Combining Evidence Across Filtrations Using Adjusters

Preprint: https://arxiv.org/abs/2402.09698
Presented in March 2024

Yo Joong "YJ" Choe
Data Science Institute
University of Chicago
yjchoe.github.io

Aaditya Ramdas
Statistics & Data Science; Machine Learning
Carnegie Mellon University
stat.cmu.edu/~aramdas

Outline

1. Motivation: Combining e-processes for testing exchangeability
2. $\mathfrak{p}$-lifting and $\mathfrak{e}$-lifting: Lifting evidence into finer filtrations via adjusters
3. Experiments & further illustrative examples
4. Implications & new characterizations of adjusters for e-processes
5. Randomized adjustments for $\mathfrak{e}$-lifting

cf. Ramdas, Grünwald, Shafer, & Vovk (*Stat. Sci.*, 2023)

Evidence Measures for Anytime-Valid Inference

- $\mathbb{G} = (\mathscr{G}_t)_{t \geq 0}$: filtration
- τ : any $\mathbb{G}$-stopping time
- P : point null
- $\mathscr{P}$: composite null

$\mathfrak{e}_t = C(\mathfrak{p}_t)$ **(P-to-E Calibration)**

Test Supermartingale $(M_t)_{t\geq 0}$ for P w.r.t. $\mathbb{G}$	**E-Process** $(\mathfrak{e}_t)_{t\geq 0}$ for $\mathscr{P}$ w.r.t. $\mathbb{G}$	**P-Process** $(\mathfrak{p}_t)_{t\geq 0}$ for $\mathscr{P}$ w.r.t. $\mathbb{G}$
1. $(M_t)_{t\geq 0}$ is **adapted** to $\mathbb{G}$.	1. $(\mathfrak{e}_t)_{t\geq 0}$ is **adapted** to $\mathbb{G}$.	1. $(\mathfrak{p}_t)_{t\geq 0}$ is **adapted** to $\mathbb{G}$.
2. $M_0 = 1$ and $M_t \geq 0, \forall t$.	2. $\mathfrak{e}_0 = 1$ and $\mathfrak{e}_t \geq 0, \forall t$.	2. $\mathfrak{p}_0 = 1$ and $\mathfrak{p}_t \in [0, 1], \forall t$.
3. $\mathbb{E}_P[M_t \mid \mathscr{G}_{t-1}] \leq M_{t-1}, \forall t.$	3. $\mathbb{E}_P[\mathfrak{e}_\tau] \leq 1, \forall P \in \mathscr{P}$, for **any** $\mathbb{G}$-stopping time τ.	3. $P(\mathfrak{p}_\tau \leq \alpha) \leq \alpha, \forall P \in \mathscr{P}, \forall \alpha$, for **any** $\mathbb{G}$-stopping time τ.

Optional Stopping

$\mathfrak{p}_t = 1/\mathfrak{e}_t^*$ **(Ville's Inequality)**

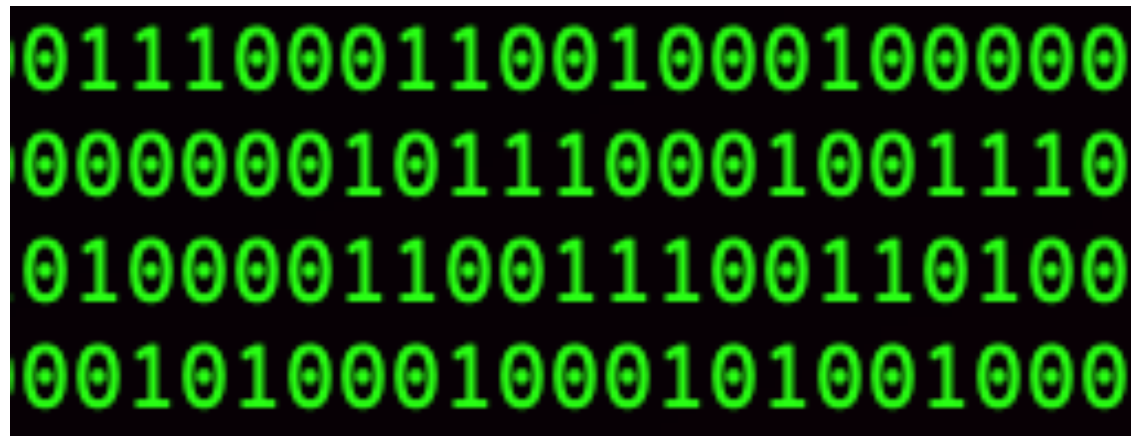
What goes wrong when combining e-processes across filtrations?

Example: Testing Exchangeability

"Is your data stream actually random?"

- We want to sequentially test whether a binary stream of data $X_1, X_2, \ldots$ is **exchangeable**:

$$\mathscr{P}^{\text{exch}} : X_1, X_2, \ldots \text{ is exchangeable.}$$



- This is a *composite* null for which no nontrivial test martingales exist in the data filtration.

- $(\mathfrak{e}_t)_{t \geq 0}$ is a nontrivial e-process for testing **randomness** ("Is the data i.i.d.?") **if and only if** it is a nontrivial e-process for testing exchangeability (Ramdas et al., *IJAR* 2022).

Example: Testing Exchangeability

"Is your data stream actually random?"

It turns out that there are two different methods to construct an e-process for $\mathscr{P}^{\text{exch}}$:

