Game-theoretic statistics & sequential anytime-valid inference (SAVI): a martingale theory of evidence

Aaditya Ramdas

Dept. of Statistics and Data Science
Machine Learning Dept.
Carnegie Mellon University

Ack support: NSF CAREER, Sloan Fellowship

Online learning
+ optimization
(Eg: Cover,
Orabona)

Statistics
(Eg: Grunwald,
Shafer, Vovk)

Gambling/betting

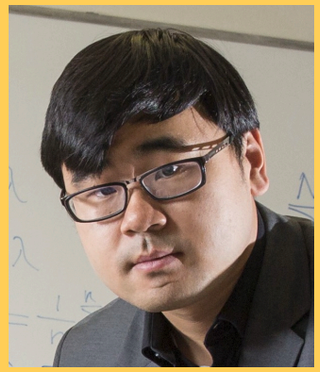
Information
Theory
(Eg: Kelly, Cover)

Probability
+ Finance
(Eg: Ville, Shafer, Vovk)

New: The E-Book

Title: *Hypothesis testing with e-values*

Aaditya
Ramdas



Ruodu
Wang

Available on my webpage and on arXiv
(just released!)

Inaugural issue of Foundations & Trends in Statistics (Aug'25)

Outline of this tutorial

A. First half: game-theoretic hypothesis testing

B. Second half: game-theoretic estimation

Slides and references at

http://www.stat.cmu.edu/~aramdas/icml25

Outline of first half

1. "Sequential anytime-valid inference (SAVI)"
2. Testing by betting yields SAVI inference (an example)
3. Kelly betting and log-optimality
4. Core SAVI concepts: e-values and e-processes
5. *Optimal* gambling strategies

Why test hypotheses? Popper's legacy

What is a scientific theory or hypothesis? Popper: a claim that is empirically *testable with data*. If it is not (plausibly) testable, it is not a scientific theory/hypothesis. Brought data, statistics, etc. bang into the center of philosophy and science.

After formulating a scientific theory/hypothesis, how does one test it?

One typically formulates *statistical* hypotheses that are implied by the scientific hypothesis, and test those. So how does one perform statistical hypothesis testing?

Stochastic proof by contradiction. You assume that your theory is *wrong* ("null hypothesis"). You then design an experiment & collect data. If data appears to contradict the null, you reject the null hypothesis ("make a discovery").

This tutorial: a simple and universal way to test a null hypothesis — *making money by betting against it*. Mathematically: e-values, an alternative to p-values.

An infamous instance of **"peeking at p-values"** is the power-posing controversy (Amy Cuddy, Dana Carney).

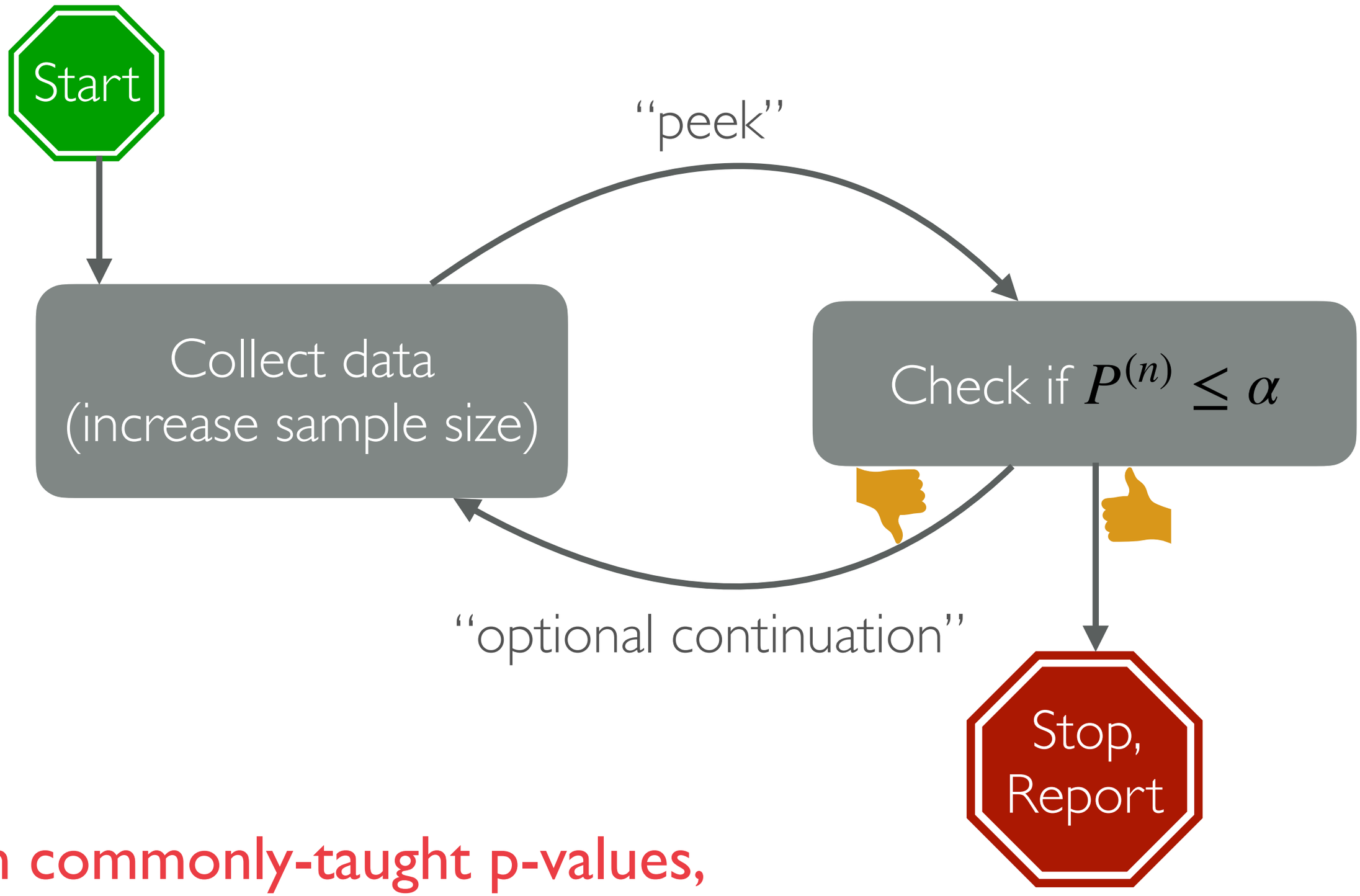
faculty.haas.berkeley.edu/dana_carney/pdf_My%20position%20on%20power%20poses.pdf

