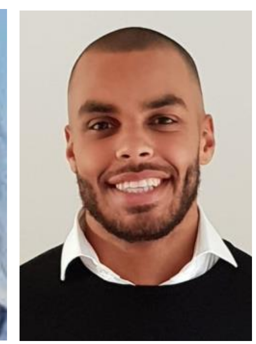
E is the new P

Peter Grünwald

Centrum Wiskunde & Informatica – Amsterdam

Mathematical Institute – Leiden University



NP

Neyman-Pearson

A Paradox of NP Hypothesis Testing

*"[all of] mathematical statistics deals with problems relating to performance characteristics of rules of inductive behavior [ie **decision rules**] based on random experiments"*

- J. Neyman (1950)

A Paradox of NP Hypothesis Testing

"[all of] mathematical statistics deals with problems relating to performance characteristics of rules of inductive behavior [ie ***decision rules****] based on random experiments"* - J. Neyman (1950)

- …when doing a level α-hypothesis test, it is common to mention p-value next to accept/reject decision

- …but then, what are the decision-theoretic consequences of this? Can we use an observation $p \ll \alpha$ to somehow get better decisions?

The Problem of

roving α`s

The Paradox of NP Hypothesis Testing

*"[all of] mathematical statistics deals with problems relating to performance characteristics of rules of inductive behavior [ie **decision rules**] based on random experiments"* - J. Neyman (1950)

- …when doing a level α-hypothesis test, it is common to mention p-value next to accept/reject decision…but then, what are the decision-theoretic consequences of this?

No simple, valid consequences can be given!

NP adherents (e.g. Lehmann, 1993) simply say p should be mentioned because it *provides more information*...

The Paradox of NP Hypothesis Testing

*"[all of] mathematical statistics deals with problems relating to performance characteristics of rules of inductive behavior [ie **decision rules**] based on random experiments"* - J. Neyman (1950)

- …when doing a level α-hypothesis test, it is common to mention p-value next to accept/reject decision…but then, what are the decision-theoretic consequences of this?

No simple, valid consequences can be given!

Paradox: NP says: statistics is about decision rules. Yet a standard practice in NP statistics has no decision-theoretic interpretation…

Fiercest critics are from within…

- p-values often “bashed” because they are interpreted as evidence against the null, whereas they have properties that any reasonable definition of evidence shouldn’t have (Cox ‘58, likelihoodists, Bayesians…)

- here we provide **Criticism from Within**: even if (like me!) one likes the decision-theoretic NP framework, p-values are better replaced by something else **which does have decision-theoretic consequences…**

E is the new P

- 1st claim: we should mention an **E**-value rather than a **P**-value next to our decision!
- Reason: E-values have valid frequentist **post-hoc** decision-theoretic implications – decision problem may be formulated after data observed, and may be data-dependent
- 2nd contribution: the **E-posterior** – alternative to confidence interval/distribution that can deal with post-hoc loss functions as well
  - **link to ISIPTA concepts…**

But first: …back to Wald (1939)

Reformulate NP testing in terms of losses/risks rather than errors/ error probabilities

- $L : \{0,1\} \times \{0,1\} \to \mathbb{R}_0^+$
- $L(\theta, a)$: loss incurred when H_θ is true, and action a is taken
- $L(0,0) = L(1,1) = 0$

accept H_0 **reject H_0**

- $L(0,1)$: Type-I loss
- $L(1,0)$: Type-II loss

NP with losses rather than errors

- $L : \{0,1\} \times \{0,1\} \rightarrow \mathbb{R}_0^+ \,; L(0,0) = L(1,1) = 0$
- We fix $\ell > 0$ and restrict ourselves to decision rules $\delta : \mathcal{Y} \rightarrow \{0,1\}$ with **Type-I risk bound**:

$$\sup_{P \in H_0} \mathbf{E}_{Y \sim P}[L(0, \delta(Y))] \leq \ell$$

(usually $\ell = L(1,0)$, cost of “maintaining status quo”)

Here’s a δ that achieves this based on p-value $\mathrm{p}(Y)$:

$$\delta(y) = 1 \text{ iff } \mathrm{p}(y) \cdot L(0,1) \leq \ell$$

Example: $\ell := 1 \,; L(0,1) = \alpha^{-1}$ (e.g. 20)

Then $\delta(y) = 1$ iff $\mathrm{p}(y) \cdot \alpha^{-1} \leq 1$ (e.g. reject if $\mathrm{p}(Y) \leq 0.05$)

NP with losses rather than errors

- $L: \{0,1\} \times \{0,1\} \to \mathbb{R}_0^+ ; L(0,0) = L(1,1) = 0$
- We fix $\ell > 0$ and restrict ourselves to decision rules $\delta: \mathcal{Y} \to \{0,1\}$ with **Type-I risk bound**:

$$\sup_{P \in H_0} \mathbf{E}_{Y \sim P}[L(0, \delta(Y))] \leq \ell$$

Set $\delta(y) = 1$ iff $\mathrm{p}(y) \cdot L(0,1) \leq \ell$

Example: $\ell := 1 ; L(0,1) = \alpha^{-1}$ (e.g. 20)

Then $\delta(y) = 1$ iff $\mathrm{p}(y) \cdot \alpha^{-1} \leq 1$ (e.g. reject iff $\mathrm{p}(Y) \leq 0.05$)

Up till now we merely reformulated standard NP theory!