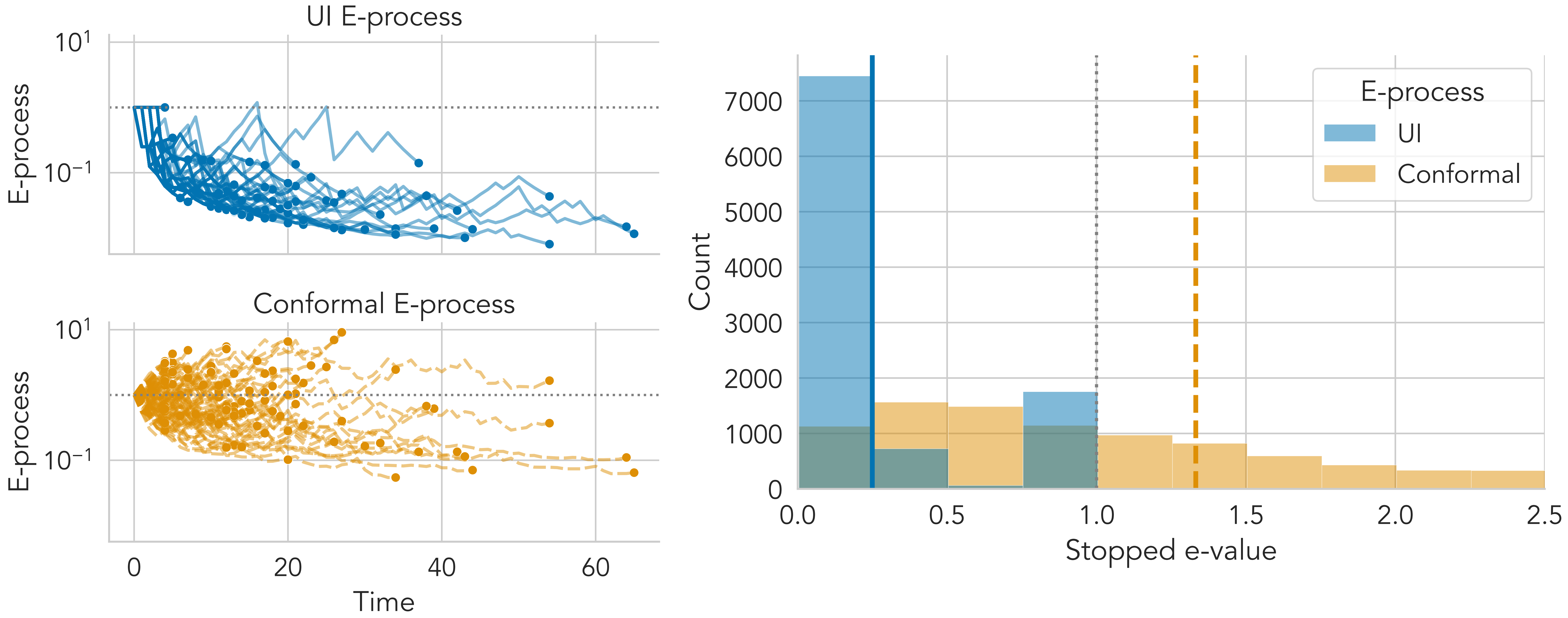
1. **Universal inference (UI) e-process** (Ramdas et al., 2022): $\mathfrak{e}_t^{\text{UI}} = \dfrac{\text{mixture over Markov alternatives}}{\text{maximum likelihood under null}}$.
    - Powerful against Markovian alternatives.
    - Anytime-valid w.r.t. the data ("full") filtration $\mathbb{F}$, $\mathscr{F}_t = \sigma(X_1, \ldots, X_t)$.
2. **Conformal test martingale** (Vovk, 2021): $\mathfrak{e}_t^{\text{conf}} = \prod_{i=1}^{t} \left[1 + \lambda\left(p_i - \frac{1}{2}\right)\right]$, where p_i are *conformal p-values.*
    - Powerful against changepoint alternatives.
    - **This e-process is ONLY anytime-valid w.r.t. a coarse filtration $\mathbb{G}$, $\mathscr{G}_t = \sigma(p_1, \ldots, p_t)$!**

E-process w.r.t. a coarse filtration is NOT anytime-valid in the data filtration

(in general)

Data: from i.i.d. Bernoulli.

$\tau^{\mathbb{F}}$ = first time we observe five consecutive 0's.



The conformal test martingale only has **"restricted" anytime-validity**, as it does NOT allow stopping w.r.t. the data filtration.

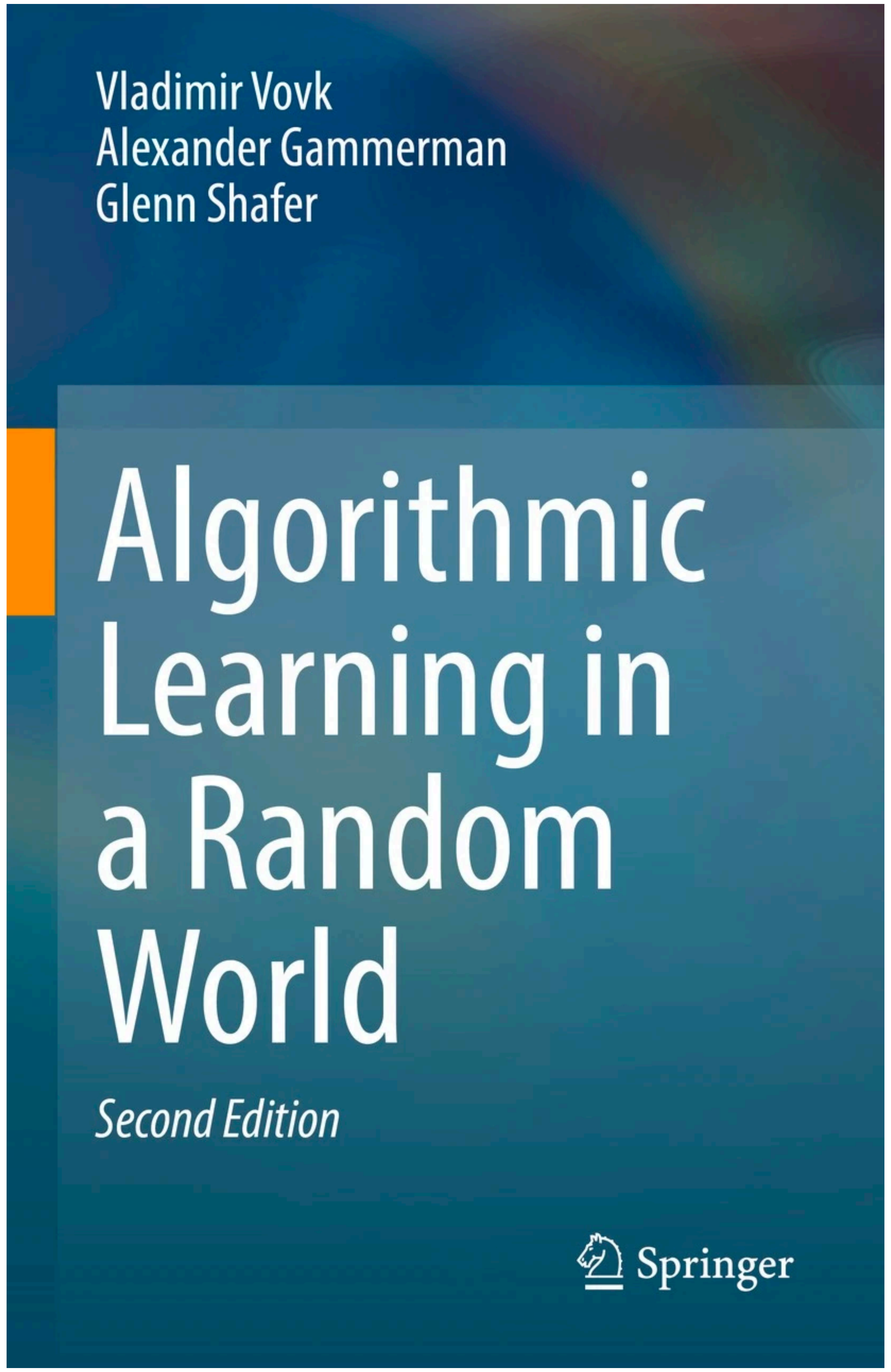
Over 10,000 repeated trials, $\hat{\mathbb{E}}_P[\mathfrak{e}^{\mathrm{conf}}_{\tau^{\mathbb{F}}}] \approx 1.33 \pm 0.02$.

