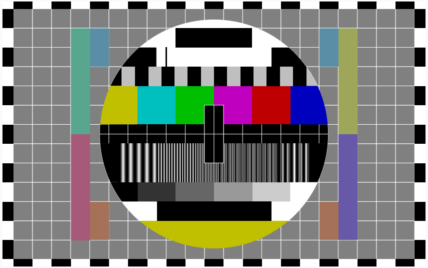
The t-test is a supermartingale after all

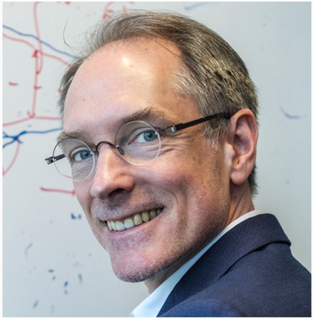
Wouter M. Koolen

E-Readers, April 4, 2025

CWI and University of Twente



Warm Thanks



Peter Grünwald

Menu

1. Introduction
2. Setup
3. Posing the Problem
4. Resolution
5. Linear Regression

Introduction

Warm up: point vs point

We want to do sequential hypothesis testing.

Warm up: point vs point

We want to do sequential hypothesis testing.

As an example, consider i.i.d. $X_1, X_2, \dots$ from either P_δ or P_0.

Warm up: point vs point

We want to do sequential hypothesis testing.

As an example, consider i.i.d. $X_1, X_2, \ldots$ from either P_δ or P_0.

A **sequential test** at confidence $\alpha \in (0,1)$ is a stopping time τ such that

- Safety: $P_0\{\tau < \infty\} \leq \alpha$.
- Power: $\mathbb{E}_{P_\delta}[\tau]$ is small.

Warm up: point vs point

We want to do sequential hypothesis testing.

As an example, consider i.i.d. $X_1, X_2, \ldots$ from either P_δ or P_0.

A **sequential test** at confidence $\alpha \in (0,1)$ is a stopping time τ such that

- Safety: $P_0\{\tau < \infty\} \leq \alpha$.
- Power: $\mathbb{E}_{P_\delta}[\tau]$ is small.

Lower bound: for any sequential test τ, sample complexity is at least $\mathbb{E}_{P_\delta}[\tau] \geq \frac{\ln \frac{1}{\alpha}}{\mathrm{KL}(P_\delta \| P_0)}$

Warm up: point vs point

We want to do sequential hypothesis testing.

As an example, consider i.i.d. $X_1, X_2, \ldots$ from either P_δ or P_0.

A **sequential test** at confidence $\alpha \in (0,1)$ is a stopping time τ such that

- Safety: $P_0 \{\tau < \infty\} \le \alpha$.
- Power: $\mathbb{E}_{P_\delta}[\tau]$ is small.

Lower bound: for any sequential test τ, sample complexity is at least $\mathbb{E}_{P_\delta}[\tau] \ge \frac{\ln \frac{1}{\alpha}}{\mathrm{KL}(P_\delta \| P_0)}$

A great idea: consider the likelihood ratio-based sequential test

$$M_n := \frac{P_\delta(X^n)}{P_0(X^n)} \qquad \text{and take} \qquad \tau := \inf\left\{ n \ge 0 \,\middle|\, M_n \ge \frac{1}{\alpha} \right\}$$

Warm up: point vs point

We want to do sequential hypothesis testing.

As an example, consider i.i.d. $X_1, X_2, \dots$ from either P_δ or P_0.

A **sequential test** at confidence $\alpha \in (0,1)$ is a stopping time τ such that

- Safety: $P_0 \{\tau < \infty\} \leq \alpha$.
- Power: $\mathbb{E}_{P_\delta}[\tau]$ is small.

Lower bound: for any sequential test τ, sample complexity is at least $\mathbb{E}_{P_\delta}[\tau] \geq \frac{\ln \frac{1}{\alpha}}{\mathrm{KL}(P_\delta \| P_0)}$

A great idea: consider the likelihood ratio-based sequential test

$$M_n := \frac{P_\delta(X^n)}{P_0(X^n)} \qquad \text{and take} \qquad \tau := \inf \left\{ n \geq 0 \middle| M_n \geq \frac{1}{\alpha} \right\}$$

As M_n is a P_0 test martingale, Ville's inequality gives us safety: $P_0 \left\{ \exists n : M_n \geq \frac{1}{\alpha} \right\} \leq \alpha$.

