Robust likelihood ratio tests for composite nulls and alternatives

Aytijhya Saha

Indian Statistical Institute, Kolkata
(MIT, from Fall 2025)

Aaditya Ramdas

Carnegie Mellon University, Pittsburgh

https://arxiv.org/abs/2408.14015

June 30, 2025

Outline

1. Huber's robust SPRT
2. Our modification to Huber's robust SPRT
3. Extensions to composite nulls and alternatives

Huber's Robust SPRT

The Classical SPRT Isn't Robust

Given a sequence of observations $X_1, X_2, \ldots$, consider testing P_0 vs P_1. The test martingale is

$$M_n = \prod_{i=1}^{n} \frac{p_1(X_i)}{p_0(X_i)}.$$

SPRT is **NOT robust**: a single factor $p_1(X_j)/p_0(X_j)$ equal or close to 0 or ∞ may upset the entire test statistic M_n.

One natural solution: (Truncated Probability Ratio Test)

Replace $p_1(X_i)/p_0(X_i)$ by $\pi(X_i) = \max\{c', \min\{c'', p_1(X_i)/p_0(X_i)\}\}$, for some $c' < c''$ and

$$S_n = \prod_{i=1}^{n} \pi(X_i).$$

Huber's Robust SPRT

- Huber (1965) showed that for a specific choice of (c', c''), the Truncated Probability Ratio Test:
  accept null if $S_n \leq a$ and reject null if $S_n \geq b$ (sequential)
  is optimal in some well defined minimax sense.

- To account for the possibility of small deviations from the idealized models $P_j, j = 0,1$; Huber (1965) expanded them into the following composite hypotheses:
  $$H_j^\epsilon = \{Q \in \mathcal{M} : Q = (1-\epsilon)P_j + \epsilon H, H \in \mathcal{M}\}$$
  But we will consider a more general total-variation model:
  $$H_j^\epsilon = \{Q \in \mathcal{M} : d_{\mathsf{TV}}(P_j, Q) \leq \epsilon\}$$

Huber's Robust SPRT (cont.)

Huber (1965) defined the distributions $Q_{j,\epsilon} \in \mathscr{H}_j^{\epsilon}$, by their densities

$$q_{0,\epsilon}(x) = (1-\epsilon)\,p_0(x) \text{ for } p_1(x)/p_0(x) < c''$$

$$= (1/c'')\,(1-\epsilon)\,p_1(x) \text{ for } p_1(x)/p_0(x) \geq c''$$

$$q_{1,\epsilon}(x) = (1-\epsilon)\,p_1(x) \text{ for } p_1(x)/p_0(x) > c'$$

$$= c'\,(1-\epsilon)\,p_0(x) \text{ for } p_1(x)/p_0(x) \leq c'\,.$$

$$\boxed{\begin{aligned} \pi_\epsilon(x) &:= q_{1,\epsilon}(x)/q_{0,\epsilon}(x) \\ &= \max\{c', \min\{c'', p_1(X_i)/p_0(X_i)\}\} \end{aligned}}$$

$0 \leq c' < c'' \leq \infty$ are determined such that $q_{0,\epsilon}, q_{1,\epsilon}$ are probability densities-

$$(1-\epsilon)\left\{P_0\left[p_1/p_0 < c''\right] + (c'')^{-1}P_1\left[p_1/p_0 \geq c''\right]\right\} = 1$$

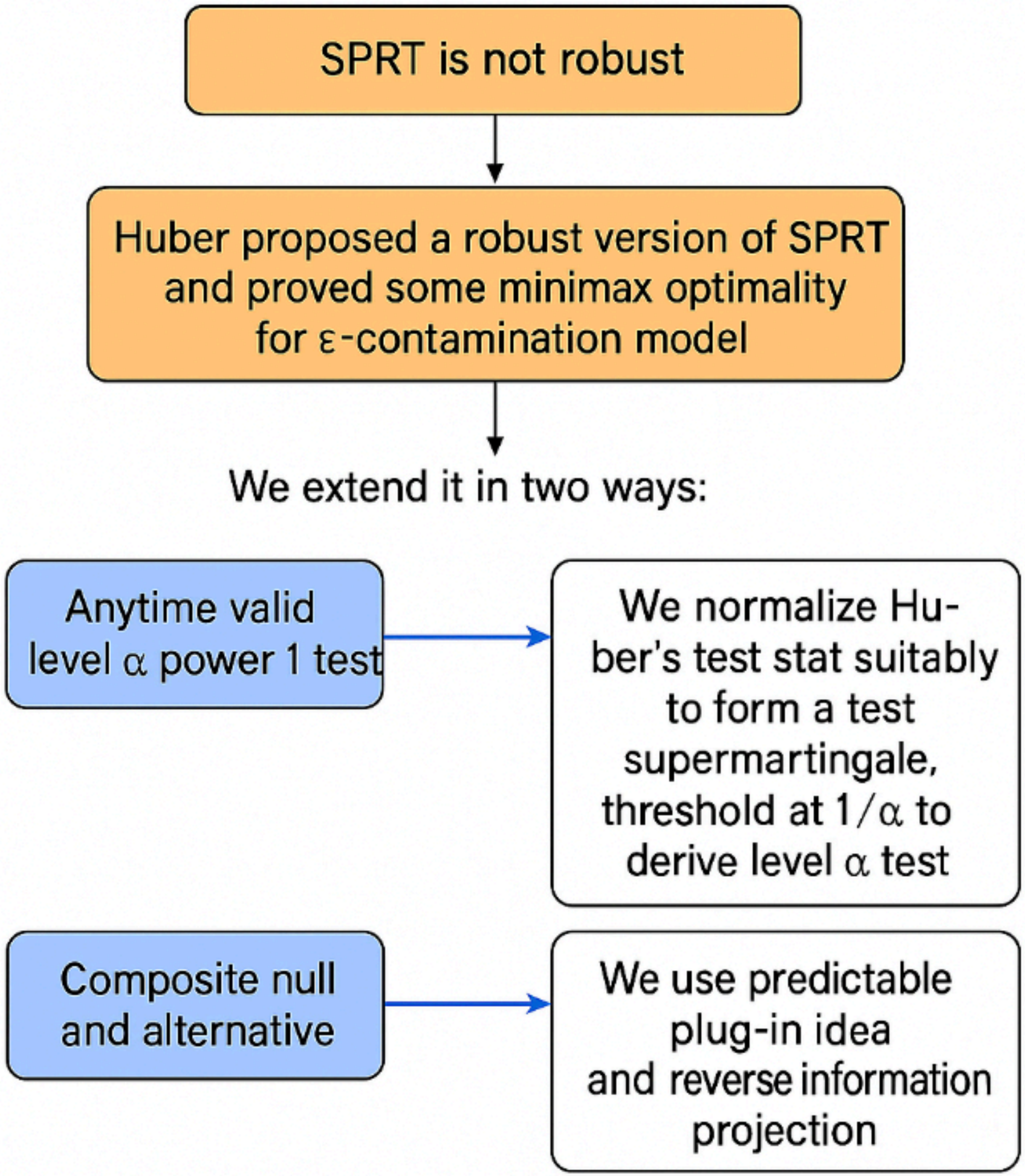
$$(1-\epsilon)\left\{P_1\left[p_1/p_0 > c'\right] + c'P_0\left[p_1/p_0 \leq c'\right]\right\} = 1$$

