Turbo Learning

Wouter M. Koolen

Friday 20th February, 2026

E-readers, CWI

CWI and University of Twente

Menu for Today

We are going to do hypothesis testing with e-values in an active, composite, sequential setting.

Menu for Today

We are going to do hypothesis testing with e-values in an active, composite, sequential setting.

And **outgrow** GROW.

Menu for Today

We are going to do hypothesis testing with e-values in an active, composite, sequential setting.

And **outgrow** GROW.

Which should be impossible by definition.

Menu for Today

We are going to do hypothesis testing with e-values in an active, composite, sequential setting.

And **outgrow** GROW.

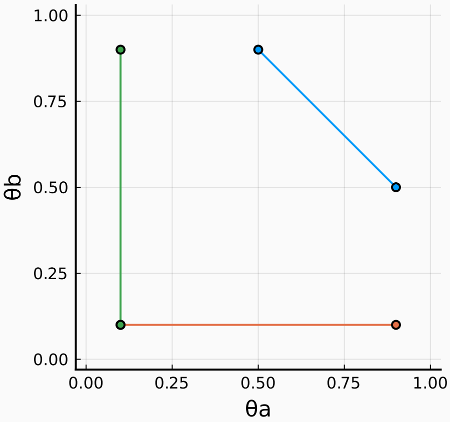
Which should be impossible by definition.

Spot the catch.

Our CWI server is under attack . . .

. . . but by who?

Enemy	Attack	Scan	
		Net (a)	CPU (b)
Nation state	(1) XSS	0.9	0.5
	(2) Spectre	0.5	0.9
Script kid	(1) Login Guess	0.1	0.1
	(2) Port Scan	0.9	0.1
Criminal	(1) Metasploit	0.1	0.1
	(2) SQL Injection	0.1	0.9



Observations

- Each hypothesis (= enemy type) generated by two attacks
  - Enemy can randomise
  - Enemy can switch attacks between rounds

  So **not IID extension** but Walley extension (fork convex).
- We control scanning : active testing

Our Job as the Statistician

Say we are convinced we are victim of a **Nation state**, not a **Script kid** or **Criminal**.

But we need to statistically prove it.

With e-values.

One Round

Protocol

Fix enemy type $I \in \{$**Nation state**, **Script kid**, **Criminal**$\}$.

1. *Independently*
   - *Learner picks a scan* $K \in \{a, b\}$, *possibly at random* $K \sim w$.
   - *Enemy picks an attack* $J \in \{1, 2\}$, *possibly at random* $J \sim q$.
2. *Learner sees scan* K *and outcome* $Y \sim \theta_{I,J,K}$
3. *Learner reports evidence* $X_{\mathsf{S}}(K, Y)$ *against* **Script kid** *and* $X_{\mathsf{C}}(K, Y)$ *against* **Criminal**.

NB: the sample space is $\{a, b\} \times \{0, 1\}$.

E-variable

Definition

We call $X : \{a, b\} \times \{0, 1\} \to [0, \infty]$ an **e-variable** against **Script kid** under sampling rule w if

$$\forall \text{ attack } j \in \{1, 2\} : \sum_{k \in \{a,b\}} w(k) \sum_{y \in \{0,1\}} \theta_{\mathbf{S},j,k}(y) X(k, y) \ \leq \ 1$$

We denote by $\mathcal{E}_{\mathbf{S},w}$ all e-variables against **S** under w.

E-variable

Definition

We call $X : \{a, b\} \times \{0, 1\} \to [0, \infty]$ an **e-variable** against **Script kid** under sampling rule w if

$$\forall \text{ attack } j \in \{1, 2\} : \sum_{k \in \{a,b\}} w(k) \sum_{y \in \{0,1\}} \theta_{\mathsf{S},j,k}(y) X(k, y) \ \leq \ 1$$

We denote by $\mathcal{E}_{\mathsf{S},w}$ all e-variables against **S** under w.