We can't just average the two to obtain an e-process…

- What happens if we just try to take the average anyway?

$$m_t = \frac{1}{2}\left(\mathbf{e}_t^{\mathrm{UI}} + \mathbf{e}_t^{\mathrm{conf}}\right), \quad \forall t.$$

- In $\mathbb{F}$, $(m_t)_{t\geq 0}$ is **not** an e-process because $(\mathbf{e}_t^{\mathrm{conf}})_{t\geq 0}$ is not $\mathbb{F}$-anytime-valid, as we just saw.
- In $\mathbb{G}$, $(m_t)_{t\geq 0}$ is also **not** an e-process because $(\mathbf{e}_t^{\mathrm{UI}})_{t\geq 0}$ is not $\mathbb{G}$-adapted.
- **So, $(m_t)_{t\geq 0}$ is not an e-process w.r.t. *either* filtration.**



In fact, the average of randomized exchangeability martingales is not guaranteed to be a randomized exchangeability martingale. Martingales in a fixed filtration form a linear space, but different runs of the simplified Bayes–Kelly martingale are martingales in different filtrations, as discussed earlier. Figure 9.14 suggests that the averaged simplified Bayes–Kelly martingale is not an exchangeability martingale any more.

From Section 9.3 (p. 296)

Example: Testing a Scale-Invariant Gaussian Mean

From Pérez-Ortiz et al. (2022)

- Suppose the data $X_1, X_2, \ldots$ is sampled from $\mathcal{N}(\mu, \sigma^2)$, and let $\delta = \mu/\sigma$. Consider testing

$$\mathcal{H}_0 : \delta = \delta_0 \quad \text{vs.} \quad \mathcal{H}_1 : \delta = \delta_1.$$

- Let $\mathbb{F}$ be the "full" data filtration, and let $\mathbb{G}$ denote the scale-invariant **coarsening** of $\mathbb{F}$:

$$\mathcal{G}_t = \sigma\left(\frac{X_1}{|X_1|}, \ldots, \frac{X_t}{|X_1|}\right), \quad \forall t.$$

- In $\mathbb{G}$, a GROW e-process $(\mathfrak{e}_t)_{t \geq 0}$ for $(\mathcal{H}_0, \mathcal{H}_1)$ can be derived.
  However, it is also shown that this e-process is not anytime-valid w.r.t. $\mathbb{F}$:

If $\tau^{\mathbb{F}} = 1 + \mathbf{1}(|X_1| \in [0.44, 1.70])$, then $\mathbb{E}[\mathfrak{e}_{\tau^{\mathbb{F}}}] \approx 1.19 > 1$.

Main Goal

How can we combine e-processes **across different filtrations**?

(Especially, if the e-process in the coarser filtration isn't valid in the finer one.)

$\mathfrak{p}$-lifting & $\mathfrak{e}$-lifting:
Lifting evidence across filtrations

$\mathfrak{p}$-lifting: P-processes can be lifted "freely"

Theorem ($\mathfrak{p}$-lifting). Suppose $\mathbb{G} \subseteq \mathbb{F}$, and let $(\mathfrak{p}_t)_{t \geq 0}$ be a p-process for $\mathscr{P}$ **w.r.t. $\mathbb{G}$.** Then,

$(\mathfrak{p}_t)_{t \geq 0}$ is a p-process for $\mathscr{P}$ **w.r.t. $\mathbb{F}$.**

- This result follows from the so-called **lifting lemma**, which is an extension of the "equivalence lemma" (Ramdas et al., 2020; Howard et al., 2021) to pairs of filtrations.
- Essentially, any "probability statements" of anytime-validity translate across filtrations.
  - In contrast, analogous "expectation statements" of anytime-validity do not translate.

The Equivalence Lemma

Ramdas et al. (2020); Howard et al. (2021)

Let $(\xi_t)_{t \geq 1}$ be a sequence of events adapted to a filtration $\mathbb{G}$. (E.g., $\xi_t = \{\mathfrak{p}_t \leq \alpha\}$.)

Given any probability P and any $\alpha \in (0, 1)$, the following statements are equivalent:

(a) **Time-uniform validity**: $P\left(\cup_{t \geq 1} \xi_t \right) \leq \alpha$.

(b) **Random time validity**: for any (possibly infinite) random time T, $P(\xi_T) \leq \alpha$.

(c) **$\mathbb{G}$-anytime-validity**: for any (possibly infinite) $\mathbb{G}$-stopping time $\tau^{\mathbb{G}}$, $P(\xi_{\tau^{\mathbb{G}}}) \leq \alpha$.

The Lifting Lemma

Adaptation of the equivalence lemma to two filtrations

Let $(\xi_t)_{t \geq 1}$ be a sequence of events adapted **to a sub-filtration $\mathbb{G} \subseteq \mathbb{F}$.**

Given any probability P and any $\alpha \in (0, 1)$, the following statements are equivalent:

(a) $\mathbb{G}$-**anytime-validity**: for any (possibly infinite) $\mathbb{G}$-stopping time $\tau^{\mathbb{G}}$, $P(\xi_{\tau^{\mathbb{G}}}) \leq \alpha$.

(b) $\mathbb{F}$-**anytime-validity**: for any (possibly infinite) $\mathbb{F}$-stopping time $\tau^{\mathbb{F}}$, $P(\xi_{\tau^{\mathbb{F}}}) \leq \alpha$.

Implication: anytime-validity of any p-process w.r.t. $\mathbb{G}$ => anytime-validity w.r.t. $\mathbb{F}$.

