E is the new P

Rianne de Heide

Vrije Universiteit Amsterdam
joint work with Peter Grünwald, Wouter Koolen, Muriel Pérez-Ortiz,
Tyron Lardy, Allard Henrdriksen

November 2, 2022

Background (math)

Education

- BSc Math (Groningen), MSc Math (Leiden)
- BMus (Groningen/Hamburg), MMus (The Hague)

Background (math)

Education

- BSc Math (Groningen), MSc Math (Leiden)
- BMus (Groningen/Hamburg), MMus (The Hague)

Work

- PhD at Centrum Wiskunde & Informatica and Leiden University, supervisors: Peter Grünwald and Wouter Koolen, promotores: Peter, Wouter and Jacqueline Meulman
- Postdoc at Otto von Guericke University Magdeburg, supervisor: Alexandra Carpentier
- Postdoc at INRIA Lille, supervisor: Emilie Kaufmann
- Rubicon grant - INRIA Lille

Background (math)

Education

- BSc Math (Groningen), MSc Math (Leiden)
- BMus (Groningen/Hamburg), MMus (The Hague)

Work

- PhD at Centrum Wiskunde & Informatica and Leiden University, supervisors: Peter Grünwald and Wouter Koolen, promotores: Peter, Wouter and Jacqueline Meulman
- Postdoc at Otto von Guericke University Magdeburg, supervisor: Alexandra Carpentier
- Postdoc at INRIA Lille, supervisor: Emilie Kaufmann
- Rubicon grant - INRIA Lille
- March 2022 - ... VU :)

Background (non-math)

Non-mathematical topics I love (to talk about during coffee)

- Classical music

Background (non-math)

Non-mathematical topics I love (to talk about during coffee)

- Classical music
- Running

Background (non-math)

Non-mathematical topics I love (to talk about during coffee)

- Classical music
- Running
- Learning languages

Background (non-math)

Non-mathematical topics I love (to talk about during coffee)

- Classical music
- Running
- Learning languages
- Chess

Work

Hypothesis testing (Stats)

- A new theory of hypothesis testing (this talk)
- Group invariance in hypothesis testing
- Optional stopping

Other topics I work on:

- Inductive logic (philosophy of science)
- Bayesian inference under model misspecification (learning theory — Stats/ML)
- Best-arm-identification (bandits — ML)
- Mathematics of explainable AI (XAI — ML)

Hypothesis testing with E-values

- A new theory of hypothesis testing
- Main notion: E-variable / E-value
- Upshots: combining evidence; interpretation; flexibility
- Main mathematical contributions: existence of non-trivial E-values for composite $\mathcal{H}_0$ and design criterion for *optimal* (GRO(W)) E-values (*Safe Testing* - Grünwald, De Heide, Koolen); group-invariance in hypothesis testing (*Optional stopping with Bayes Factors* - Hendriksen, De Heide, Grünwald; *E-Statistics, Group Invariance and Anytime Valid Testing* - Pérez-Ortiz, Lardy, De Heide, Grünwald; and *Why optional stopping can be a problem for Bayesians* - De Heide, Grünwald).

Menu

- Why do we need a new theory for hypothesis testing?
- E-values
  - A lady tasting coffee
  - Highlights 1: interpretations
  - Highlights 2: RIPr and JIPr
  - Highlights 3: Combining experiments
  - Highlights 4: T-test simulations

Menu

- Why do we need a new theory for hypothesis testing?
- E-values
  - A lady tasting coffee
  - Highlights 1: interpretations
  - Highlights 2: RIPr and JIPr
  - Highlights 3: Combining experiments
  - Highlights 4: T-test simulations

Why do we need a new theory for hypothesis testing?

Reproducibility crisis in social and medical science

Why do we need a new theory for hypothesis testing?

Reproducibility crisis in social and medical science

- Medicine: J. Ioannidis, Why most published research findings are false , PLoS Medicine 2(8) (2005).
- Social Science: 270 authors, Estimating the reproducibility of psychological science , Science 349 (6251), 2015.

