Two general frameworks for sequential change detection with composite, nonparametric classes

Aaditya Ramdas

Dept. of Statistics and Data Science (75%)
Machine Learning Dept. (25%)
Carnegie Mellon University

Amazon Research (20%)

Outline of this lecture series

- ✓ Yesterday: game-theoretic testing
- ✓ Today morning: game-theoretic estimation

3. Now: game-theoretic change detection

Super high-level problem setup

You observe data (scalars or vectors or objects) one at a time.

You calculate **some statistic** of the data as you go along. ("CUSUM", "SR")

When the statistic crosses **a threshold**, you proclaim a change.

Aim: Minimize detection delay while controlling false alarms.

Metrics of success

Average run length (ARL) or Frequency of False Alarms: this is the analog of "Type-I error" from testing

Definition: When there is no change point, the ARL is the expected number of steps before we (falsely) proclaim a change point.

Eg: maybe we are ok with a false alarm roughly every 1000 steps.

Detection delay (DD): analog of "Type-II error"

Definition: When there is a change point at some time ν, DD is the expected number of steps after ν that we need to (correctly) proclaim a change point.

Critical issue in practice

Controlling the ARL is very hard!

If the pre-change distribution is perfectly known,
then controlling ARL is easy: set threshold via offline **simulation**,
or use **analytical calculations** (asymptotics) to find threshold.

Cannot do this if the pre-change data is only known to lie
in some *set* of distributions
i.e. we only know some partial aspects of pre-change distribution,
(Eg: we know variance < 10, but not sure what the exact value is)

If we do not know how to model some aspects of the pre-change
distribution (i.e. "nonparametric")
(Eg: we know the data is bounded, but we don't know much else)
In such cases, it is much harder to control the ARL.
Simulation and math both don't easily work.

Two recent reductionist works

Using "E-detectors"

(NEJSDS'23)

When pre- and post-change distributions are known to lie in separate classes $\mathscr{P}, \mathscr{Q}$ (eg: mean change from <0 to >0)

Key idea: reduce change detection to repeated sequential testing of $\mathscr{P}$ vs. $\mathscr{Q}$

Method: at every time t, start a new level-α **sequential test** based on $X_t, X_{t+1}, \ldots$ and declare change when accumulated evidence crosses $1/\alpha$

Main theorem: ARL $\geq 1/\alpha$

Main theorem2: "Optimal" detection delays

Via "confidence sequences"

(arXiv)

When pre- and post-change distributions are unknown but from the same class (eg: the mean changed from a to b)

Key idea: reduce change detection to repeated sequential estimation

Method: at every time t, start a new level-α **confidence sequence** based on $X_t, X_{t+1}, \ldots$ and declare change when their intersection is empty.

Main theorem: ARL $\geq 1/\alpha$

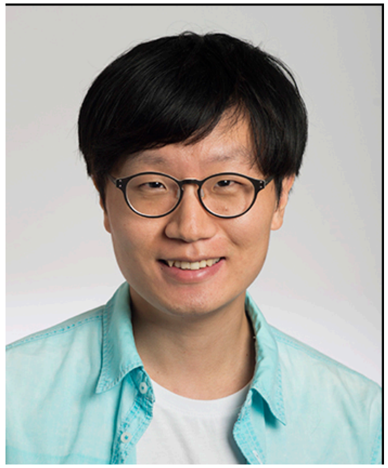
Main theorem2: "Optimal" detection delays

Advantage: sequential testing/estimation are basic problems, increasingly well studied.

Today's talk

1. Reducing partitioned change detection to sequential testing
2. Reducing non-partitioned change detection to sequential estimation
3. High level summary of game-theoretic statistics

E-detectors



Jaehyeok
Shin
(Google)

Alessandro
Rinaldo
(CMU / UT Austin)

TLDR: Reducing sequential (nonparametric) change detection to (modern) sequential testing

Recap: Lorden's reduction to testing (1970)

What is a "one-sided" level-α **sequential test** of $\mathscr{P}$ vs. $\mathscr{Q}$?

$H_0 : X_1, X_2, \ldots \sim P \in \mathscr{P}$ vs. $H_1 : X_1, X_2, \ldots \sim Q \in \mathscr{Q}$

For each $t \geq 1$, after seeing X_t we output 0 or 1.

0 means "not enough evidence to reject $\mathscr{P}$, collect more data".

1 means "we have enough evidence to reject $\mathscr{P}$".

A level-α sequential test ϕ satisfies

$P(\exists t \in \mathbb{N} : \phi(X_1, \ldots, X_t) = 1) \leq \alpha$ for all $P \in \mathscr{P}$.

Lorden's Reduction: at every time t, start a *new* level α sequential test based on $X_t, X_{t+1}, \ldots$ (so there are t tests running after t steps). We declare a change when any of them rejects the null.

Key theorem: This method controls ARL (FFA) at level $1/\alpha$.

TLDR of our paper: more sophisticated reduction to sequential testing.

Idea: accumulate the evidence across tests.

Quantifying evidence using e-processes

Consider sequential testing $H_0 : P \in \mathscr{P}$

An e-process for $\mathscr{P}$ is a nonnegative process $\Lambda \equiv (\Lambda_n)_{n \geq 1}$ such that $\mathbb{E}_P[\Lambda_\tau] \leq 1$ for all $P \in \mathscr{P}, \tau \in \mathscr{T}$.

We reject the null whenever the "evidence" Λ_t exceeds $1/\alpha$.

E-processes are nonparametric, composite generalizations of likelihood ratios and have a nice game-theoretic foundation.

Vovk, Shafer, Grunwald, R,
Wang, Larsson, Koolen, Ruf, Howard, etc. (last 5 years)

Define the SR e-detector as $M_n^{SR} := \sum_{j=1}^{n} \Lambda_n^{(j)}$,
where $\Lambda^{(j)}$ is a $\mathscr{P}$-e-process started at time j.

E-detector

$\mathcal{P}$: pre-change set of distributions over infinite sequence

$\mathcal{F} \equiv (\mathcal{F}_n)_{n\geq 1}$: filtration, and $\mathcal{F}_n$ need not be $\sigma(X_1, \ldots, X_n)$.

An e-detector for $\mathcal{P}$ is a nonnegative process $M \equiv (M_n)_{n\geq 1}$

such that $\mathbb{E}_P[M_\tau] \leq \mathbb{E}_P[\tau]$ for all $P \in \mathcal{P}, \tau \in \mathcal{T}$.