$\mathfrak{e}$-lifting: Lifting e-processes via adjusters

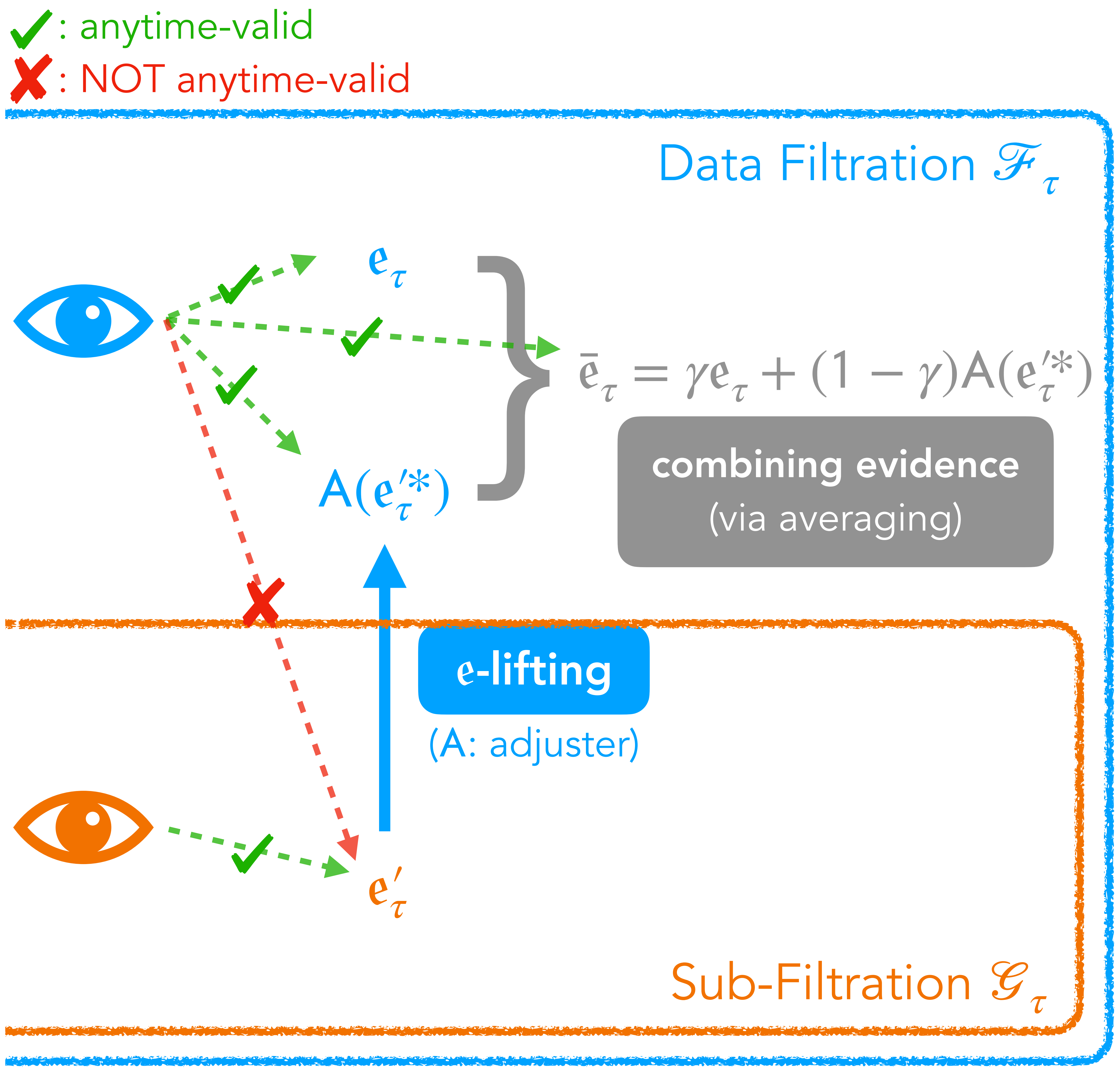
Theorem ($\mathfrak{e}$-lifting). Let A be an **adjuster** (to be defined soon). Suppose $\mathbb{G} \subseteq \mathbb{F}$, and let $(\mathfrak{e}_t)_{t\geq 0}$ be an e-process for $\mathscr{P}$ **w.r.t. $\mathbb{G}$.** Then

$$\left(A(\mathfrak{e}_t^*)\right)_{t\geq 0} \text{ is an e-process for } \mathscr{P} \textbf{ w.r.t. } \mathbb{F}. \qquad (\mathfrak{e}_t^* = \sup_{i\leq t} \mathfrak{e}_i)$$

Proof:

1. By Ville's inequality, $\mathfrak{p}_t = 1/\mathfrak{e}_t^*$ is a p-process for $\mathscr{P}$ **w.r.t. $\mathbb{G}$.**
2. By **$\mathfrak{p}$-lifting**, $\mathfrak{p}_t = 1/\mathfrak{e}_t^*$ is also a p-process for $\mathscr{P}$ **w.r.t. $\mathbb{F}$.**
3. The adjuster has a corresponding **p-to-e calibrator** C, such that $A(\mathfrak{e}) = C(1/\mathfrak{e})$, $\forall \mathfrak{e} \geq 1$.
4. Thus, $\mathfrak{e}_t^{\text{adj}} = C(\mathfrak{p}_t) = A(\mathfrak{e}_t^*)$ is an e-process for $\mathscr{P}$ **w.r.t. $\mathbb{F}$.**

Combining evidence across filtrations via e-lifting



What are adjusters?

a.k.a. lookback adjusters & martingale calibrators

- An increasing, right-continuous function $A : [1, \infty] \to [0, \infty]$ is an **adjuster** if it satisfies:

$$\int_1^\infty \frac{A(e)}{e^2} de \leq 1.$$

- It is *admissible* if the above holds with equality and $A(\infty) = \infty$.

- Adjusters allow betting on the **running maximum** of a test supermartingale (or a capital process).

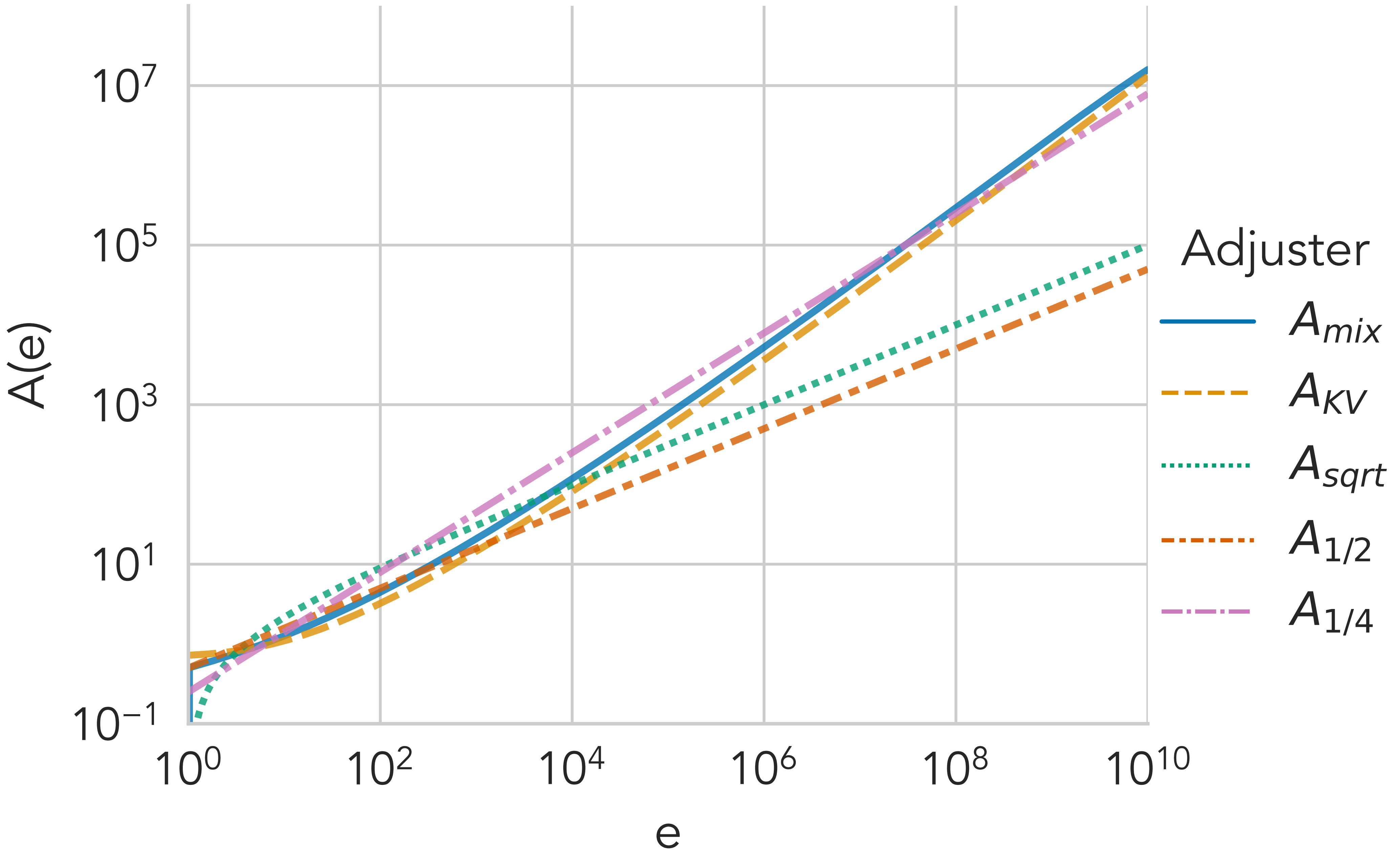
Admissible adjusters:

