Sequential Inference and Decision Making

Jiri Hron and Wessel Bruinsma

CBL, University of Cambridge

25 Nov 2020

Introduction and Motivation

An (in)famous example: Power poses

A well cited paper by Carney, Cuddy & Yap (2010).

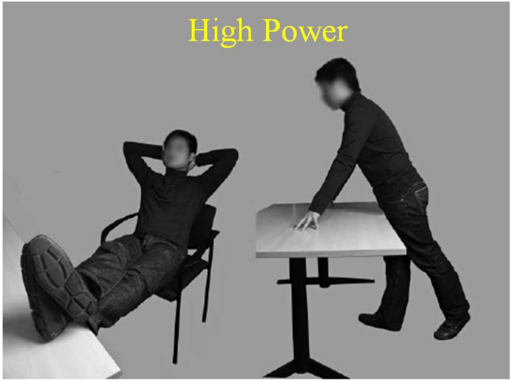


Fig. 1. The two high-power poses used in the study. Participants in the high-power-pose condition were posed in expansive positions with open limbs.

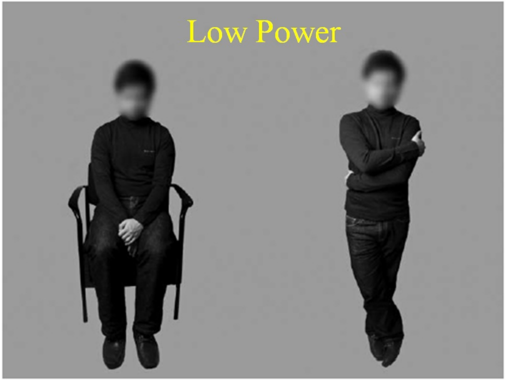


Fig. 2. The two low-power poses used in the study. Participants in the low-power-pose condition were posed in contractive positions with closed limbs.

An (in)famous example: Power poses (2)

Dana Carney's (first author) retraction of her name:

There are a number of methodological comments regarding Carney, Cuddy & Yap (2010) paper that I would like to articulate here.

Here are some facts

1. There is a dataset posted on dataverse that was posted by Nathan Fosse. It is posted as a replication but it is, in fact, merely a "re-analysis." I disagree with one outlier he has specified on the data posted on dataverse (subject # 47 should also be included—or none since they are mostly 2.5 SDs from the mean. However the cortisol effect is significant whether cortisol outliers are included or not). I have posted data on my website that replicates all effects in a re-analysis except the cortisol one (although it is still significant).
2. The data are real.
3. The sample size is tiny.
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 $N = 42$.

Peak at your peril

Why does Carney speak about *p-hacking*? Let's review.

Confidence sets: Must satisfy $\mathbb{P}(\theta \in \mathrm{CI}_n^\alpha) \geq 1-\alpha$. For example, $\hat{\theta}_n := \frac{1}{n}\sum_{i=1}^n X_i \sim \mathcal{N}(\theta, \frac{\sigma^2}{n})$ for $n \gg 0$ by CLT, so the classical *z-interval* $\mathrm{CI}_n^\alpha = [\hat{\theta}_n - z_{\frac{\alpha}{2}}\frac{\hat{\sigma}_n}{\sqrt{n}}, \hat{\theta}_n + z_{\frac{\alpha}{2}}\frac{\hat{\sigma}_n}{\sqrt{n}}]$.

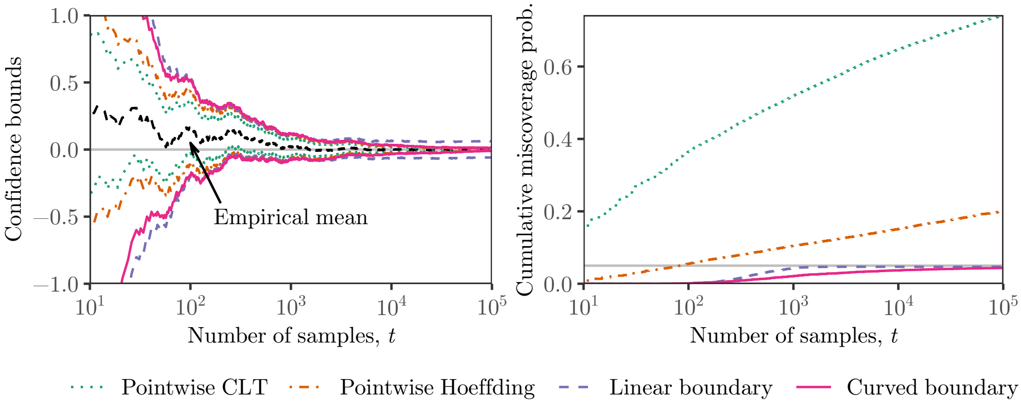
Duality between p-values and confidence sets:

1. a p-value for $H_0 \colon \theta = \theta_0$ based on $(\mathrm{CI}_n^\alpha)_\alpha$ is $P_{\theta_0} = \sup\{\alpha \in [0,1] \colon \theta_0 \in \mathrm{CI}_n^\alpha\}$.
2. a confidence set based on $(P_{\theta_0})_{\theta_0}$ is $\mathrm{CI}_n^\alpha = \{\theta_0 \in \mathbb{R} \colon P_{\theta_0} > \alpha\}$.