Courses | Wshops | UCB | Sports | Vovk | Group Ex | TRL | Jonathan Pon Mem... | Classes | Charities | Podcasts | CausalInf | Other Bookmarks

4. The data are flimsy. The effects are small and barely there in many cases.
5. Initially, the primary DV of interest was risk-taking. We ran subjects in chunks and checked the effect along the way. It was something like 25 subjects run, then 10, then 7, then 5. Back then this did not seem like p-hacking. It seemed like saving money (assuming your effect size was big enough and p-value was the only issue).
6. Some subjects were excluded on bases such as "didn't follow directions." The total number of exclusions was 5. The final sample size was $N = 42$.
7. The cortisol and testosterone data (in saliva at that point) were sent to Salimetrics (which was in State College, PA at that time). The hormone results came back and data were analyzed.
8. For the risk-taking DV: One p-value for a Pearson chi square was .052 and for the Likelihood ratio it was .05. The smaller of the two was reported despite the Pearson being the more ubiquitously used test of significance for a

"Sampling to a foregone conclusion" — Anscombe (1950s)

What is the problem with continuous monitoring?



With commonly-taught p-values, false positive rate $\gg \alpha$.

Let $P^{(n)}$ be a classical p-value (eg: t-test), calculated using the first n samples.

Under the null hypothesis (no treatment effect),

$$\forall n \geq 1, \qquad \underbrace{\Pr(P^{(n)} \leq \alpha)}_{\text{prob. of false positive}} \qquad \leq \alpha \,.$$