$$A_{mix}(e) = \frac{e - 1 - \log(e)}{\log^2(e)}$$

$$A_{KV}(e) = \frac{e^2 \log(2)}{(1+e)\log^2(1+e)}$$

$$A_{sqrt}(e) = \sqrt{e} - 1$$

$$A_\kappa(e) = \kappa e^{1-\kappa}, \ \kappa \in (0, 1)$$



Dawid et al. (2011a;b), Shafer et al. (2011); Koolen & Vovk (2014)

Adjusters $\Longleftrightarrow$ P-to-E Calibrators

- A decreasing, left-continuous function $C : [0,1] \to [0,\infty]$ is a **(p-to-e) calibrator** if

$$\int_0^1 C(\mathfrak{p})d\mathfrak{p} \leq 1.$$

- It is *admissible* if the above holds with equality.

- There is a straightforward **1-to-1 correspondence between calibrators and adjusters.** Setting $A(\mathfrak{e}) = C(1/\mathfrak{e})$, and by change-of-variables ($\mathfrak{p} = 1/\mathfrak{e}$),

$$\int_1^\infty \frac{A(\mathfrak{e})}{\mathfrak{e}^2}d\mathfrak{e} = \int_1^\infty \frac{C(1/\mathfrak{e})}{\mathfrak{e}^2}d\mathfrak{e} = \int_0^1 C(\mathfrak{p})d\mathfrak{p} \underset{(=)}{\leq} 1.$$

cf. Shafer et al. (2011); Vovk & Wang (2021)

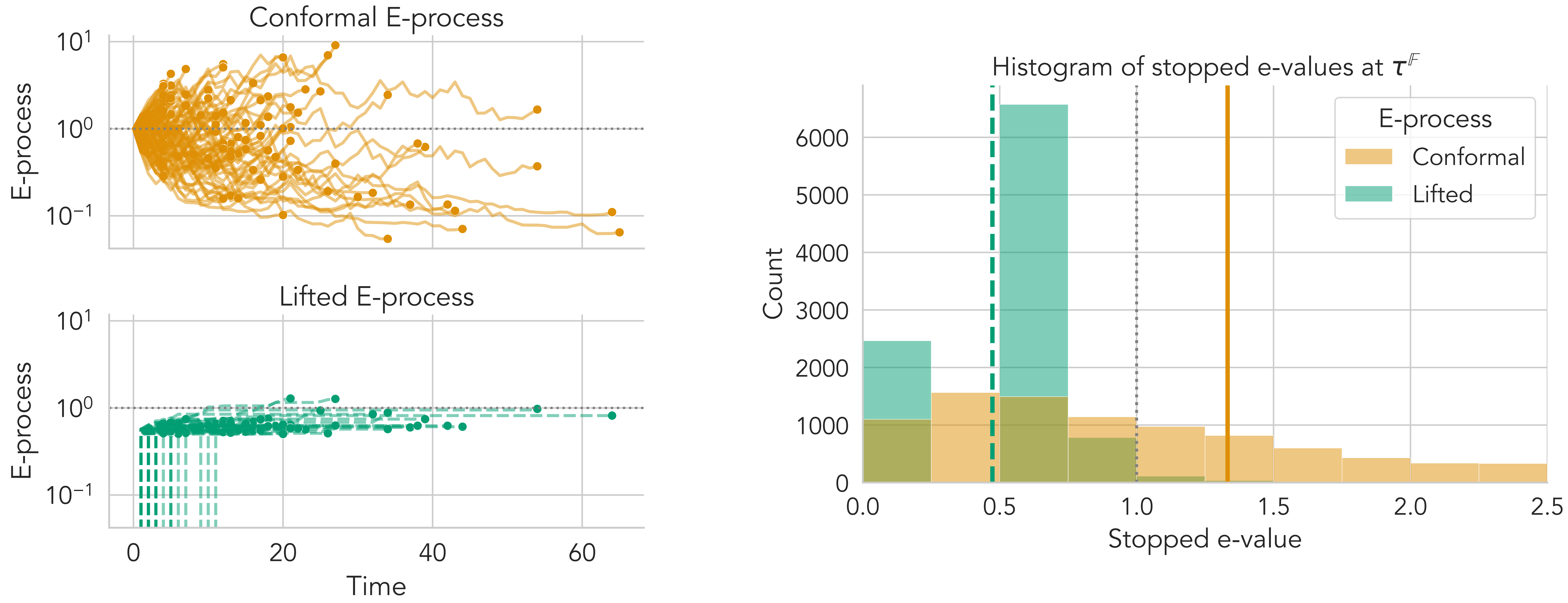
Experiments & illustrative examples

Testing Exchangeability: Null Case

The lifted e-process is $\mathbb{F}$-anytime-valid, so we can now combine it with the UI e-process.

Data: from i.i.d. Bernoulli.

