

## Various Projects on E-Values, Always-Valid Confidence Sequences and "Safe Testing"

**Supervisor:** Peter Grünwald

**Keywords:** Testing; uncertainty quantifications; foundations of statistics and machine learning

How much evidence do the data give us about one hypothesis versus another? The standard way to measure evidence is still the p-value, despite a myriad of problems surrounding it – problems which are one of the reasons for the ongoing *replicability crisis* in the applied sciences such as medicine and psychology: the fraction of published results that are irreproducible (which often just means ‘wrong’) is much higher than one would hope.

The e-value is a recently popularized notion of evidence which overcomes some of the issues with p-values. While e-values have lain dormant until 2019, interest in them has recently exploded with papers in the world’s top machine learning conferences and statistics journals. In June 2022 we held a first international workshop on e-values with attendees from the areas of clinical trial design and meta-analysis but also from some of the large tech companies interested in A/B testing.

Unlike p-values, E-values allow for tests with strict 'classical' Type-I error control under *optional continuation and combination of data from different sources*. They are also easier to interpret than p-values, having a straightforward interpretation in terms of sequential betting.

E-values are also the basic building blocks of *anytime-valid confidence intervals* that remain valid under optional stopping and that are crucial for gaining trustworthy uncertainty quantification in e.g. A/B testing and bandit settings. In simple cases, inference based on e-values coincides with a particular Bayesian method, the *Bayes factor*. But if the null is composite or nonparametric, or an alternative cannot be explicitly formulated, e-values and Bayes factors become distinct and e-processes can be seen as a generalization of nonnegative supermartingales, a central topic in stochastic process theory.

The theory of E-Values is still very young, so many types of projects are possible. Here are a few examples:

- Design and implementation (in R or Python) of E-Values for Cox Regression, standard Regression, Mixed Models, Confidence Sequences for Effect Size in stratified contingency tables
- Comparison of different existing E-Values for the 'Model-X' Conditional Independence Tests. Investigating the claim that the Model-X assumption is unavoidable
- Comparing the GRAPA and the REGROW design principles for e-variables
- (more theoretical) Investigating the relation between E-variables and the Likelihood Principle