Universiteit Leiden
The Netherlands

# E-values for anytime-valid inference with exponential families

Hao, Y.

**Citation**
Hao, Y. (2025, February 18). *E-values for anytime-valid inference with exponential families*. Retrieved from https://hdl.handle.net/1887/4195433

| | |
|---|---|
| Version: | Publisher's Version |
| License: | Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden |
| Downloaded from: | https://hdl.handle.net/1887/4195433 |

**Note:** To cite this publication please use the final published version (if applicable).

# 总　结

本论文主要研究统计假设检验，这是一个在多个学术领域和行业中具有广泛重要性的关键研究领域。一个能够突出假设检验重要性的例子来自医学领域，特别是在药物开发中。假设研究人员正在测试一种新型降血压药物。我们设定原假设是这种新药对血压没有影响，而备择假设则表明这种药物确实能降低血压。

为验证上述假设，研究人员进行了一项临床试验，其中一组人服用新药，另一组人服用安慰剂。数据收集完成后，假设检验方法可以帮助研究人员分析观察到的血压降低是否具有统计显著性，或者是否纯属偶然。如果检验结果提供了拒绝原假设的强有力证据，研究人员就可以有信心地得出结论：该药物是有效的。假设检验通过提供一个控制错误（例如假阳性）的方法框架，确保只有存在强有力统计证据表明药物有效的情况下才会批准该药物，这对于公共安全和医学科学的发展至关重要。

然而，大多数经典的假设检验方法通常要求研究人员在检验前预先确定一个固定的样本量。一旦收集了预先确定样本量的数据，就会执行检验，并得出是否拒绝原假设的结论。

固定样本量的方法存在以下几个限制：

1. **预先确定样本量**：研究人员必须提前决定收集多少数据点，如果所选样本量不够理想，可能会导致研究的统计功效不足或过高。

2. **无法进行中间分析**：在经典的假设检验中，研究人员通常不被允许在数据收集中途查看数据（以避免决策偏差），必须等到整个数据集可用后才能进行分析。

3. **缺乏灵活性**：如果出现意外结果，或者数据收集后样本量被证明不足，研究人员无法轻松调整样本量，否则可能增加一类错误（假阳性率）的风险。

尽管存在这些限制，固定样本量的方法仍然是统计检验的基础，已被广泛应用了数十年。然而，更新的方法（如序贯分析和任意时间有效检验，例如e-value 方法和e-process）提供了更大的灵活性，使研究人员能够在数据收集中连续评估证据，而无需预先指定样本量。e-value 甚至不需要在实验开始前确定停止检验的规则。

本论文主要探讨了在指数族分布下的e-value 和e-process 方法。

第 2章研究如何判断一个样本是否符合特定的指数族模型，本质上是检验模型是否被正确地指定。例如，我们可能希望评估一个样本是否符合高斯分布。在这种情况下，原假设包括所有高斯分布的集合，使其成为复合假设。目标是当样本偏离高斯分布时拒绝原假设。本章的重点是GRO（增长率最优）e-variable，其通常对应于特定的贝叶斯因子，并具有最高的e-power（即当备择假设为真时选择备择假设的能力）。然而，找到GRO e-variable的先验分布可能在计算上非常耗费资源。本章展示了在某些情况下（被称为“简单情况”），GRO e-variable简化为似然比，并提供了多种等价条件，用以描述在指数族原假设下，何时这种似然比存在。本章还创造了一种e-process 方法，证明了GRO e-variable是一种任意有效的检验方法。

第 3章在第 2章的基础上进一步扩展，为多个e-variables 在模型正确性检验中的应用提供了更一般的理论结果，涵盖了简单和复合假设。本章表明，在“非简单情况”（简单情况的对立情况）中，条件e-value 可以实现渐近意义上的最高e-power。这一点尤为重要，因为在这些情况下，虽然GRO e-variable 难以计算，但是条件e-variable 在计算上非常简便。此外，本章介绍并比较了多种类型的e-variables，包括GRO e-variable、条件e-variable、UI（一致推断）e-variable和序贯e-variable，并对其e-power 进行了详细分析，并且给出渐近表达式。我们还证明了其中一些e-variables是任意有效检验方法，即e-process。

第 4章聚焦于多组样本检验，使用e-value方法检验$k$ 组样本是否按照指数族中的相同元素分布。例如，在两样本检验($k = 2$) 时，运用e-value方法比较治疗组和对照组的结果分布，从而来评估一种新的医疗治疗对某种疾病是否有效。本章提出了四种$k$ 样本检验的e-variables：GRO e-variable、条件e-variable、混合e-variable和伪e-variable。在每组样本服从相同指数族中的不同但固定的分布时，我们对这些e-variables的e-power进行了比较。本章提供的理论结果表明，在不同的分布之间的差异较小时，这些e-variables表现类似。同时，在某些特定分布下，某种e-variable等价为GRO e-variable，从而降低计算复杂性。在更复杂的情况下，我们提供了一种算法用来估计GRO e-variable。

第 5章重点研究了最差情况下增长率最优（GROW）e-variable。我们分析了GROW e-variable在多元分布假设检验框架中的应用。在此框架中，原假设分布$P_0$ 的均值为零，而各种备择假设$\mathcal{H}_1$ 则由不同的均值集合定义。值得注意的是，我们发现GROW e-variable与一种新的集中不等式相关联，我们将其称为*Csiszár-Sanov-Chernoff (CSC)* 不等式。该CSC 不等式将先前的研究扩展到处理多元情况，并支持围绕0 的凸或有界的备择参数区域。这种不等式在实际应用中可能具有重要价值，例如在线学习领域，尤其是Bandit 算法。

总结：本论文通过创立和分析多种e-variables扩展了指数族中的e-value和e-process理论，并为传统$p$值检验可能失效的情境提供了可行的新一套方法论。