Why do we need a new theory for hypothesis testing?

Reproducibility crisis in social and medical science

Causes:

Why do we need a new theory for hypothesis testing?

Reproducibility crisis in social and medical science

Causes:

- Publication bias

Why do we need a new theory for hypothesis testing?

Reproducibility crisis in social and medical science

Causes:

- Publication bias
- Fraud

Why do we need a new theory for hypothesis testing?

Reproducibility crisis in social and medical science

Causes:

- Publication bias
- Fraud
- Lab environment vs. natural environmnent

Why do we need a new theory for hypothesis testing?

Reproducibility crisis in social and medical science

Causes:

- Publication bias
- Fraud
- Lab environment vs. natural environmnent
- use of P-values

Why do we need a new theory for hypothesis testing?

We wish to test a *null hypothesis* $\mathcal{H}_0$ in contrast with an *alternative hypothesis* $\mathcal{H}_1$.

Definition

Fix some $\alpha \in (0,1)$. A P-value is a function mapping data $X^n = X_1, \ldots, X_n$ to $[0,1]$, such that for all $P \in \mathcal{H}_0$

$$P\left(\mathrm{P}(X^n) \leq \alpha\right) \leq \alpha.$$

Why do we need a new theory for hypothesis testing?

We wish to test a *null hypothesis* $\mathcal{H}_0$ in constrast with an *alternative hypothesis* $\mathcal{H}_1$.

Type-I guarantee α:

$$P(\text{reject } \mathcal{H}_0) \leq \alpha.$$

Why do we need a new theory for hypothesis testing?

Problems with P-values

- Limited applicability: unknown probabilities

Consider two weather forecasters A and B. On sunny days, $P_A(\text{RAIN}) \geq P_B(\text{RAIN})$. Is B better than A?

P-values rely on *counterfactuals*. See also:

A.P. Dawid, Present position and potential developments: Some personal views, statistical theory, the prequential approach, Journal of the Royal Statistical Society, Series A 147(2) (1984), 278–292.

P. Grünwald, The Minimum Description Length Principle, MIT Press, Cambridge, MA, 2007.

Why do we need a new theory for hypothesis testing?

Problems with P-values

- Limited applicability: unknown probabilities
- Limited applicability: unknown stopping rules

Many practitioners don't know that *optional stopping* is forbidden with P-values, so they do it.

Many practitioners DO know that *optional stopping* is forbidden with P-values, and they still do it!

55% of psychologists admits to doing it — John et. al. (2012)

Why do we need a new theory for hypothesis testing?

Problems with P-values

- Limited applicability: unknown probabilities
- Limited applicability: unknown stopping rules
- Interpretational problems: combining evidence from different experiments

Hospitals A and B perform similar trials, and they report P-values P_A and P_B. How to combine the evidence?

Neyman/Pearson: *significance tests*. Only report *reject* or *accept*.
Fisher: P-values as measure of evidence, not for testing.

Why do we need a new theory for hypothesis testing?

Problems with P-values

- Limited applicability: unknown probabilities
- Limited applicability: unknown stopping rules
- Interpretational problems: combining evidence from different experiments
- Interpretational problems: misunderstanding (hence misuse) of P-values

What do Doctors know about statistics?

A controlled trial of a new treatment led to the conclusion that it is significantly better than placebo: $\mathrm{P} < 0.05$. Which of the following statements do you prefer?

Go to menti.com and use the code 3419 1778.

1. It has been proved that the treatment is better than placebo.
2. If the treatment is not effective, there is less than a 5 per cent chance of obtaining such results.
3. The observed effect of the treatment is so large that there is less than a 5 per cent chance that the treatment is no better than placebo.
4. I do not really know what a p-value is and do not want to guess.

What do Doctors know about statistics?

A controlled trial of a new treatment led to the conclusion that it is significantly better than placebo: $\mathrm{P} < 0.05$. Which of the following statements do you prefer?