Drawbacks of Huber's Robust SPRT

Despite its elegance, it has two important limitations that we address:

- Huber did not provide a way to get an anytime valid, *level* α *power-one* test. Our approach yields *level* α *power-one* test, which is valid at arbitrary stopping times.

- Huber's test was a robust version of point nulls and alternatives. We extend Huber's robust test for dealing with general composite nulls and alternatives.


SPRT is not robust
Huber proposed a robust version of SPRT and proved some minimax optimality for ε-contamination model
We extend it in two ways:
Anytime valid level α power 1 test
We normalize Huber's test stat suitably to form a test supermartingale, threshold at 1/α to derive level α test
Composite null and alternative
We use predictable plug-in idea and reverse information projection

(Super)Martingales and Anytime Valid Tests

$\{M_n\}_n$ is called a test (super)martingale for H_0 if it is a (super) martingale for every $\mathbb{P} \in H_0$, and if it is non-negative with $M_0 = 1$.

Ville's inequality applied to test (super)martingale implies $P(\exists n \in \mathbb{N} : M_n \geq 1/\alpha) \leq \alpha, \forall P \in H_0$ which is equivalent to

$$\boxed{P(M_\tau \geq 1/\alpha) \leq \alpha, \text{for every stopping time } \tau, P \in H_0}$$

Reject null at $\tau_\alpha = \inf \left\{ n \geq 1 : M_n \geq 1/\alpha \right\}$ to obtain a level α test.

Ville's inequality ensures that the above test is valid at arbitrary data-dependent stopping times, accommodating optional stopping or continuation.

Our Modification to Huber's Robust SPRT

Our Adaptive Contamination Model

Supermartingale tools allow us to deal with time-varying adaptive - potentially adversarial - deviations from the null, where

$$X_n \mid X_1, \cdots, X_{n-1} \sim Q_n \in H_0^{\epsilon}$$

Set of all possible joint distributions of the sequence $X_1, X_2, \cdots$ under the adaptive contamination model as $H_0^{\epsilon,\infty}$.

Our Test Supermartingale

Define, $R_t^{\epsilon} = \prod_{i=1}^{t} \frac{\pi_{\epsilon}(X_i)}{\mathbb{E}_{P_0}\pi_{\epsilon}(X) + (c'' - c')\epsilon} . t = 1,2,\cdots, R_0 = 1.$

Total variation distance is an integral probability metric: for any $c_1 < c_2$,

$$d_{\mathsf{TV}}(P, Q) = \frac{1}{c_2 - c_1} \sup_{c_1 \le f \le c_2} \left| E_{X \sim P} f(X) - E_{X \sim Q} f(X) \right| .$$

$X_n \mid X^{n-1} \sim Q_n \in \mathscr{H}_0^{\epsilon}, d_{\mathsf{TV}}(Q_n, P_0) < \epsilon$, which implies

$$\mathbb{E}_{X_n | X^{n-1} \sim Q_n} \left[\pi_{\epsilon}(X_n) \mid X^{n-1}\right] \le \mathbb{E}_{P_0} \left[\pi_{\epsilon}(X_n)\right] + (c'' - c')\epsilon \implies \mathbb{E}_{X_n | X^{n-1} \sim Q_n} \left[\frac{\pi_{\epsilon}(X_n)}{\mathbb{E}_{P_0} \left[\pi_{\epsilon}(X)\right] + (c'' - c')\epsilon} \mid X^{n-1} \right] \le 1$$

So, R_t^{ϵ} is a test supermartingale for $H_0^{\epsilon,\infty}$.