Issue: These are valid only for a **fixed a priori selected** n!

A concrete example

Setup: X_i i.i.d. Rademacher, $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^n X_i$, $\theta = \mathbb{E}[\hat{\theta}_n] = 0$.



Francis J. Anscombe

'Sampling to reach a foregone conclusion.'

Fixed-sample-size analysis of sequential observations.
Biometrics, 10 (1954), 89-100.

Rest of the presentation

Goal: Introduce *sequential analysis* tools which are valid under arbitrary *stopping rules*, and control for *multiple hypothesis testing* under *arbitrary dependencies*.

Applications: Statistical analysis in natural and social sciences, randomised trials in medicine, causal inference, A/B testing, . . .

Rest of the presentation (2)

Time-Uniform Chernoff Bounds

EXAMPLE: Evaluation of a Classifier

- Consider data set of labelled i.i.d. observations:

$$(Z_i, Y_i)_{i=1}^{n_{\mathsf{train}}} \subseteq \underbrace{\mathcal{Z}}_{\text{some feature space}} \times \{❶, ❷\}.$$

- Consider a trained classifier

$$F\colon \mathcal{Z} \to \{❶, ❷\}.$$

- We wish to test whether the classifier learned *anything*:

$$\mathbb{P}(F(Z^*) = Y^*) \overset{?}{>} \tfrac{1}{2}.$$

new observation

EXAMPLE: Evaluation of a Classifier (2)

- We will collect *new* i.i.d. samples $(Z_1^*, Y_1^*), (Z_2^*, Y_2^*), \ldots$, set

$$X_i = \begin{cases} +1 & \text{if } F(Z_i^*) = Y_i^*, \\ -1 & \text{if } F(Z_i^*) \neq Y_i^*, \end{cases}$$

  and test $H_0 \colon \mathbb{E}[X] = 0$ against $H_1 \colon \mathbb{E}[X] > 0$.

- Standard approach:
  1. Collect n such samples.
  2. Estimate the score on this new data: $\frac{1}{n} S_n := \frac{1}{n} \sum_{i=1}^{n} X_i$.
  3. Compute a one-sided α-confidence interval $(-\infty, U_n)$:
  $$\mathbb{P}_0(\tfrac{1}{n} S_n \in (-\infty, U_n)) \geq 1 - \alpha.$$
  4. Reject H_0 if $\frac{1}{n} S_n \notin (-\infty, U_n)$.

EXAMPLE: Evaluation of a Classifier (3)

- Collecting samples can be expensive!

✗ Cannot prematurely stop the collection process.

✗ Cannot collect more samples if you failed to reject H_0.

- Mathematical statement of these observations:

$$\mathbb{P}_0\left(\tfrac{1}{\tau} S_\tau \in (-\infty, U_\tau)\right) \not\geq 1 - \alpha \quad \text{for a stopping rule } \tau.$$

$$\boxed{\text{Can we modify the standard approach to allow any stopping rule?}}$$

- Yes! Need uniform guarantee (Lem 1, Ramdas et al., 2020):

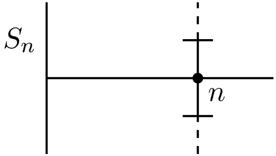
$$\mathbb{P}_0\left(\forall\, n : \tfrac{1}{n} S_n \in (-\infty, U_n)\right) \geq 1 - \alpha.$$

Uniform Confidence Sequences

Pointwise confidence interval:

$$\mathbb{P}_0(S_n \in \mathrm{CI}_n) \geq 1 - \alpha$$

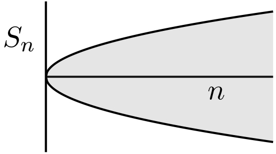
- A single CI_n
- ✗ Valid for a single $n \in \mathbb{N}$
- ✓ Tighter



Uniform confidence sequence:

$$\mathbb{P}_0(\forall\, n : S_n \in \mathrm{CI}_n) \geq 1 - \alpha$$

- A sequence $(\mathrm{CI}_n)_{n \geq 1}$
- ✓ Valid for all $n \in \mathbb{N}$
- ✗ Looser (no free lunch!)