Example

These are e-variables under w uniform on $\{a, b\}$:

k	y	X_{S}	X_{C}
a	0	1	5/9
	1	1	5
b	0	5/9	1
	1	5	1

Growth rate

The GROW e-variable against **S** under w with the help of alternative **N** is the maximiser of

$$GROW(\mathbf{S}, w) \ := \ \max_{X \in \mathcal{E}_{\mathbf{S},w}} \min_{j \in \{1,2\}} \sum_{k \in \{a,b\}} w(k) \sum_{y \in \{0,1\}} \theta_{\mathbf{N},j,k}(y) \ln X(k, y)$$

Growth rate

The GROW e-variable against **S** under w with the help of alternative **N** is the maximiser of

$$GROW(\mathbf{S}, w) \;:=\; \max_{X \in \mathcal{E}_{\mathbf{S},w}} \min_{j \in \{1,2\}} \sum_{k \in \{a,b\}} w(k) \sum_{y \in \{0,1\}} \theta_{\mathbf{N},j,k}(y) \ln X(k, y)$$

Fact

The two e-variables in the example are each GROW against their enemy for uniform sampling. Their growth rate is 0.2554 *each.*

The active part

How to pick the sampling rule w? Idea: maximise the per-round growth rate:

$$AGROW := \max_{w \in \triangle_{\{a,b\}}} \min_{I \in \{\mathbf{S},\mathbf{C}\}} GROW(I, w)$$

Great: we will stop and claim **N** in about $\frac{\ln \frac{1}{\delta}}{AGROW}$ rounds on average.

The active part

How to pick the sampling rule w? Idea: maximise the per-round growth rate:

$$AGROW := \max_{w \in \triangle_{\{a,b\}}} \min_{I \in \{\mathbf{S},\mathbf{C}\}} GROW(I, w)$$

Great: we will stop and claim **N** in about $\frac{\ln \frac{1}{\delta}}{AGROW}$ rounds on average.

Fact

The AGROW sampling rule for our setup is uniform on scans $\{a, b\}$. The AGROW is 0.2554.

But is AGROW actually optimal?

Two things suggest YES

- For **IID extensions** (not Walley/fork-convex), AGROW is optimal. In fact, all instance-optimal sample complexity results in the fixed confidence bandit testing world hinge on this fact. (Garivier and Kaufmann, 2016).
- Every e-process against a fork-convex null is dominated by a **supermartingale** for that null. (Ramdas, Ruf, Larsson, and Koolen, 2021, Theorem 14).
  And we maximised the per-round growth rate by AGROW. So what else can you do?

Really

?

Two Rounds

Protocol

Fix enemy type $I \in \{$**Nation state**, **Script kid**, **Criminal**$\}$.

For $t = 1, 2$

1. *Independently*
   - *Learner picks a scan* $K_t \in \{a, b\}$, *possibly at random* $K_t \sim w(\cdot|history)$.
   - *Enemy picks an attack* $J_t \in \{1, 2\}$, *possibly at random* $J_t \sim q(\cdot|history)$.
2. *Learner sees scan* K_t *and outcome* $Y_t \sim \theta_{I,J_t,K_t}$

Learner reports evidence $X_{\mathbf{S}}(K_1, Y_1, K_2, Y_2)$ *against* **Script kid** *and* $X_{\mathbf{C}}(K_1, Y_1, K_2, Y_2)$ *against* **Criminal**.

NB: the sample space is $(\{a, b\} \times \{0, 1\})^2$.