Reject null at $\tau_\alpha = \inf \left\{ n \ge 1 : R_n^{\epsilon} \ge 1/\alpha \right\}$ to obtain a level α sequential test for $H_0^{\epsilon,\infty}$.

Growth Rate of the test

The "growth rate" of a test supermartingale R_t is defined as $\inf_{\mathbb{P}\in H_1} \lim_{t\to\infty} \frac{\log R_t}{t}$. A positive growth rate implies consistency of the test.

Theorem 1:

Suppose that $\epsilon > 0$ and $X_1, X_2, \cdots \sim Q \in H_1^\epsilon$ are iid. Then, as $t \to \infty$, $(\log R_t^\epsilon)/t \to r_Q^\epsilon$ almost surely, for some constant r_Q^ϵ and the growth rate,

$$r^\epsilon = \inf_{Q\in H_1^\epsilon} r_Q^\epsilon \geq d_{\mathsf{KL}}(Q_{1,\epsilon}, Q_{0,\epsilon}) - 2(\log c'' - \log c')\epsilon - \log(1 + 2(c'' - c')\epsilon)$$

Asymptotic Optimality of Growth Rate

Note that likelihood ratio test (SPRT) is the log-optimal test for testing P_0 vs P_1 and hence the optimal growth rate is $\mathbb{E}_{P_1}[\log p_1(X)/p_0(X)] = d_{\mathsf{KL}}(P_1, P_0)$.

Theorem 3:

The growth rate of our test $r^\epsilon \to d_{\mathsf{KL}}(P_1, P_0),$ as $\epsilon \to 0.$

Robust Predictable Plug-in for Composite Alternatives

Simple Null (P_0) vs Composite Alternative ($\mathscr{P}_1$)

Robust Predictable Plug-in: Obtain a robust estimate $\hat{p}_n$ based on past observations $X_1, \cdots, X_{n-1}$. (e.g., for testing Gaussian mean, we can use sample median as an estimate of the parameter)

$$\hat{\pi}_{n,\epsilon}(x) = \max\{c_n', \min\{c_n'', \hat{p}_n(x)/p_0(x)\}\}.$$

$$(1-\epsilon)\left\{P_0\left[\hat{p}_n/p_0 < c_n''\right] + \frac{1}{c_n''}\hat{P}_n\left[\hat{p}_n/p_0 \geq c_n''\right]\right\} = 1,$$

$$(1-\epsilon)\left\{\hat{P}_n\left[\hat{p}_n/p_0 > c_n'\right] + c_n' P_0\left[\hat{p}_n/p_0 \leq c_n'\right]\right\} = 1.$$

Simple Null (P_0) vs Composite Alternative ($\mathscr{P}_1$)

Test Supermartingale: $R_{n,\epsilon}^{\text{plug-in}} = R_{n-1,\epsilon}^{\text{plug-in}} \times \hat{E}_{n,\epsilon}(X_n)$,

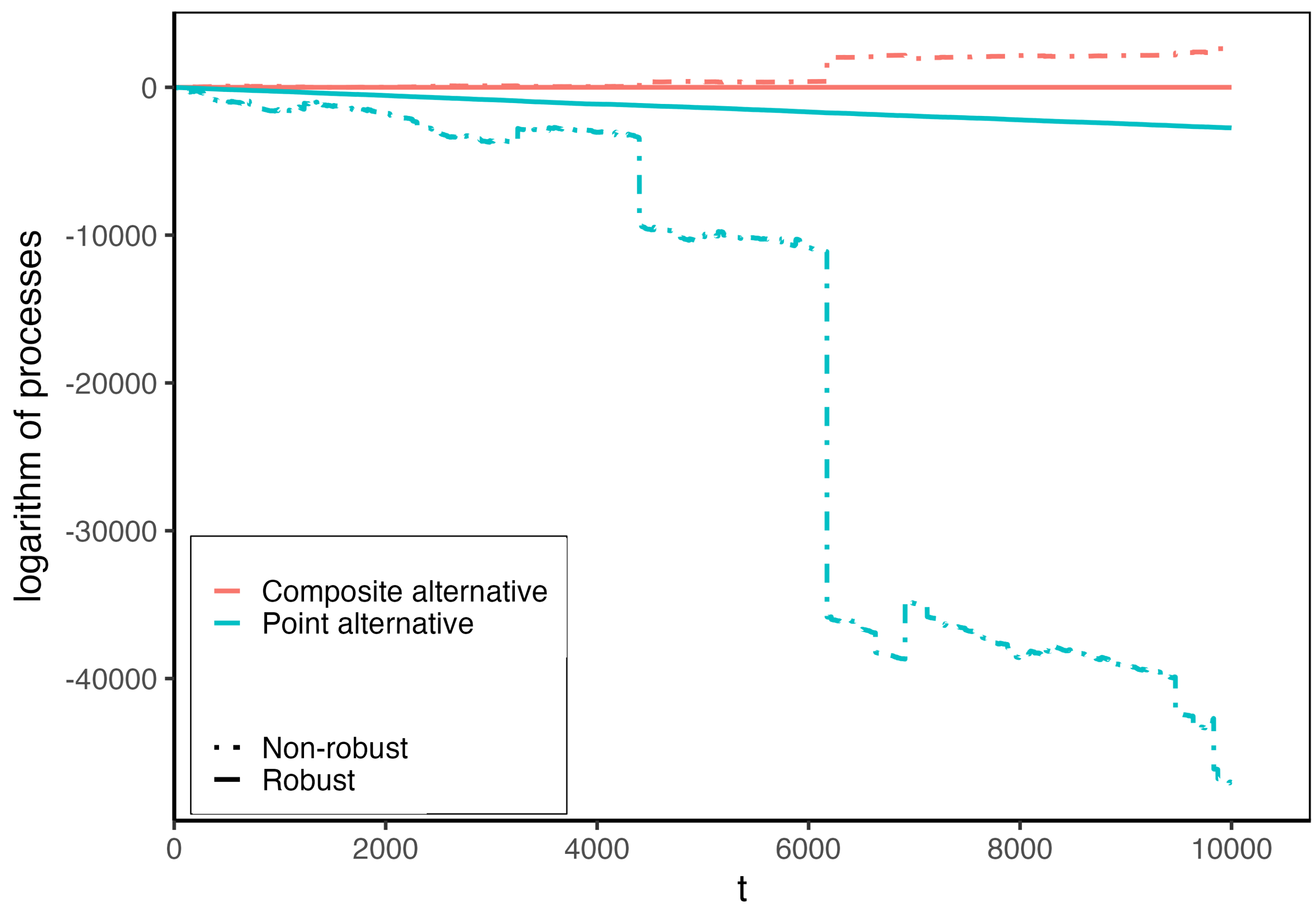
$$\hat{E}_{n,\epsilon}(X_n) := \begin{cases} \dfrac{\hat{\pi}_{n,\epsilon}(X_n)}{\mathbb{E}_{X_n|X^{n-1}\sim P_0}\left[\hat{\pi}_{n,\epsilon}(X_n) \mid X^{n-1}\right] + (c_n'' - c_n')\epsilon}, & \text{if } c_n' < c_n'', \\ 1, & \text{otherwise.} \end{cases}$$