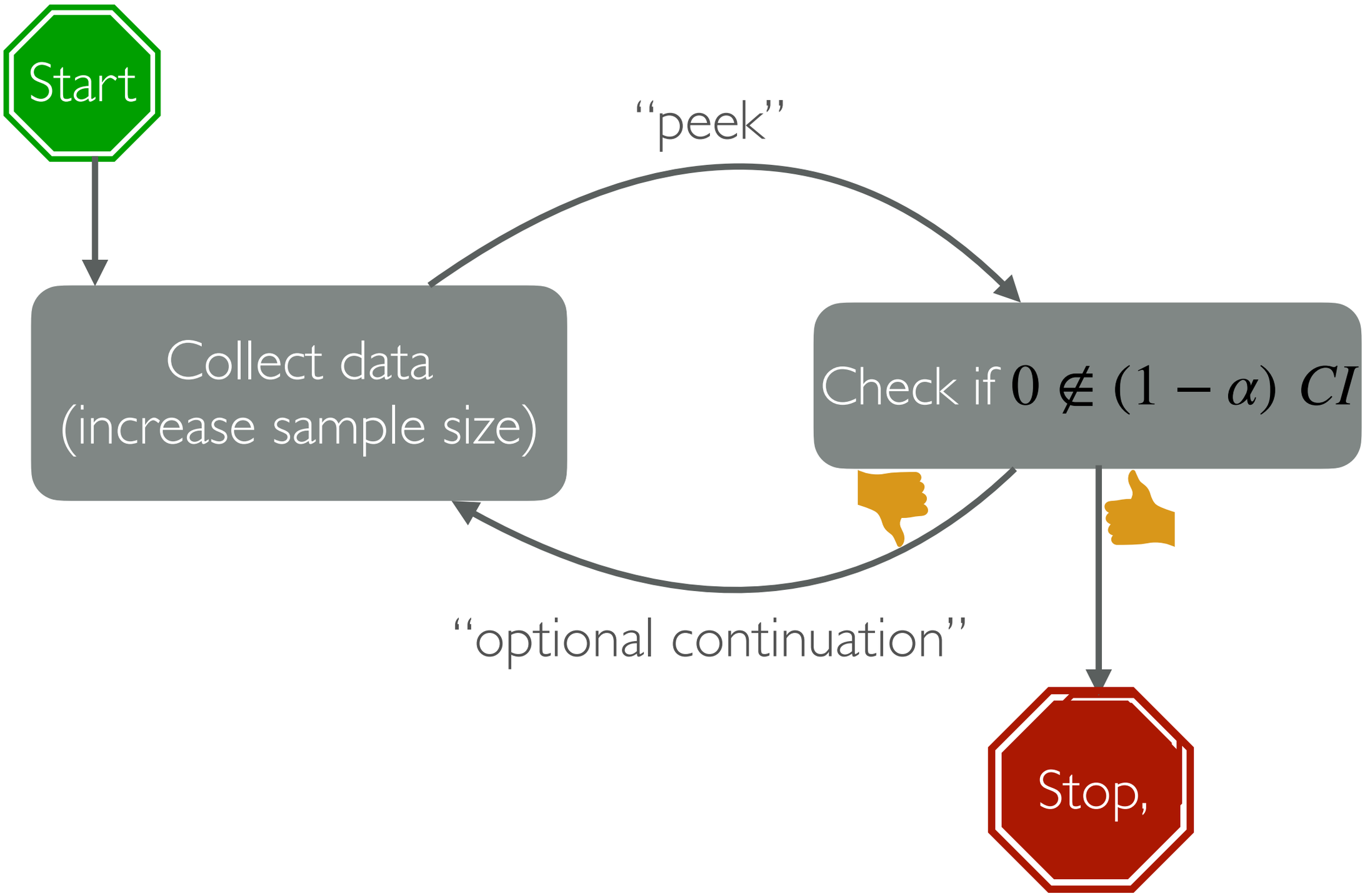
Let τ be the stopping time of the experiment.

Often, τ depends on data, eg: $\tau := \min\{n \in \mathbb{N} : P^{(n)} \leq \alpha\}$.

Unfortunately, $\Pr(P^{(\tau)} \leq \alpha) \not\leq \alpha$. usually $= 1$.

Not special to p-values. Same holds for confidence intervals.

Same issue with confidence intervals



Again, false positive rate $\gg \alpha$.

Let $(L^{(n)}, U^{(n)})$ be any classical $(1-\alpha)$ CI, calculated using the first n samples.

When trying to estimate the treatment effect θ,

$$\forall n \geq 1,\ \underbrace{\Pr(\theta \in (L^{(n)}, U^{(n)}))}_{\text{prob. of coverage}} \geq 1-\alpha\,.$$

Let τ be the stopping time of the experiment.

Again, τ may depend on data, eg: $\tau := \min\{n \in \mathbb{N} : L^{(n)} > 0\}$.

Unfortunately, $\Pr(\theta \in (L^{(\tau)}, U^{(\tau)})) \not\geq 1-\alpha\,.$

usually $= 0$.

We want "safe, anytime-valid inference" (SAVI) methods

SAVI methods are those that yield valid inference at
arbitrary stopping times, possibly not specified or anticipated in advance.

SAVI methods allow for continuous monitoring and analysis of data, adaptive decisions to halt or continue experiments (for any reason), all without violating the validity of the claims.

Provides a lot of flexibility to the statistician ("peeking"),
useful for a lot of exploratory settings, or those without oversight
(like university labs and tech industry).

Outline of this talk

"Sequential anytime-valid inference (SAVI)"

2. Testing by betting yields SAVI inference (an example)

3. Kelly betting and log-optimality

4. Core SAVI concepts: e-values and e-processes

5. *Optimal* gambling strategies

The lady tasting tea (1920s)

Ronald Fisher

Would you like some tea?