$\mathcal{T}$: set of all stopping times with respect to $\mathcal{F}$

Nonasymptotic ARL control

Define stopping time $N^*_{1/\alpha} := \inf\{t \geq 1 : M_t \geq 1/\alpha\}$.

Theorem: If M is an e-detector for $\mathcal{P}$, then
the ARL is $\mathbb{E}_P[N^*_{1/\alpha}] \geq 1/\alpha \ \forall P \in \mathcal{P}$

Computationally efficient e-detectors using "baseline" e-processes

Often, our e-process can be written as $\Lambda_n^{(j)} := \prod_{i=j}^{n} L_i$

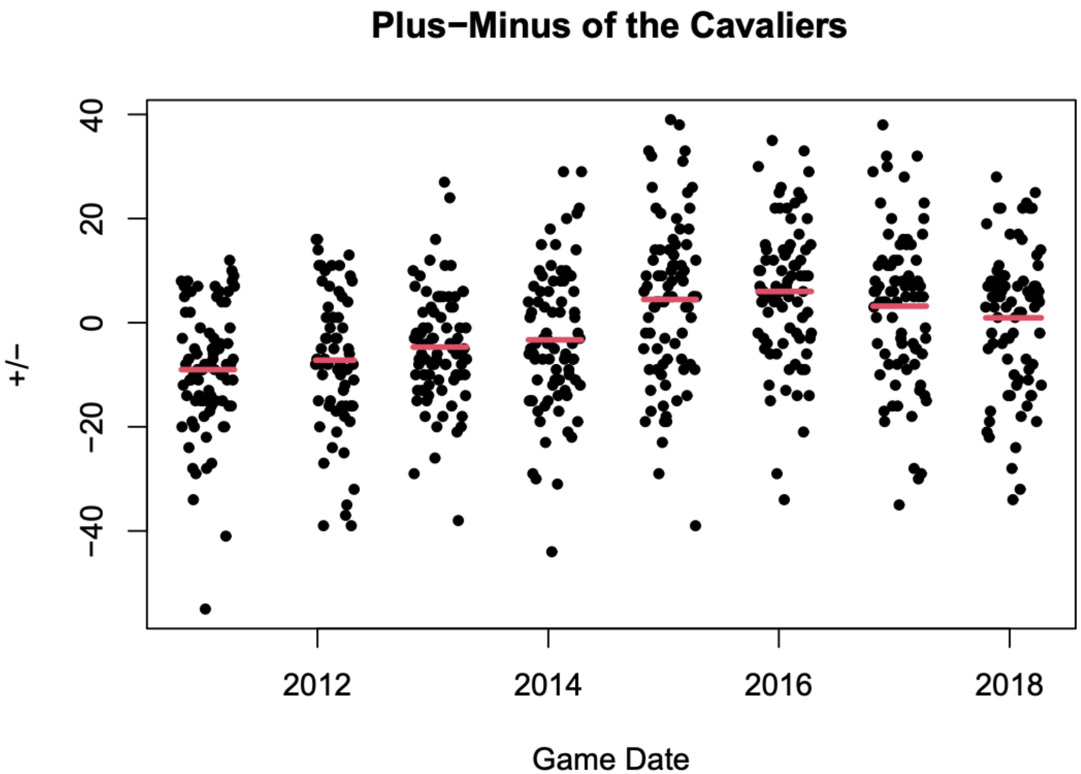
where $\mathbb{E}_P[L_n \mid \mathcal{F}_{n-1}] \leq 1$ for all $P \in \mathcal{P}$.

Then, $M_n^{SR} := L_n(M_{n-1}^{SR} + 1)$ can be computed online, like the Shiryaev-Roberts procedure.

Example: likelihood ratios, if $\mathcal{P} = \{P\}$ is a point null, but also any setting with a nonnegative supermartingale for composite $\mathcal{P}$.

Cleveland Cavaliers (NBA, 2011-18)

Plus-minus = Points scored - Points conceded



Is there a changepoint from a negative to a positive plus-minus?

E-detector: a nonparametric framework to answer such questions.

The paper: related work (Lorden, Pollak, Siegmund, Tartakovsky, Veeravalli, Xie, Harchaoui, and many others)

Why is it hard?

Challenges: the plus-minus stat of a game is a bounded r.v. between [-100, +100] (let's rescale to [0,1]) and its mean varies over time (with form, injuries, etc), and we know nothing else about the distribution.

$X_1, X_2, \ldots \sim P$ for some $P \in \mathscr{P}$

where $\mathscr{P} :=$ $\{P$ on $[0,1]^\infty$: $\mathbb{E}[X_n \,|\,$ past$] \leq 1/2$ for all $n \geq 1\}$

If there is a change at time ν, then $X_\nu, X_{\nu+1}, \ldots \sim Q$ for some $Q \in \mathscr{Q}$

where $\mathscr{Q} :=$ $\{Q$ on $[0,1]^\infty$: $\mathbb{E}[X_n \,|\,$ past$] > 1/2$ for all $n \geq 1\}$

Here P, Q are distributions on *infinite* sequences of observations, so the data are not iid (neither i nor id), and are only restricted by their conditional means

Method

$X_1, X_2, \ldots \sim P$ for some $P \in \mathscr{P}$
where $\mathscr{P} := \{P$ on $[0,1]^\infty : \mathbb{E}[X_n \mid \text{past}] \leq 1/2$ for all $n \geq 1\}$

If there is a change at time ν, then $X_\nu, X_{\nu+1}, \ldots \sim Q$ for some $Q \in \mathscr{Q}$
where $\mathscr{Q} := \{Q$ on $[0,1]^\infty : \mathbb{E}[X_n \mid \text{past}] > 1/2$ for all $n \geq 1\}$