Actions of varying intensity

- Suppose there are more than 2 actions. The more extreme an action, the larger Type-I loss and the higher Type-II gain:

- $L\colon \{0,1\} \times \mathcal{A} \to \mathbb{R}_0^+$ with $\mathcal{A} \subset \mathbb{R}$ is loss function such that $L(0, a)$ increasing in a, and $L(1, a)$ decreasing in a

Actions of varying intensity

- The more extreme an action, the larger Type-I loss and the higher Type-II gain:

- $L: \{0,1\} \times \mathcal{A} \to \mathbb{R}_0^+$ is a loss function such that $L(0, a)$ increasing in a, and $L(1, a)$ decreasing in a

- Intuitive **extension** of p-value based decision rule: set $\delta(y)$ to largest a such that $\mathrm{p}(y) \cdot L(0, a) \leq \ell$

Actions of varying intensity

- $L: \{0,1\} \times \mathcal{A} \to \mathbb{R}_0^+$ with $L(0, a)$ increasing in a, $L(1, a)$ decreasing in a
- Intuitive extension of p-value based decision rule:

  set $\delta(y)$ to largest a such that $\mathrm{p}(y) \cdot L(0, a) \leq \ell$

Ex.: $L(0,0) = 0, L(0,1) = 20, L(0,2) = 100, L(0,3) = 500$

$L(1,0) \ > L(1,1) > L(1,2) > L(1,3), \ell = \mathbf{1}$

If p uniform under H_0 then Type-I risk is: $E_{P_0}\left[L\big(0, \delta(Y)\big)\right] =$

$$\left(\frac{1}{20} - \frac{1}{100}\right) \cdot 20 + \left(\frac{1}{100} - \frac{1}{500}\right) \cdot 100 + \frac{1}{500} \cdot 500 = \mathbf{2.6}$$

Actions of varying intensity

- $L: \{0,1\} \times \mathcal{A} \to \mathbb{R}_0^+, L(0,a)$ increasing, $L(1,a)$ decreasing

- Intuitive extension of p-value based decision rule:
  set $\delta(y)$ to largest a such that $\mathrm{p}(y) \cdot L(0,a) \leq \ell$ (*)

This does not give desired Type-I risk $\leq \ell$ guarantee

…and it is not easy to ‘repair’ the rule (*)

Actions of varying intensity

- $L: \{0,1\} \times \mathcal{A} \to \mathbb{R}_0^+$ is a loss function such that $L(0, a)$ increasing in a, and $L(1, a)$ decreasing in a
- Intuitive extension of p-value based decision rule:

  set $\delta(y)$ to largest a such that $\mathrm{p}(y) \cdot L(0, a) \leq \ell$ (*)

This does not give desired Type-I risk $\leq \ell$ guarantee

…and it is not easy to 'repair' the rule (*)

Clue #1 of this Talk: if we instead use decision rule

  set $\delta(y)$ to largest a such that $S^{-1}(y) \cdot L(0, a) \leq \ell$

where $S(y)$ is any E-value, then we always have the Type-I risk guarantee.

but it gets worse:
Post-Hoc Decision Tasks

- Now suppose loss function L_b only determined after Y is observed. It may depend, in unknown ways, on Y:

- i.e. there is a collection $\{L_b: \{0,1\} \times \mathcal{A}_b \to \mathbb{R}_0^+ : b \in \mathcal{B}\}$ $\forall b \in \mathcal{B}, \mathcal{A}_b \subset \mathbb{R}$, $L_b(0, a)$ increasing, $L_b(1, a)$ decreasing

- We are presented with L_B where B is a RV, and are asked to play $a \in \mathcal{A}_B$. Intuitive p-value based δ is still

  $\delta_B(y) =$ largest $a \in \mathcal{A}_B$ such that $\mathrm{p}(y) \cdot L_B(0, a) \leq \ell$

doesn't give desired Type-I risk bound, **even if $|\boldsymbol{\mathcal{A}_B}| = \mathbf{2}$**

Post-Hoc Decision Tasks

Type-I risk now defined as

$$\sup_{P_0 \in H_0, Q} \mathbf{E}_{Y \sim P_0, B \sim Q|Y} \left[L_B(0, \delta_B(Y)) \right]$$

…where supremum is over **all** conditional distributions Q that map each $Y = y$ to distribution on $\mathcal{B}$

Ex.: $\mathcal{B} = \{1,2,3\}, \mathcal{A}_b = \{0, b\}$,

$L_b(0,0) = 0, L_1(0,1) = 20, L_2(0,2) = 100, L_3(0,3) = 500$

$B(Y) =$ largest b such that $\mathrm{p}(Y) \cdot L_b(0, b) \leq 1$

(Type-I risk of our p-value based δ with this B is 3)

First it gets worse: Post-Hoc Decision Tasks

- There is some collection $\{L_b : \{0,1\} \times \mathcal{A}_b \to \mathbb{R}_0^+ : b \in \mathcal{B}\}$ with $\forall b \in \mathcal{B}$, $L_b(0, a)$ increasing, $L_b(1, a)$ decreasing

- We are presented with $L_B ; \mathcal{A}_B$

  $\delta_B(y) =$ largest $a \in \mathcal{A}_B$ such that $\mathrm{p}(Y) \cdot L_B(0, a) \leq \ell$

doesn't give desired Type-I risk bound

…but again, replacing p by an inverse e-value does give the Type-I guarantee

Why care about post-hoc data-dependent decision problems?