1. It has been proved that the treatment is better than placebo. 20%
2. If the treatment is not effective, there is less than a 5 per cent chance of obtaining such results. 13%
3. The observed effect of the treatment is so large that there is less than a 5 per cent chance that the treatment is no better than placebo. 51%
4. I do not really know what a p-value is and do not want to guess. 16%

Menu

- Why do we need a new theory for hypothesis testing?
- E-values
  - A lady tasting coffee
  - Highlights 1: interpretations
  - Highlights 2: RIPr and JIPr
  - Highlights 3: Combining experiments
  - Highlights 4: T-test simulations

Testing by betting

Hypothesis testing with e-values and martingales

Rianne de Heide

A lady tasting tea

A lady tasting tea

Null hypothesis: the lady has no ability to distinguish the teas.

A lady tasting tea

Null hypothesis: the lady has no ability to distinguish the teas.

$$\binom{8}{4} = \frac{8!}{4!(8-4)!} = 70$$

Safe Testing

e-values in stead of p-values

- intuitive interpretation: betting
- sequential testing possible

A guy tasting coffee…


No!!!

Aaditya Ramdas (CMU)

Leila Wehbe (CMU)

Safe Testing - a lady tasting coffee

Safe Testing - a lady tasting coffee

Safe Testing - a lady tasting coffee

M C

Safe Testing - a lady tasting coffee

$B_1 = -1$

M C

Safe Testing - a lady tasting coffee

$B_1 = -1$

M C

Safe Testing - a lady tasting coffee

$B_1 = -1$

M C

Safe Testing - a lady tasting coffee

$B_1 = -1$

M C

M C

Safe Testing - a lady tasting coffee

$B_1 = -1$

M C

$B_2 = +1$

M C

A lady tasting coffee: guessing

$B_1 = -1$

M C

$B_2 = +1$

M C

A lady tasting coffee: guessing

$B_1 = -1$  

M C

$B_2 = +1$  

M C

$$S_t = \sum_{s=1}^{t} B_s$$

A lady tasting coffee: guessing

$B_1 = -1$

M C

$B_2 = +1$

M C

$$S_t = \sum_{s=1}^{t} B_s$$

$\mathscr{H}_0$: There is no difference between MC and CM.

A lady tasting coffee: guessing

$B_1 = -1$

M C

$B_2 = +1$

M C

$$S_t = \sum_{s=1}^{t} B_s ,$$

$\mathcal{H}_0$: There is no difference between MC and CM.

Under $\mathcal{H}_0$, $(S_t)_{t\in\mathbb{N}}$ is a martingale: $\mathbb{E}[S_t \,|\, S_1, \ldots, S_{t-1}] = S_{t-1}$.

A lady tasting coffee: guessing

$B_1 = -1$

M C

$B_2 = +1$

M C

$S_t = \sum_{s=1}^{t} B_s$,

$\mathscr{H}_0$: There is no difference between MC and CM.

Under $\mathscr{H}_0$, $(S_t)_{t\in\mathbb{N}}$ is a martingale: $\mathbb{E}[S_t \mid S_1, \ldots, S_{t-1}] = S_{t-1}$.

Reject $\mathscr{H}_0$ if $|S_n| \geq \sqrt{\frac{1}{n}\left(1+\frac{1}{n}\right)\log\left(\frac{n+1}{\alpha^2}\right)}$

A lady tasting coffee: betting

$$L_0 = 1$$

A lady tasting coffee: betting

$L_0 = 1$

A lady tasting coffee: betting

$L_0 = 1$

$\lambda_1 = 0.2$ (on heads)

A lady tasting coffee: betting

$L_0 = 1$ 

$B_1 = -1$ 

$\lambda_1 = 0.2$ (on heads)

A lady tasting coffee: betting

$L_0 = 1$

$B_1 = -1$

$\lambda_1 = 0.2$ (on heads)

$$L_1 = L_0 \cdot (1 + \lambda_1 B_1) = 0.8$$

A lady tasting coffee: betting

$L_0 = 1$

$B_1 = -1$

$\lambda_1 = 0.2$ (on heads)

$$L_1 = L_0 \cdot (1 + \lambda_1 B_1) = 0.8$$

$B_2 = +1$

$\lambda_2 = 0.4$ (on heads)

$$L_2 = L_1 \cdot (1 + \lambda_2 B_2) = 1.12$$

A lady tasting coffee: betting

$L_0 = 1$

$B_1 = -1$

$\lambda_1 = 0.2$ (on heads)