Constructing CIs and UCSs

Pointwise confidence interval:

$$\mathbb{P}_0(S_n \geq b) \leq g(b)$$

- Cramér–Chernoff method
- Markov's inequality:

$$\mathbb{P}(X \geq b) \leq \frac{1}{b}\mathbb{E}[X]$$

- Underlying martingale:

$$(L_n)_{n\geq 1}$$

Uniform confidence sequence:

$$\mathbb{P}_0(\exists\, n : S_n \geq an + b) \leq g(a, b)$$

- Cramér–Chernoff method

$\Rightarrow$ Ville's inequality:

$$\mathbb{P}(\sup_{n\geq 1} L_n \geq b) \leq \frac{1}{b}\mathbb{E}[L_1]$$

Enables uniform guarantee!

all information
up to time n

L_n = capital at time n,
fair betting game

- Definition: $(L_n)_{n\geq 1}$ is a martingale if

$$\mathbb{E}[L_{n+1} \,|\, \mathcal{F}_n] = L_n.$$

- Intuition: martingales are increasingly finer averages:

$$L_n = \mathbb{E}[L_\infty \,|\, \mathcal{F}_n].$$

✓ Constant expectation: $\mathbb{E}[L_n] = \mathbb{E}[L_1]$.

✓ Ville's inequality:

$$\mathbb{P}(\sup_{n\geq 1} L_n \geq b) \leq \tfrac{1}{b}\mathbb{E}[L_1]$$

RECAP: MGF, CFG, and Convex Conjugate

- Moment-generating function (MGF):

$$\varphi_X(\lambda) := \mathbb{E}[e^{\lambda X}], \qquad \varphi_{S_n}(\lambda) = \varphi_X(\lambda)^n.$$

- Cumulant-generating function (CGF):

$$\psi_X(\lambda) := \log \varphi_X(\lambda), \quad \psi_{S_n}(\lambda) = n\psi_X(\lambda).$$

- Convex conjugate of CGF:

$$\psi_X^*(b) := \sup_{\lambda \in \mathbb{R}} (b\lambda - \psi_X(\lambda)).$$

- ψ_X^* determines how $\frac{1}{n}S_n$ fluctuates around $\mathbb{E}[X]$ (*e.g.*, Cramér's Theorem).

Pointwise Concentration of S_n

- Cramér–Chernoff method:

$$\begin{aligned}
\mathbb{P}_0(S_n \geq b) &= \mathbb{P}_0(e^{\lambda S_n} \geq e^{\lambda b}) && (\lambda > 0) \\
&\leq e^{-\lambda b}\,\mathbb{E}[e^{\lambda S_n}] && \text{(Markov's inequality)} \\
&\overset{\text{(i)}}{=} \exp\left[-n\left(\tfrac{b}{n}\lambda - \psi_X(\lambda)\right)\right], && \text{(def. of } \psi_X\text{)}
\end{aligned}$$

$$\mathbb{P}_0(S_n \geq b) \leq \exp\left[-n\psi_X^*\left(\tfrac{b}{n}\right)\right]. \qquad (\text{inf over } \lambda > 0)$$

- Equivalently, (i) uses that L_n is a martingale:

$$L_n := e^{\lambda S_n - n\psi_X(\lambda)}, \quad \underbrace{\mathbb{E}[L_{n+1} \mid \mathcal{F}_n] = L_n}_{\text{def. of martingale}}, \quad \mathbb{E}[L_n] = 1.$$

$\Rightarrow$ Use this to generalise to UCS!

Line Crossing Inequality for S_n

- Choose $\lambda > 0$ such that $a \geq \frac{\psi(\lambda)}{\lambda}$:

$$\begin{aligned}\mathbb{P}_0(\exists\, n : S_n \geq an + b) &\leq \mathbb{P}_0(\exists\, n : S_n \geq \tfrac{\psi(\lambda)}{\lambda} n + b) \\ &= \mathbb{P}_0(\exists\, n : e^{\lambda S_n - n\psi(\lambda)} \geq e^{\lambda b}) \\ &= \mathbb{P}_0(\,\sup_{n\geq 1} e^{\lambda S_n - n\psi(\lambda)} \geq e^{\lambda b}) \\ &\leq e^{-\lambda b}\, \mathbb{E}[L_1] = e^{-\lambda b}, && \text{(Ville's inequality)}\end{aligned}$$

$$\mathbb{P}_0(\exists\, n : S_n \geq an + b) \leq \exp[-D(a) b], \qquad (\inf \text{ over } \lambda)$$

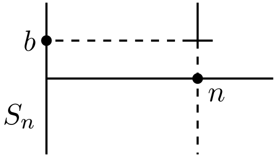
where $D(a) = \sup\{\lambda > 0 : a \geq \frac{\psi(\lambda)}{\lambda}\}$ (inverse of $\lambda \mapsto \frac{\psi(\lambda)}{\lambda}$).