Wald's equation gives us $\mathbb{E}_{P_\delta}[\tau] \approx \frac{\ln \frac{1}{\alpha}}{\mathrm{KL}(P_\delta \| P_0)}$

Composite case

Now consider rejecting a composite null $\mathcal{H}_0$ with the help of a **composite alternative** $\mathcal{H}_1$.

Composite case

Now consider rejecting a composite null $\mathcal{H}_0$ with the help of a **composite alternative** $\mathcal{H}_1$.

Idea: find a measure of evidence that behaves like a likelihood ratio.

Composite case

Now consider rejecting a composite null $\mathcal{H}_0$ with the help of a **composite alternative** $\mathcal{H}_1$.

Idea: find a measure of evidence that behaves like a likelihood ratio.

Approach	M_n		

Composite case

Now consider rejecting a composite null $\mathcal{H}_0$ with the help of a **composite alternative** $\mathcal{H}_1$.

Idea: find a measure of evidence that behaves like a likelihood ratio.

Approach	M_n		
Classical statistics	$\frac{\max_{P \in \mathcal{H}_1} P(X^n)}{\max_{P \in \mathcal{H}_0} P(X^n)}$	well-understood asymptotics	hard to calibrate

Composite case

Now consider rejecting a composite null $\mathcal{H}_0$ with the help of a **composite alternative** $\mathcal{H}_1$.

Idea: find a measure of evidence that behaves like a likelihood ratio.

Approach	M_n		
Classical statistics	$\frac{\max_{P \in \mathcal{H}_1} P(X^n)}{\max_{P \in \mathcal{H}_0} P(X^n)}$	well-understood asymptotics	hard to calibrate
Bayesian approach	$\frac{\int_{\mathcal{H}_1} P(X^n) w_1(dP)}{\int_{\mathcal{H}_0} P(X^n) w_0(dP)}$	symmetry	no martingale

Composite case

Now consider rejecting a composite null $\mathcal{H}_0$ with the help of a **composite alternative** $\mathcal{H}_1$.

Idea: find a measure of evidence that behaves like a likelihood ratio.

Approach	M_n		
Classical statistics	$\frac{\max_{P \in \mathcal{H}_1} P(X^n)}{\max_{P \in \mathcal{H}_0} P(X^n)}$	well-understood asymptotics	hard to calibrate
Bayesian approach	$\frac{\int_{\mathcal{H}_1} P(X^n) w_1(dP)}{\int_{\mathcal{H}_0} P(X^n) w_0(dP)}$	symmetry	no martingale
Universal Inference	$\frac{\int_{\mathcal{H}_1} P(X^n) w_1(dP)}{\max_{P \in \mathcal{H}_0} P(X^n)}$	Ville (e-process)	twice conservative

Composite case

Now consider rejecting a composite null $\mathcal{H}_0$ with the help of a **composite alternative** $\mathcal{H}_1$.

Idea: find a measure of evidence that behaves like a likelihood ratio.

Approach	M_n		
Classical statistics	$\frac{\max_{P \in \mathcal{H}_1} P(X^n)}{\max_{P \in \mathcal{H}_0} P(X^n)}$	well-understood asymptotics	hard to calibrate
Bayesian approach	$\frac{\int_{\mathcal{H}_1} P(X^n) w_1(dP)}{\int_{\mathcal{H}_0} P(X^n) w_0(dP)}$	symmetry	no martingale
Universal Inference	$\frac{\int_{\mathcal{H}_1} P(X^n) w_1(dP)}{\max_{P \in \mathcal{H}_0} P(X^n)}$	Ville (e-process)	twice conservative
GRO(W)	$\ldots$	log-optimal	fixed sample size n

Composite case

Now consider rejecting a composite null $\mathcal{H}_0$ with the help of a **composite alternative** $\mathcal{H}_1$.

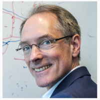
Idea: find a measure of evidence that behaves like a likelihood ratio.