Scientific results go ***on record****, for* ***later*** *use.*

… sometimes years later, when e.g. costs of mass-producing a medication have decreased dramatically…

e-variables/e-values: General Definition

An **e-variable** is a **nonnegative** function $S = S(Y)$ such that for **all** $P_0 \in H_0$, we have

$$\mathbf{E}_{P_0}[S(Y)] \leq 1$$

Meet The E-Team:

Aaditya Ramdas & co., CMU (2018-)

Johanna Ziegel,
& co. Bern (2021-)

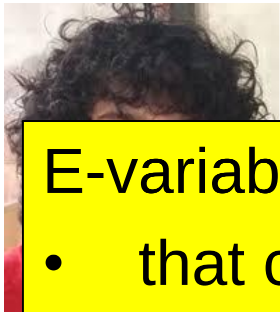
Glenn Shafer & Volodya Vovk (1993, 2011,2019-)

Johannes Ruf,
LSE (2020-)

G., Wouter Koolen and
the CWI Team (2019-)

Ruodu Wang,
U. Waterloo
(2020-)

Meet The E-Team:



Aaditya Ramdas & co., CMU (2018-)

E-variables have been proposed as alternative for p-values

- that can deal more easily with optional continuation, stopping and combining of studies
- that have a clearer interpretation (in terms of betting)
- that are mathematically easier to handle in multiple-testing problems.

…cf. first ***SAVI*** *meeting* last June in the Netherlands…

Today I provide quite a different motivation.

G., Wouter Koolen and the CWI Team (2019-)

U. Waterloo (2020-)

1st Interpretation: p-value after all

- Proposition: Let S be an e-variable. Then $S^{-1}(Y)$ is a conservative p-value, i.e. p-value with **wiggle room**:
- for all $P \in H_0$, all $0 \leq \alpha \leq 1$,

$$P\left(\frac{1}{S(Y)} \leq \alpha \right) \leq \alpha$$

- Proof: just Markov's inequality!

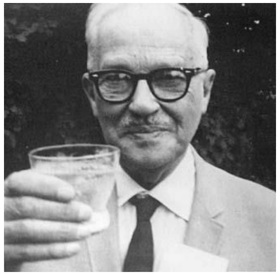
$$P\left(S(Y) \geq \alpha^{-1} \right) \leq \frac{\mathbf{E}[S(Y)]}{\alpha^{-1}} \leq \alpha$$

“Safe” Tests (G. et al, ‘21)

- The test against H_0 at level α based on e-variable S is defined as the test which rejects H_0 if $S(Y) \geq \frac{1}{\alpha}$

- Since S^{-1} is a conservative p-value...

- ....the test which rejects H_0 iff $S(Y) \geq 20$, i.e. $S^{-1}(Y) \leq 0.05$, has **Type-I Error** Bound of 0.05



E-variables and LRs/Bayes factors

- **Bayes factor hypothesis testing (Jeffreys '39)**
  with **simple** $H_0 = \{P_0\}$ vs $H_1 = \{ p_\theta | \theta \in \Theta_1\}$:
  Evidence in favour of H_1 measured by Bayes factor:

$$S = \frac{p_{W_1}(Y)}{p_0(Y)} \quad \text{where} \quad p_{W_1}(Y) := \int_{\theta \in \Theta_1} p_\theta(Y) dW_1(\theta)$$

E-variables and LRs/Bayes factors

- **Bayes factor hypothesis testing (Jeffreys '39)**
  with **simple** $H_0 = \{P_0\}$ vs $H_1 = \{ p_\theta | \theta \in \Theta_1\}$:
  Evidence in favour of H_1 measured by Bayes factor:

$$S = \frac{p_{W_1}(Y)}{p_0(Y)} \quad \text{where} \quad p_{W_1}(Y) := \int_{\theta \in \Theta_1} p_\theta(Y) dW_1(\theta)$$

$$\mathbf{E}_{Y \sim P_0}[S(Y)] = \int p_0(y) \cdot \frac{p_{W_1}(y)}{p_0(y)} dy = \int p_{W_1}(y) dy = 1$$

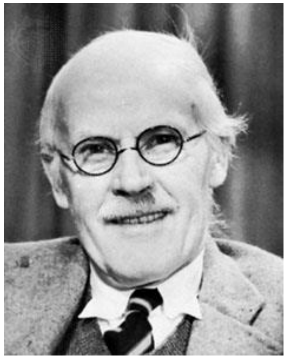
E-variables and LRs/Bayes factors

- **Bayes factor hypothesis testing** **(Jeffreys '39)**
  with **simple** $H_0 = \{P_0\}$ vs $H_1 = \{ p_\theta | \theta \in \Theta_1\}$:
  Evidence in favour of H_1 measured by Bayes factor:

$$S = \frac{p_{W_1}(Y)}{p_0(Y)} \quad \text{where} \quad p_{W_1}(Y) := \int_{\theta \in \Theta_1} p_\theta(Y) dW_1(\theta)$$

$$\mathbf{E}_{Y \sim P_0}[S(Y)] = \int p_0(y) \cdot \frac{p_{W_1}(y)}{p_0(y)} dy = \int p_{W_1}(y) dy = 1$$

No matter what prior W_1 we choose,
Bayes Factor for simple H_0 is an E-variable



E-variables and LRs/Bayes factors

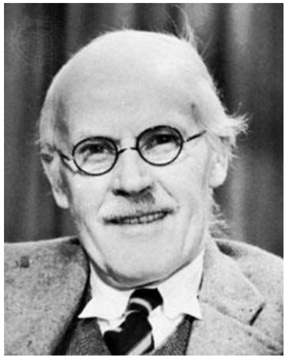
- Side-Remark: even if H_0 is composite or nonparametric, nontrivial E-variables invariably exist but relation to likelihood/ Bayes factors is much trickier.