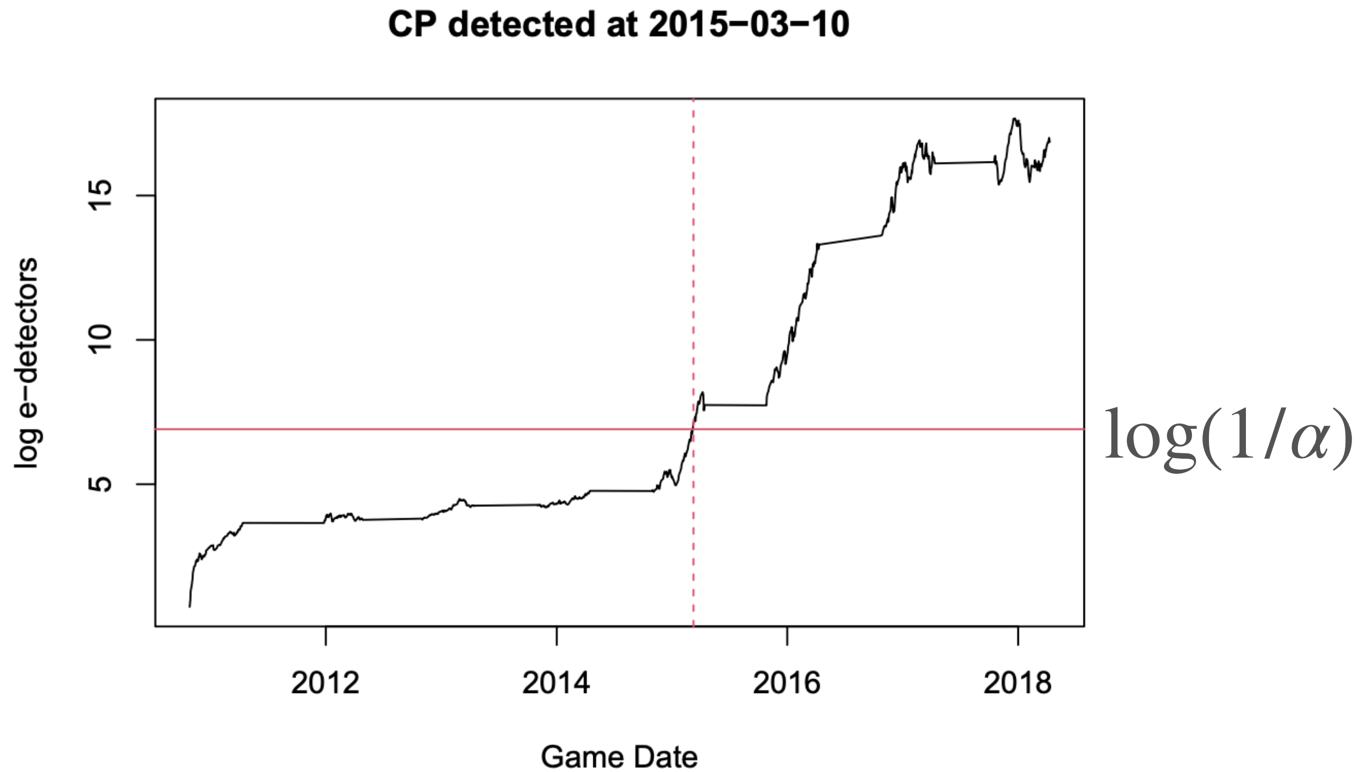
Fix Q for now: we have seen that we can mix over Q, or plug-in.

For a fixed Q, recall the previous two lectures:

$$\prod_{i=1}^{t} (1 + \lambda_Q^*(X_t - 1/2))$$

Is the optimal supermartingale. So plug this into the e-detector.

Result



Our e-detector announces a change point during the 2014-15 season.
It controls the average run length (ARL) at $1/\alpha = 1000$,
which is more than twelve seasons of games.

More broadly: the e-detector can help detect changes when the pre-change and post-change distributions are composite, non-stationary and nonparametrically specified (eg: no common reference measure, no likelihoods, infinite dimensional nuisance parameters, etc.)

(Simulation, asymptotics are intractable for ARL.)

A short list of e-processes for other settings

- ANY $\mathscr{P}$ for which max-likelihood (after smoothing or profiling) is feasible (eg: shape constraints like monotone, log-concave)

Wasserman, Balakrishnan, Ramdas'20
Universal inference

- $\mathscr{P} = \{P : X_1, X_2, \ldots$ are exchangeable$\}$, $\mathscr{Q}$ unrestricted

Vovk'21 (Testing randomness online)
Ramdas, Ruf, Larsson, Koolen'21
(Testing exchangeability: fork-convex hulls, supermartingales and e-processes)

- $\mathscr{P} := \{(P \times P)^\infty :$ for any $P\}$, $\mathscr{Q} := \{(P \times Q)^\infty :$ for any $P \neq Q\}$

Shekhar, Ramdas'21
(Nonparametric two-sample testing by betting)

Some more e-processes for other settings

- $$\mathscr{P} := \{P^\infty : \mathbb{E}[X] \le \mu, \mathrm{Var}(P) \le \sigma^2\}$$
  $$\mathscr{Q} := \{P^\infty : \mathbb{E}[X] > \mu, \mathrm{Var}(P) \le \sigma^2\}$$

Wang, Ramdas'23 (Catoni-style CSs for heavy-tailed mean estimation)

- $\mathscr{P} := \{P^\infty : \mathbb{E}[X] \le \mu, \mathrm{Var}(P) \le \sigma^2, \text{adversary corrupts } \epsilon \text{ fraction of data}\}$
  $\mathscr{Q} := \{P^\infty : \mathbb{E}[X] > \mu, \mathrm{Var}(P) \le \sigma^2, \text{adversary corrupts } \epsilon \text{ fraction of data}\}$

Wang, Ramdas'23 (Huber-robust confidence sequences)

- $$\mathscr{P} := \{P_{XY}^\infty : \text{where } P_{XY} = P_X \times P_Y\},$$
  $$\mathscr{Q} := \{P_{XY}^\infty : \text{where } P_{XY} \neq P_X \times P_Y\}$$

Shekhar, Ramdas'23 and Podkopaev et al.' 23
(Nonparametric independence testing by betting)

Detection delay: suppose Q is known

For a changepoint stopping rule N,
Define worst detection delay

$$\mathscr{D}(N) := \sup_{\nu \geq 0} \mathbb{E}_{P,\nu,Q}[(N - \nu)_+ \,|\, N > \nu]$$

If the baseline increment L_n is function of only X_n
and the post-change observations from Q are strongly stationary,

then $$\mathscr{D}(N^{SR}) \leq \frac{\log(1/\alpha)}{\mathbb{E}_Q \log L_1} + \frac{\mathbb{E}_Q \log^2 L_1}{(\mathbb{E}_Q \log L_1)^2} + 1,$$

where $N^{SR} := \inf\{t \geq 1 : M_t^{SR} \geq 1/\alpha\}$.

Minimizing delay corresponds to maximizing $K := \mathbb{E}_Q \log L_1$,
but we usually don't know Q .

Let $(L^\lambda)_{\lambda \in \Pi}$ be a family of baseline processes indexed by λ .
$\lambda^* := \arg\max_{\lambda \in \Pi} \mathbb{E}_Q \log L_1^\lambda$ depends on Q.

