Safe Testing: S-Values and Optional Continuation

Peter Grünwald PDG@CWI.NL
Rianne de Heide HEIDE@CWI.NL
Centrum Wiskunde & Informatica, Amsterdam and Leiden University, The Netherlands

Wouter M. Koolen WMKOOLEN@CWI.NL
Centrum Wiskunde & Informatica, Amsterdam, The Netherlands

One of the many problems surrounding p-value based null hypothesis testing is the following: if our test result is promising but nonconclusive (say, $p = 0.07$) we cannot simply decide to gather a few more data points. While this "optional continuation" is ubiquitous in science, it invalidates frequentist error guarantees. Here we propose an alternative hypothesis testing methodology based on 'S-values'. Its primary interpretation is in terms of sequential gambling with re-investment, a high S-value corresponding to a large gain in a gamble in which one would not expect to gain any money under the null hypothesis. S-values can generically handle optional continuation: if we reject when $S > b$ (say 20), we get a frequentist Type-I error guarantee of $1/b$ (say 0.05) that *remains valid under optional continuation*. For simple null hypotheses, every Bayes factor is also an S-value, different priors on the alternative corresponding to different definitions of the S-value. However, if the null is *composite* then the Bayes factor is usually not an S-value and indeed violates Type-I error guarantees when used in a frequentist test.

Here we provide a generic solution to this issue — we show that, for arbitrary composite nulls, there exist special priors under which Bayes factors become S-values. In general, these priors are unlike any of the priors usually encountered in Bayesian practice; however, for the special case where all parameters in the null satisfy a group invariance, using the improper right Haar priors one does get an S-value. Remarkably, **?** Bayesian t-test, which uses the right Haar on σ and a Cauchy on μ/σ thus gives an S-value and can handle optional continuation; however, we show that there exists an alternative prior on μ/σ with which we can get substantially higher frequentist power, while still handling optional continuation.

We work out the case of composite null hypotheses, which allows us to formulate safe, nonasymptotic versions of the most popular tests such as the t-test and the χ^2 square tests. Safe tests for composite $\mathscr{H}_0$ are not always Bayesian, but rather based on the 'reverse information projection', an elegant concept with roots in information theory rather than statistics.

Log-Optimal S-Values We now define S-values, and describe a powerful procedure for obtaining them in general. Let $\mathscr{H}_0$ be a composite null model, represented as a set of probability distributions on a common space of outcomes $\mathscr{X}$. We call a non-negative variable $S : \mathscr{X} \to \mathbb{R}_+$ an *S-value* if $\mathbb{E}_P[S] \leq 1$ for every $P \in \mathscr{H}_0$. S-values are *safe* in the following sense: for any significance level α, rejecting $\mathscr{H}_0$ when $S \geq 1/\alpha$ guarantees a Type-I error bounded by α per Markov's inequality. Yet one may wonder whether there are any interesting and useful S-values beyond the trivial constant $S \equiv 1$. Ideally, we would like high $S \gg 1$ whenever $\mathscr{H}_0$ is false. In the following, we propose an approach to formalise this. First, we specify an "alternative" by picking a single auxiliary distribution Q that represents what we hope will happen whenever $\mathscr{H}_0$ does not obtain. For multiple reasons, we champion the S-value that maximises the logarithmic growth under Q:

$$S^* \;=\; \underset{\substack{S:\mathscr{X}\to\mathbb{R}_+ \\ \forall P\in\mathscr{H}_0:\mathbb{E}_P[S]\leq 1}}{\operatorname{argmax}} \quad \mathbb{E}_Q\left[\log S\right].$$

To provide more insight into the makeup of such S^*, we provide the following dual representation. For any prior probability distribution w on $\mathscr{H}_0$, let us denote the Bayesian mixture by $P_w(x) = \int_{\mathscr{H}_0} P(x)w(P)\mathrm{d}P$. Then S^* equals the Bayes factor

$$S^*(x) \;=\; \frac{Q(x)}{P_{w^*}(x)} \qquad \text{where} \qquad w^* \;=\; \underset{\text{prior } w \text{ on } \mathscr{H}_0}{\operatorname{argmin}} \; D_{KL}(Q\|P_w).$$

At the poster we present more interpretation, generalisations and computation for several practical cases of interest.

© P. Grünwald, R. de Heide & W.M. Koolen.