  The paper ‘Safe Testing’ (G. et al. ‘21) is entirely devoted to composite nulls

$$\mathbf{E}_{Y\sim P_0}[S(Y)] = \int p_0(y)\cdot\frac{p_{W_1}(y)}{p_0(y)}dy = \int p_{W_1}(y)dy = 1$$

No matter what prior W_1 we choose,

Bayes Factor for simple H_0 is an E-variable



Back to the post-hoc losses

GNP Lemma (G. '22), Part I: for every *Generalized* NP problem, and every E-variable S, the **Type-I risk** of the **maximal** S-based decision rule

$\delta_b(y) =$ largest $a \in \mathcal{A}_b$ such that $S^{-1}(y) \cdot L_b(0, a) \leq \ell$

is bounded by ℓ .

Generalized NP means: Type I risk defined as

$$\sup_{P_0 \in H_0, Q} \mathbf{E}_{Y \sim P_0, B \sim Q|Y} \left[L_B(0, \delta_B(Y)) \right]$$

Back to the post-hoc losses

GNP Lemma (G. '22), Part I: for every *Generalized* NP problem, and every E-variable S, the **Type-I risk** of the **maximal** S-based decision rule

$\delta_b(y) =$ largest $a \in \mathcal{A}_b$ such that $S^{-1}(y) \cdot L_b(0, a) \leq \ell$

is bounded by ℓ .

Back to the post-hoc losses

GNP Lemma (G. '22), Part I: for every generalized NP problem, and every E-variable S, the **Type-I risk** of the maximal S-based decision rule

$\delta_b(y) =$ largest $a \in \mathcal{A}_b$ such that $S^{-1}(y) \cdot L_b(0, a) \leq \ell$

is bounded by ℓ .

Proof: for all $P \in H_0$, arbitrary conditional distributions Q for random variable B , we have:

$$\mathbf{E}_{Y \sim P_0, B \sim Q|Y} \left[L_B(0, \delta_B(Y)) \right]$$

$$\leq \mathbf{E}_{Y \sim P_0} \left[\sup_{b \in \mathcal{B}} L_b(0, \delta_b(Y)) \right] \leq \mathbf{E}_{Y \sim P_0} \left[S(Y) \ell \right] \leq 1 \cdot \ell.$$

A Complete Class-Like Theorem

GNP Lemma, Part I : for every generalized NP problem, and every E-variable S for which it exists, the **Type-I risk** of the maximal S-based decision rule

$\delta_b(y) =$ largest $a \in \mathcal{A}_b$ such that $S^{-1}(y) \cdot L_b(0, a) \leq \ell$

is bounded by ℓ .

Part II, about Type-II risk: under reasonable regularity conditions, every δ that has Type-I risk bound $\leq \ell$ and that is "**Type-II risk admissible**" is a maximal S-based decision rule relative to some E-variable S

The **BIND** Assumption

- **If** for each $b \in \mathcal{B}$ the Type-I loss functions $L_b(0,\cdot)$ can take on only two values $\{0, c_b\}$ and B is probabilistically independent of Y **then** we get valid Type-I risk bounds with p- rather than E-vals. after all

- Thus, we might say that classical NP testing works under a **BIND** (Binary+Independence) assumption…

Second Contribution: The E-Posterior

- **BIND** assumption also underlies the valid use of **confidence intervals**, **confidence distributions** and ‘matching’ Objective Bayes posteriors.
- It often cannot be justified…and then we should replace the confidence distribution by something more conservative...: the E-Posterior

E-posteriors…

- …have been implicitly used for several years, to provide anytime-valid confidence intervals

- i.e. to define confidence sequences that are valid under optional stopping, but still for **fixed $\boldsymbol{\alpha}$!**
  - p-values $\Rightarrow$ standard Cis
  - e-values $\Rightarrow$ e-CIs

E-posteriors…

- …have been implicitly used for several years, to provide anytime-valid confidence intervals

- i.e. to define confidence sequences that are valid under optional stopping, but still for **fixed $\boldsymbol{\alpha}$!**

Instead we use them

- to make **risk-based confidence statements** (a bit like data-dependent α) …
- (make predictions and uncertainty assessments for general loss functions (like the Bayes posterior))

The E-Posterior

- Let $\{P_\theta : \theta \in \Theta\}$ represent a statistical model, and let $\{S_\theta : \theta \in \Theta\}$ be an e-collection
- i.e. each S_θ is E-variable for null hypothesis $\{P_\theta\}$
- E-Posterior relative to $\mathcal{S}$ defined simply as

$$\bar{P}(\,\theta \mid y\,) := S_\theta^{-1}(Y)$$

Prime example: each S_θ is Bayes factor,

$$S_\theta = \frac{p_W\,(Y)}{p_\theta\,(Y)} = \frac{\int p_\theta(Y)\, w(\theta) d\theta}{p_\theta(Y)} \quad \text{so that } \bar{P}(\theta|y) = \frac{w\,(\theta|y)}{w\,(\theta)}$$

In this special case a.k.a. Savage-Dickey ratio

The E-Posterior

Prime example: each S_θ is Bayes factor,

$$S_\theta = \frac{p_W\,(Y)}{p_\theta\,(Y)} = \frac{\int p_\theta(Y)\, w(\theta) d\theta}{p_\theta(Y)} \text{ so that } \bar{P}(\theta|y) = \frac{w\,(\theta|y)}{w\,(\theta)}$$