No, T in M ≠ M in T

Can you *really* tell them apart?

Indeed, yes!

Muriel Bristol

Scientific claim (Muriel): big difference between MT and TM.

Statistical null hypothesis (negation of Muriel's claim):
Muriel unlikely to make many correct guesses.

The lady tasting tea (1920s)

T M T T M M M T

What's the probability that a *chance guess* would be *perfect*? 1/70

This is a p-value for H_0 : there is no difference between MT and TM.

Randomization-based causal inference, design of experiments…

However, the odds were stacked against Muriel from the start!

The probability that a *chance guess* would yield *at most one error* is 17/70 $\approx \mathbf{0.24}$, which is not so impressive.

With the benefit of 100 years of hindsight, I would have (and did) run the experiment quite differently…

The lady keeps tasting coffee (2020)

Let's play a game

Umm…sure…?

It involves coffee

Sure!

(self)

Leila Wehbe

Statistical null hypothesis: there is no difference between "espresso in milk" and "milk in espresso" (not mixed).

We will set up a *betting* game in which,
if the null is true, Leila should not be able to make money.

The lady keeps tasting coffee (2020, betting)

Result?

$L_0 = 1$

$R_1 = -1$

$\lambda_1 = 0.2$ (on heads)

$L_1 = L_0 \cdot (1 + \lambda_1 R_1) = 0.8$

$R_2 = +1$

$\lambda_2 = 0.4$ (on heads)

$L_2 = L_1 \cdot (1 + \lambda_2 R_2) = 1.12$

...

$L_t := \prod_{i=1}^{t} (1 + \lambda_i R_i)$, where (λ_i) are "predictable" bets in $[0,1]$.

Under the null, $(L_t)_{t \in \mathbb{N}}$ is a nonnegative martingale ("fair game").

The lady keeps tasting coffee (2020, betting)

At any stopping time τ, $\mathbb{E}_{H_0}[L_\tau] \leq 1$ — optional stopping theorem.

Ville's inequality (time-uniform Markov's for nonneg. supermartingales)

$\Pr(\exists t \in \mathbb{N} : L_t \geq 1/\alpha) \leq \alpha$.

If the null holds, then Leila is unlikely to turn one pound into fifty

- L_t directly measures evidence against H_0 ("e-process").
- $\inf_{s \leq t} 1/L_s$ is an "anytime-valid p-value" or "p-process".
- $1\{L_t \geq 1/\alpha\}$ is a level-α sequential test for H_0 .

Wald, Robbins, Shafer, Vovk, Grunwald, Ramdas, …

Why is this interesting?

(a) simple and clean approach to sequential experimental design

(b) can express *doubt* naturally

(c) *cooperation between subject and statistician allowed between rounds*

(d) flexible (can design *many* games for each problem)

(e) make up the game (and extend the game) on the fly

(f) evidence only depends on what did occur, not on hypothetical worlds

Testing by betting

Shafer & Vovk
(+ Ville, Robbins, Cover)

In order to test a hypothesis, one sets up a game such that:
if the null is true, no strategy can systematically make (toy) money,
but if the null is false, then a good betting strategy can make money.

Wealth in the game is directly a measure of evidence against the null.

Each strategy of the statistician = a different estimator or test statistic.
So there are "good" and "bad" strategies for betting,
just as there are good and bad estimators or test statistics.

Testing (and estimation) == game and strategy design.

Kelly's game corresponds to H_0 : fair coin against H_1 : bias p

Outline of this talk

1. "Sequential anytime-valid inference (SAVI)"

2. Testing by betting yields SAVI inference (an example)

3. Kelly betting and log-optimality

4. Core SAVI concepts: e-values and e-processes

5. *Optimal* gambling strategies

A New Interpretation of Information Rate

reproduced with permission of AT&T

By J. L. Kelly, jr.

(Manuscript received March 21, 1956)

If the input symbols to a communication channel represent the outcomes of a chance event on which bets are available at odds consistent with their probabilities (i.e., "fair" odds), a gambler can use the knowledge given him by the received symbols to cause his money to grow exponentially. The maximum exponential rate of growth of the gambler's capital is equal to the rate of transmission of information over the channel. This result is generalized to include the case of arbitrary odds.

Kelly's game

Suppose we observe iid coin flips B_i of bias $p > 1/2$ (for known p).

We start with one dollar, and can make "double or nothing" bets.
In each round, we bet some fraction λ of our wealth on heads.
Then we observe the coin toss.
If H, we earn that amount, and if T, we lose that amount.

$W_t(\lambda) := \prod_{i=1}^{t} (1 + \lambda(2B_i - 1))$ is wealth after t rounds.