Objective

Let's optimise the evidence over two rounds.

$$AGROW_2 := \max_{\text{sampling rule } w} \max_{\substack{(X_1^I, X_2^I)_{I \in \{\mathbf{S},\mathbf{C}\}} \\ \text{cond. evars (1)}}} \min_{I \in \{\mathbf{S},\mathbf{C}\}} \min_{\text{mixing rule } q}$$

$$\sum_{\substack{k_1, j_1, y_1 \\ k_2, j_2, y_2}} \underbrace{w(k_1) q(j_1) \theta_{\mathbf{N}, j_1, k_1}(y_1)}_{\text{round 1}} \underbrace{w(k_2 | k_1, y_1) q(j_2 | k_1, y_1) \theta_{\mathbf{N}, j_2, k_2}(y_2)}_{\text{round 2}} \underbrace{\ln\left(X_1^I(k_1, y_1) \cdot X_2^I(k_2, y_2 | k_1, y_1)\right)}_{\text{objective}}$$

Objective

Let's optimise the evidence over two rounds.

$$AGROW_2 := \max_{\text{sampling rule } w} \max_{\substack{(X_1^I, X_2^I)_{I \in \{\mathbf{S},\mathbf{C}\}} \\ \text{cond. evars (1)}}} \min_{I \in \{\mathbf{S},\mathbf{C}\}} \min_{\text{mixing rule } q}$$

$$\sum_{\substack{k_1, j_1, y_1 \\ k_2, j_2, y_2}} \underbrace{w(k_1) q(j_1) \theta_{\mathbf{N}, j_1, k_1}(y_1)}_{\text{round 1}} \underbrace{w(k_2 | k_1, y_1) q(j_2 | k_1, y_1) \theta_{\mathbf{N}, j_2, k_2}(y_2)}_{\text{round 2}} \underbrace{\ln\Big(X_1^I(k_1, y_1) \cdot X_2^I(k_2, y_2 | k_1, y_1)\Big)}_{\text{objective}}$$

Observations:

- Conditional E-variable constraint in every context:

$$X_1^I \in \mathcal{E}_{I, w(\cdot|\epsilon)} \qquad \text{and} \qquad X_2^I(\cdot | k_1, y_1) \in \mathcal{E}_{I, w(\cdot | k_1, y_1)} \qquad \text{for both } I \in \{\mathbf{S}, \mathbf{C}\}. \qquad (1)$$

- Single sampling rule w needs to ensure growth of evidence against both **S** and **C**.
- **Active**: Sampling rule w in round 2 depends on round one
- **Worst-case**: Mixture q in round 2 depends on round one (fork-convex)

Answer is NO: You can outperform AGROW

k_1	y_1	k_2	y_2	$X_1^{\mathbf{S}}$	$X_2^{\mathbf{S}}$	$X_1^{\mathbf{S}} \cdot X_2^{\mathbf{S}}$	$X_1^{\mathbf{C}}$	$X_2^{\mathbf{C}}$	$X_1^{\mathbf{C}} \cdot X_2^{\mathbf{C}}$
a	0	a	0	1	1	1	5/9	5/9	25/81
			1		1	1		5	25/9
	1	b	0	1	5/9	5/9	5	1	5
			1		5	5		1	5
b	0	b	0	5/9	5/9	25/81	1	1	1
			1		5	25/9		1	1
	1	a	0	5	1	5	1	5/9	5/9
			1		1	5		5	5

This results in $AGROW_2 = 0.6130$. This outgrows $2 \times AGROW_1 = 2 \times 0.2554 = 0.5108$

Questions

When does this phenomenon happen?

Is $AGROW_1 < AGROW_2 < \cdots$?

What is $\lim_{n \to \infty} AGROW_n$ and how to achieve it?

Can we sample to behave like $AGROW_n$ for every n?

Conclusion

Let's talk!

References

- Garivier, A. and E. Kaufmann (2016). **“Optimal Best arm Identification with Fixed Confidence”.** In: *Proceedings of the 29th Conference On Learning Theory (COLT)*.
- Ramdas, A., J. Ruf, M. Larsson, and W. M. Koolen (July 2021). **“Testing exchangeability: fork-convexity, supermartingales and e-processes”.** In: *International Journal of Approximate Reasoning*.