Mixtures of e-detectors

A convex combination (mixture) of e-detectors is also an e-detector. (ARL control maintained)

This fact can be used to take statistically and computationally efficient mixtures over Π (implicitly over $\mathcal{Q}$), in order to derive e-detectors that adapt to unknown Q.

We derive e-detectors where the number of mixture components increases (logarithmically) with time.

Detection delay for the mixture e-SR procedure

$$\mathscr{D}(N^{mSR}) \le \frac{g_\alpha}{D(Q\|\mathscr{P})} + \frac{\mathbb{V}_Q[\log L_1^{\lambda^*}]}{D^2(Q\|\mathscr{P})} + 1$$

Defining $\Delta^* = \nabla\psi(\lambda^*)$ as the "signal strength", lying in $[\Delta_L, \Delta_U]$,

we have $g_\alpha \le \inf_{\eta>1} \eta \left[\log(1/\alpha) + \log\left(1 + \lceil \log_\eta \frac{\psi^*(\Delta_U)}{\psi^*(\Delta_L)} \rceil\right)\right]$

Qualitatively similar results in the non-separated case.

Takeaway messages

A $\mathscr{P}$-e-detector is a nonnegative adapted process $M \equiv (M_n)_{n\geq 1}$ such that $\mathbb{E}_P[M_\tau] \leq \mathbb{E}_P[\tau]$ for all $P \in \mathscr{P}, \tau \in \mathscr{T}$.

A $\mathscr{P}$-e-process is a nonnegative adapted process $\Lambda \equiv (\Lambda_n)_{n\geq 1}$ such that $\mathbb{E}_P[\Lambda_\tau] \leq 1$ for all $P \in \mathscr{P}, \tau \in \mathscr{T}$.

Eg: the SR e-detector as $M_n^{SR} := \sum_{j=1}^{n} \Lambda_n^{(j)}$,
where $\Lambda^{(j)}$ is a $\mathscr{P}$-e-process started at time j.

Thresholding at $1/\alpha$ controls ARL at $1/\alpha$.
Mixtures of e-detectors are also e-detectors.
Baseline e-processes (eg: exponential) enable online computation.

Now, we can perform changepoint detection in a slew of nonparametric settings, sometimes good bounds on detection delay.

Today's talk

✓ Reducing partitioned change detection to sequential testing

2. Reducing non-partitioned change detection to sequential estimation

3. High level summary of game-theoretic statistics

E-detectors (NEJSDS'23)

When pre- and post-change distributions are known to lie in separate classes (eg: mean change from < 0 to > 0) ✓

Confidence sequences arXiv'23

When pre- and post-change distributions are simply two different distributions in one class (eg: the mean changed from something to something else)

Via confidence sequences

(arXiv)

Shubhanshu
Shekhar
(CMU)

TLDR: Reducing sequential (nonparametric) change detection to (modern) sequential estimation

Sequential Changepoint Detection

- Stream of independent $\mathcal{X}$-valued observations: $X_1, X_2, \ldots$
- For some $T \in \mathbb{N} \cup \{\infty\}$:
  - $X_t \sim P_0$ for $t \leq T$
  - $X_t \sim P_1 \neq P_0$ for $t > T$
- Mild requirements on the distributions:
  - Both P_0, P_1 are unknown
  - $P_0, P_1 \in \mathcal{P}$ for some known class of distributions $\mathcal{P}$
- Decide between

$$H_0 : T = \infty, \quad \text{versus} \quad H_1 : T < \infty.$$

- **Objective:** Define a stopping time τ to declare a detection, that
  - minimizes false alarms under H_0
  - has a small detection delay, $(\tau - T)^+$, under H_1

A "confidence sequence (CS)" for a parameter θ is a sequence of confidence intervals (L_n, U_n) that are constructed from the first n samples, and have a uniform (simultaneous) coverage guarantee.

$$\mathbb{P}(\forall n \geq 1 : \theta \in (L_n, U_n)) \geq 1 - \alpha \,.$$

Equivalently, for any stopping time τ : $\mathbb{P}(\theta \in (L_\tau, U_\tau)) \geq 1 - \alpha$.

Darling, Robbins '67, '70s
Lai '76, '84
Robbins, Siegmund '70s

Much stronger than the pointwise (fixed-sample) confidence interval (CI) guarantee:

$$\forall n \geq 1, \mathbb{P}(\theta \in (\tilde{L}_n, \tilde{U}_n)) \geq 1 - \alpha \,.$$

Overview of our results

If we can construct a CS for θ $\Rightarrow$ We can detect changes in θ

Our Reduction: at every time t, start a *new* level-α confidence sequence based on $X_t, X_{t+1}, \dots$ (there are t active CSs after t steps). We declare a change when their intersection is empty.

Key theorem 1: This method controls ARL (FFA) at level $1/\alpha$.

Key theorem 2: The detection delay is minimax optimal for many problems for which the confidence sequence shrinks like $\widetilde{O}(1/\sqrt{t})$.

Main Assumptions

- We work with distribution class $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$.
- Possibly infinite dimensional Θ endowed with metric d.
- $P_0 = P_{\theta_0}$ and $P_1 = P_{\theta_1}$ for θ_0, θ_1 such that $d(\theta_0, \theta_1) > 0$.

Assumptions

1. Uniformly decaying width: We can construct a CS $\{C_t(\theta) : t \geq 1\}$ for all $\theta \in \Theta$, satisfying

$$\sup_{\theta \in \Theta} \sup_{\theta', \theta'' \in C_t(\theta)} d(\theta', \theta'') \leq w_t \equiv w_t(\Theta, \alpha),$$

   such that $\lim_{t \to \infty} w_t = 0$.