- This is an exponential line crossing inequality.
- Cannot be improved: equality for BM ($D(a) = 2a$).

Line Crossing Inequality for S_n (2)

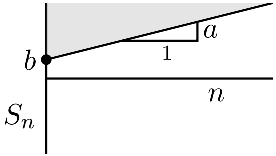
Pointwise confidence interval:

$$\mathbb{P}_0(S_n \geq b) \leq \exp\left[-n\psi_X^*\left(\tfrac{b}{n}\right)\right]$$



Uniform confidence sequence:

$$\mathbb{P}_0(\exists\, n : S_n \geq an + b) \leq \exp[-D(a)b]$$

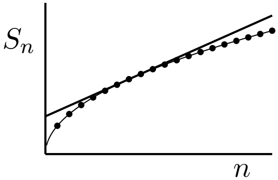


- How these compare?
- Towards CI and UCS: set RHS to α and solve for b.

Line Crossing Inequality for S_n (3)

Pointwise confidence interval:

$$\mathbb{P}_0(S_n \geq n\psi_X^{*-1}[\tfrac{1}{n}\log(\tfrac{1}{\alpha})]) \leq \alpha$$

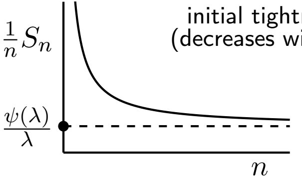


Uniform confidence sequence:

$$\mathbb{P}_0(\exists n : S_n \geq \frac{\psi(\lambda)}{\lambda} n + \frac{\log(1/\alpha)}{\lambda}) \leq \alpha$$

asymptotic tightness (increases with λ)

initial tightness (decreases with λ)



- Linearisation of pointwise bound gives uniform bound!

✗ Uniform bound fails to produce UCS for $\frac{1}{n}S_n$ that goes to zero...

Mixture Boundaries

typically $S_n = O_p(\sqrt{n})$

- We require a sublinear-boundary crossing inequality for S_n.
- Heart of argument of line crossing inequality:

$$S_n \geq \frac{\psi(\lambda)}{\lambda} n + \frac{\log(1/\alpha)}{\lambda} \implies L_n(\lambda) = e^{\lambda S_n - n\psi(\lambda)} \geq \frac{1}{\alpha}.$$

- Condition on S_n can be written in a more direct way:

$$S_n \geq \sup\{s \in \mathbb{R} : e^{\lambda s - n\psi(\lambda)} < \tfrac{1}{\alpha}\} := \mathcal{M}_\alpha(n \mid \lambda).$$

- Observation: $\lambda > 0$ optimally restricts s for one $n \in \mathbb{N}$.
- Idea: average over $\lambda > 0$ to get compromise for all $n \geq 1$.

Mixture Boundaries (2)

- Mixture boundary:

$$\mathcal{M}_\alpha(n) = \sup\{s \in \mathbb{R} : \mathbb{E}_{\lambda>0}[e^{\lambda s - n\psi(\lambda)}] < \tfrac{1}{\alpha}\},$$

  which guarantees $\mathbb{P}_0(\exists\, n : S_n \geq \mathcal{M}_\alpha(n)) \leq \alpha$.
- Distribution F over λ determines around which n the boundary $\mathcal{M}_\alpha$ is tightest. (Knob to tune in practice!)
- Exploit conjugacy to obtain convenient $\mathcal{M}_\alpha(n) = O(\sqrt{n \log n})$.
- Optimise F to approach optimal $\mathcal{M}_\alpha(n) = O(\sqrt{n \log\log n})$.
- Reveals $\mathcal{M}_\alpha(n)$ as nonasymptotic analogue of LIL:

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n \log\log n}} = 1 \quad \text{almost surely.}$$

Sub-ψ Processes and the Master Theorem

- Howard et al. (2018a,b) generalise story and provide much more detail. Fantastic read. Highly recommended!
- Key definition: $(S_t)_{t \in \mathcal{T} \cup \{0\}}$ is ℓ_0-sub-ψ if

any function *like* CGF

$$\exp(\lambda S_t - \psi(\lambda) V_t) \leq L_t(\lambda) \quad \text{almost surely.}$$

variance process, measures time

supermartingale, $L_0(\lambda) \leq \ell_0$

- Theorem 1 by Howard et al. (2018a): weaker assumptions and stronger results!
- Results generalise to continuous time and processes taking values in Banach spaces (vectors, matrices, ...).
- Cool applications: empirical Bernstein UCS to estimate ATE in Neyman–Rubin model, matrix LIL, ...