Example: normal location family $\mathrm{Y} \sim N(\theta, 1)$, $w(\theta)$ normal with mean 0, variance 1

$$\bar{P}(\theta|y) = \sqrt{n+1} \cdot e^{-\frac{n}{2}\left(\theta - \hat{\theta}\right)^2 + \frac{1}{2} \cdot \frac{n}{n+1} \hat{\theta}^2}$$

- luckiness term depending on how well data aligns with prior

risk confidence intervals

- Actions are intervals.
- Loss: $L_B(\theta, [\theta_L, \theta_R]) = B \cdot 1_{\theta \notin [\theta_L, \theta_R]}$

(linear importance weighting ; could also do non-linear, but not in this talk!)

risk confidence intervals

- Actions are intervals.
- Loss: $L_B(\theta, [\theta_L, \theta_R]) = B \cdot 1_{\theta \notin [\theta_L, \theta_R]}$
- Decision rule $\mathrm{CI}_B(Y) = \left[\hat{\theta}_{L,B}(Y), \hat{\theta}_{R,B}(Y)\right]$ is called a **risk-confidence interval** if for all $\theta \in \Theta$: $\mathbf{E}_{P_\theta}[L_B(\theta, \mathrm{CI}_B(Y)\ )] \leq \ell$

risk confidence intervals

- Actions are intervals.
- Loss: $L_B(\theta, [\theta_L, \theta_R]) = B \cdot 1_{\theta \notin [\theta_L, \theta_R]}$
- Decision rule $\mathrm{CI}_B(Y) = \left[\hat{\theta}_{L,B}(Y), \hat{\theta}_{R,B}(Y)\right]$ is called a **risk-confidence interval** if for all $\theta \in \Theta$:
  $\mathbf{E}_{P_\theta}[L_B(\theta, \mathrm{CI}_B(Y)\,)] \leq \ell$
- If $\ell = 1$ and B would be constant > 1 then this would simply be a $1 - 1/B$ standard confidence interval
- $B = 20 \Rightarrow 95\%\ \mathrm{CI}$

risk confidence intervals

- Actions are intervals.
- Loss: $L_B(\theta, [\theta_L, \theta_R]) = B \cdot 1_{\theta \notin [\theta_L, \theta_R]}$
- Decision rule $\mathrm{CI}_B(Y) = \left[\hat{\theta}_{L,B}(Y), \hat{\theta}_{R,B}(Y)\right]$ is called a **risk-confidence interval** if for all $\theta \in \Theta$:
  $\mathbf{E}_{P_\theta}[L_B(\theta, \mathrm{CI}_B(Y)\,)] \leq \ell$
- If $\ell = 1$ and B would be constant > 1 then this would simply be a $1 - 1/B$ standard confidence interval
- B indicates 'how bad it is to be wrong'.
- Again we allow it to depend on data!
- CI' is **better** than CI if it is "always" narrower

Post-Hoc Trouble

Goal: design best (narrowest) confidence interval st for all $\theta \in \Theta$: $\text{risk}(\theta) \coloneqq \mathbf{E}_{P_\theta}\left[L_B\left(\theta, \left[\hat{\theta}_{L,B}, \hat{\theta}_{R,B}\right]\right)\right] \leq 1$

- If B were constant we'd simply pick $1 - B^{-1}$ standard CI
- In normal location family, this is $\left[\hat{\theta} - A/\sqrt{n}, \hat{\theta} + A/\sqrt{n}\right]$ with e.g. if $B = 20$ then $A = 1.96$ (95% CI)
- If we treat confidence distributions or O'Bayes posteriors as subjective posteriors, then we **condition on B and would output the same thing**

…may again fail **dramatically** if B is determined post-hoc

Post-Hoc Trouble

We **expect** $\mathrm{risk}(\theta)$ of this procedure bounded by ℓ

But if in reality B can depend on data in unknown ways, actual $\mathrm{risk}(\theta)$ may be much larger, **even** ∞

(see paper)

Post-Hoc Fun

- Loss: $L_B(\theta, [\theta_L, \theta_R]) = B \cdot 1_{\theta \notin [\theta_L, \theta_R]}$

e-posterior approach: for any Y, pick $\mathrm{CI}_B(Y)$ such that

$$\max_\theta \bar{P}(\theta|Y) \cdot L_B(\theta, \mathrm{CI}_B(Y)) \leq \ell$$

(and then of course, the smallest such $\mathrm{CI}_B(Y)$)

$\mathrm{risk}(\theta)$ of this procedure is indeed bounded by ℓ, for all θ, as long as $\bar{P}(\,\theta \mid Y\,)$ is a valid e-posterior

The Capped E-Posterior

- We call the importance B **nontrivial**, if $B > \ell$

(if presented with trivial B, simply output empty interval!)

