Safe Testing with optimally GRO(W)ing e-values

Peter Grünwald[1,2]; Rianne de Heide*[1,2]; Wouter M. Koolen[2]

[1] Leiden University, Leiden, The Netherlands
[2] Centrum Wiskunde & Informatica Amsterdam, The Netherlands
e-mail: {pdg, r.de.heide, w.m.koolen-wijkstra}@cwi.nl
session code: IPS76

Abstract:

There has been significant recent progress on a new theory of hypothesis testing. The core of the new approach is the abstract concept of e-values, also known as betting scores. In this talk we focus on the question: "what is a good e-value?", for completely general testing problems with composite null and alternatives. Surprisingly, optimal Growth-Rate Optimal in Worst case "GROW" e-values, which lead to fastest capital growth, are fully characterised by the joint information projection (JIPr) between the set of all Bayes marginal distributions on H0 and H1. We also discuss the related notion of Growth-Rate Optimal "GRO" (without the W), in case there is additional information about H1, or this can be learned from the data. As examples we show the GRO(W) e-value equivalent of the classical t-test and Fisher's exact test for 2x2 contingency tables.

Keywords:
hypothesis testing; e-values; betting scores; information projection

1. Introduction:

There has been significant recent progress on a new theory of hypothesis testing. The core of the new approach is the abstract concept of e-values, also known as betting scores. E-values are easier to interpret than classical p-values, behave well under optional continuation, and are easy to combine. E-values are closely connected to non-negative martingales and Bayes factors.

In this talk, part of the session on e-values, we focus on the question "what is a good e-value?". We introduce the GRO(W) criterion to characterize e-values that are in some sense optimal.

2. Methodology:

Surprisingly, the GROW e-values can be fully characterized by the joint information projection (JIPr) between the set of all Bayes marginal distributions on H0 and H1.

3. Result:

We develop GRO(W) e-values for completely general testing problems with composite null and alternatives.

4. Discussion and Conclusion:

We introduce the GRO(W) criterion for characterising e-values that are in some sense optimal. We discuss the GRO(W) e-value equivalents of the classical t-test and Fishers exact test for 2x2 contingency tables. We show that they perform competatively.

References:

1. Safe Testing - Peter Grünwald, Rianne de Heide, Wouter M. Koolen.

2.
3.

NOTE: THE MAXIMUM NUMBER OF PAGES FOR THE PAPER IS SIX PAGES