Bets and confidence sequences

A boundary of uniform confidence set can be constructed via

$$\mathcal{M}_\alpha(v) = \sup\left\{s \in \mathbb{R} \colon \int \exp\{\lambda s - \psi(\lambda) v\} \, \mathrm{d}F(\lambda) < \frac{1}{\alpha}\right\}$$

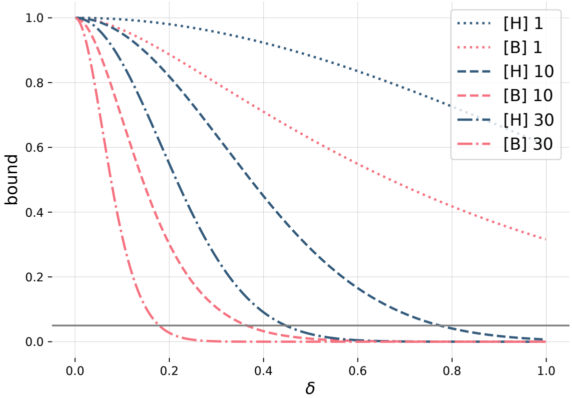
yielding $\mathbb{P}(\exists n \colon S_n \geq \mathcal{M}_\alpha(V_n)) \leq \mathbb{P}(\exists n \colon L_n \geq \frac{1}{\alpha}) \leq \alpha$ (Ville).

$\mathcal{M}_\alpha$ is unimprovable in the sub-Gaussian case (tight for Brownian motion), but can be loose or computationally demanding in other cases!

Can we do better in some special case?

Bounds and variances

If $|X_i| \leq b$ and $\sigma^2 \ll b^2$, Bernstein $\mathbb{P}(S_n \geq n\delta) \leq e^{-\frac{n\delta^2}{2(\sigma^2 + b\delta/3)}}$ much tighter than the (sub-Gaussian) Hoeffding $\mathbb{P}(S_n \geq n\delta) \leq e^{-\frac{n\delta^2}{2b^2}}$!



Looseness of $\mathcal{M}_\alpha$ for bounded r.v.s

$$\mathcal{M}_\alpha(v) = \sup\left\{ s \in \mathbb{R} \colon \int \exp\{\lambda s - \underbrace{\psi(\lambda) v}_{\text{(II)}}\} \underbrace{\mathrm{d}F(\lambda)}_{\text{(I)}} < \frac{1}{\alpha} \right\}$$

(I): Ville valid for any $L_n(\lambda) = e^{\lambda S_n - \psi(\lambda) V_n}$. λ determines where $S_n \geq \frac{\psi(\lambda)}{\lambda} V_n + \frac{\log(1/\alpha)}{\lambda}$ tightest, but $\mathcal{M}_\alpha$ compromises by mixing rather than optimising over λ.

Idea: Replace $\mathrm{d}F(\lambda)$ by a *predictable* sequence $(\lambda_n)_{n \geq 1}$. For $V_n = n$

$$\begin{aligned}
\mathbb{E}[L_{n+1} \mid X_{1:n}, \lambda_{1:n}] &= \mathbb{E}[e^{\sum_{i=1}^{n+1} \lambda_i X_i - \psi(\lambda_i)} \mid X_{1:n}, \lambda_{1:n}] \\
&= L_n \mathbb{E}[e^{\lambda_{n+1} X_{n+1} - \psi(\lambda_{n+1})} \mid X_{1:n}, \lambda_{1:n}] \leq L_n
\end{aligned}$$

which allows estimating λ_n closer to an optimal $\lambda_n^\star$.

Looseness of $\mathcal{M}_\alpha$ for bounded r.v.s (2)

$$\mathcal{M}_\alpha(v) = \sup\left\{ s \in \mathbb{R} \colon \int \exp\{\lambda s - \underbrace{\psi(\lambda) v}_{\text{(II)}}\} \underbrace{\mathrm{d}F(\lambda)}_{\text{(I)}} < \frac{1}{\alpha} \right\}$$

(II): If $\psi_X < \psi$, L_n 'strict' supermartingale $\implies$ Ville loose. Why? Recall in Ville, we define a stopping time $\tau := \inf\{n \geq 1 \colon L_n \geq \delta\}$

$$\mathbb{E}[L_0] \overset{(\star)}{\geq} \mathbb{E}[L_\tau] \geq \mathbb{E}[L_\tau \mathbb{1}_{\tau < \infty}] \geq \delta \mathbb{P}(\exists n \colon L_n \geq \delta)$$

with $(\star)$ an equality if L is a martingale (OST), i.e., $\psi = \psi_X$.

Idea: Use $(L_n)_{n \geq 1}$ which is always a martingale. (Comes next!)