Approach	M_n		
Classical statistics	$\frac{\max_{P \in \mathcal{H}_1} P(X^n)}{\max_{P \in \mathcal{H}_0} P(X^n)}$	well-understood asymptotics	hard to calibrate
Bayesian approach	$\frac{\int_{\mathcal{H}_1} P(X^n) w_1(dP)}{\int_{\mathcal{H}_0} P(X^n) w_0(dP)}$	symmetry	no martingale
Universal Inference	$\frac{\int_{\mathcal{H}_1} P(X^n) w_1(dP)}{\max_{P \in \mathcal{H}_0} P(X^n)}$	Ville (e-process)	twice conservative
GRO(W)	$\ldots$	log-optimal	fixed sample size n

The theory of **e-variables** aims to find better likelihood-ratio-like quantities.

The birth of an open problem

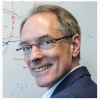






The birth of an open problem

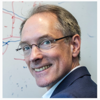
Regarding your evidence process for the t-test



The birth of an open problem

Regarding your evidence process for the t-test

Lovely! What about it?

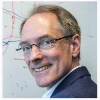


The birth of an open problem

Regarding your evidence process for the t-test

Lovely! What about it?

It is martingale against **zero** effect [GdHK24]



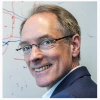
The birth of an open problem


Regarding your evidence process for the t-test

Lovely! What about it?

It is martingale against **zero** effect [GdHK24]

and e-variable against **negative** effect [Pér+24]





The birth of an open problem

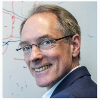

Regarding your evidence process for the t-test

Lovely! What about it?

It is martingale against **zero** effect [GdHK24]

and e-variable against **negative** effect [Pér+24]

But supermartingale against **negative** effect?





The birth of an open problem


Regarding your evidence process for the t-test

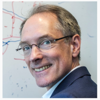
Lovely! What about it?

It is martingale against **zero** effect [GdHK24]

and e-variable against **negative** effect [Pér+24]

But supermartingale against **negative** effect?

Surely! Hold on . . .





The birth of an open problem


Regarding your evidence process for the t-test

Lovely! What about it?

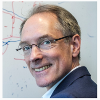
It is martingale against **zero** effect [GdHK24]

and e-variable against **negative** effect [Pér+24]

But supermartingale against **negative** effect?

Surely! Hold on . . .

— several months later —





The birth of an open problem

Regarding your evidence process for the t-test

Lovely! What about it?

It is martingale against **zero** effect [GdHK24]

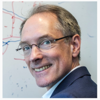
and e-variable against **negative** effect [Pér+24]

But supermartingale against **negative** effect?

Surely! Hold on . . .

— several months later —

Not trivial, eh?



The birth of an open problem

Regarding your evidence process for the t-test

Lovely! What about it?

It is martingale against **zero** effect [GdHK24]

and e-variable against **negative** effect [Pér+24]

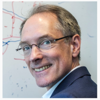
But supermartingale against **negative** effect?

Surely! Hold on . . .

— several months later —

Not trivial, eh?

. . . only mild partial progress . . .



The birth of an open problem

Regarding your evidence process for the t-test

Lovely! What about it?

It is martingale against **zero** effect [GdHK24]

and e-variable against **negative** effect [Pér+24]

But supermartingale against **negative** effect?

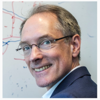
Surely! Hold on . . .

— several months later —

Not trivial, eh?

. . . only mild partial progress . . .

— one eternity later —



The birth of an open problem

Regarding your evidence process for the t-test

Lovely! What about it?

It is martingale against **zero** effect [GdHK24]

and e-variable against **negative** effect [Pér+24]

But supermartingale against **negative** effect?

Surely! Hold on . . .

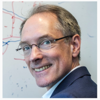
— several months later —

Not trivial, eh?

. . . only mild partial progress . . .

— one eternity later —

Ok, supermartingale after all.



Setup

Simplest composite vs composite example

Consider data stream $X_1, X_2, \ldots$

Simplest composite vs composite example

Consider data stream $X_1, X_2, \ldots$

We assume throughout that X_i are i.i.d. $\mathcal{N}(\delta\sigma, \sigma^2)$ for some effect size δ and **variance** σ^2.

