E-values for anytime-valid inference with exponential families

Hao, Y.

Summary

This dissertation primarily focuses on statistical hypothesis testing, a critical area of study with widespread importance across various academic disciplines and industries. An example that highlights the importance of hypothesis testing comes from the field of medicine, specifically in drug development. Suppose researchers are testing a new drug designed to lower blood pressure. The null hypothesis might state that the new drug has no effect on blood pressure, while the alternative hypothesis suggests that the drug does reduce blood pressure. The researchers conduct a clinical trial, giving one group the new drug and another group a placebo. After collecting the data, the method of *statistical hypothesis testing* allows them to analyze whether the observed reduction in blood pressure is significant or could have occurred by chance. If the test provides strong evidence against the null hypothesis, the researchers can confidently conclude that the drug is effective. Without statistical hypothesis testing, they cannot rigorously assess whether the drug truly works or if the results are just random variations. By providing a framework to control for errors (like false positives), hypothesis testing ensures that the drug is only approved if there is strong statistical evidence of its efficacy, which is critical for public safety and the advancement of medical science.

However, most classical hypothesis testing methods require researchers to collect a fixed sample size in advance before conducting the test. Once the data from the predetermined sample size is collected, the test is performed, and conclusions are drawn about whether to reject or fail to reject the null hypothesis.

The fixed-sample approach has several limitations:

1. **Pre-determined sample size**: Researchers must decide in advance how many data points to collect, which may lead to underpowered or overpowered studies if the chosen sample size is not optimal.

2. **No intermediate analysis**: In classical hypothesis testing, researchers are usually not allowed to look at the data as it comes in (to prevent biased decisions) and

must wait until the full dataset is available.

3. **Inflexibility**: If unexpected results occur or if the sample size turns out to be inadequate after data collection, researchers cannot easily adjust the sample size without risking inflating the Type I error (false positive rate).

Despite these limitations, the fixed-sample approach has been the foundation of statistical testing for many decades and remains widely used. However, newer methods, such as anytime-valid tests (e.g., e-value- and e-process-based methods), offer more flexibility, allowing researchers to evaluate the evidence continuously as data is collected, without needing a pre-specified sample size. With e-values, one does not even need to determine the rules for stopping the experiment before it starts.

This dissertation primarily explores e-values and e-processes within the context of exponential families.

Chapter 2 addresses the problem of determining whether a sample conforms to a specific exponential family model, essentially testing whether a model is well-specified. For example, we may want to assess whether a sample follows a Gaussian distribution. In this case, the null hypothesis includes the entire set of Gaussian distributions, making it composite. The objective is to reject the null hypothesis if the sample deviates from a Gaussian distribution. A key focus is on the GRO (Growth-Rate Optimal) e-variable, which typically corresponds to a specific Bayes factor and has the highest e-power (i.e., the ability to detect alternatives) among all e-variables. However, finding the prior for the GRO e-variable can be computationally expensive. This chapter demonstrates that in certain scenarios, termed "simple cases," the GRO e-variable simplifies to a simple-vs.-simple likelihood ratio, and it provides several equivalent conditions under which such a likelihood ratio exists for exponential family null hypotheses.

Chapter 3 extends the work of Chapter 2 by offering more general theoretical results for several e-variables in the context of testing model specification, covering both simple and composite hypotheses. It is shown that in the "anti-simple case" (the opposite of the simple case), the conditional e-value achieves asymptotically highest e-power. This is particularly valuable because, while the GRO e-variable is hard to compute in such cases, the conditional e-variable is computationally more straightforward. Additionally, this chapter introduces and compares various types of e-values, including the GRO e-variable, the conditional e-variable, the UI (Universal Inference) e-variable, and sequential e-variables, with a detailed analysis of their e-power. Some of these e-values also lead to the development of anytime-valid tests, known as e-processes.

Chapter 4 focuses on developing methods for testing whether k groups of samples are distributed according to the same element of an exponential family, using e-values.

For instance, in the case of two-sample tests ($k = 2$), these methods can be applied to assess whether a new medical treatment is effective for a particular disease by comparing the distribution of outcomes between the treatment and control groups. We introduce four types of e-variables for the k-sample test: the GRO e-variable, a conditional e-variable, a mixture e-variable, and a pseudo-e-variable. These e-variables are compared in terms of their growth rates under alternative hypotheses, where each group has a different, but fixed, distribution from the same exponential family. The paper provides theoretical results showing that, under small effect sizes, the e-variables behave similarly. It also identifies cases where one e-variable simplifies to the GRO e-variable, reducing computational complexity. In more complex settings, algorithms for estimating the reverse information projection are suggested.

Chapter 5 focuses on *growth-rate optimal in the worst-case* (GROW) e-variables. We analyze the application of GROW e-variables within a hypothesis testing framework for multivariate distributions. In this framework, the null hypothesis distribution P_0 has a mean of zero, and various alternative hypotheses $\mathcal{H}_1$ are defined by different sets of means. Interestingly, we show that the GROW e-variable connects to a new concentration inequality we call the *Csiszár-Sanov-Chernoff* (CSC) bound. This CSC inequality extends earlier work to handle multivariate cases with either a convex or bounded alternative parameter region around 0. Such an inequality is likely to be valuable in practical applications, such as online learning, and especially in the field of *bandit algorithms*.

In summary, this dissertation expands e-variable theory within exponential families by developing and analyzing several e-variables, providing practical insights for situations where traditional p-value-based testing may fall short.