$L_1 = L_0 \cdot (1 + \lambda_1 B_1) = 0.8$

$B_2 = +1$

$\lambda_2 = 0.4$ (on heads)

$L_2 = L_1 \cdot (1 + \lambda_2 B_2) = 1.12$

$$L_t := \prod_{s=1}^{t} (1 + \lambda_s B_s)$$

A lady tasting coffee: betting

$L_0 = 1$

$B_1 = -1$

$\lambda_1 = 0.2$ (on heads)

$$L_1 = L_0 \cdot (1 + \lambda_1 B_1) = 0.8$$

$B_2 = +1$

$\lambda_2 = 0.4$ (on heads)

$$L_2 = L_1 \cdot (1 + \lambda_2 B_2) = 1.12$$

$L_t := \prod_{s=1}^{t} (1 + \lambda_s B_s)$; Under $\mathscr{H}_0$, $(L_t)_{t \in \mathbb{N}}$ is a non-negative martingale.

A lady tasting coffee: betting

$$L_t := \prod_{s=1}^{t} (1 + \lambda_s B_s) \; ;$$ Under $\mathscr{H}_0$, $(L_t)_{t \in \mathbb{N}}$ is a non-negative martingale.

At any stopping time τ, we have $\mathbb{E}_{\mathscr{H}_0}[L_\tau] = 1$ (optional stopping theorem).

A lady tasting coffee: betting

$L_t := \prod_{s=1}^{t} (1 + \lambda_s B_s)$; Under $\mathcal{H}_0$, $(L_t)_{t \in \mathbb{N}}$ is a non-negative martingale.

At any stopping time τ, we have $\mathbb{E}_{\mathcal{H}_0}[L_\tau] = 1$ (optional stopping theorem).

Ville's inequality:

$$\mathbb{P}(\exists t \in \mathbb{N} : L_t > 1/\alpha) \leq \alpha$$

p-value equivalent:

$$\mathbb{P}(\exists t \in \mathbb{N} : p_t > 1/\alpha) = 1$$

A lady tasting coffee: betting

$L_t := \prod_{s=1}^{t} (1 + \lambda_s B_s)$; Under $\mathscr{H}_0$, $(L_t)_{t \in \mathbb{N}}$ is a non-negative martingale.

At any stopping time τ, we have $\mathbb{E}_{\mathscr{H}_0}[L_\tau] = 1$ (optional stopping theorem).

Ville's inequality:

$$\mathbb{P}(\exists t \in \mathbb{N} : L_t > 1/\alpha) \leq \alpha$$

p-value equivalent:

$$\mathbb{P}(\exists t \in \mathbb{N} : p_t > 1/\alpha) = 1$$

L_t is called an e-value

L_t measures evidence against $\mathscr{H}_0$

Safe Testing: e-values

- e-value: non-negative random variable E satisfying

$$\text{for all } P \in \mathscr{H}_0 : \quad \mathbb{E}_P[E] \leq 1.$$

Safe Testing: e-values

- e-value: non-negative random variable E satisfying

$$\text{for all } P \in \mathcal{H}_0 : \quad \mathbb{E}_P[E] \leq 1.$$

- We can define hypothesis tests based on e-values.

Safe Testing: e-values

- e-value: non-negative random variable E satisfying

$$\text{for all } P \in \mathscr{H}_0 : \quad \mathbb{E}_P[E] \leq 1.$$

- But what is a good e-value?

Safe Testing: e-values

- e-value: non-negative random variable E satisfying

$$\text{for all } P \in \mathscr{H}_0 : \quad \mathbb{E}_P[E] \leq 1.$$

- But what is a good e-value?