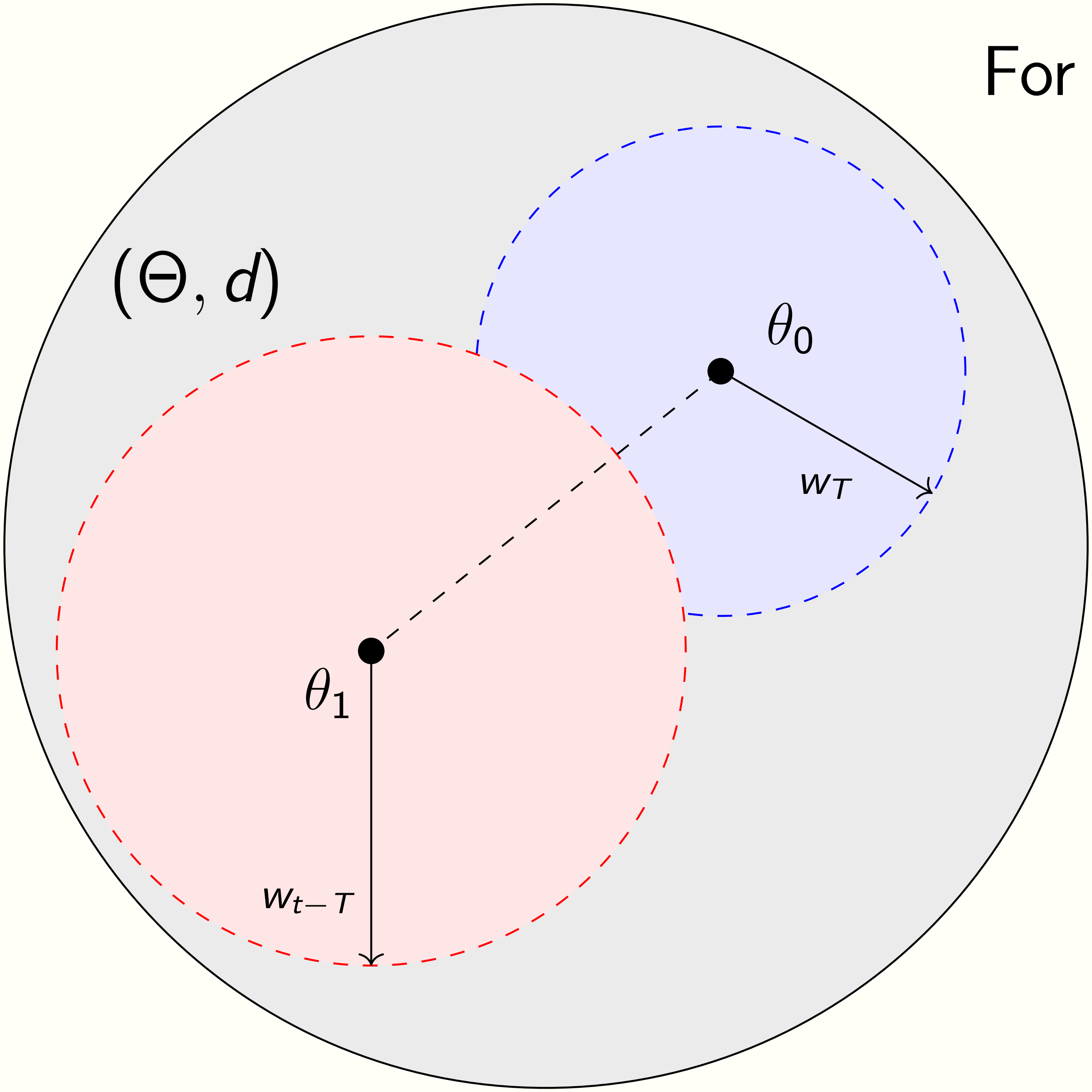
2. Enough pre-change data: Under H_1, the changepoint T is large enough to ensure $w_T < \Delta := d(\theta_1, \theta_0)$.

- Control over the detection delay under H_1:
  - Introduce the "good event": $\mathcal{E} = \{\forall t \leq T : \theta_0 \in C_t\}$
  - If $w_t = \mathcal{O}\left(\sqrt{\log\log t / t}\right)$, then

$$\mathbb{E}[(\tau - T)^+ | \mathcal{E}] = \mathcal{O}\left(\frac{\log\log(1/\Delta)}{\Delta^2}\right) \text{ where } \Delta = d(\theta_0, \theta_1).$$

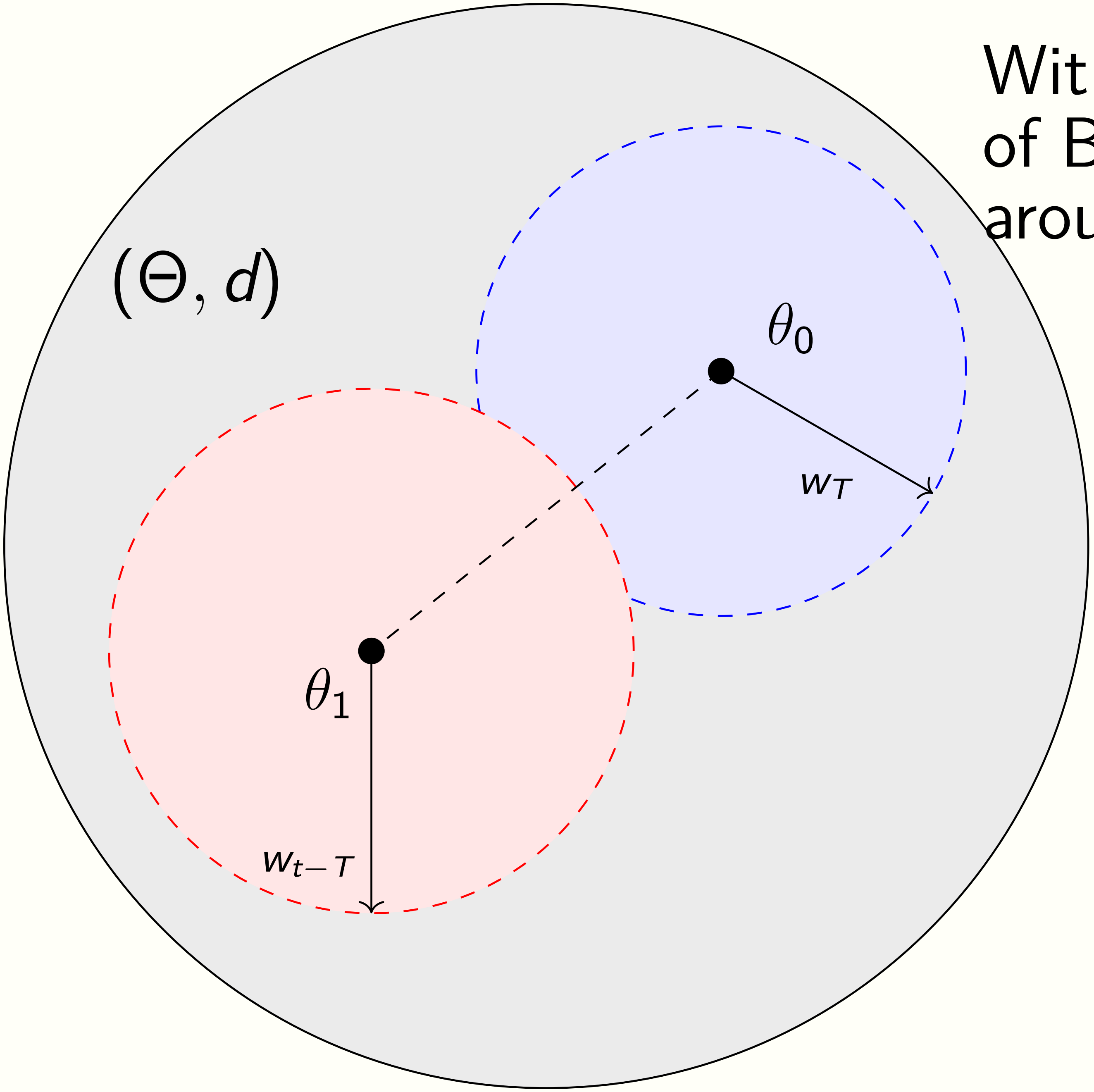
  - Can generalize to arbitrary $w_t \to 0$ (next slide).