What fraction λ of our wealth should we bet at each step?
(Think of the extremes of 0,1.)

Kelly's solution: log-optimality

$$W_t(\lambda) := \prod_{i=1}^{t} (1 + \lambda(2B_i - 1))$$ is wealth after t rounds.

$$W_t = \exp\left(\sum_{i=1}^{t} \log(1 + \lambda B_i) \right) = \exp\left(t\mathbb{E}[\log(1 + \lambda B)] + o(t) \right)$$

Kelly: choose λ to maximize $\lim_{t\to\infty} \frac{\mathbb{E} \log W_t(\lambda)}{t} = \mathbb{E}[\log(1 + \lambda B)].$

Solution: bet $\lambda^* = 2(p - 1/2)$ on heads.

Optimal Wealth $W_t(\lambda^*) = \exp(t \cdot H(p \,|\, 0.5) + o(t))$,
where H is the relative entropy (KL divergence)

$$H(Q \,|\, P) := \mathbb{E}_Q \left[\log \frac{dQ}{dP} \right] \text{ if } Q \ll P, \infty \text{ o.w}$$

Equivalently, $\mathbb{E}[\log W_t]/t = H(p \,|\, 0.5)$.

OPTIMAL GAMBLING SYSTEMS FOR FAVORABLE GAMES

L. BREIMAN

UNIVERSITY OF CALIFORNIA, LOS ANGELES

1. Introduction

Assume that we are hardened and unscrupulous types with an infinitely wealthy friend. We induce him to match any bet we wish to make on the event that a coin biased in our favor will turn up heads. That is, at every toss we have probability $p > 1/2$ of doubling the amount of our bet. If we are clever, as well as unscrupulous, we soon begin to worry about how much of our available fortune to bet at every toss. Betting everything we have on heads on every toss will lead to almost certain bankruptcy. On the other hand, if we bet a small, but fixed, fraction (we assume throughout that money is infinitely divisible) of our available fortune at every toss, then the law of large numbers informs us that our fortune converges almost surely to plus infinity. What to do?

Generalizes Kelly betting to other settings.

Proves that the Kelly criterion also asymptotically optimizes
a) Expected time to reach a threshold wealth
b) Expected wealth at some threshold time

Outline of this talk

1. "Sequential anytime-valid inference (SAVI)"

2. Testing by betting yields SAVI inference (an example)

3. Kelly betting and log-optimality

4. Core SAVI concepts: e-values and e-processes

5. *Optimal* gambling strategies

Hypothesis testing formalism

The "null hypothesis" H_0 is a set of distributions $\mathscr{P}$
defined on some *filtered* measurable space $(\Omega, \mathscr{F})$.
A *filtration* is a sequence of nested σ-algebras $\mathscr{F}_1 \subset \mathscr{F}_2 \ldots$
representing the accumulation of information over time.
Eg: $\mathscr{F}_t = \sigma(X_1, \ldots, X_t)$.
The "alternative hypothesis" H_1 is a set of distributions $\mathscr{Q} \subseteq \mathscr{P}^c$.

We observe $X_1, X_2, \ldots \sim P$.
When we are "testing $\mathscr{P}$ against $\mathscr{Q}$", we are asking whether
$H_0 : P \in \mathscr{P}$ or $H_1 : P \in \mathscr{Q}$.

In statistical practice, the null has a special role (eg: "no effect").
Rejecting the null may correspond to an interesting scientific phenomenon (described by the alternative).
Thus the first goal is to calibrate/control errors under $\mathscr{P}$.

An e-value for $H_0 : P \in \mathscr{P}$ is a $[0,\infty]$-valued r.v. e s.t. $\forall P \in \mathscr{P}$, $\mathbb{E}_P(e) \leq 1$. (e for evidence or expectation)

An e-process for $\mathscr{P}$ is a sequence of e-values $(e_t)_{t \geq 1}$ s.t. for any stopping time τ, $P \in \mathscr{P} : \mathbb{E}_P(e_\tau) \leq 1$.

M is a *test supermartingale* for $\mathscr{P}$ if $M \geq 0, M_0 = 1$, $\mathbb{E}_P[M_t \mid M_1, \ldots, M_{t-1}] \leq M_{t-1}$, P-a.s. $\forall P \in \mathscr{P}, t \geq 1$.

Test martingale for $\mathscr{P}$: replace $\leq$ by $=$

Howard, Ramdas, et al. (2018-2021)
Grunwald et al. (2019-2021)
Shafer (2020), Vovk & Wang (2021)

Sequential testing with e-processes

If L is an e-process for $\mathscr{P}$, Ville's inequality implies that
$P(\exists t \geq 1 : L_t \geq 1/\alpha) \leq \alpha$ for all $P \in \mathscr{P}$.