- GROW: Growth-Rate Optimal in Worst case: the e-value E^* that achieves

$$\max_{E:E \text{ is an e-value}} \min_{P \in \mathscr{H}_1} \mathbb{E}_P[\log E]$$

Safe Testing with e-values: Main Theorem

- The GROW e-value $E^*_{W_1}$ exists (for composite $\mathscr{H}_0$), and satisfies $\mathbb{E}_{Z \sim P_{W_1}}[\log E^*_{W_1}] = \sup_{E \in \mathscr{E}} \mathbb{E}_{Z \sim P_{W_1}}[\log E] = \inf_{W_0 \in \mathscr{W}_0} D(P_{W_1} \| P_{W_0})$
- if the inf is achieved by some W_0°, the GROW e-value takes a simple form: $E^*_{W_1} = p_{W_1}(Z)/p_{W_0^\circ}(Z)$
- GROW e-values $E^*_{\mathscr{W}_1} = p_{W_1^*}(Z)/p_{W_0^*}(Z)$ can be found by a double KL-minimization problem $\min_{W_1 \in \mathscr{W}_1} \min_{W_0 \in \mathscr{W}_0} D(P_{W_1} \| P_{W_0})$ and they satisfy

$$\inf_{W \in \mathscr{W}_1} \mathbb{E}_{Z \sim P_W}[\log E^*_{\mathscr{W}_1}] = \sup_{E \in \mathscr{E}} \inf_{W \in \mathscr{W}_1} \mathbb{E}_{Z \sim P_W}[\log E] = D(P_{W_1^*} \| P_{W_0^*})$$

Menu

- Why do we need a new theory for hypothesis testing?
- E-values
  - A lady tasting coffee
  - Highlights 1: interpretations
  - Highlights 2: RIPr and JIPr
  - Highlights 3: Combining experiments
  - Highlights 4: T-test simulations

Highlights: 1. Interpretations

1. **Kelly Gambling**
2. P**-value, Type I error probability**
3. **Bayes Factors**

$$\mathrm{BF} := \frac{p_{W_1}(Z)}{p_{W_0}(Z)} \tag{1}$$

Simple $\mathcal{H}_0 = \{P_0\}$: Bayes factor is also an E-test statistic, since

$$\mathbf{E}_P[\mathbf{B}] := \int p_0(z) \cdot \frac{p_{W_1}(z)}{p_0(z)} dz = 1. \tag{2}$$

(and e-values for more complicated problems can also be interpreted as Bayes factors (but not always vice versa), see the main theorem)

Menu

- Why do we need a new theory for hypothesis testing?
- E-values
  - A lady tasting coffee
  - Highlights 1: interpretations
  - Highlights 2: RIPr and JIPr
  - Highlights 3: Combining experiments
  - Highlights 4: T-test simulations

Highlights 2. The JIPr - Main Theorem (1)

1. The GROW E-value $E^*_{W_1}$ exists, and satisfies

$$\mathbf{E}_{Z \sim P_{W_1}}[\log E^*_{W_1}] = \sup_{E \in \mathcal{E}(\Theta_0)} \mathbf{E}_{Z \sim P_{W_1}}[\log E] = \inf_{W_0 \in \mathcal{W}(\Theta_0)} D(P_{W_1} \| P_{W_0})$$

2. Suppose that the inf is achieved by some W_0°, i.e. $\inf_{W_0 \in \mathcal{W}(\Theta_0)} D(P_{W_1} \| P_{W_0}) = D(P_{W_1} \| P_{W_0^\circ})$. Then the minimum is achieved uniquely by this W_0° and the GROW E-value takes a simple form: $E^*_{W_1} = p_{W_1}(Z)/p_{W_0^\circ}(Z)$.

