Hypothesis Testing

Rianne de Heide

CWI & Leiden University

April 6, 2018

Me

PhD-student
Machine Learning Group

Peter Grünwald

Study coordinator
M.Sc. Statistical Science

Jacqueline Meulman

Projects

- Bayesian inference under model misspecification (SafeBayes)
- Foundations of Bayesianism
- Hypothesis testing (Frequentist, Bayesian, new methods)

Peter

Tom

Wouter

Allard

Frequentist vs. Bayesian probability

Different views on **probability**:

- Frequentists: frequency in a long run
- Bayesians: degree of belief

Hypothesis Testing

Example:

- Null Hypothesis $\mathcal{H}_0$: $X^n \sim N(0, \sigma)$
- Alternative Hypothesis $\mathcal{H}_1 : X^n \sim N(\theta, \sigma)$, $\theta \neq 0$

Hypothesis Testing

Example:

- Null Hypothesis $\mathcal{H}_0$: $X^n \sim N(0, \sigma)$
- Alternative Hypothesis $\mathcal{H}_1 : X^n \sim N(\theta, \sigma)$, $\theta \neq 0$

Different types of hypothesis tests

- p-value based null hypothesis significance tests (frequentist)

Hypothesis Testing

Example:

- Null Hypothesis $\mathcal{H}_0$: $X^n \sim N(0, \sigma)$
- Alternative Hypothesis $\mathcal{H}_1 : X^n \sim N(\theta, \sigma)$, $\theta \neq 0$

Different types of hypothesis tests

- p-value based null hypothesis significance tests (frequentist)
- Bayes Factor hypothesis tests (about probability of $\mathcal{H}_0$ given the data)
  Limitations: see De Heide and Grünwald (2018)

Hypothesis Testing

Example:

- Null Hypothesis $\mathcal{H}_0$: $X^n \sim N(0, \sigma)$
- Alternative Hypothesis $\mathcal{H}_1 : X^n \sim N(\theta, \sigma)$, $\theta \neq 0$

Different types of hypothesis tests

- p-value based null hypothesis significance tests (frequentist)
- Bayes Factor hypothesis tests (about probability of $\mathcal{H}_0$ given the data)
  Limitations: see De Heide and Grünwald (2018)
- Test martingales and S-tests (current work)

Optional stopping

Optional stopping: 'peeking at the results so far to decide whether or not to gather more data'

$\rightarrow$ reproducibility crisis in life and behavioural sciences (far too many false positives)

Frequentist Optional Stopping and p-values (1)

- Data $X^n = X_1, X_2, \ldots, X_n$; $X_i \in \mathcal{X}$
- Hypothesis test $T_n : \mathcal{X}^n \mapsto \{0, 1\}$
- Significance level α

Frequentist Optional Stopping and p-values (1)

- Data $X^n = X_1, X_2, \ldots, X_n$; $X_i \in \mathcal{X}$
- Hypothesis test $T_n : \mathcal{X}^n \mapsto \{0, 1\}$
- Significance level α

Frequentist optional stopping
A sequence of hypothesis tests $T_n : \mathcal{X}^n \to \{0, 1\}$ with significance level α is said to be *robust under frequentist optional stopping* if

$$\mathbf{P}_{H_0}\left(\exists n : T_n(X^n) = 1\right) \leq \alpha.$$

Frequentist Optional Stopping and p-values (2)

Example:

Data $X^n \stackrel{\text{i.i.d.}}{\sim} N(\theta, 1)$

Frequentist Optional Stopping and p-values (2)

Example:

Data $X^n \overset{\text{i.i.d.}}{\sim} N(\theta, 1)$
$\mathcal{H}_0$: $\theta = 0$
$\mathcal{H}_1$: $\theta \neq 0$

Frequentist Optional Stopping and p-values (2)

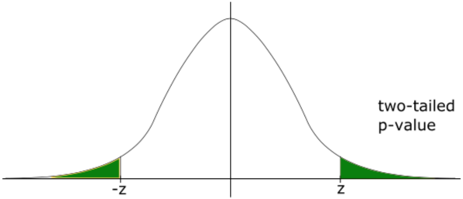
Example:

Data $X^n \overset{\text{i.i.d.}}{\sim} N(\theta, 1)$
$\mathcal{H}_0$: $\theta = 0$
$\mathcal{H}_1$: $\theta \neq 0$
Test statistic: $Z_n = \overline{X}\sqrt{n}$

Frequentist Optional Stopping and p-values (2)

Example:

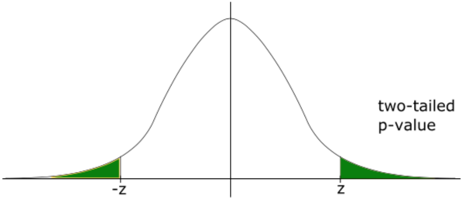
Data $X^n \overset{\text{i.i.d.}}{\sim} N(\theta, 1)$
$\mathcal{H}_0$: $\theta = 0$
$\mathcal{H}_1$: $\theta \neq 0$
Test statistic: $Z_n = \overline{X}\sqrt{n}$
p-value: $p = 2(1 - \Phi(|Z_n|)$



Frequentist Optional Stopping and p-values (2)

Example:

Data $X^n \overset{\text{i.i.d.}}{\sim} N(\theta, 1)$
$\mathcal{H}_0$: $\theta = 0$
$\mathcal{H}_1$: $\theta \neq 0$
Test statistic: $Z_n = \overline{X}\sqrt{n}$
p-value: $p = 2(1 - \Phi(|Z_n|)$

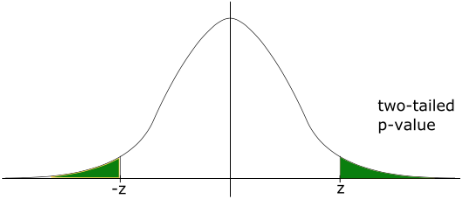


Stopping rule: Continue until $|Z_n| > k$, then stop (with $\alpha = 0.05, k = 1.96$).