Thus, thresholding an e-process (for $\mathscr{P}$) at $1/\alpha$,
i.e. defining $\tau_{\mathscr{P}} = \inf\{t \geq 1 : L_t \geq 1/\alpha\}$,
yields a level α test (for $\mathscr{P}$).

Theorem: *Every* level-α sequential test (for any $\mathscr{P}$) can be obtained by thresholding some e-process (for $\mathscr{P}$) at $1/\alpha$.

Thus, e-processes are fundamental objects,
worthy of independent study.

Nontrivial e-processes exist

* : no reference measure

Any $\mathscr{P}$ that is sequentially testable

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)

Outline of this talk

1. "Sequential anytime-valid inference (SAVI)"

2. Testing by betting yields SAVI inference (an example)

3. Kelly betting and log-optimality

4. Core SAVI concepts: e-values and e-processes

5. *Optimal* gambling strategies

Simple nulls and alternatives

$$H_0 : X_i \sim P \text{ versus } H_1 : X_i \sim Q$$

What is the game?

Initial capital $W_0 = 1$

For each $t = 1,2,\ldots$

Statistician declares "bet" $S_t : \mathcal{X} \to [0,\infty)$ s.t. $\mathbb{E}_P[S_t(X) \mid X_1, \ldots, X_{t-1}] \leq 1$

Reality reveals X_t

Statistician's wealth becomes $W_t = W_{t-1} \cdot S_t(X_t)$

What is the log-optimal betting strategy?

Answer: likelihood ratio of Q to P

The **log-optimal bet** is $S_t(x) = \dfrac{q(x)}{p(x)}$.

$W_T^* = \displaystyle\prod_{i=1}^{T} \frac{q(X_i)}{p(X_i)}$ is the log-optimal wealth process:

— it is a positive test martingale under P, $\mathbb{E}_P[W_\tau] \leq 1, \mathbb{E}_P[\log W_\tau] \leq 0.$

— $\mathbb{E}_Q[\log W_T] > 0$ is maximized by these bets, equals $T \cdot H(Q \mid P)$

Shafer (2021, JRSSA discussion paper)

Next setting: Composite null vs. Simple alternative

- We have a composite null hypothesis $\mathscr{P}$ and a point alternative hypothesis Q. The data is either drawn from some P in $\mathscr{P}$ (the null is true), or from Q (the null is false).

- A *valid* bet is an "e-variable", which is a $X \geq 0$ such that $\mathbb{E}_P[X] \leq 1$ for every $P \in \mathscr{P}$. Think of X as being the multiplier of your wealth in each round of a multi-round game.

- Question: What is the optimal one-round bet X^*? Is it unique? Can we characterize/derive it?

- Answer: It is the likelihood ratio of Q to a special element P^*, which we call the Reverse Information Projection (RIPr).

Coming up: a complete story about (X^*, P^*).

Larsson et al. (2025, AoS)

Introducing X^*, the "numeraire" e-variable

Theorem: Under no assumptions on null $\mathscr{P}$ and alternative Q, there *always* exists a special e-variable (bet) X^* which satisfies two properties:

A. First, $X^* \geq 0$ and $\mathbb{E}_P[X^*] \leq 1, \forall P \in \mathscr{P}$ (the e-variable or fair bet property)
B. Second, for any e-variable X, we have $\mathbb{E}_Q[X/X^*] \leq 1$ (the "numeraire property")

Further, X^* is unique up to Q-nullsets. In fact, X^* is the numeraire if and only if it is log-optimal.

Applying Jensen's inequality, we get two other interpretable implications: for any e-variable X, we have $\mathbb{E}_Q[X^*/X] \geq 1$ and $\mathbb{E}_Q[\log(X/X^*)] \leq 0$ (log-optimality!)

Introducing P^*, the reverse information projection

Definition: Define a measure P^* by defining its likelihood ratio (Radon-Nikodym derivative) with respect to Q:

$$dP^*/dQ := 1/X^*$$

- This is understood to be zero on $\{X^* = \infty\}$.
- $P^* \ll Q$ by definition. Also $X^* = dQ/dP^*$ by definition.
- P^* is not a probability measure in general, it is a sub-probability measure, meaning that $\int dP^* \leq 1$.
- P^* lies in the *bipolar* of $\mathscr{P}$, which is defined as follows.
  A. The polar is $\mathscr{P}^\circ := \{X \geq 0 : \mathbb{E}_P[X] \leq 1 \text{ for all } P \in \mathscr{P}\}$, which is simply the set of all e-variables.
  B. The bipolar is $\mathscr{P}^{\circ\circ} := \{P \geq 0 : \mathbb{E}_P[X] \leq 1 \text{ for all } X \in \mathscr{P}^\circ\}$, which we also call "the effective null hypothesis".