Detection Delay Analysis: general w_t



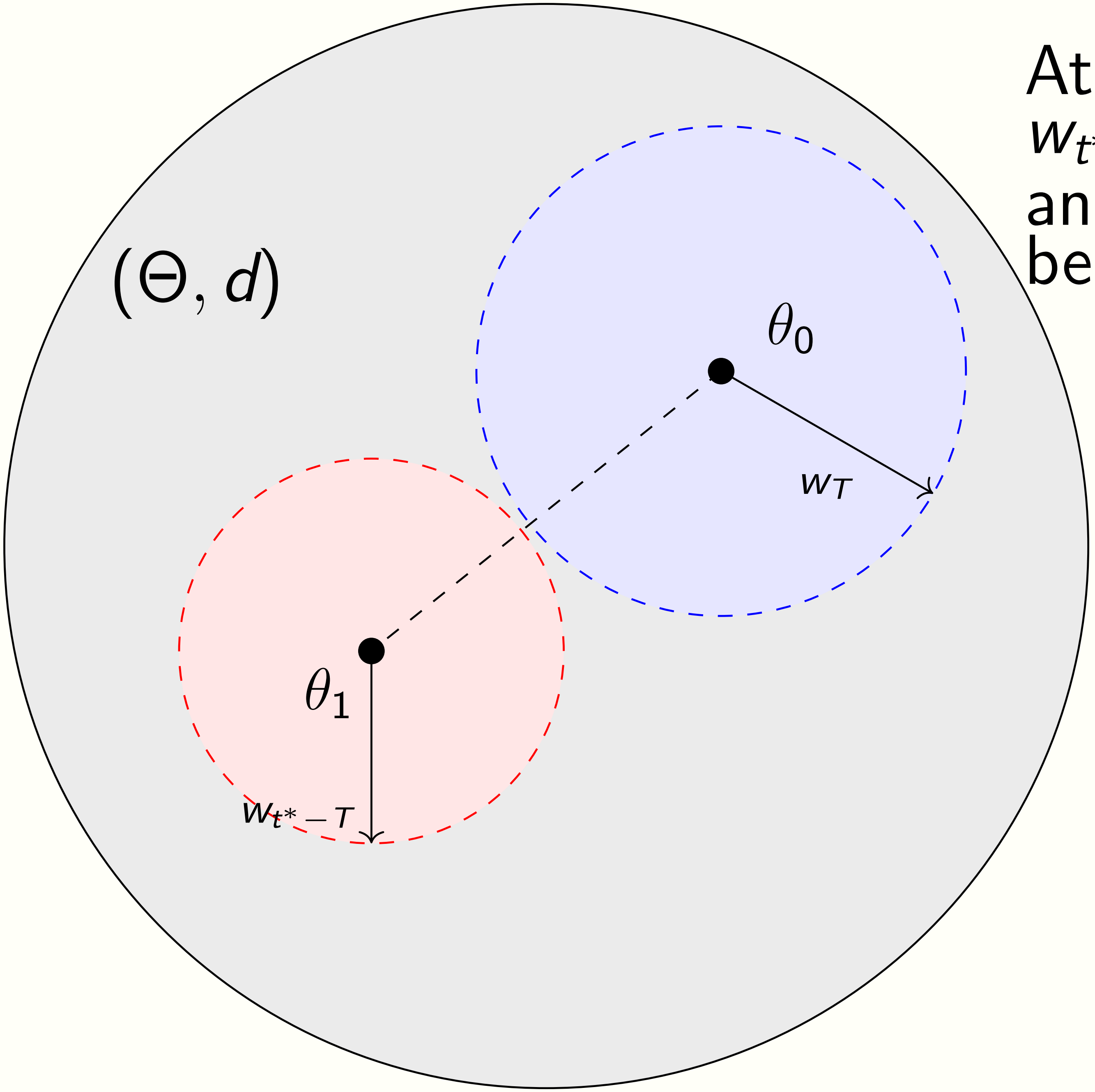
For t just after T

Detection Delay Analysis: general w_t



With increasing t, width of BCS at T (i.e., w_{t-T}) around θ_1 shrinks.

Detection Delay Analysis: general w_t



At t^*:
$w_{t^*-T} + w_T < d(\theta_1, \theta_0)$,
and the two balls
become disjoint.

Applications

- Mean-shift detection with univariate Gaussians.
  - $P_{\theta_0} = N(\theta_0, 1)$, and $P_{\theta_1} = N(\theta_1, 1)$, with $\Delta = |\theta_1 - \theta_0|$.
- Mean-shift detection with bounded observations
  - P_{θ_0} and P_{θ_1}, supported on $[0, 1]$ with $\Delta = |\theta_1 - \theta_0|$.
- Changes in CDF
  - $\Delta = d_{KS}(\theta_0, \theta_1)$, with θ_i = CDF of P_{θ_i}.
- Two-sample changepoint detection
  - $\theta_0 = P \times P$, $\theta_1 = P \times Q$, and $\Delta = \mathrm{MMD}(P, Q)$.
- *Several other tasks:* distribution shifts in ML, nonparametric regression, exponential family.

Summary

Using "E-detectors"

arXiv: 2203.03532 (under review)

When pre- and post-change distributions are known to lie in separate classes $\mathscr{P}, \mathscr{Q}$ (eg: mean change from < 0 to > 0)

Key idea: reduce change detection to repeated sequential testing of $\mathscr{P}$ vs. $\mathscr{Q}$

Method: at every time t, start a new level-α **sequential test** based on $X_t, X_{t+1}, \ldots$ and declare change when accumulated evidence crosses $1/\alpha$

Main theorem: ARL $\geq 1/\alpha$

Using "confidence sequences"

ICML 2023
+ to-be-arXived

When pre- and post-change distributions are unknown but from the same class (eg: the mean changed from a to b)

Key idea: reduce change detection to repeated sequential estimation

Method: at every time t, start a new level-α **confidence sequence** based on $X_t, X_{t+1}, \ldots$ and declare change when their intersection is empty.