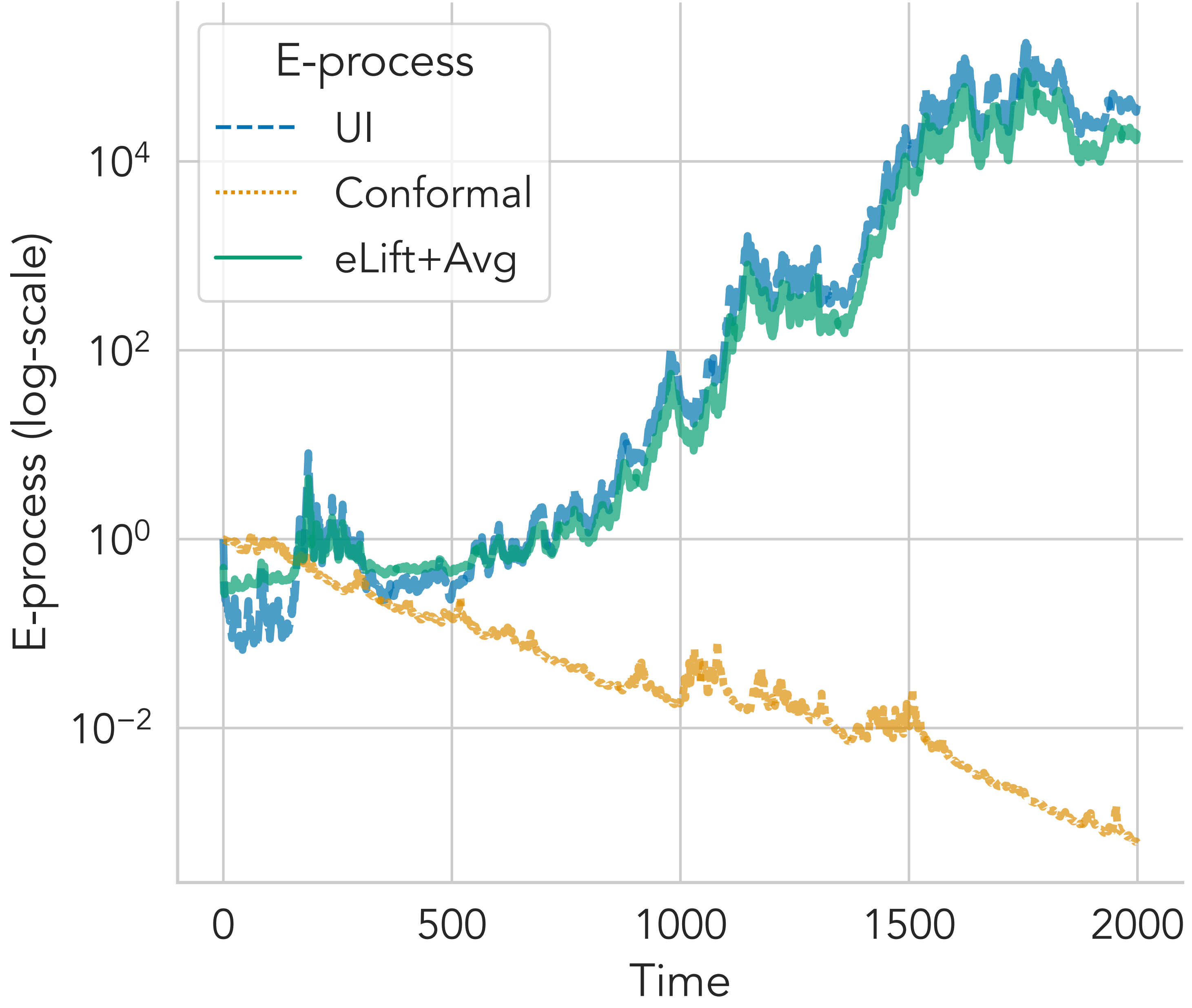
$\tau^{\mathbb{F}}$ = first time we observe five consecutive 0's.



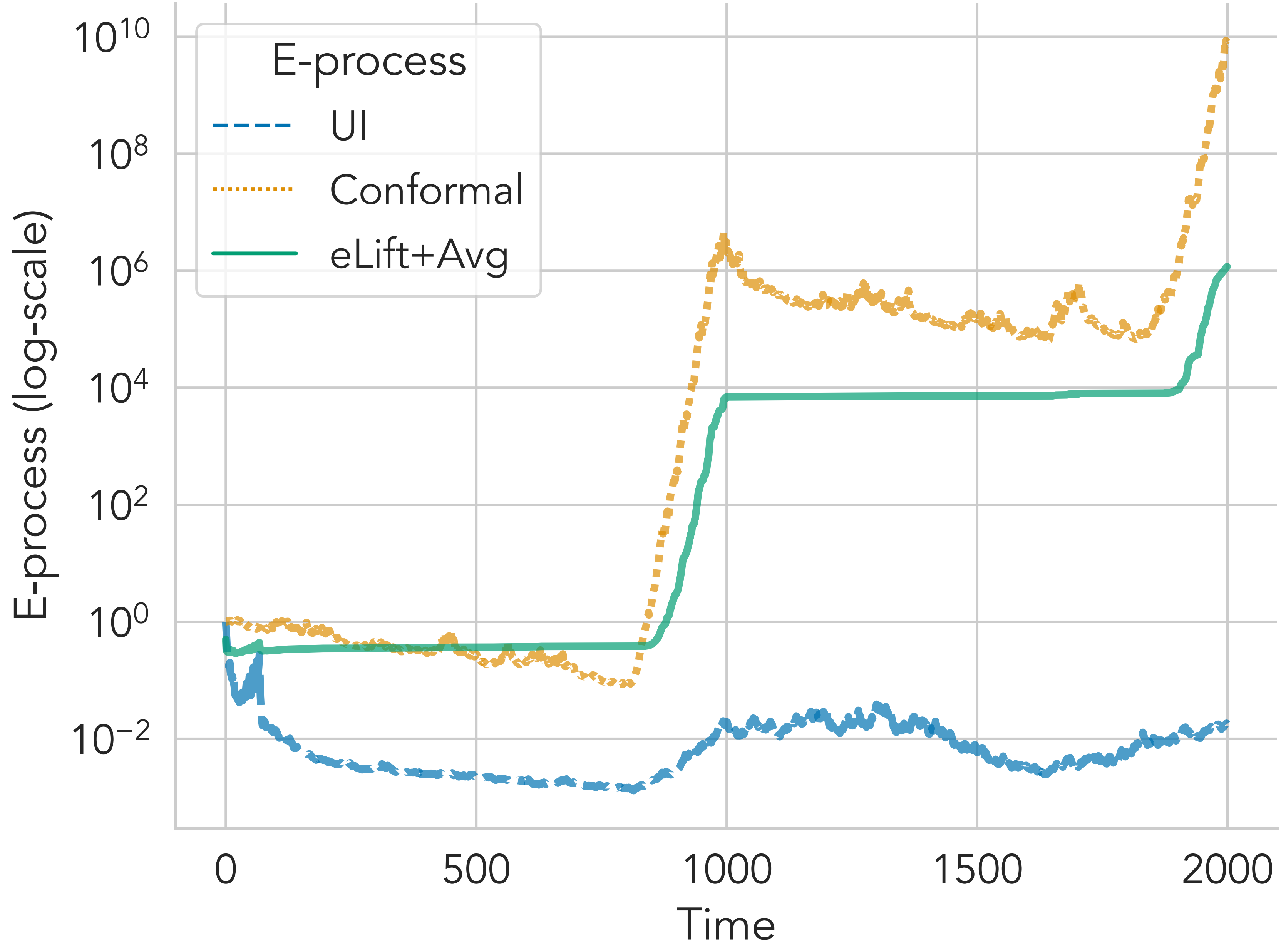
Across 10,000 simulations, $\mathbb{E}_P[A(e^*_{\tau^{\mathbb{F}}})] \approx 0.47 \leq 1.$

Testing Exchangeability: Alternative Case

Combined ("eLift+Avg"): $\bar{\mathfrak{e}}_t = \frac{1}{2}\left[\mathfrak{e}_t^{\mathrm{UI}} + A((\mathfrak{e}_t^{\mathrm{conf}})*)\right]$



Alternative #1: First-order Markov



Alternative #2: Two changepoints

The combined e-process achieves power against both alternatives.