Capital processes

Setup: Initial capital $L_0 = 1$. We are tasked with placing a series of *predictable* bets $\lambda_n \in [-1, 1]$ on a trial outcome $X_n \in \{-1, +1\}$

$$\begin{aligned} L_1 &:= 1 + \lambda_1 X_1 && \leftarrow X_1 = +1 \\ L_2 &:= (1 + \lambda_2 X_2)(1 + \lambda_1 X_1) && \leftarrow X_2 = -1 \\ &\vdots \\ L_n &:= (1 + \lambda_n X_n) \prod_{i=1}^{n-1} (1 + \lambda_i X_i) && \leftarrow X_n = -1 \end{aligned}$$

where $\operatorname{sign}(\lambda_n)$ encodes belief about direction, $|\lambda_n|$ confidence.

$(L_n)_{n \geq 1}$ is called capital process in game-theoretic probability. Each round we can loose all or double our *capital* depending on λ_n.

Gambling with Ville

If $(X_n) \subset [-1, 1]$ a zero-mean martingale (null), then

$$L_n = \prod_{i=1}^{n} (1 + \lambda_i X_i)$$

is a non-negative martingale for $(\lambda_n) \subset [-1, 1]$.

$L_n \gg 0$ *evidence against null* as $\mathbb{P}_0(\exists n \colon L_n \geq \frac{1}{\alpha}) \leq \alpha$ (Ville) $\implies$ invert the hypothesis test to get UCS!

$X_n \in [0, 1]$, $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = \mu$. Play simultaneously for all $m \in [0, 1]$ with $|\lambda_n(m)| \leq 1$

$$\mathcal{K}_n(m) \coloneqq \prod_{i=1}^{n} [1 + \lambda_i(m)(X_i - m)]$$

which makes $\mathcal{K}(\mu)$ a martingale. The UCS is then $\mathrm{CI}_n = \{m \in [0, 1] \colon \mathcal{K}_n(m) < \frac{1}{\alpha}\}$.

A betting scheme

Approximate *hindsight optimal constant* bet $\lambda_n^\star(m)$ for each $m \in [0,1]$

$$\frac{1}{n}\frac{\mathrm{d}\log \mathcal{K}_n(m)}{\mathrm{d}\lambda} = \underbrace{\frac{1}{n}\sum_{i=1}^{n}\frac{X_i - m}{1+\lambda(X_i - m)}}_{\text{not predictable}} \approx \underbrace{\frac{1}{n-1}\sum_{i=1}^{n-1}\frac{X_i - m}{1+\lambda(X_i - m)}}_{\text{predictable}} \overset{\text{set}}{=} 0$$

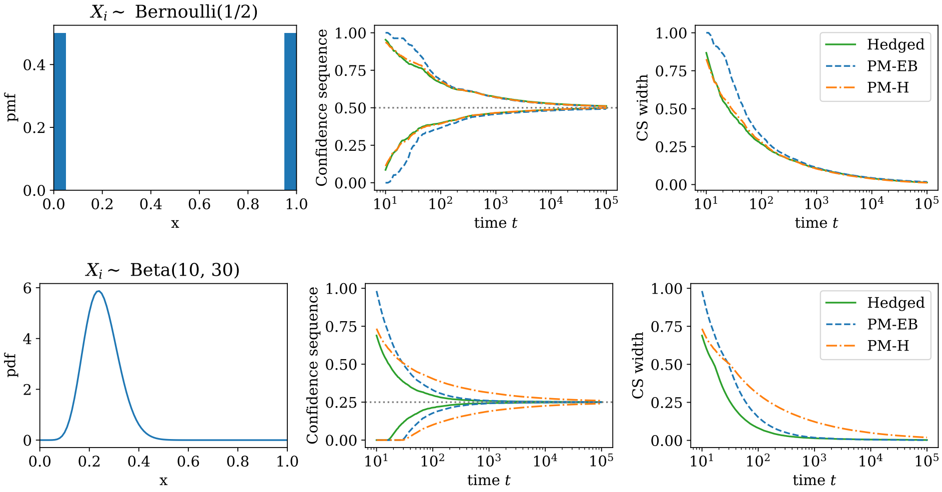
Approximating $(1+z)^{-1} \approx 1 - z$ for $z \approx 0$ (Taylor)

$$\frac{1}{n-1}\sum_{i=1}^{n-1}(X_i - m)[1 - \lambda(X_i - m)] = \hat{\mu}_{n-1} - m - \lambda[\hat{\sigma}_{n-1}^2 + (\hat{\mu}_{n-1} - m)^2] = 0$$

