E-values for anytime-valid inference with exponential families

Hao, Y.

Chapter 6

Discussion

In this dissertation, we presented several key mathematical results related to e-values and e-processes within exponential families. Specifically, we demonstrated the theoretical foundations for anytime-valid testing with well-specified exponential family models and developed four types of e-variables for k-sample tests, comparing their respective e-powers. We also analyzed the GROW e-variable under a composite $\mathcal{H}_1$ and a simple $\mathcal{H}_0$, introducing a novel multivariate concentration inequality. In this discussion, we summarize the main contributions of the dissertation and highlight potential directions for future research.

6.1 Well-specified model tests

We can interpret Chapters 2 and 3, as providing a comprehensive examination of testing whether an exponential family model is well-specified. This involves determining if the data fit the hypothesized exponential family distribution under a specific null hypothesis $\mathcal{H}_0$.

Grünwald et al. [42] showed that the growth-rate-optimal (GRO) e-variable can be obtained via a specific Bayes factor between $\mathcal{H}_1$ and $\mathcal{H}_0$. When both $\mathcal{H}_0$ and $\mathcal{H}_1$ are simple hypotheses, the Bayes factor reduces to a likelihood ratio, often referred to as a "simple e-variable." However, when $\mathcal{H}_0$ is composite, the GRO e-variable does not always correspond to a simple e-variable, though it does in some cases.

In Chapter 2, we presented a theorem, under certain regularity conditions, providing a general sufficient condition for the existence of a simple e-variable when testing a simple alternative against a composite regular exponential family null. This condition can be expressed as "$\Sigma_p(\boldsymbol{\mu}) - \Sigma_q(\boldsymbol{\mu})$ is positive semidefinite for all $\boldsymbol{\mu} \in \mathtt{M}_q$." We also explored the possibility of constructing GRO or close-to-GRO e-variables when this condition does not hold. We found that for some $\boldsymbol{\mu}$ within a specific parameter range, $q_{\boldsymbol{\mu}}/p_{\boldsymbol{\mu}}$ still provides a global simple e-variable; for others, it provides a local but not global e-variable, and for some, it does not provide an e-variable at all. An interesting direction for future work involves extending these results to *curved exponential families* [33].

While we have no general results in this area yet, the work of Liang [62] suggests that this may be feasible. Liang's variation of the Cochran-Mantel-Haenszel test involves a null hypothesis that can be reframed in terms of a curved exponential family, where a local e-variable exists by considering the second derivative of a specific function. This local e-variable is shown to be global by a different method than what we used in our construction, suggesting the potential for unifying these approaches.

In Chapter 3, we investigated the 'opposite' scenario, where "$\Sigma_p(\boldsymbol{\mu}) - \Sigma_q(\boldsymbol{\mu})$ is negative semidefinite for all $\boldsymbol{\mu} \in \mathtt{M}_q$". In addition, we mainly studied various types of e-variables and e-processes ($S_{\text{RIP}}, S_{\text{UI}}, S_{\text{COND}}, S_{\text{SEQ-RIP}}$) for multivariate exponential family null hypotheses and compared their e-power for i.i.d. data. In this context, we observed that in certain scenarios, the e-power of the "conditional" e-variable S_{COND} is asymptotically equal to the e-power of GRO e-variables in the "opposite" scenario. Additionally, we discovered an interesting phenomenon when considering composite alternative hypotheses, both in the Gaussian and general case, particularly regarding the relationship between conditional and RIPr e-variables, suggesting a near "approximate optional stopping" result.

Future work We also highlighted two e-variables that have not been extensively analyzed: the *sequential conditional e-variable* and a certain *weighted average of e-variables*. The former is a sequentialized version of the conditional e-variable, used in classical sequential testing and applicable to k-sample tests with exponential families [44], which we study in the next chapter. The latter is a weighted average of RIPr e-variables across different priors on the alternative, which, though an e-variable, behaves differently from the e-variables we focused on in this thesis.

Finally, future research could focus on relaxing the assumption that the distribution of the sufficient statistics X must have exponentially small tails under the alternative hypothesis. This regularity condition underpins most of our results, but its relaxation could broaden the applicability of e-variables in exponential family settings.