Highlights 2. The JIPr -Main Theorem (2)

3. Now let $\Theta'_1 \subset \Theta_1$ and let $\mathcal{W}'_1$ be a convex subset of $\mathcal{W}(\Theta'_1)$ such that for all $\theta \in \Theta_0$, all $W_1 \in \mathcal{W}'_1$, P_θ is absolutely continuous relative to P_{W_1}. Suppose that $\min_{W_1 \in \mathcal{W}'_1} \min_{W_0 \in \mathcal{W}_0} D(P_{W_1} \| P_{W_0}) = D(P^*_{W_1} \| P^*_{W_0}) < \infty$ is achieved by some (W^*_1, W^*_0) such that $D(P_{W_1} \| P_{W^*_0}) < \infty$ for all $W_1 \in \mathcal{W}'_1$. Then the minimum is achieved uniquely by (W^*_1, W^*_0), and the GROW E-value $E^*_{\mathcal{W}'_1}$ relative to $\mathcal{W}'_1$ exists, is essentially unique, and is given by

$$E^*_{\mathcal{W}'_1} = \frac{p_{W^*_1}(Z)}{p_{W^*_0}(Z)}, \tag{3}$$

and it satisfies

$$\inf_{W \in \mathcal{W}'_1} \mathbf{E}_{Z \sim P_W}[\log E^*_{\mathcal{W}'_1}] = \sup_{E \in \mathcal{E}(\Theta_0)} \inf_{W \in \mathcal{W}'_1} \mathbf{E}_{Z \sim P_W}[\log E] = D(P_{W^*_1} \| P_{W^*_0}). \tag{4}$$

If $\mathcal{W}'_1 = \mathcal{W}(\Theta'_1)$, then by linearity of expectation we further have $E^*_{\mathcal{W}'_1} = E^*_{\Theta'_1}$.

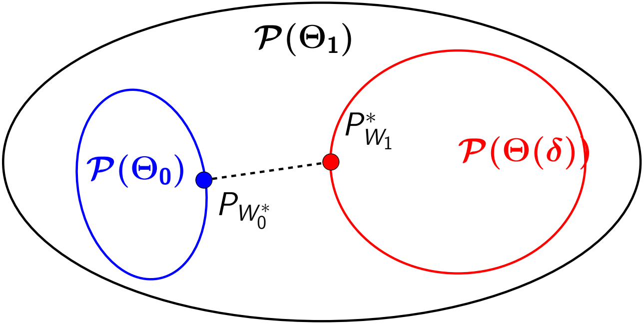
Highlights 2. The JIPr - The RIPr and the JIPr

- $\inf_{W_0 \in \mathcal{W}(\Theta_0)} D(P_{W_1} \| P_{W_0}) = D(P_{W_1} \| P_{W_0^\circ})$
- We call P_{W° the *Reverse Information Projection (RIPr)* of P_{W_1} on $\{P_W : W \in \mathcal{W}(\Theta_0)\}$.

Highlights 2. The JIPr - The RIPr and the JIPr

- $\min_{W_1 \in \mathcal{W}_1'} \min_{W_0 \in \mathcal{W}_0} D(P_{W_1} \| P_{W_0}) = D(P^*_{W_1} \| P^*_{W_0}) < \infty$
- We call $(P_{W_1^*}, P_{W_0^*})$ the *Joint Information Projection (JIPr)* of $\{P_W : W \in \mathcal{W}_1'\}$ and $\{P_W : W \in \mathcal{W}(\Theta_0)\}$ onto each other.

Highlights: 2. The JIPr



Menu

- Why do we need a new theory for hypothesis testing?
- E-values
  - A lady tasting coffee
  - Highlights 1: interpretations
  - Highlights 2: RIPr and JIPr
  - Highlights 3: Combining experiments
  - Highlights 4: T-test simulations

Highlights 3.: Optional Continuation Proposition

Suppose that P satisfies the assumptions. Let $E_{(0)} := 1$ and let, for $k = 1, \ldots, k_{\max}$, $E_{(k)} = e_k(Z_{(k)})$ be a function of $Z_{(k)}$ that is an E-value, i.e. $\mathbf{E}_{Z \sim P}[E_{(k)}] \leq 1$. Let $E^{(K)} := \prod_{k=0}^{K} E_{(k)}$, and let $K_{\text{STOP}} := K - 1$ where $K \geq 1$ is the smallest number for which $B_{(K)} = \text{STOP}$. Then