Frequentist Optional Stopping and p-values (2)

Example:

Data $X^n \overset{\text{i.i.d.}}{\sim} N(\theta, 1)$
$\mathcal{H}_0$: $\theta = 0$
$\mathcal{H}_1$: $\theta \neq 0$
Test statistic: $Z_n = \overline{X}\sqrt{n}$
p-value: $p = 2(1 - \Phi(|Z_n|)$



Stopping rule: Continue until $|Z_n| > k$, then stop (with $\alpha = 0.05, k = 1.96$).

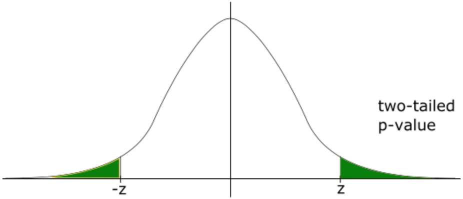
LIL for $X_1, X_2, \ldots$ i.i.d. standard normal:

$$\limsup_{n \to \infty} \frac{\sum_{i=1}^{n} x_i}{\sqrt{2n \log \log n}} = 1 \text{ a.s.}$$

Frequentist Optional Stopping and p-values (2)

Example:

Data $X^n \overset{\text{i.i.d.}}{\sim} N(\theta, 1)$
$\mathcal{H}_0$: $\theta = 0$
$\mathcal{H}_1$: $\theta \neq 0$
Test statistic: $Z_n = \overline{X}\sqrt{n}$
p-value: $p = 2(1 - \Phi(|Z_n|)$



Stopping rule: Continue until $|Z_n| > k$, then stop (with $\alpha = 0.05, k = 1.96$).

LIL for $X_1, X_2, \ldots$ i.i.d. standard normal:

$$\limsup_{n \to \infty} \frac{\sum_{i=1}^{n} x_i}{\sqrt{2n \log \log n}} = 1 \text{ a.s.}$$

This gives us: $Z_n > \lambda\sqrt{2 \log \log n}$ i.o. w.p. 1, for $\lambda < 1$.

Frequentist Optional Stopping and Bayes Factors

Can Bayes Factor hypothesis tests handle frequentist optional stopping? (Bayesian statisticians suggest they do...)

- Not always (see Sanborn and Hills (2014); De Heide and Grünwald (2018))
- Yes, if $\mathcal{H}_0$ is simple
- Yes, with composite $\mathcal{H}_0$ and certain group invariance structure (see Hendriksen, De Heide and Grünwald (2018))
  Example: Bayesian t-test

Test Martingales (1)

The Test Martingale (Vovk et al., 2011):

A non-negative martingale $\{M_n\}_{n\in\mathbb{N}}$ with initial value $M_0 = 1$.

Test Martingales (1)

The Test Martingale (Vovk et al., 2011):

A non-negative martingale $\{M_n\}_{n\in\mathbb{N}}$ with initial value $M_0 = 1$.

Test Martingales (2)

Example:
$\mathcal{H}_0$: the coin is fair, i.e. the outcomes $X_1, X_2, \ldots$ are i.i.d. Bernoulli(0.5)
Start with capital $K_0 = 1$
At each round i

- We invest a fraction q_i of our capital on the outcome $\{X_i = 0\}$ and $1 - q_i$ on $\{X_i = 1\}$
- The outcome is revealed
- We get a pay-off: twice our stakes on the correct outcome, nothing on the other

After N rounds: $K_N = 2^N \prod_{t=1}^{N} q(X_N | X^{N-1})$.

Test Martingales (3)

- If $\mathcal{H}_0$ is true, the bet is *fair*, i.e. $\mathbf{E}[K_N] \leq K_0$
- If K_N is large, it is an indication that $\mathcal{H}_0$ is not true.
- Test martingales can handle frequentist optional stopping.

S-test statistics

Test martingales can handle frequentist optional stopping and have additional advantages over p-values (interpretability)

Problem: how to construct test martingales for composite $\mathcal{H}_0$.

Current work: S-value

References

P. Grünwald, R. de Heide, and W. Koolen. Safe tests. *work in progress*, 2018.

R. de Heide and P. Grünwald. Why optional stopping is a problem for Bayesians. *arXiv preprint arXiv:1708.08278*, 2018.

A. Hendriksen, R. de Heide, and P. Grünwald. Folklore results on optional stopping of Bayesian methods made precise. *work in progress*, 2018.

A. Sanborn and T. Hills. The frequentist implications of optional stopping on Bayesian hypothesis tests. *Psychonomic Bulletin & Review*, 21(2):283–300, 2014.

V. Vovk, N. Vereshchagin, A. Shen, and G. Shafer. Test martingales, Bayes factors and p-values. *Statistical Science*, 26 (1):84–101, 2011.