$$\mathbb{E}_{X_n|X^{n-1}\sim Q_n}[\hat{E}_{n,\epsilon}(X_n) \mid X^{n-1}] = I_{c_n' \geq c_n''} + \frac{\mathbb{E}_{X_n|X^{n-1}\sim Q_n}\left[\hat{\pi}_{n,\epsilon}(X_n) \mid X^{n-1}\right] I_{c_n' < c_n''}}{\mathbb{E}_{X|X^{n-1}\sim P_0}\left[\hat{\pi}_{n,\epsilon}(X_n) \mid X^{n-1}\right] + (c_n'' - c_n')\epsilon} \leq 1$$

Reject null at $\tau_\alpha = \inf\left\{n \geq 1 : R_{n,\epsilon}^{\text{plug-in}} \geq 1/\alpha\right\}$ to obtain a level α sequential test for $H_0^{\epsilon,\infty}$.

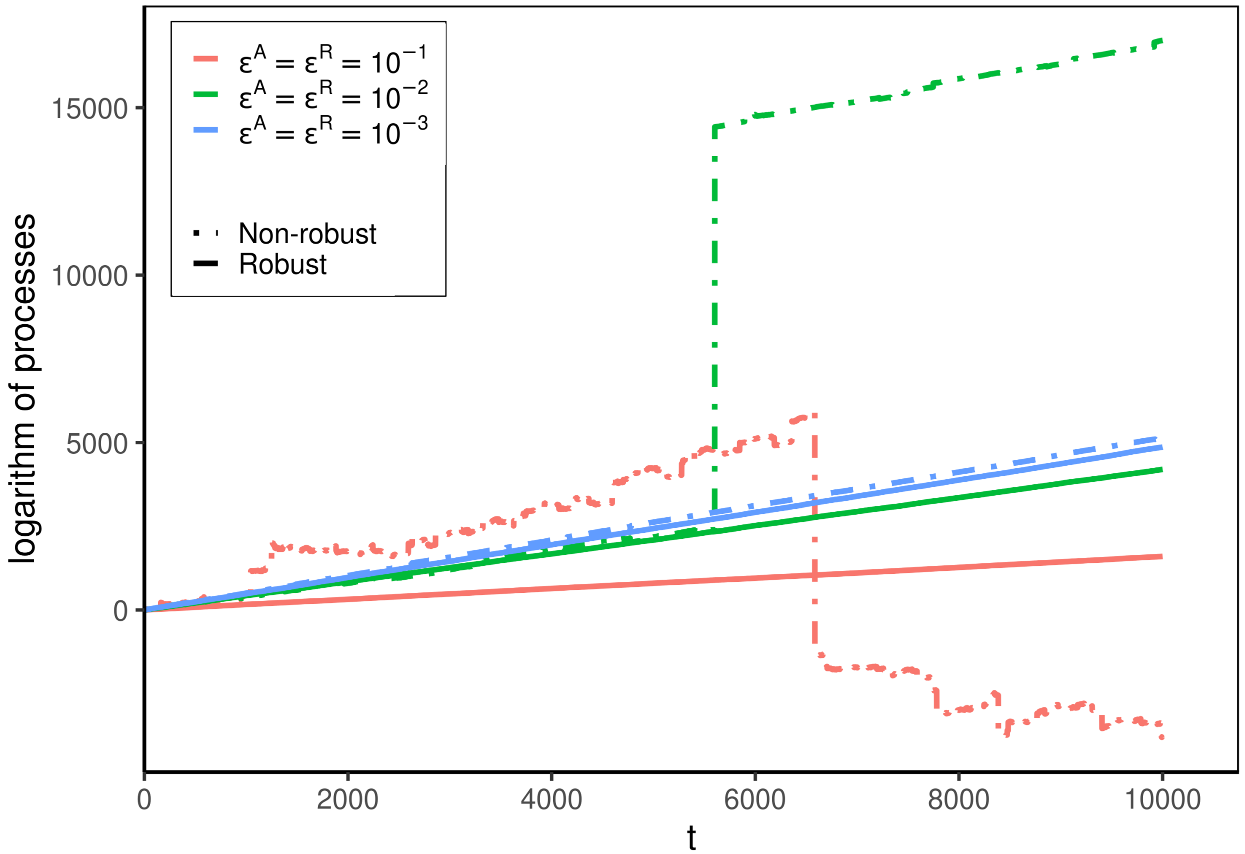
- We can show results on the growth rate and consistency assuming $\hat{p}_n$ to be a consistent estimator for p_1.