- For any nontrivial B, $\mathrm{CI}_B(Y) = [\hat{\theta}_{L,B}, \hat{\theta}_{R,B}]$ satisfies constraint

  $\max_{\theta} \bar{P}(\theta|Y) \cdot B \cdot 1_{\theta \notin \mathrm{CI}_B(Y)} \leq \ell$

  only if $\bar{P}(\theta \mid Y) \leq 1$ for all $\theta \notin [\hat{\theta}_{L,B}, \hat{\theta}_{R,B}]$

...but this means that we may replace the E-posterior by its capped version: $\bar{P}^{\mathrm{CAP}}(\theta \mid y) = \min\{1, \bar{P}(\theta|Y)\}$

Post-Hoc Fun

- Loss: $L_B(\theta, [\theta_L, \theta_R]) = B \cdot 1_{\theta \notin [\theta_L, \theta_R]}$

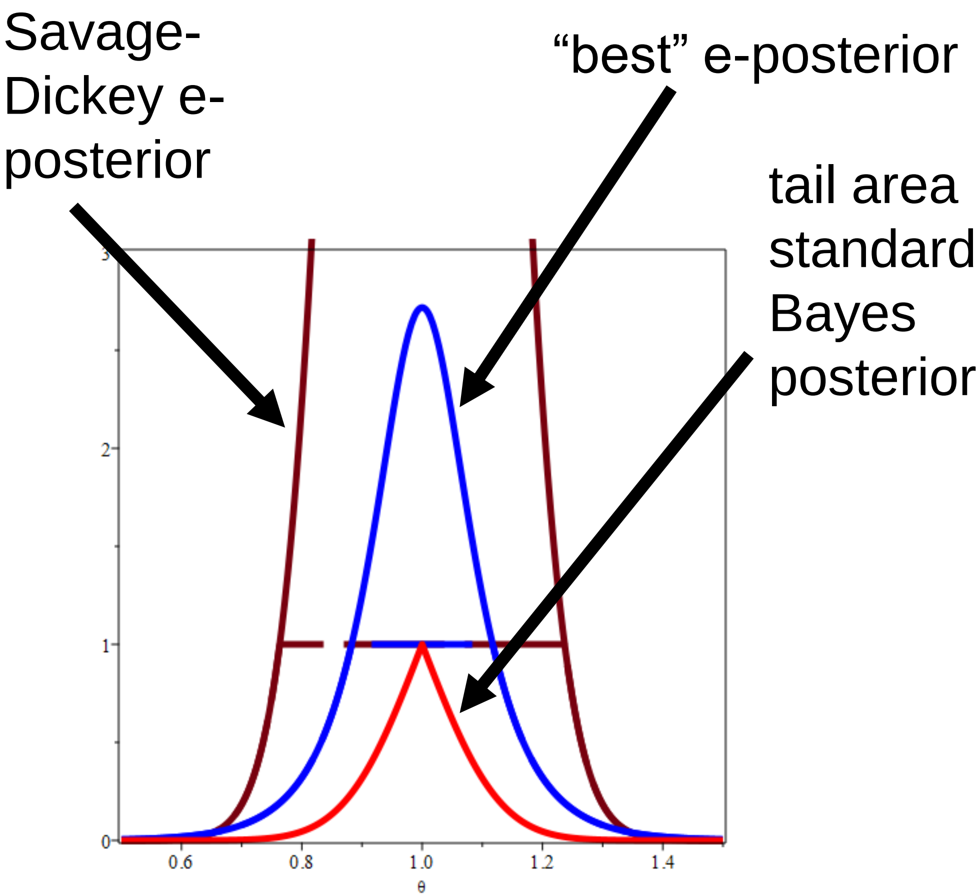
capped e-posterior approach: for any Y, pick $\mathrm{CI}_B(Y)$ s.t.

$$\max_{\theta} \overline{\boldsymbol{P}}^{\mathbf{CAP}}(\theta|Y) \cdot L_B(\theta, \mathrm{CI}_B(Y)) \leq \ell$$

(and then of course, the smallest such $\mathrm{CI}_B(Y)$)

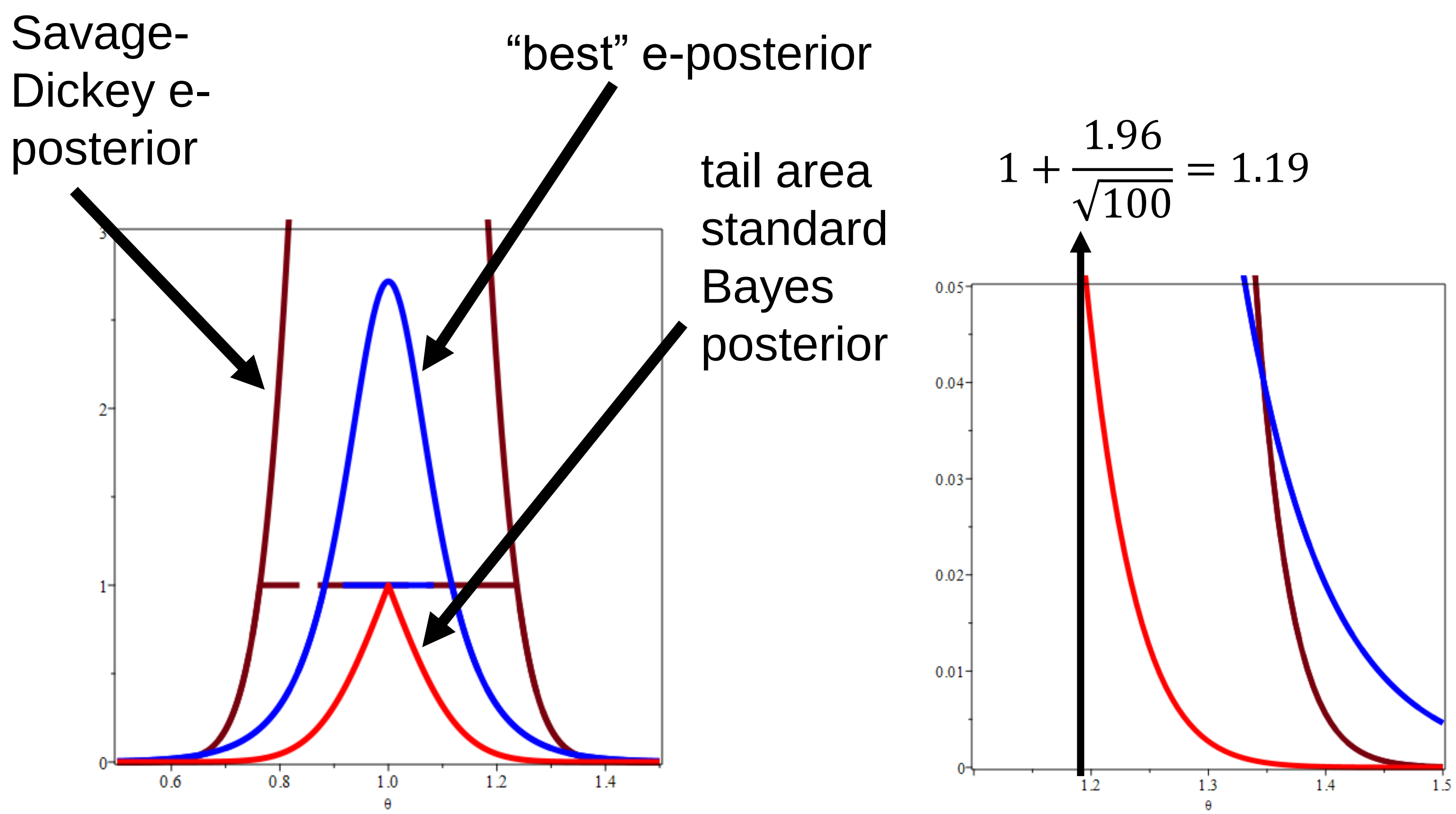
$\mathrm{risk}(\theta)$ of this procedure is indeed bounded by ℓ, for all θ, as long as $\bar{P}(\theta \mid Y)$ is a valid e-posterior