Simplest composite vs composite example

Consider data stream $X_1, X_2, \ldots$

We assume throughout that X_i are i.i.d. $\mathcal{N}(\delta\sigma, \sigma^2)$ for some effect size δ and **variance** σ^2.

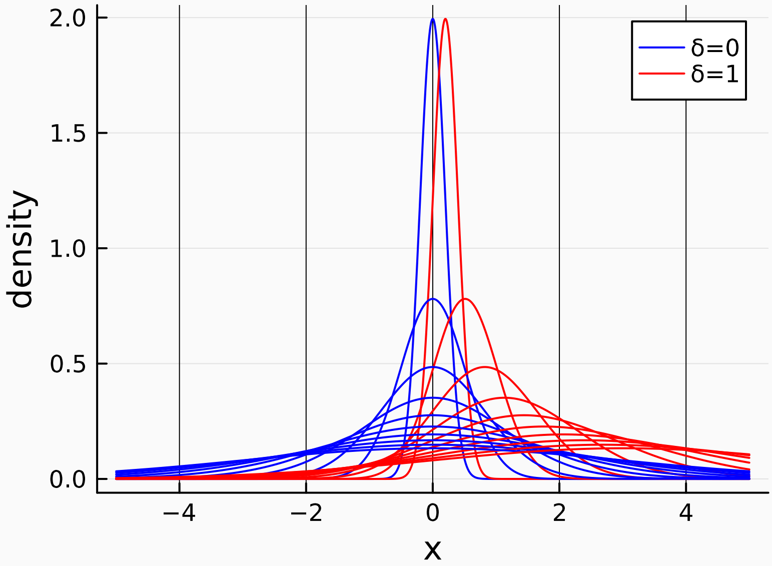
Aim: to disqualify the composite null of **no effect**

$$\mathcal{H}_0 \; = \; \left\{\delta = 0, \sigma^2 > 0\right\}$$

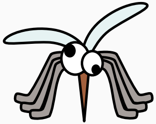
with the help of the composite alternative that the **effect size** is a given $\delta_+ > 0$

$$\mathcal{H}_+ \; = \; \left\{\delta = \delta_+, \sigma^2 > 0\right\}$$

Is that hard?

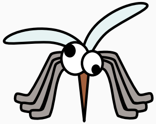


Nuisance



The variance/scale σ^2 is a nuisance parameter.

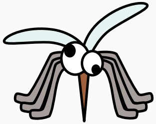
Nuisance



The variance/scale σ^2 is a nuisance parameter.

The nuisance is a **group** (here: positive scaling)

Nuisance



The variance/scale σ^2 is a nuisance parameter.

The nuisance is a **group** (here: positive scaling)

We can quotient it out

- Coarsen the data
- Work with orbits

Coarsening the data

Define the coarsening $(Z_i)_{i\geq 1}$ of the data $(X_i)_{i\geq 1}$ by

$$Z_i = \frac{X_i}{|X_1|}$$

Coarsening the data

Define the coarsening $(Z_i)_{i \geq 1}$ of the data $(X_i)_{i \geq 1}$ by

$$Z_i \;=\; \frac{X_i}{|X_1|}$$

Everyone in $\mathcal{H}_0$ agrees about the distribution of $Z_1, Z_2, \ldots$

And everyone in $\mathcal{H}_{\delta_+}$ agrees about the distribution of $Z_1, Z_2, \ldots$

yet they don't agree with each other.

Coarsening the data

Define the coarsening $(Z_i)_{i\geq1}$ of the data $(X_i)_{i\geq1}$ by

$$Z_i = \frac{X_i}{|X_1|}$$

Everyone in $\mathcal{H}_0$ agrees about the distribution of $Z_1, Z_2, \ldots$

And everyone in $\mathcal{H}_{\delta_+}$ agrees about the distribution of $Z_1, Z_2, \ldots$

yet they don't agree with each other.

Both null and alternative collapse to a **point**. NB: both see $(Z_i)_{i\geq1}$ as dependent (not i.i.d.)!