Other topical examples in the literature

1. **Multi-step forecast evaluation/comparison.** When evaluating sequential forecasters making their forecasts $h > 1$ days ahead of time, we'd want to evaluate them **conditioned on the information that they had at the time of forecasting.**

    - For each offset $k = 1, \ldots, h$, there exists an e-process $(e_t^{[k]})_{t \geq 0}$ w.r.t. *different coarsenings of the data filtration*, say $\mathbb{G}^{[k]} \subsetneq \mathbb{F}$.

    - To obtain an evaluation across all offsets, we would need to $\mathfrak{e}$-lift all h e-processes!

2. **Sequential independence testing.** For testing independence sequentially, there is no nontrivial test martingale w.r.t. the data filtration (Henzi & Law, 2023).

    - When combining e-processes for this null, existing e-processes have to be $\mathfrak{e}$-lifted.

[1]Henzi & Ziegel (2022); Arnold et al., (2022); Choe & Ramdas (2023)

[2]Balasubramani & Ramdas (2016); Shekhar & Ramdas (2023); Podkopaev et al. (2023); Henzi & Law (2023)

Implications & new characterizations of adjusters for e-processes

Implications on coarsening the filtration

1. Are there testing problems for which there is no powerful **e-process** in the data filtration, but there exist ones in a coarser filtration?

   - *Unlike in the case of test supermartingales,* $\mathfrak{e}$-lifting implies that the answer is **no**!
   - If there exists a powerful e-process $(\mathfrak{e}_t)_{t\geq 0}$ for $\mathscr{P}$ w.r.t. *some* $\mathbb{G} \subseteq \mathbb{F}$, then we can $\mathfrak{e}$-lift it to $\mathbb{F}$.
     - If $\limsup_{t\to\infty} \mathfrak{e}_t = \infty$ under some $\mathscr{Q}\backslash\mathscr{P}$, then $\limsup_{t\to\infty} A(\mathfrak{e}_t^*) = \infty$ under $\mathscr{Q}\backslash\mathscr{P}$ (for any admissible A), so the adjusted e-process is also powerful.

2. There appears to be an **unavoidable cost** to coarsening the filtration to obtain an e-process.

   - The original e-process is not *truly* immune to "data peeking."
   - It appears that it is necessary to sacrifice some of the evidence (via adjusters).

Is adjusting the e-process the only way?

- Suppose you claim to have a function that, if I give you some e-process w.r.t. a coarse filtration, then the function can transform it into an e-process w.r.t. the data filtration. (The e-process can be for any null w.r.t. any filtration.)

- Is the function necessarily an adjuster?

- **<u>Theorem (informal):</u>** The function is ***necessarily*** an adjuster, as long as it is an increasing function that maps the running maximum $\mathfrak{e}_t^*$ to some $\mathfrak{e}_t'$ for each t.

A game-theoretic definition of adjusters

How can we make betting on the running maximum a "fair game"?

- An increasing function A is an adjuster **if and only if**, for every test supermartingale $(M_t)_{t\geq 0}$ for some P, there exists a test supermartingale $(M'_t)_{t\geq 0}$ for P s.t.

$$A(M^*_t) \leq M'_t, \quad \forall t.$$

A is an "adjuster for test supermartingales"

- Game-theoretically, adjusters allow betting with the **running maximum** of the gambler's wealth.
  - A is an adjuster if and only if, in Protocol 1, Rival Skeptic has a betting strategy to ensure that

$$A(\mathscr{K}^*_t) \leq \mathscr{K}'_t.$$

Protocol 1 Competitive scepticism

$\mathcal{K}_0 := 1$ and $\mathcal{K}'_0 := 1$
for $n = 1, 2, \ldots$ **do**
  Forecaster announces $\mathcal{E}_n \in \mathbf{E}$
  Sceptic announces $f_n \in [0, \infty]^{\mathcal{X}}$ such that $\mathcal{E}_n(f_n) \leq \mathcal{K}_{n-1}$
  Rival Sceptic announces $f'_n \in [0, \infty]^{\mathcal{X}}$ such that $\mathcal{E}_n(f'_n) \leq \mathcal{K}'_{n-1}$
  Reality announces $x_n \in \mathcal{X}$
  $\mathcal{K}_n := f_n(x_n)$ and $\mathcal{K}'_n := f'_n(x_n)$
end for

Dawid et al. (2011a;b); Shafer et al. (2011)

A Characterization Theorem for Adjusters

Theorem. Let $A : [1, \infty] \to [0, \infty]$ be an increasing function. The following are equivalent:

(a) A is an adjuster, i.e., it satisfies $\displaystyle\int_1^\infty \frac{A(\mathfrak{e})}{\mathfrak{e}^2} d\mathfrak{e} \leq 1$.

(b) A is an "adjuster for test supermartingales" (previous slide).

(c) A is an **"adjuster for e-processes"**: for any e-process $(\mathfrak{e}_t)_{t\geq 0}$ for some $\mathscr{P}$ w.r.t. $\mathbb{G}$, there exists another e-process $(\mathfrak{e}'_t)_{t\geq 0}$ for $\mathscr{P}$ w.r.t. $\mathbb{G}$ such that, for all t, $A(\mathfrak{e}^*_t) \leq \mathfrak{e}'_t$.

(d) A is an **"$\mathfrak{e}$-lifter"**: for any e-process $(\mathfrak{e}_t)_{t\geq 0}$ for some $\mathscr{P}$ w.r.t. $\mathbb{G}$, **and *any* finer filtration $\mathbb{F} \supseteq \mathbb{G}$,** $(A(\mathfrak{e}^*_t))_{t\geq 0}$ is an e-process for $\mathscr{P}$ **w.r.t.** $\mathbb{F}$.

(e) For any e-process $(\mathfrak{e}_t)_{t\geq 0}$ for some $\mathscr{P}$ w.r.t. $\mathbb{G}$, $(A(\mathfrak{e}^*_t))_{t\geq 0}$ is an e-process for $\mathscr{P}$ w.r.t. $\mathbb{G}$.

Takeaways

Takeaways

1. E-processes constructed on a coarse filtration are **not** anytime-valid in the data filtration, so they cannot be combined seamlessly with other e-processes.
   - Examples: testing exchangeability; independence; comparing multi-step forecasters
2. P-processes can be lifted freely across filtrations and retain their anytime-validity.
3. For e-processes, we can use **adjusters** to achieve validity in the data filtration.
4. In a sense, any function that lifts the anytime-validity of an e-process **must be** an adjuster.
5. U-randomization can be applied *after* adjustment, but *not* before (w/o hurting validity).