Simulations (Under the null)

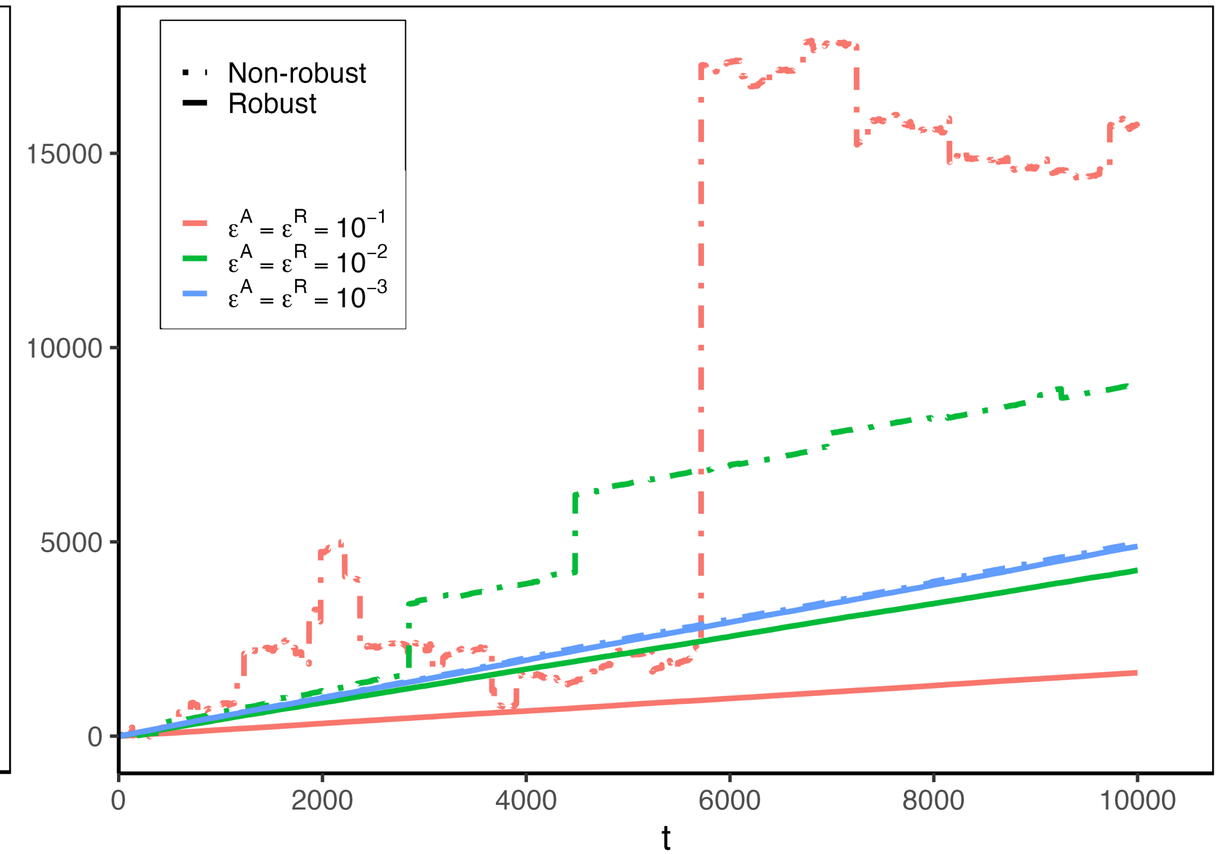


Data is drawn from $(1-\epsilon^R) \times N(0,1) + \epsilon^R \times \text{Cauchy}(-1,10)$ and $\epsilon^A = \epsilon^R = 0.01$, $P_0 = N(0,1)$, the simple and the composite alternative to be $P_1 = N(1,1)$ and $\mathscr{P}_1 = \{N(\mu,1) : \mu \neq 0\}$ respectively. Our robust tests are safe, but the non-robust tests exhibit unstable and unreliable behaviour.

Simulations



(A) $P_0 = N(0,1), P_1 = N(1,1)$



(B) $P_0 = N(0,1), \mathscr{P}_1 = \{N(\mu,1) : \mu \neq 0\}$

Data is drawn from $(1-\epsilon^R) \times N(1,1) + \epsilon^R \times \text{Cauchy}(-1,10)$, $\epsilon^A = \epsilon^R = 0.1, 0.01, 0.001$. The growth rate of our robust tests increases as ϵ decreases. As anticipated, The growth rates for our robust tests based on simple and composite alternatives almost overlap.