Coarsened Likelihood Ratio

Let $p_0(Z^n)$ and $p_{\delta_+}(Z^n)$ be the density of the coarsening Z^n under the null and the alternative.

Let's look at the process $(M_n)_{n \geq 0}$

$$M_n \;:=\; \frac{p_{\delta_+}(Z^n)}{p_0(Z^n)}$$

Representations of the coarsened likelihood ratio

Let $S_n = \sum_{i=1}^n X_i$ and $V_n = \sum_{i=1}^n X_i^2$ and $R_n = \frac{S_n}{\sqrt{V_n}}$.

We have the Hypergeometric form

$$M_n = \frac{\Gamma\left(\frac{n}{2}\right)\, {}_1F_1\left(\frac{n}{2}; \frac{1}{2}; \frac{\delta_+^2 R_n^2}{2}\right) + \sqrt{2}\delta_+ R_n \Gamma\left(\frac{n+1}{2}\right)\, {}_1F_1\left(\frac{n+1}{2}; \frac{3}{2}; \frac{\delta_+^2 R_n^2}{2}\right)}{\Gamma\left(\frac{n}{2}\right) e^{\frac{n}{2}\delta_+^2}}$$

the Pochhammer form

$$M_n = \frac{1}{\Gamma\left(\frac{n}{2}\right) e^{\frac{n}{2}\delta_+^2}} \sum_{k=0}^{\infty} \frac{\Gamma\left(\frac{k+n}{2}\right)}{k!} \left(\sqrt{2}\delta_+ R_n\right)^k$$

the Haar forms

$$M_n = \frac{\int p_{\mathcal{N}(\delta_+\sigma, \sigma^2)}(X^n) \frac{1}{\sigma}\, \mathrm{d}\sigma}{\int p_{\mathcal{N}(0, \sigma^2)}(X^n) \frac{1}{\sigma}\, \mathrm{d}\sigma} = \frac{2}{\Gamma\left(\frac{n}{2}\right) e^{\frac{n}{2}\delta_+^2}} \int_0^{\infty} e^{w\sqrt{2}\delta_+ R_n - w^2} w^{n-1}\, \mathrm{d}w$$

the non-central Student-t form

$$M_n = \frac{P\left(T_n; n-1, \delta_+\sqrt{n}\right)}{P\left(T_n; n-1, 0\right)} \qquad \text{where} \qquad T_n = R_n \sqrt{\frac{n-1}{n - R_n^2}}$$

Martingale

Is the coarsened likelihood ratio a martingale for $\mathcal{H}_0$?

$$M_n \;=\; \frac{p_{\delta_+}(Z^n)}{p_0(Z^n)}$$

NB: for every $P \in \mathcal{H}_0$

$$\mathbb{E}_P\left[M_{n+1} | Z^n\right] \;=\; M_n$$

But

$$\mathbb{E}_P\left[M_{n+1} | X^n\right] \;\neq\; M_n$$

e-variable

Recall our null of zero effect was

$$\mathcal{H}_0 \;=\; \{\delta = 0, \sigma^2 > 0\}.$$

Let's look at the **much larger** null of negative effect

$$\mathcal{H}_{\leq 0} \;:=\; \{\delta \leq 0, \sigma^2 > 0\}$$

e-variable

Recall our null of zero effect was

$$\mathcal{H}_0 \;=\; \{\delta = 0, \sigma^2 > 0\}.$$

Let's look at the **much larger** null of negative effect

$$\mathcal{H}_{\leq 0} \;:=\; \{\delta \leq 0, \sigma^2 > 0\}$$

[Pér+24] show that for every $P \in \mathcal{H}_{\leq 0}$ and fixed n,

$$\mathbb{E}_P\left[M_n\right] \;\leq\; 1$$

We say "M_n is an e-variable against $\mathcal{H}_{\leq 0}$".