Main theorem: ARL $\geq 1/\alpha$

Advantage: sequential testing/estimation are basic problems, increasingly well studied.

Today's talk

1. ✓ Reducing partitioned change detection to sequential testing
2. ✓ Reducing non-partitioned change detection to sequential estimation
3. High level summary of game-theoretic statistics

What is game-theoretic statistics?

A subfield whose quest is to quantify uncertainty in statistical inference tasks like

— **hypothesis testing**
— **estimation (confidence sets)**
— **probabilistic forecasting**
— **model selection**
— **change detection**

by using game-theoretic intuition, language and formalism.

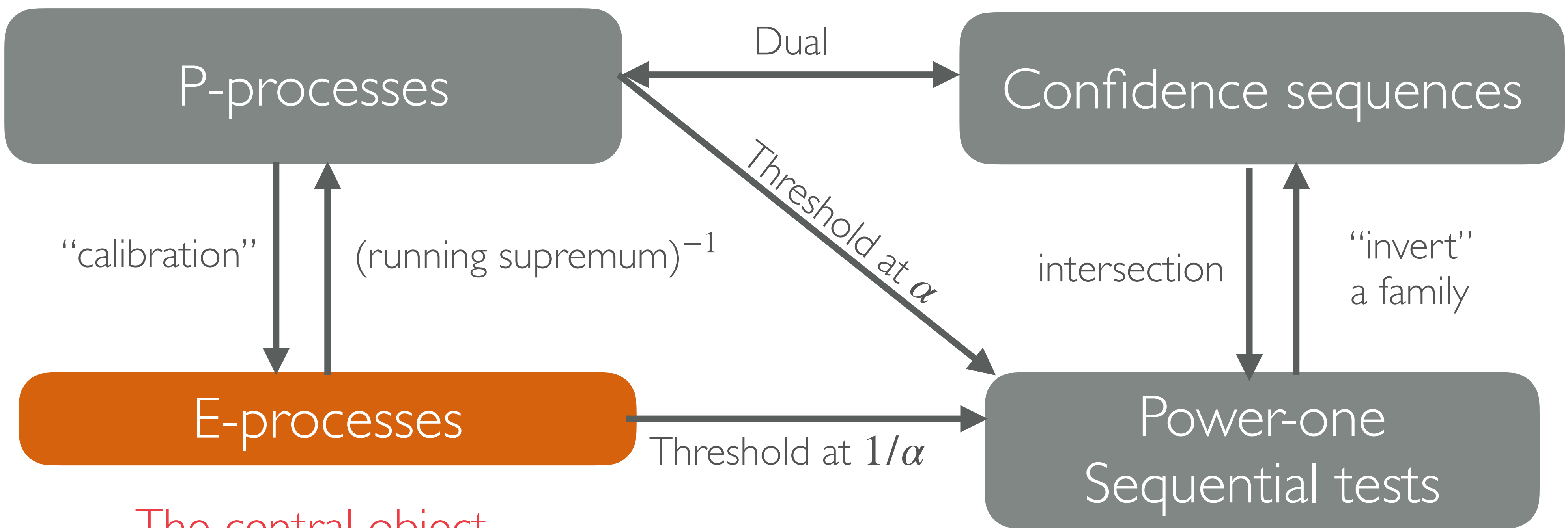
Helps design **new nonparametric inference procedures** which

- Have strong *theoretical guarantees under weaker assumptions*,
- Provide *greater flexibility in applications*,
- Combine the use of *prior knowledge with frequentist guarantees*,
- Are designed to be adaptive to the underlying data.

The 4 characters in SAVI

Real-valued measures of evidence

Associated with a level $\alpha \in (0,1)$



The central object.
Understanding their properties and designing "good" ones has been the focus of all my research for many years.

E-processes are nonparametric, composite generalizations of likelihood ratios

Every likelihood ratio is a test martingale.
Further, every test martingale is (implicitly) a likelihood ratio.

If (M_t) is a test martingale for $\mathscr{P}$, then for every $P \in \mathscr{P}$, we can write $M_t = \frac{Q(X_1, \ldots, X_t)}{P(X_1, \ldots, X_t)}$ for some Q — it is a "simultaneous likelihood ratio", despite there being no dominating reference measure for $\mathscr{P}$.

Test martingales are at the *heart* of parametric inference, (likelihood ratio) — but also nonparametric inference.

E-processes generalize test martingales: *they exist more generally*.

Ramdas, Ruf, Larsson, Koolen
(arXiv:2009.03167)

* : no dominating reference measure
for the set of distributions

Nontrivial (e-power) test martingales exist

Testing symmetry*

Two-sample testing*

Bounded means*

T-test (in shrunk filtration)

Exchangeability* (in shrunk filtration)

Independence testing* (in shrunk filtration)

Nontrivial test supermartingales exist

SubGaussian distributions* (or any bounded MGF)