Combining Robust Predictable Plug-in and Numeraire for Composite Nulls and Alternatives

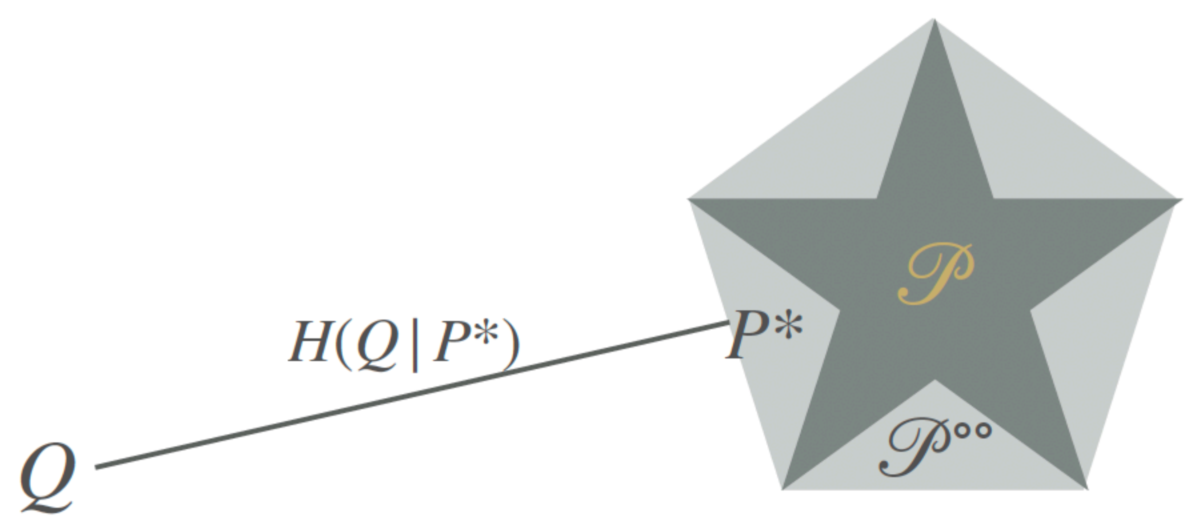
Reverse Information Projection (RIPr)

RIPr of Q onto $\mathscr{P}$:"closest" element of $\mathscr{P}$ to Q.

For any null $\mathscr{P}$ and simple alternative Q, there always exists a unique and strictly positive e-variable B^* called the numeraire, such that for any e-variable B for $\mathscr{P}$,

$$\mathbb{E}_Q[B/B^*] \leq 1 \ .$$

Define a measure P^* by defining its likelihood ratio $\frac{dP^*}{dQ} := \frac{1}{B^*}$. This P^* is called the **Reverse Information Projection (RIPr)** of Q onto $\mathscr{P}$.



Composite Null ($\mathscr{P}_0$) vs Composite Alternative ($\mathscr{P}_1$)

TV neighborhoods: $\mathscr{H}_i^\epsilon = \bigcup_{P \in \mathscr{P}_i} \{Q : d_{TV}(P, Q) \leq \epsilon\}, i = 0,1.$

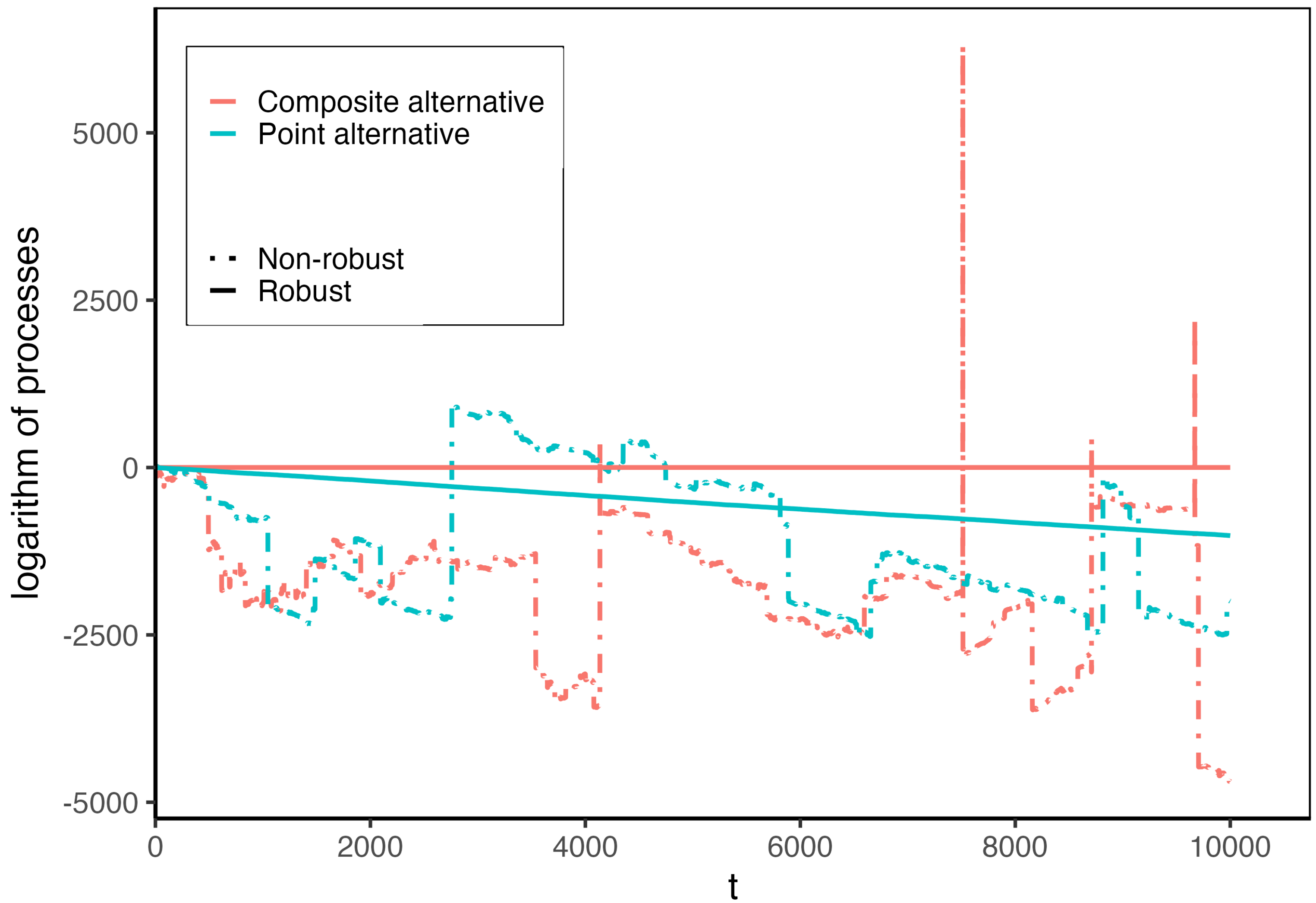
Let $\hat{p}_{1,n}$ be some robust estimate of the density based on past observations, which belongs to the alternative. Let $\hat{P}_{0,n}$ be the reverse information projection (RIPr) of $\hat{P}_{1,n}$ on the null $\mathscr{P}_0$.