The Engine of Safety for One-Sided Null

Definition

Random variable T has monotone likelihood ratio (MLR) if, whenever $\delta' \geq \delta$,

$$\frac{p^T_{\delta'}(t)}{p^T_{\delta}(t)} \quad \text{is increasing in } t$$

The Engine of Safety for One-Sided Null

Definition

Random variable T has monotone likelihood ratio (MLR) if, whenever $\delta' \geq \delta$,

$$\frac{p^T_{\delta'}(t)}{p^T_{\delta}(t)} \quad \text{is increasing in } t$$

Lemma

For MLR T, for all $\delta_- \leq 0 \leq \delta_+$,

$$\mathbb{E}_{\delta_-}\left[\frac{p_{\delta_+}(T)}{p_0(T)}\right] \leq 1.$$

Proof runs via stochastic dominance.

The Engine of Safety for One-Sided Null

Definition

Random variable T has monotone likelihood ratio (MLR) if, whenever $\delta' \geq \delta$,

$$\frac{p^T_{\delta'}(t)}{p^T_{\delta}(t)} \quad \text{is increasing in } t$$

Lemma

For MLR T, for all $\delta_- \leq 0 \leq \delta_+$,

$$\mathbb{E}_{\delta_-}\left[\frac{p_{\delta_+}(T)}{p_0(T)}\right] \leq 1.$$

Proof runs via stochastic dominance.

In our case, the t-statistic T_n at sample size n has MLR. So $\frac{p_{\delta_+}(T_n)}{p_{\delta_0}(T_n)}$ is an e-variable for $\mathcal{H}_{\leq 0}$.

Summary so far

We have that M_n is a **martingale** against every $P \in \mathcal{H}_0$ on the filtration $(\sigma(Z^n))_{n \geq 0}$.

$$\mathbb{E}_{P_0}[M_{n+1} | Z^n] \;=\; M_n$$

So we can stop and reject $\mathcal{H}_0$ when $M_n \geq \frac{1}{\alpha}$.

Summary so far

We have that M_n is a **martingale** against every $P \in \mathcal{H}_0$ on the filtration $(\sigma(Z^n))_{n\geq 0}$.

$$\mathbb{E}_{P_0}[M_{n+1}|Z^n] \;=\; M_n$$

So we can stop and reject $\mathcal{H}_0$ when $M_n \geq \frac{1}{\alpha}$.

We get Type-1 error control by Ville's inequality: for every $P \in \mathcal{H}_0$:

$$P_0\left\{\exists n : M_n \geq \frac{1}{\alpha}\right\} \;\leq\; \alpha$$

Summary so far

We have that M_n is a **martingale** against every $P \in \mathcal{H}_0$ on the filtration $(\sigma(Z^n))_{n \geq 0}$.

$$\mathbb{E}_{P_0}[M_{n+1}|Z^n] \;=\; M_n$$

So we can stop and reject $\mathcal{H}_0$ when $M_n \geq \frac{1}{\alpha}$.

We get Type-1 error control by Ville's inequality: for every $P \in \mathcal{H}_0$:

$$P_0 \left\{ \exists n : M_n \geq \frac{1}{\alpha} \right\} \;\leq\; \alpha$$

And we stop at the right moment: for every $P \in \mathcal{H}_+$,

$$\mathbb{E}_P\left[\tau\right] \;\approx\; \frac{2 \ln \frac{1}{\alpha}}{\ln(1 + \delta_+^2)} \;=\; \frac{\ln \frac{1}{\alpha}}{\mathrm{KL}\left(\mathcal{H}_+ \| \mathcal{H}_0\right)}$$

Posing the Problem

Open problem

Recall the null of negative effect

$$\mathcal{H}_{\leq 0} := \{\delta \leq 0, \sigma^2 > 0\}.$$

Question

Is $(M_n)_{n \geq 0}$ a **supermartingale** against negative effect (i.e. for every $P \in \mathcal{H}_{\leq 0}$)?

Resolution

Monotone Likelihood Ratio

For

$$M_n = \frac{p_{\delta_+}(Z^n)}{p_0(Z^n)}$$

and $\delta_- \leq 0$, we aim to show

$$\mathbb{E}_{\delta_-}[M_{n+1}|Z^n] \leq M_n$$

Monotone Likelihood Ratio

For

$$M_n \;=\; \frac{p_{\delta_+}(Z^n)}{p_0(Z^n)}$$

and $\delta_- \leq 0$, we aim to show

$$\mathbb{E}_{\delta_-}\left[M_{n+1}|Z^n\right] \;\leq\; M_n$$

Attempt 1: Fix Z^n. Then $\frac{M_{n+1}}{M_n} = \frac{p_{\delta_+}(Z_{n+1}|Z^n)}{p_0(Z_{n+1}|Z^n)}$. Does Z_{n+1} have the monotone likelihood ratio property under the **conditional** model $P_\delta\left(\cdot|Z^n\right)$?

