}_{t-1}$.
- $(M_t)_{t\geq 0}$ from (i) is log-optimal among e-processes built on $(E_t)_{t\geq 1}$
- $(M_t)_{t\geq 0}$ from (ii) has good e-power & power if $(E_t)_{t\geq 1}$ are roughly iid under $\mathbb{Q}$.

Next: Going back to our iid Bernoulli example...

Example:

- $(X_t)_{t\geq 1}$ iid from Bern(p), with $\mathcal{H}_0 : p \leq 0.5$ vs. $\mathcal{H}_1 : p \geq 0.55$.
- For $t \geq 1$, let

$$E_t = \begin{cases} 2, & \text{if } X_t = 1 \\ 0, & \text{if } X_t = 0 \end{cases}$$

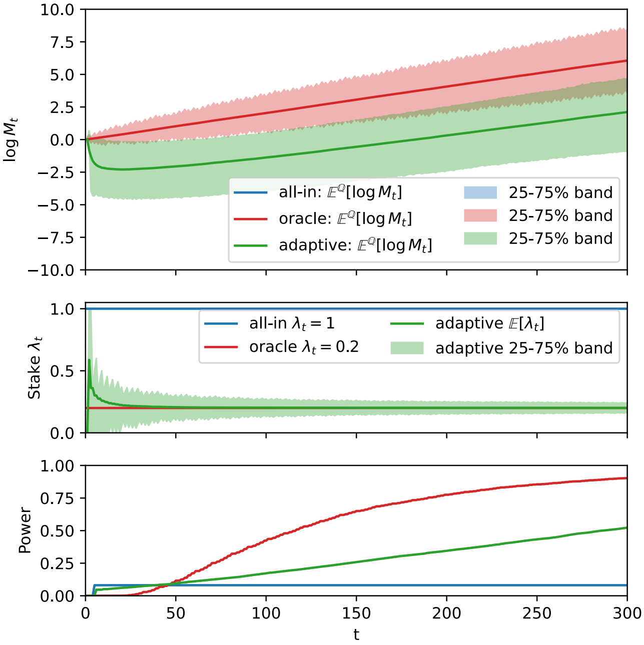
$\implies \mathbb{E}^{\mathbb{P}}[E_t(X_t) \mid \mathcal{F}_{t-1}] \leq 0.5 \cdot 2 + 0.5 \cdot 0 = 1$ for $p \leq 0.5$.

- Nature picks $\mathbb{Q} = \text{Bern}(0.6)$.
- Oracle e-process built on $(E_t)_{t\geq 1}$ bets with

$$\lambda_t = 0.2 \in \operatorname*{arg\,max}_{\lambda \in [0,1]} \mathbb{E}^{\mathbb{Q}}\Big[\log\big((1-\lambda) + \lambda E_t\big)\Big].$$

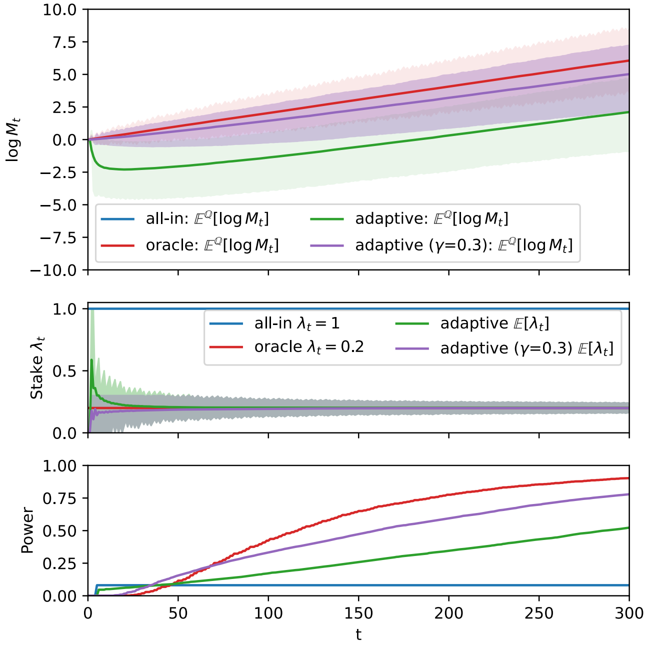
- Empirically adaptive e-process bets with

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



Remarks:

- oracle e-process is log-optimal among e-processes built on $(E_t)_{t\geq 1}$
- empirically-adaptive e-process
  - lies between oracle and all-in
  - stakes concentrate as $t \to \infty$
  - more aggressive at the start
  - has good e-power & power when $(E_t)_{t\geq 1}$ iid under $\mathbb{Q}$



Remarks:

- oracle e-process is log-optimal among e-processes built on $(E_t)_{t\geq 1}$
- empirically-adaptive e-process
  - lies between oracle and all-in
  - stakes concentrate as $t \to \infty$
  - more aggressive at the start
  - has good e-power & power when $(E_t)_{t\geq 1}$ iid under $\mathbb{Q}$

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

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

with the oracle 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) due to for $E \geq 0$, $\mathbb{E}^{\mathbb{Q}}[E] > 1 \iff \exists \lambda \in [0,1]$ s.t. $\mathbb{E}^{\mathbb{Q}}[\log((1-\lambda) + \lambda E)] > 0$.