Strong duality of (X^*, P^*)

Theorem: Assume $Q \ll \mathscr{P}$ for simplicity. Let X^* be the numeraire and let P^* be the RIPr. Then, one has the strong duality:

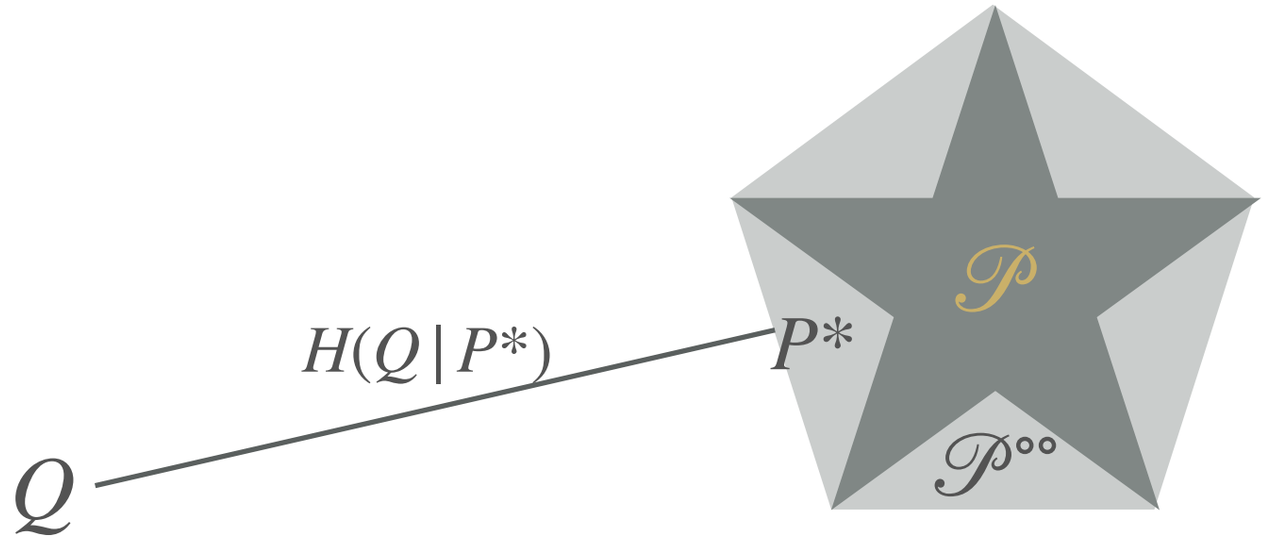
$$\mathbb{E}_Q[\log X^*] = \sup_{X \in \mathscr{P}^\circ} \mathbb{E}_Q[\log X] = \inf_{P \in \mathscr{P}^{\circ\circ}} H(Q \,|\, P) = H(Q \,|\, P^*),$$

where these quantities may equal $+\infty$.

- $Q \ll \mathscr{P}$: whenever $P(A) = 0$ for every $P \in \mathscr{P}$, we also have $Q(A) = 0$. Very weak assumption! Not required (see paper).

- One can also write a *benchmarked* strong duality theorem, with all quantities finite. The numeraire is benchmark-invariant.

Generalizes Li (1999), Grünwald et al. (2024), etc.



Theorem: X^* is the only e-variable which can be written as the likelihood ratio of Q to some element in $\mathcal{P}^{\infty\infty}$.

Thus, the numeraire is a "composite likelihood ratio",
it is log-optimal, and is the *only* e-variable that is a likelihood ratio.

Example 1: Symmetric distributions

Let Z denote the data, in this case real-valued.

$\mathscr{P} := \{P \in M_1 : Z \text{ and } -Z \text{ have the same distribution under } P\}$

Note that $\mathscr{P}$ has no dominating reference measure.

Suppose Q has a Lebesgue density q.

(In the paper, we generalize further.)

Older theory does not apply in this case. But we can easily show

$p^*(z) = \dfrac{q(z) + q(-z)}{2} 1\{q(z) > 0\}$ is the RIPr density.

It is a probability density iff Q has symmetric support.

Thus, we get that $X^* = \dfrac{2q(Z)}{q(Z) + q(-Z)}$ is the numeraire.

Example 2: 1-Sub-Gaussian distributions

$$\mathscr{P} := \{P \in M_1 : \mathbb{E}_P[e^{\lambda Z - \lambda^2/2}] \leq 1 \text{ for all } \lambda \geq 0\}$$

Above condition implies that $\mathbb{E}_P[Z] \leq 0$ for all $P \in \mathscr{P}$.