Monotone Likelihood Ratio

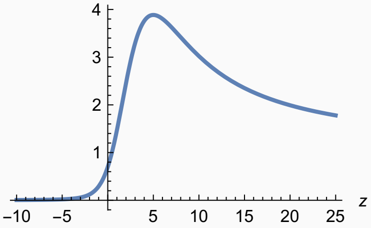
For

$$M_n = \frac{p_{\delta_+}(Z^n)}{p_0(Z^n)}$$

and $\delta_- \leq 0$, we aim to show

$$\mathbb{E}_{\delta_-}[M_{n+1}|Z^n] \leq M_n$$

Attempt 1: Fix Z^n. Then $\frac{M_{n+1}}{M_n} = \frac{p_{\delta_+}(Z_{n+1}|Z^n)}{p_0(Z_{n+1}|Z^n)}$. Does Z_{n+1} have the monotone likelihood ratio property under the **conditional** model $P_\delta(\cdot|Z^n)$?



NO!

Sufficiency to the Rescue

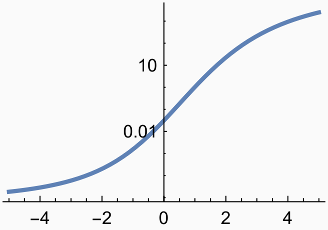
Recall that

$$M_n = \frac{P\left(T_n; n-1, \delta_+\sqrt{n}\right)}{P\left(T_n; n-1, 0\right)}$$

where T_n is the t-statistic. We then have

$$\frac{M_{n+1}}{M_n} = \frac{p_{\delta_+}(Z_{n+1}|Z^n)}{p_0(Z_{n+1}|Z^n)} \overset{\text{sufficiency}}{=} \frac{p_{\delta_+}(T_{n+1}|Z^n)}{p_0(T_{n+1}|Z^n)} \overset{\text{sufficiency}}{=} \frac{p_{\delta_+}(T_{n+1})}{p_0(T_{n+1})}\frac{p_0(Z^n)}{p_{\delta_+}(Z^n)}$$

and so the conditional likelihood ratio is increasing.



YES!

The general case

Theorem

Fix $\delta_0 \leq \delta_+$. Let $(T_n)_{n\in\mathbb{N}}$ be a sequence of **sufficient statistics** *satisfying the monotone likelihood ratio property. Then the process $\left(\prod_{i=1}^{n} \frac{p_{\delta_+}^{T_i}(T_i|U^{i-1})}{p_{\delta_0}^{T_i}(T_i|U^{i-1})}\right)_{n\in\mathbb{N}}$ is identical to the likelihood ratio process $\left(\frac{p_{\delta_+}(U^n)}{p_{\delta_0}(U^n)}\right)_{n\in\mathbb{N}}$ and both are "test" (positive, starting at 1)* **supermartingales** *relative to the one-sided null $\mathcal{H}_{\leq 0}$.*

Linear Regression

Linear Regression with Nuisance Covariates

Consider i.i.d. observations (X_i, Y_i, Z_i) from the linear regression model

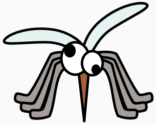
$$Y_i \;=\; \delta\sigma X_i + \beta^\intercal Z_i + \sigma\varepsilon_i,$$

where $\delta \in \mathbb{R}$, $\beta \in \mathbb{R}^d$ and $\sigma \in \mathbb{R}^+$ are the parameters, and $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. $\mathcal{N}(0, 1)$.