normal location, $\widehat{\theta} = 1, n = 100$



with cleverly chosen e-vars, the E-posterior 95% confidence is $\hat{\theta} \pm 2.72/\sqrt{n}$, i.e. a factor 1.4 wider than the standard NP CI/Bayesian credible interval. This is the best we can do.

normal location, $\hat{\theta} = 1, n = 100$



with cleverly chosen e-vars, the E-posterior 95% confidence is $\hat{\theta} \pm 2.72/\sqrt{n}$, i.e. a factor 1.4 wider than the standard NP CI/Bayesian credible interval. This is the best we can do.

The ‘Best’ E-Posterior

$$\bar{P}_{\text{Savage−Dickey}}(\theta|y) = \frac{p_\theta(y)}{\int p_\theta(y)\, d\, W(\theta)}$$

$$\bar{P}_{\text{clever}}(\, \theta \mid y\,) := \frac{p_\theta(y)}{\int p_\theta(y)\, d\boldsymbol{W}_{\leftsquigarrow\boldsymbol{\theta}}(\theta)}$$

normal location: $\bar{P}_{\text{best}}(\, \theta \mid y\,) := \frac{p_\theta(y)}{\frac{1}{2} p_{\theta^-}(y) + \frac{1}{2} p_{\theta^+}(y)}$

with θ^-, θ^+ such that $D(\theta || \theta^-) = D(\theta || \theta^+) = \frac{\log \frac{2}{\boldsymbol{\alpha}^*}}{n}$

Questions for ISIPTA

- What is relation between capped e-posterior and
  - Martin-Liu inferential models?

  (I think the capped posterior is one – with a substantially stronger notion of validity)
  - Possibility distributions/contour functions?

- If there is a relation, can it be used to derive an **e-posterior calculus!?**

Another Isipta Issue

- Let $\bar{P}(\ \theta \mid y\ )$ be Savage-Dickey e-posterior for prior W
- Bayes posterior risk assessment for general losses:

$$\mathbf{E}_{\theta\sim W|Y}\ [L_B(\theta,\delta_B(Y))] = \mathbf{E}_{\theta\sim W}[\ \bar{P}(\ \theta \mid Y\ ) \cdot L_B(\theta,\delta_B(Y))\ ]$$

- E-posterior risk assessment for general losses:

$$\max_{\theta} \bar{P}(\theta|Y) L_B\big(\theta,\delta_B(Y)\big) = \max_{W'} \mathbf{E}_{\theta\sim W'}\ [\bar{P}(\theta|Y) \cdot L_B\big(\theta,\delta_B(Y)\big)]$$