Let $Q = N(\mu, 1)$ for some known $\mu > 0$.

Once more, $\mathscr{P}$ has no reference measure.

So older theory does not apply.

But we can easily show that $\exp(\mu Z - \mu^2/2)$ is the numeraire and $N(0,1)$ is the RIPr.

Simple null vs. Composite alternative

$$H_0 : X_i \sim P \quad \text{versus} \quad H_1 : X_i \sim \{Q_\theta\}_{\theta \in \Theta}$$

Option 1: Mix (hedge your bets) with "prior" π

$$W_T = \int_\Theta \prod_{i=1}^{T} \frac{q_\theta(X_i)}{p(X_i)} d\pi(\theta)$$

Option 2: Plug-in a representative $\hat{\theta}_i \equiv \theta_i(X_1, \ldots, X_{i-1})$ in each round

$$W_T = \prod_{i=1}^{T} \frac{q_{\hat{\theta}_i}(X_i)}{p(X_i)}$$

Typically, $\lim_{T \to \infty} \mathbb{E}_{Q^*}[\log W_T]/T = \mathbb{E}_{Q^*}\left[\log \frac{q^*(X)}{p(X)}\right]$,
which is the best possible "growth rate", even without knowing Q^*.

Outline of this talk

1. "Sequential anytime-valid inference (SAVI)"
2. Testing by betting yields SAVI inference (an example)
3. Kelly betting and log-optimality
4. Core SAVI concepts: e-values and e-processes
5. *Optimal* gambling strategies

To end, a few miscellaneous slides.

Multiple testing under arbitrary dependence and optional stopping

The e-BH procedure: Given e-values $E_1, \ldots, E_K$ for K hypotheses, define

$$k^* := \max \left\{ k : E_{[k]} \geq \frac{K}{k\alpha} \right\}.$$

Reject the k^* hypotheses with largest e-values.

Theorem: The e-BH procedure controls the FDR at level α under arbitrary dependence between the e-values.

In fact, $\mathbb{E}\left[\sup_{\alpha \in (0,1)} \frac{\mathrm{FDR}_\alpha}{\alpha} \right] \leq 1$, allowing post-hoc choice of α.

For e-processes, the FDR guarantee holds at any stopping time.

In fact, *every* FDR procedure can be written as an application of e-BH on a set of "compound" e-values!

Wang, Ramdas (JRSSB'22), Ignatiadis et al. (arXiv)

E-*processes* for composite null vs. composite alternative?

$$\frac{\text{Likelihood of alternative } P_\nu}{\text{Likelihood of null } P_\theta}$$

"Frequentist"

$$\frac{\text{Maximum (alt) likelihood}}{\text{Maximum (null) Likelihood}}$$

Bayes

$$\frac{\text{Mixture (alt) likelihood}}{\text{Mixture (null) Likelihood}}$$

Game-theoretic
(wealth in a game)

$$\frac{\text{Mixture or plug-in (alt) likelihood}}{\text{Maximum (null) Likelihood}}$$

Can use "prior" information to bet
on the alternative.
But evidence is compared to best null.

Only the last option is an e-process (the "universal inference" e-process).
It has the asymptotically optimal growth rate (Dixit-Martin'23).

Universal Inference (PNAS 2020) and Testing exchangeability (IJAR'22)

(Continued: "universal inference")

$$W_T = \frac{\text{Mixture/Plug-in (alt) likelihood}}{\text{Maximum (null) Likelihood}} = \prod_{i=1}^{T} \frac{q_{\hat{\theta}_i}(X_i)}{p_{\hat{\theta}_T}(X_i)} \text{ is an e-process.}$$

Also if the numerator is nonparametrically chosen smartly, then universal inference (above) is also asymptotically growth rate optimal!

Under mild conditions, $\mathbb{E}[\log W_T]/T \to K(Q^*, \mathscr{P})$

Dixit and Martin (2023, arXiv)

As an e-value, it is always worse than the numeraire,
but the numeraire is an e-value, while universal inference is an e-process.

Open problem: determine when the sequence of numeraires (at increasing sample sizes) does or does not yield an e-process.

Summary

1. E-processes are sufficient for sequential testing. Every sequential test can be recovered by thresholding an e-process.

2. E-processes are necessary for sequential testing composite nulls. Nonnegative martingales and supermartingales do not suffice.

3. Testing by betting is a practical and powerful approach for statistical testing, even in nonparametric settings.

4. Log-optimal betting strategies always exist without any assumptions, and are intimately connected to coding/info theory.

5. The e-BH procedure is a universal procedure for controlling the false discovery rate, when used with "compound" e-values.