$$\hat{\pi}_{n,\epsilon}(x) = \max\{c_n', \min\{c_n'', \hat{p}_{1,n}(x)/\hat{p}_{0,n}(x)\}.$$

Test Supermartingale: $R_{n,\epsilon}^{\text{RIPr,plug-in}} = R_{n-1,\epsilon}^{\text{RIPr,plug-in}} \times \hat{B}_{n,\epsilon}(X_n),$

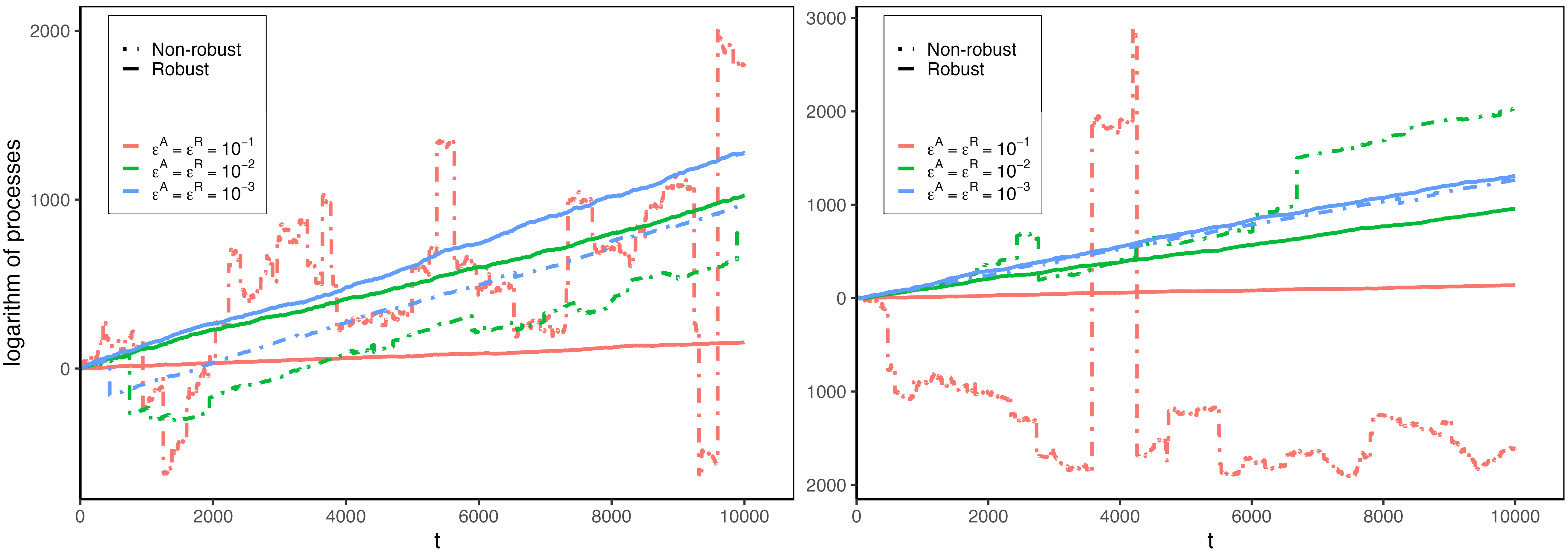
$$\hat{B}_{n,\epsilon}(x) := \begin{cases} \dfrac{\hat{\pi}_{n,\epsilon}(X_n)}{\sup_{P \in \mathscr{P}_0} \mathbb{E}_{X|X^{n-1}\sim P}\left[\hat{\pi}_{n,\epsilon}(X_n) \mid X^{n-1}\right] + (c_n'' - c_n')\epsilon}, & \text{if } c_n' < c_n'', \\ 1, & \text{otherwise.} \end{cases}$$

Simulations (Under the null)