6.2 k-sample tests

In Chapter 4, we introduced and analyzed four types of e-variables for testing whether k groups of data are distributed according to the same element of an exponential family. These e-variables include the GRO e-variable (S_{RIP}), a conditional e-variable (S_{COND}), a mixture e-variable (S_{MIX}), and a pseudo-e-variable (S_{PSEUDO}).

Our analysis focused on comparing the growth rates of these e-variables under a simple alternative where each of the k groups has a distinct, but fixed, distribution within the same exponential family. We demonstrated that for any pair of e-variables $S, S' \in \{S_{\text{RIP}}, S_{\text{COND}}, S_{\text{MIX}}, S_{\text{PSEUDO}}\}$, the difference in their expected log-growth rates is $O(\delta^4)$, where δ represents the ℓ_2 distance between the alternative distribution's parameters and the null parameter space. This result indicates that when the effect size is small, the performance of all four e-variables is remarkably similar. For more substantial effect sizes, S_{RIP} has the highest growth rate by definition, making it the most powerful e-variable. However, calculating S_{RIP} requires determining the reverse information projection of the alternative distribution onto the null, which is computationally challenging. We provided theoretical results showing that for certain exponential families, one of the following equalities holds: $S_{\text{PSEUDO}} = S_{\text{RIP}}$, $S_{\text{COND}} = S_{\text{RIP}}$, or $S_{\text{MIX}} = S_{\text{RIP}}$. These cases significantly reduce the computational complexity of identifying the most effective e-variable. In instances where such equalities do not hold, algorithms can approximate the reverse information projection, and we verified numerically that these approximations lead to near-optimal values for S_{RIP}. Despite this, the choice of using S_{COND} or S_{MIX} might still be preferable due to their computational efficiency. Our simulations revealed that the optimal choice between S_{COND} and S_{MIX} depends on the specific exponential family under consideration, and in some cases, no clear ordering between them emerges.

These results provide practical insights into the trade-offs between different e-variables in terms of their theoretical properties and computational demands, guiding the selection of appropriate e-variables in real-world applications.

6.3 GROW e-variables and concentration inequality

Chapter 5 demonstrated how GROW e-variables, relative to an alternative hypothesis $\mathcal{H}_1$ defined by a set of means $\mathtt{M}_1$, connect to a *Csiszár-Sanov-Chernoff* (CSC) probability bound on events determined by the same set $\mathtt{M}_1$. Initially, we focused on cases where $\mathtt{M}_1$ is convex, largely involving reformulations and reinterpretations of known results. Subsequently, we developed results for nonconvex, surrounding $\mathtt{M}_1$, showing that both GROW and a form of *relative* GROW, based on individual-sequence regret, relate to a modified CSC theorem. For cases where $\mathtt{M}_1^{\text{C}}$ is defined as a fixed-radius KL ball for sample size 1, we also derived results that hold as the actual sample size n increases.

To our knowledge, the CSC bounds we derived for surrounding $\mathtt{M}_1$ characterized by KL balls are the best available for this context. It is, however, interesting to consider a different approach: for sample size n, using a KL ball with a radius, in terms of the Euclidean distance in parameter space, that is decreasing as $O(\frac{1}{n})$ or $O(\frac{f(n)}{n})$ with $f(n)$ with very slowly increasing. Firstly, we consider the case that the KL ball is decreasing as $O(\frac{1}{n})$. Since the boundary $\text{BD}(\mathtt{M}_1^{\text{C}})$ now varies with n, our asymptotic results for the CSC bound derived earlier in Chapter 5 no longer apply in the same form. While the CSC theorem itself remains valid, evaluating the bound may present additional challenges.

Future work Now, let us explore the case that $f(n) = a \log(b + c \log n)$ for some suitable constants a, b, and c. Kaufmann and Koolen [51] provide an *anytime-valid bound* for this setting, where the bound's right-hand side also stabilizes to a nontrivial constant (i.e., less than 1) for all sufficiently large n.

It remains an open question whether our approach could yield similar bounds; addressing this is left as a potential direction for future work. Additionally, further analysis is needed to understand the relationship between anytime-valid bounds and those derived here. Although our bounds are related to e-values and thus indirectly connected to anytime-validity, they are not anytime-valid themselves.