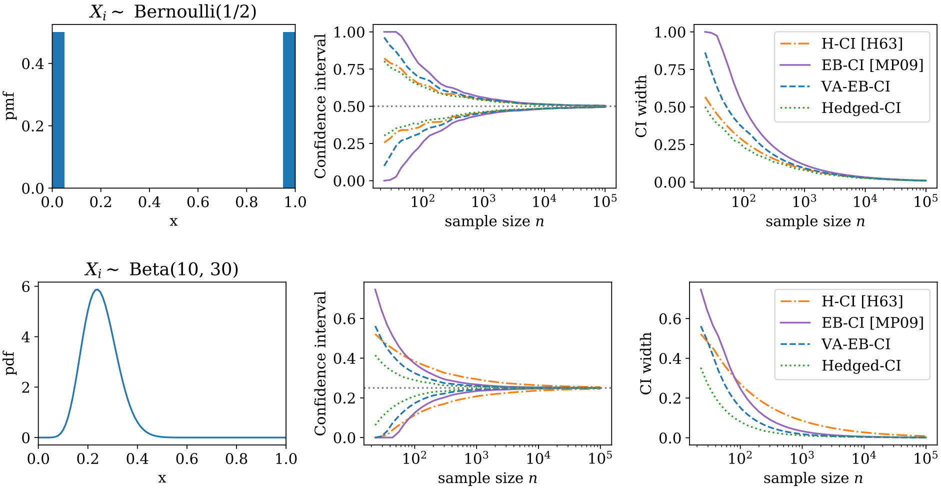
$$\implies \lambda_n^\star(m) \approx \lambda_n(m) = \frac{\hat{\mu}_{n-1} - m}{\hat{\sigma}_{n-1}^2 + (\hat{\mu}_{n-1} - m)^2}$$

where $\lambda_n(m)$ is clipped to $[-1,1]$ when substituting to $\mathcal{K}_n(m)$.

Uniform confidence sequences



Fixed sample confidence intervals!



E-values and false discovery rate control

Betting scores are e-values

We saw $\mathcal{K}_n(m) = \prod_{i=1}^n (1 + \lambda_i(X_i - m))$ can be used to construct

$$\mathsf{CI}_n = \left\{ m \colon \mathcal{K}_n(m) < \frac{1}{\alpha} \right\} = \left\{ m \colon \frac{1}{\mathcal{K}_n(m)} > \alpha \right\}$$

Compare: *p-values P_{θ_0} ($H_0 \colon \theta = \theta_0$) can be inverted* to obtain $\mathsf{CI}_n = \{\theta_0 \colon P_{\theta_0} > \alpha\}$.

$\mathcal{K}_n(m)$ is an example of a class of random variables called e-values.

- Pointwise: $[0, \infty]$ random variable E with $\mathbb{E}_0[E] \leq 1$.
- Uniform: $[0, \infty]$ random variables $(E_n)_{n \geq 1}$ with $\mathbb{E}_0[E_\tau] \leq 1$ for any stopping time τ.

Are e-values just p-values rebranded?

Not really! Both measure evidence against some null hypothesis but:

- p-value P is $[0,1]$ random variable satisfying $\mathbb{P}_0(P \leq \alpha) \leq \alpha$
- e-value E is $[0,\infty]$ random variable satisfying $\mathbb{E}_0[E] \leq 1$

E-values only require information about expectation, p-values about the CDF!

An e-value can be converted to p-value via Markov

$$\mathbb{P}_0(\tfrac{1}{E} \leq \alpha) = \mathbb{P}_0(E \geq \tfrac{1}{\alpha}) \leq \alpha \mathbb{E}_0[E] \leq \alpha$$

Conversions in the opposite direction exist as well, but neither are statistically efficient!

When we might prefer e-values over p-values?

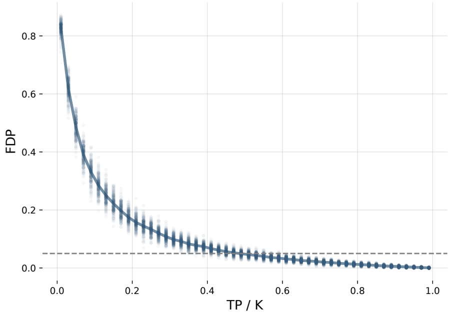
1. e-values occur naturally in (sequential) inference
   - test supermartingales $\prod_{i=1}^{n} e^{\lambda X_i - \psi(\lambda)}$
   - capital processes $\prod_{i=1}^{n}(1 + \lambda_i X_i)$
   - likelihood ratios $\prod_{i=1}^{n} \frac{p_\theta(X_i)}{p_0(X_i)}$
2. p-values need tail but e-values only expectation control
   - p-values typically more sensitive to misspecification
   - many popular p-values based on asymptotics (z/t-statistic, Wilks, etc.)
   - example: $e^{\lambda X - \frac{\lambda^2 X^2}{2}}$ valid e-value for any symmetric r.v. X
3. p-values often depend on independence
   - especially asymptotic arguments (CLT, Wilks, etc.)
   - e-values more flexible as we will see next

False discovery rate control

Task: Test multiple hypothesis $H_1, \ldots, H_K$.