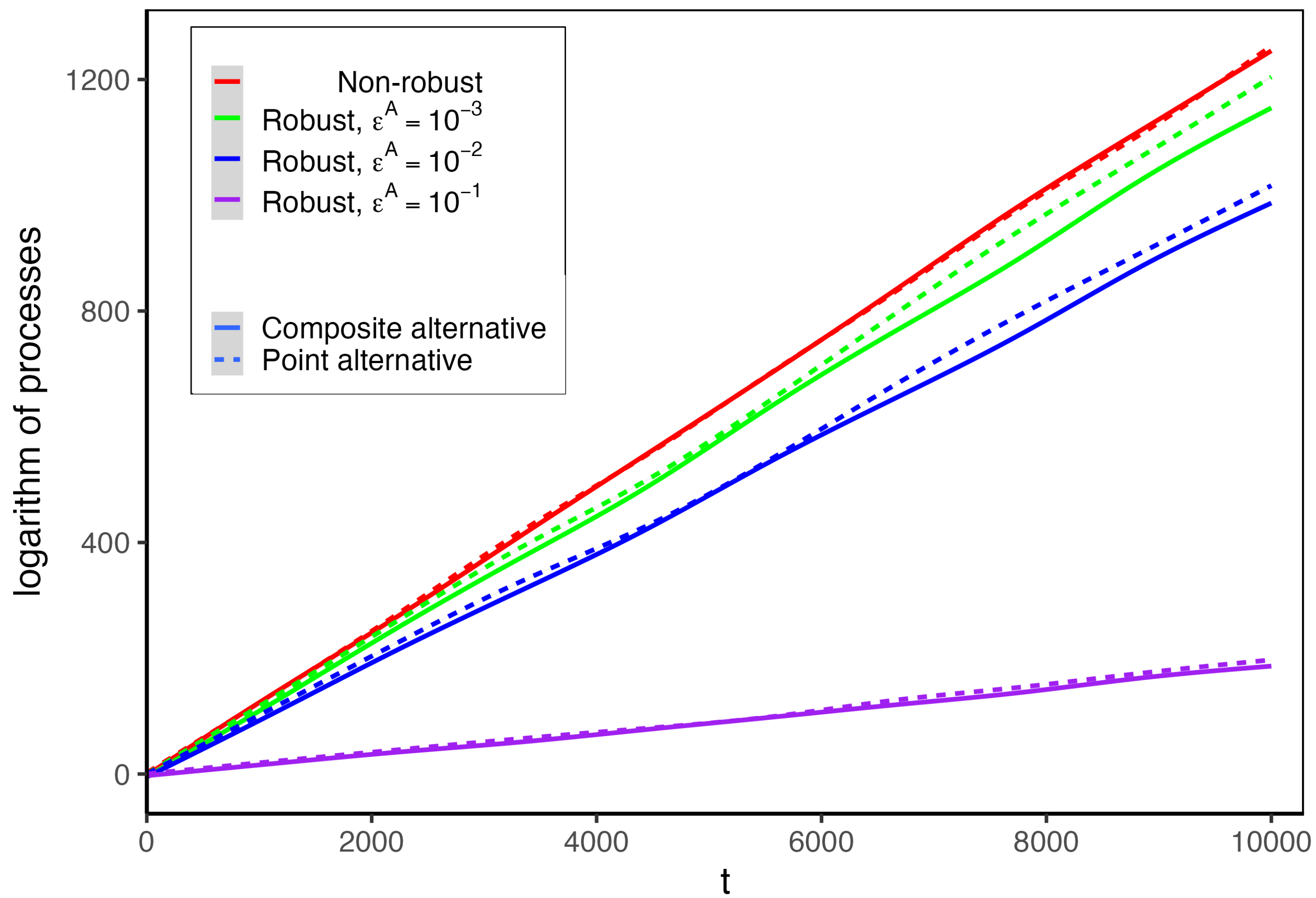
Data is drawn from $(1-\epsilon^R) \times N(0,1) + \epsilon^R \times \text{Cauchy}(-1,10)$ and $\epsilon^A = \epsilon^R = 0.01$. The null is $\mathscr{P}_0 = \{N(\mu,1) : -0/5 \leq \mu \leq 0/5\}$. Our robust tests are safe, but the non-robust tests exhibit unstable and unreliable behavior.

Simulations



Data is drawn from $(1-\epsilon^R) \times N(1,1) + \epsilon^R \times \text{Cauchy}(-1,10)$, $\epsilon^A = \epsilon^R = 0.1, 0.01, 0.001$. $\mathscr{P}_0 = \{N(\mu,1) : -0/5 \leq \mu \leq 0/5\}$. The growth rate of our robust tests increases as ϵdecreases. The growth rates for our robust tests based on simple and composite alternatives almost overlap.

Simulations



Data is drawn from $N(0,1)$ (WITH NO CONTAMINATION) and $P_0 = N(1,1), P_1 = N(0,1)$. $\mathscr{P}_0 = \{N(\mu,1) : -0/5 \leq \mu \leq 0/5\}$. Here, the growth rate of our robust tests approaches that of the non-robust test, as ϵ decreases. The growth rates for our robust tests based on simple alternatives and composite alternatives almost overlap.

Summary

- By integrating the *plug-in* and *RIPr* techniques, we propose a robust method for testing composite nulls vs composite alternatives.
- Growth rate of our tests approaches the optimal growth rates of the non-robust tests as $\epsilon \to 0$.
- Our tests are inherently sequential, being valid at arbitrary data-dependent stopping times, but they are new even for fixed sample sizes, giving type-I error control without any regularity conditions.
- Simulations validate the theory and demonstrate excellent practical performance.

THANK YOU