We aim to test the effect size δ. We treat the **coefficients** β and **scale** σ as nuisance.

$$\mathcal{H}_0 \;=\; \left\{\delta = 0, \sigma^2 > 0, \beta \in \mathbb{R}^d\right\} \qquad \text{vs} \qquad \mathcal{H}_+ \;=\; \left\{\delta = \delta_+, \sigma^2 > 0, \beta \in \mathbb{R}^d\right\}.$$

Result for Linear Regression



In fact, here the nuisance is again a **group** (scaling and general linear).

We can quotient it out, e.g. by coarsening the labels Y^n to

$$U_n := \frac{\boldsymbol{A}_n Y^n}{\|\boldsymbol{A}_n Y^n\|} \in S^{n-d-1} \subseteq \mathbb{R}^{n-d}$$

where $\boldsymbol{A}_n^\intercal \boldsymbol{A}_n Y^n$ is residual of linear regression of labels Y^n onto nuisance covariates $Z_1, \ldots, Z_n$.

Linear Regression ctd

Let us define

$$T_n := \frac{\frac{b_n^\intercal}{\|b_n\|} U_n}{\frac{1}{\sqrt{n-d-1}} \|\boldsymbol{P}_n U_n\|}$$

where

$$b_n := \boldsymbol{A}_n X^n \in \mathbb{R}^{n-d} \qquad \text{and} \qquad \boldsymbol{P}_n := \boldsymbol{I}_n - \frac{b_n b_n^\intercal}{b_n^\intercal b_n}$$

Linear Regression ctd

Let us define

$$T_n \;:=\; \frac{\frac{b_n^\mathsf{T}}{\|b_n\|} U_n}{\frac{1}{\sqrt{n-d-1}}\,\|\boldsymbol{P}_n U_n\|}$$

where

$$b_n \;:=\; \boldsymbol{A}_n X^n \in \mathbb{R}^{n-d} \qquad \text{and} \qquad \boldsymbol{P}_n \;:=\; \boldsymbol{I}_n - \frac{b_n b_n^\mathsf{T}}{b_n^\mathsf{T} b_n}$$

Then

- T_n is a sufficient statistic for the data $U_1, U_2, \ldots, U_n$.
- T_n has non-central Student-t distribution with $n-d-1$ degrees of freedom and non-centrality parameter $\delta\|b_n\|$
- T_n has the MLR property

Linear Regression ctd

Let us define

$$T_n \;:=\; \frac{\frac{b_n^\intercal}{\|b_n\|} U_n}{\frac{1}{\sqrt{n-d-1}} \|\boldsymbol{P}_n U_n\|}$$

where

$$b_n \;:=\; \boldsymbol{A}_n X^n \in \mathbb{R}^{n-d} \qquad \text{and} \qquad \boldsymbol{P}_n \;:=\; \boldsymbol{I}_n - \frac{b_n b_n^\intercal}{b_n^\intercal b_n}$$

Then

- T_n is a sufficient statistic for the data $U_1, U_2, \ldots, U_n$.
- T_n has non-central Student-t distribution with $n-d-1$ degrees of freedom and non-centrality parameter $\delta\|b_n\|$
- T_n has the MLR property

So $M_n := \frac{f_{\mathrm{T}(n-d-1,\delta_+\|b_n\|)}(T_n)}{f_{\mathrm{T}(n-d-1,0)}(T_n)}$ is a test supermartingale under the entire null $\mathcal{H}_{\leq 0}$.

Conclusion

- The t-test is a supermartingale after all
- Due the **monotone likelihood ratio** property of a sufficient statistic
- This upgrades to many cases: χ^2, linear regression, ...
- Building block for all sorts of anytime-valid testing and inference

Conclusion

- The t-test is a supermartingale after all
- Due the **monotone likelihood ratio** property of a sufficient statistic
- This upgrades to many cases: χ^2, linear regression, ...
- Building block for all sorts of anytime-valid testing and inference

Let's talk!

References i

[GdHK24] P. D. Grünwald, R. de Heide, and W. M. Koolen. **"Safe Testing".** In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 86.5 (2024). With Discussion.

[Pér+24] M. F. Pérez-Ortiz, T. Lardy, R. de Heide, and P. Grünwald. **"E-Statistics, Group Invariance and Anytime Valid Testing".** In: *The Annals of Statistics* 52.4 (2024).