Issue: $\text{FDP} = \frac{\text{FP}}{\text{TP}+\text{FP}}$ can be much higher than α even if each true null has FP probability $\leq \alpha$.

Idea: Control $\text{FDR} = \mathbb{E}[\text{FDP}]$.



Benjamini-Hochberg: Order p-values from lowest to largest $p_{(1)}, \ldots, p_{(K)}$

$$k_\star := \max\left\{k \in [K] \colon p_{(k)} \leq \alpha \frac{k}{K}\right\}$$

and reject hypotheses associated with the p-values $p_{(1)}, \ldots, p_{(k_\star)}$.

Benjamini-Hochberg with e-values

BH ensures FDR $= \mathbb{E}[\text{FPR}] \leq \alpha$ but not if tests are **dependent**!

Benjamini-Yekutieli: $k_\star = \sup\{k \in [K] \colon p_{(k)} \leq \alpha \frac{k}{K c_k}\}$, $c_k = \sum_{i=1}^{k} \frac{1}{i}$. BY works for any dependence structure but looses power!

e-BH: Order e-values from largest to lowest $e_{[1]}, \ldots, e_{[K]}$

$$k_\star = \max\left\{k \in [K] \colon e_{[k]} \geq \frac{1}{\alpha}\frac{K}{k}\right\}$$

and reject hypothesis associated with the e-values $e_{[1]}, \ldots, e_{[k_\star]}$. e-BH controls FDR $\leq \alpha$ even if $e_1, \ldots, e_K$ are arbitrarily **dependent**!

Proving e-BH controls FDR

1. Recall $K_\star = \max\left\{k \in [K] \colon E_{[k]} \geq \frac{1}{\alpha}\frac{K}{k}\right\}$.

2. With $N \subseteq [K]$ the true nulls, $G \subseteq [K]$ the rejects, and $\frac{0}{0} = 0$

$$\mathsf{FDP} = \frac{|N \cap G|}{|G|} = \sum_{k \in N} \frac{\mathbb{1}_{k \in G}}{|G|} \overset{(\star)}{\leq} \sum_{k \in N} \frac{\alpha E_k}{K} \mathbb{1}_{k \in G}$$

where $(\star)$ is by $E_k \geq E_{[K_\star]} \geq \frac{1}{\alpha}\frac{K}{K_\star} = \frac{1}{\alpha}\frac{K}{|G|}$ for all $k \in G$.

3. Since $\mathbb{E}[E_k] \leq 1$ for $k \in N$

$$\mathsf{FDR} = \mathbb{E}[\mathsf{FDP}] \leq \frac{\alpha}{K} \sum_{k \in N} \mathbb{E}[E_k \mathbb{1}_{k \in G}] \overset{E_k \geq 0}{\leq} \alpha \frac{|N|}{K} \leq \alpha$$

Wrap-Up

Wrap-Up

Time-uniform Chernoff bounds: (Howard et al., 2018a; Howard et al., 2018b)

- UCSs key to flexible sequential inference.
- Often a martingale behind pointwise concentration bound.

$\Rightarrow$ Enables generalisation to uniform bound.

Bets and confidence sequences: (Waudby-Smith and Ramdas, 2020)

- Links between betting strategies, and test power maximisation.
- Tighter bounds (even fixed-sample) for bounded random variables.

E-values and false discovery rate control: (Wang and Ramdas, 2020)

- E-values trade-off power for validity relative to p-values.
- Many uses including FDR control under arbitrary dependencies.

References

Howard, S. R., Ramdas, A., McAuliffe, J., & Sekhon, J. (2018a). Time-uniform Chernoff bounds via nonnegative supermartingales. *arXiv preprint arXiv:1808.03204*. eprint: https://arxiv.org/abs/1808.03204

Howard, S. R., Ramdas, A., McAuliffe, J., & Sekhon, J. (2018b). Time-uniform, nonparametric, nonasymptotic confidence sequences. *arXiv preprint arXiv:1810.08240*. eprint: https://arxiv.org/abs/1810.08240

Ramdas, A., Ruf, J., Larsson, M., & Koolen, W. (2020). Admissible anytime-valid sequential inference must rely on nonnegative martingales. *arXiv preprint arXiv:2009.03167*. eprint: https://arxiv.org/abs/2009.03167

Wang, R., & Ramdas, A. (2020). False discovery rate control with E-values. *arXiv preprint arXiv:2009.02824*. eprint: https://arxiv.org/abs/2009.02824

Waudby-Smith, I., & Ramdas, A. (2020). Variance-adaptive confidence sequences by betting. *arXiv preprint arXiv:2010.09686*. eprint: https://arxiv.org/abs/2010.09686