Thank You

ArXiv: https://arxiv.org/abs/2402.09698
Code: to be available soon
YJ: https://yjchoe.github.io/
Aaditya: https://www.stat.cmu.edu/~aramdas/

Questions?

Appendix

Example: Comparing Multi-Step Sequential Forecasters

- Suppose we compare two sequential forecasters with lag h using some scoring rule S w.r.t. $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$:

$$\Delta_t^{[k]} = \frac{1}{|I_t^{[k]}|} \sum_{i \in I_t^{[k]}} \mathbb{E}\left[S(p_i, y_{i+h-1}) - S(q_i, y_{i+h-1}) \mid \mathcal{F}_{i-1}\right], \quad \forall k \in [h].$$

  - If $h = 2$, $\Delta_t^{[0]}/\Delta_t^{[1]}$ measures the **average forecast score difference** on **even/odd** days.

- When testing for the null $\mathcal{H}_0^{[k]} : \Delta_t^{[k]} \leq 0, \forall t$, for each offset k, we need to construct an e-process $(\mathfrak{e}_t^{[k]})_{t \geq 0}$ **under different coarsening of the filtration** $\mathbb{F}$ **for each** k (updates on every even/odd days).

Each $(\mathfrak{e}_t^{[k]})_{t \geq 0}$ is an e-process for $\mathcal{H}_0^{[k]}$, but only w.r.t. the sub-filtration $\mathbb{G}^{[k]} \subsetneq \mathbb{F}$.

- To test for **the combined null** $\mathcal{H}_0 : \Delta_t^{[k]} \leq 0, \forall t, \forall k$ (an intersection), we want to **e-lift** all h e-processes into the data filtration $\mathbb{F}$ before combining them:

$$\bar{\mathfrak{e}}_t = \frac{1}{h} \sum_{k=1}^{h} A\left((\mathfrak{e}_t^{[k]})^*\right), \forall t.$$

Henzi & Ziegel (2022)
Arnold et al. (2022)
Choe & Ramdas (2023)

Example: Testing Independence

- Given an i.i.d. stream of paired data $Z_t = (X_t, Y_t) \sim P_{XY}$, suppose we test if the joint distribution factorizes:

$$\mathscr{H}_0 : P_{XY} = P_X \times P_Y \quad \text{vs.} \quad \mathscr{H}_1 : P_{XY} \neq P_X \times P_Y.$$

- Similar to the exchangeability null, there exist no nontrivial test martingale adapted to the data filtration $\mathbb{F}$. Two known e-processes include:
    - **Pairwise betting** (SR'23; PBKR'23; SR'24): adapted to the filtration w/ pairs of data.
    - **Rank-based test martingale** (HL'23): adapted to the filtration w/ rank stats of data.
- In this case, BOTH e-processes are constructed w.r.t. their own, non-overlapping sub-filtrations. So we should lift both of them before taking the average.

cf. Balasubramani & Ramdas (2016); Shekhar & Ramdas (2023); Podkopaev et al. (2023); Henzi & Law (2023)

Randomized adjusters for $\mathfrak{e}$-lifting

Motivation: Randomized Calibrators

- In the case of (non-sequential) **e-to-p calibration**, it is known that $p = 1/e$ is the only admissible deterministic mapping.

- Recent papers show that there is the following **"U-randomized" e-to-p calibrator** can dominate the deterministic variant (almost surely):

$$\tilde{p} = U/e, \text{ for some } U \gtrsim \text{Unif}[0,1].$$

- Can we leverage this idea to develop a **U-randomized adjuster**?

cf. Ignatiadis et al. (2023); Ramdas & Manole (2023)

Strategy #1: Lift-then-randomize ("ltr")

This is possible via UMI!

***UMI:**
for any $X \geq 0$ and $U \gtrsim \mathrm{Unif}[0, 1]$,
$$P\left(X \geq \frac{U}{\alpha}\right) \leq \alpha \cdot \mathbb{E}[X].$$

$$\mathbf{e}_\tau \xrightarrow[\text{(e-to-e)}]{\text{e-lifting}} A(\mathbf{e}^*_\tau) \xrightarrow[\text{(e-to-p)}]{\text{U-rand.}} \tilde{p}^{\mathrm{ltr}}_\tau = \frac{U}{A(\mathbf{e}^*_\tau)} \wedge 1$$

- Once we lift an e-process by adjustment, we have an e-process in $\mathbb{F}$.
- So the randomization strategy still works: $\tilde{p}^{\mathrm{ltr}}_\tau$ **is a valid p-value for any $\mathbb{F}$-stopping time τ,** due to "UMI" (uniformly-randomized Markov's inequality; Ramdas & Manole, 2023).
- Since we end up with a p-value, this is only practically useful if the stopping rule $\tilde{p}^{\mathrm{ltr}}_\tau \leq \alpha$ is more lenient than the $\mathfrak{p}$-lifted stopping rule of $\mathfrak{p}_\tau \leq \alpha$ (w/o adjustment).

Strategy #2: Randomize-then-lift ("rtl")

This does not guarantee anytime-validity (empirically, at least)

$$\mathfrak{e}_\tau \xrightarrow[\mathsf{U} \gtrsim \mathrm{Unif}[0,1]]{\text{rand. e-to-p}} \tilde{\mathfrak{p}}_\tau^{\mathrm{rtl}} := \frac{\mathsf{U}}{\mathfrak{e}_\tau} \wedge 1 \xrightarrow{\text{p-to-e}} \tilde{\mathfrak{e}}_\tau^{\mathrm{rtl}} := C\left(\frac{\mathsf{U}}{\mathfrak{e}_\tau} \wedge 1\right)$$

- Once we add the external r.v. U, the resulting sequence is **not adapted** to $\mathbb{G}$!
- <u>**The U-lifting "lemma":**</u> If $(\mathfrak{e}_t)_{t\geq 0}$ is an e-process for $\mathscr{P}$ w.r.t. $\mathbb{G} \subseteq \mathbb{F}$, $\mathsf{U} \gtrsim \mathrm{Unif}[0,1]$, and $\alpha \in (0,1)$, we have, for any $\mathbb{F}$-stopping time τ and any $\mathsf{P} \in \mathscr{P}$,

$$\mathsf{P}\left(\tilde{\mathsf{p}}_\tau^{\mathrm{rtl}} \leq \alpha\right) = \mathsf{P}\left(\mathfrak{e}_\tau \geq \frac{\mathsf{U}}{\alpha}\right) \leq \alpha\left[\underbrace{\mathbb{E}_{\mathsf{P}}[\mathfrak{e}_\tau]} \wedge \underbrace{\left(1 + \log(1/\alpha)\right)}\right].$$

The UMI bound (Empirically observed; ≈ 0.065 for conformal mtg.)

A loose type I error control using (non-randomized) Markov (≈ 0.20 if $\alpha = 0.05$)

End of Slides