- **Do something inbetween, with sets of priors!?!?!?**

The Quasi-Conditional Paradigm

With E-posteriors, decision task can depend on data in arbitrary & unknown ways

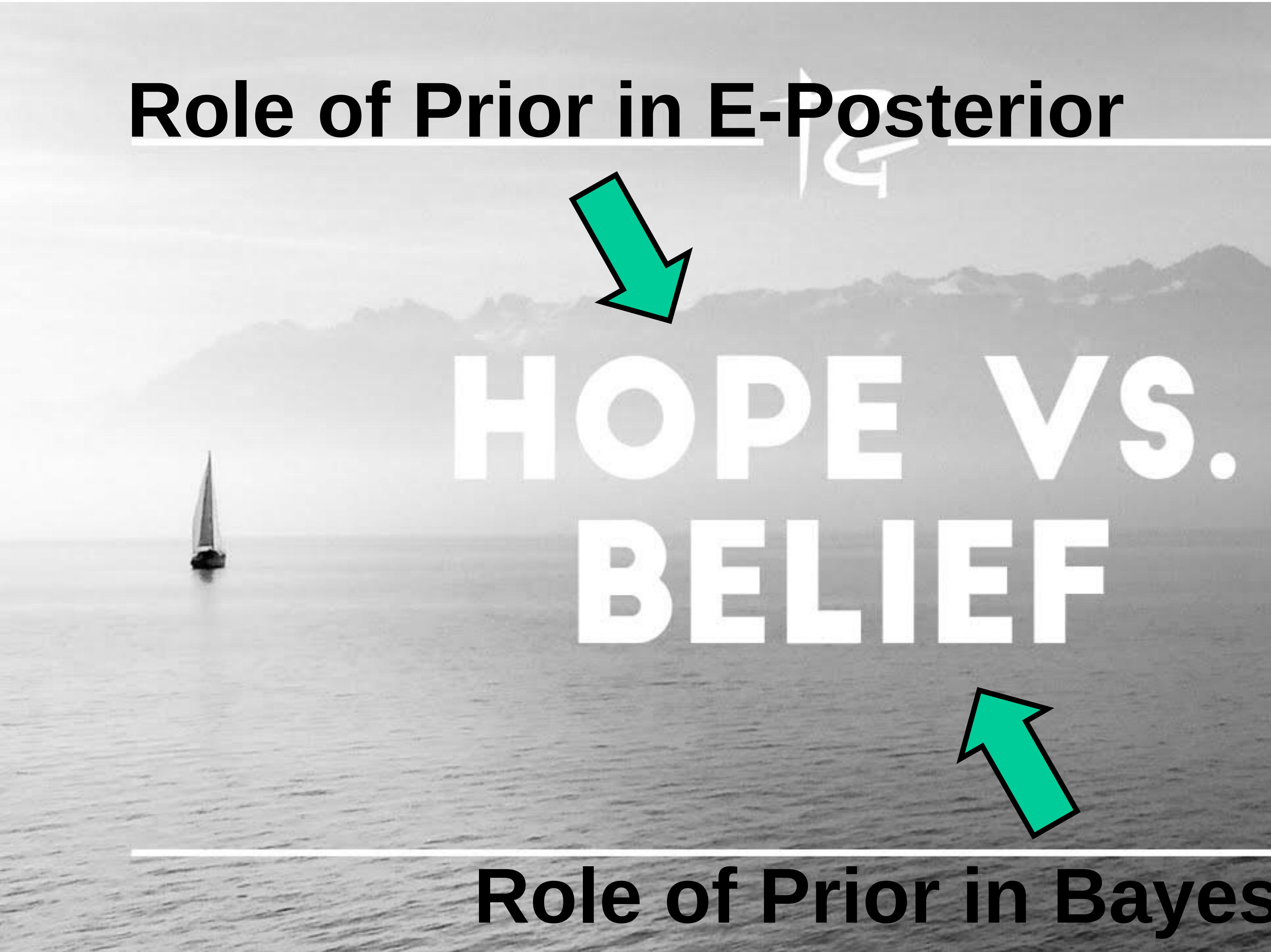
- can be chosen ‘conditional on data’, just like in Bayes

…but one evaluates procedures unconditionally, in expectation under the unknown true distribution

- Fully frequentist
- **If prior chosen badly, e-posterior becomes wide rather than wrong**

We call this a quasi-conditional approach. Inspired by, but quite different from, conditionalist frequentist inference as developed by (mainly) Kiefer and Berger

Role of Prior in E-Posterior



Role of Prior in Bayes

Read more about this *Quasi-Conditional* Paradigm?

- G. ***Beyond Neyman-Pearson*** http://arxiv.org/abs/2205.00901
- G., De Heide, Koolen. ***Safe Testing,*** Arxiv ‘19/’21

E-Variables and the NP Lemma

Suppose $H_0 = \{P_0\}, H_1 = \{P_1\}$ are simple.

The test based on LR E-variable $S^{\text{lr}} = p_1(Y)/p_0(Y)$ looks a bit like, but is *not* the maximum power NP test.

Safe Test: reject if $S(Y) \geq 1/\alpha$ **more conservative**

NP: reject if $S(Y) \geq B$ with B s.t. $P_0(S(Y) \geq B) = \alpha$

E-Variables and the NP Lemma

The test based on LR E-variable $S^{\mathrm{lr}} = p_1(Y)/p_0(Y)$ looks a bit like, but is *not* the maximum power NP test.

Yet:

$$S^{\mathsf{np}(\alpha)}(Y) := \begin{cases} 0 & \text{if } \mathsf{p}(Y) > \alpha \\ \alpha^{-1} & \text{otherwise} \end{cases}$$

...is also an E-variable: $E_{P_0}\left[S^{\mathrm{np}(\alpha)}\right] \leq \alpha \cdot \alpha^{-1} = 1$ and its corresponding test **coincides** with the NP test

E-Variables and the NP Lemma

The test based on LR E-variable $S^{\mathrm{lr}} = p_1(Y)/p_0(Y)$ looks a bit like, but is *not* the maximum power NP test.

Yet:

$$S^{\mathsf{np}(\alpha)}(Y) := \begin{cases} 0 & \text{if } \mathsf{p}(Y) > \alpha \\ \alpha^{-1} & \text{otherwise} \end{cases}$$

...is also an E-variable: $E_{P_0}\left[S^{\mathrm{np}(\alpha)}\right] \leq \alpha \cdot \alpha^{-1} = 1$ and its corresponding test **coincides** with the NP test

...coincidentally this shows that nothing lost by using E instead of p: we can mimick all of NP testing theory