Robust, heavy-tailed mean estimation*

Nontrivial test martingales exist

Testing symmetry*

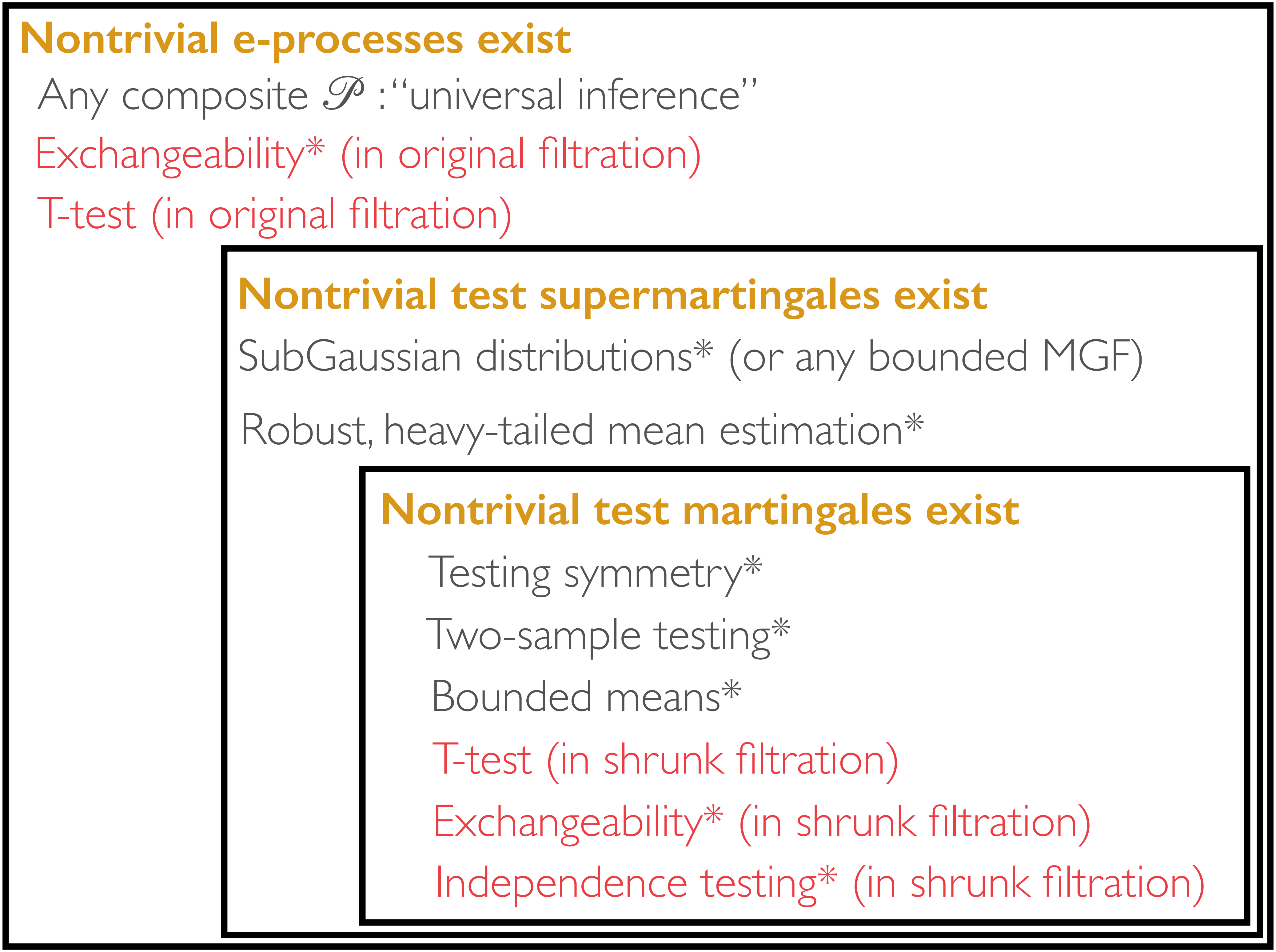
Two-sample testing*

Bounded means*

T-test (in shrunk filtration)

Exchangeability* (in shrunk filtration)

Independence testing* (in shrunk filtration)


Nontrivial e-processes exist
Any composite $\mathscr{P}$: "universal inference"
Exchangeability* (in original filtration)
T-test (in original filtration)
Nontrivial test supermartingales exist
SubGaussian distributions* (or any bounded MGF)
Robust, heavy-tailed mean estimation*
Nontrivial test martingales exist
Testing symmetry*
Two-sample testing*
Bounded means*
T-test (in shrunk filtration)
Exchangeability* (in shrunk filtration)
Independence testing* (in shrunk filtration)

Some current and future directions

1. For a new (nonparametric) problem, how do we *design* the game and *learn* to bet?)

2. When do *nontrivial* test martingales (not) exist?
   When do *nontrivial* test supermartingales (not) exist?
   When do *nontrivial* e-processes (not) exist?

3. How do we move beyond this tutorial's topics to, say, game-theoretic model selection?

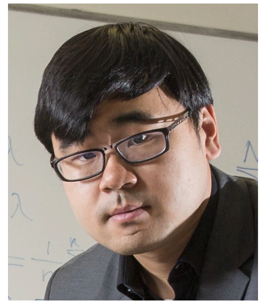
4. Everything today was nonasymptotic. What about anytime-valid asymptotics? (Few arXiv preprints)

5. How do we tie together game-theoretic statistics with game-theoretic probability?

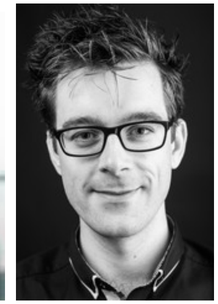

Peter Grunwald


Glenn Shafer


Volodya Vovk


Ruodu Wang


Johannes Ruf


Wouter Koolen


Martin Larsson


Johanna Ziegel

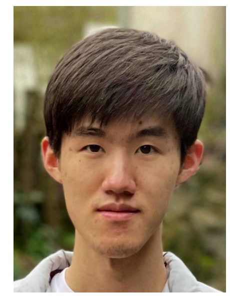

Steve Howard


Akshay Balsubramani


Ian Waudby-Smith

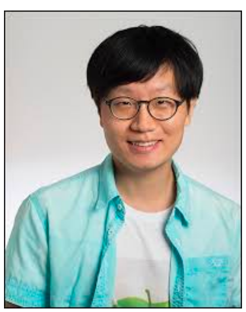

Tudor Manole

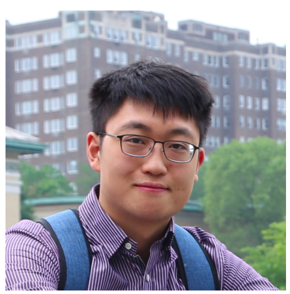

Shubhanshu Shekhar

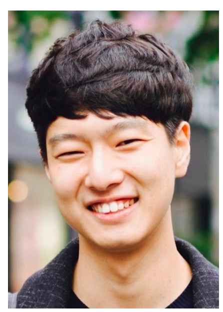

Hongjian Wang


Robin Dunn


Justin Whitehouse

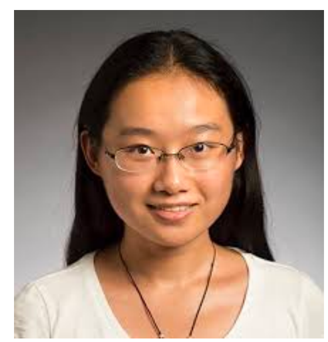

Jaehyeok Shin


Neil Xu


YJ Choe


Ben Chugg


Sasha Podkopaev


Boyan Duan

Two general frameworks for sequential change detection with composite, nonparametric classes

Aaditya Ramdas

Dept. of Statistics and Data Science (75%)
Machine Learning Dept. (25%)
Carnegie Mellon University

Amazon Research (20%)