1. For all $k \geq 1$, $E^{(k)}$ is an E-value.
2. $E^{(K_{\text{STOP}})}$ is an E-value.

Corollary: $P_0 \in \mathcal{H}_0$, for every $0 \leq \alpha \leq 1$,

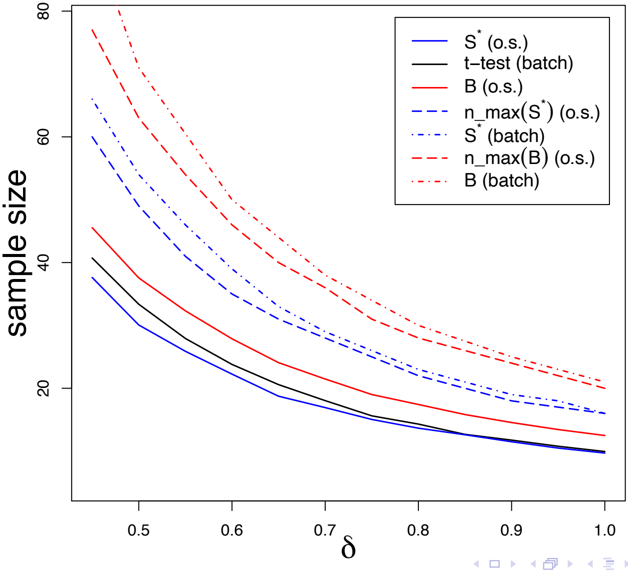
$$P_0(\mathrm{T}_\alpha(E^{(K_{\text{STOP}})}) = \text{REJECT}_0) \; (= P_0(E^{(K_{\text{STOP}})} \geq \alpha^{-1}) \;) \; \leq \alpha,$$

i.e. Type *I*-error guarantees are preserved under optional continuation, even for the most aggressive continuation rule which continues until the first K is reached such that either $\prod_{k=1}^{K} E_{(k)} \geq \alpha^{-1}$ or $K = k_{\max}$.

Menu

- Why do we need a new theory for hypothesis testing?
- E-values
  - A lady tasting coffee
  - Highlights 1: interpretations
  - Highlights 2: RIPr and JIPr
  - Highlights 3: Combining experiments
  - Highlights 4: T-test simulations

Highlights 4: The GRO(W) in practice: the *t*-test (1)



Highlights 4: The GRO(W) in practice: the t-test (2)

- Our default GRO(W) t-test E-value preserves Type I error probabilities under optional stopping,
- it needs more data than the classical t-test in the worst-case, but
- but not more on average under $\mathcal{H}_1$!

Papers

- Safe Testing - P.D. Grünwald, R. de Heide, W.M. Koolen (arXiv 1906.07801). Forthcoming in JRSS-B.
- Why optional stopping can be a problem for Bayesians - R. de Heide, P.D. Grünwald (Psychonomic Bulletin & Review 28(3):795-812, 2021)
- Optional stopping with Bayes factors - A. Hendriksen, R. de Heide, P.D. Grünwald (Bayesian Analysis, 16(3):961–989, 2021)
- E-statistics, group invariance and any time valid testing - M.F. Pérez-Ortiz, T. Lardy, R. de Heide, P.D. Grünwald (arXiv 2208.07610, submitted)

Time for questions!

References

- P.D. Grünwald, R. de Heide, W.M. Koolen - Safe Testing (2019)
- R. de Heide - Bayesian learning: Challenges, Limitations and Pragmatics (2021)
- J. Ioannidis - Why Most Published Research Findings Are False (2005)
- L.K. John, G. Loewenstein, D. Prelec - Measuring the prevalence of questionable research practices with incentives for truth telling (2012)
- A. Ramdas, L. Wehbe - The lady keeps tasting coffee (preprint)
- A. Ramdas - Lecture: http://stat.cmu.edu/~aramdas/betting/Feb11-class.pdf
- H.R. Wulff, B. Andersen, P. Brandenhoff, F. Guttler - What do doctors know about statistics? (1987)

about statistics? (1987)