此页单面打印

涉密论文 □　公开论文 □

浙 江 大 学

本科生毕业论文

题目采用华文仿宋三号加粗打印，不得手写

题目　在线错误发现率控制的共形推断方法

姓名只有两个字的，中间空出一个汉字符，e.g. 张　三；
如果是留学生，请括号注明国别，e.g.（韩　国）

姓名与学号　顾润哲 3200103625

指导教师　孙文光

年级与专业　2020级统计学

所在学院　数学科学学院

“姓名、学号、指导教师、年级与专业、年月日”均用三号华文仿宋打印，不得手写，各栏目下划线需统一长度

递交日期　2024 年 5 月 27 日

请注意：自2026届起，《承诺书》页面已替换为 《附件1.浙江大学本科生毕业论文原创性声明及使用授权书》,请注意替换使用新版文档。

浙江大学本科生毕业论文（设计）承诺书

1. 本人郑重地承诺所呈交的毕业论文（设计），是在指导教师的指导下严格按照学校和学院有关规定完成的。

2. 本人在毕业论文（设计）中除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得浙江大学或其他教育机构的学位或证书而使用过的材料。

3. 与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

4. 本人承诺在毕业论文（设计）工作过程中没有伪造数据等行为。

5. 若在本毕业论文（设计）中有侵犯任何方面知识产权的行为，由本人承担相应的法律责任。

6. 本人完全了解浙江大学有权保留并向有关部门或机构送交本论文（设计）的复印件和磁盘，允许本论文（设计）被查阅和借阅。本人授权浙江大学可以将本论文（设计）的全部或部分内容编入有关数据库进行检索和传播，可以采用影印、缩印或扫描等复制手段保存、汇编本论文（设计）。

作者签名：顾润哲　　导师签名：

签字日期：2024 年 5 月 24 日　　签字日期 2024 年 5 月 24 日

正文前使用罗马数字编页码



此页单面打印

标题：仿宋，三号字体，加粗，居中对齐

致 谢

时间滴滴答，随着落笔至此，四年的本科故事也将在这个盛夏画上句号。在数院的这四年，怀疑纠结内耗失望时常侵扰我的内心，难题考试比赛未来时常牵绊我的思绪。所幸我走了下来，并没有辜负四年前对自己的期待。感谢那些刻在我心底里的名字，由衷地对你们的关心与帮助说一声感谢。

首先，我要感谢我的导师，浙大数据科学研究中心的孙文光老师。孙文光老师既是我统计学道路上的启蒙导师，又是引导我不断探索、不断前进的科研领袖。在完成毕业设计的过程中，尽管工作繁忙，孙老师还会每周抽出固定时间给我答疑解惑。仰之弥高，钻之弥坚，在此谨向孙老师致以最诚挚的感谢。

其次，我要感谢我的父母和家人，你们的支持与鼓励是我成长道路上最坚固的后盾。然后，我要感谢这四年来遇到的所有老师与朋友。从大一时莫群、韩刚老师引导我进入大学数学的世界，到大二时赵敏智、黄炜老师让我对统计学有了初步的入门，王成波老师让我见识到了微分方程的复杂与美丽，再到大三大四郭正初、刚博文和 Brad Rava 老师耐心解答我在学习和科研过程中遇到的疑惑，四年间遇到的每个老师的悉心付出，我都记在心里。还有与我同行的朋友们：在这里特别感谢石宇轩、秦桢杰、吴佳昱、崔栋禹和汪珂学长在学习过程中对我的指引与帮助，感谢身边的林昊润、黄奕铮、陈楚文、宋文巍、孙言笑和许亦琦同学与我共享的快乐时光。感谢大四期间参与我组织的两个讨论班的所有同学，正是你们的加入让我完成了大学阶段一直想做的事。感谢文化中国的所有朋友，你们带给我了数学之外的另一种生活。感谢数院学生会与数院足球队，让我在这四年有了一份归属感。感谢 SRTP 的队友和数模比赛的队友，我很享受我们一起齐心协力克服困难的过程。感谢在香港大学暑研过程中认识的所有朋友，这一个月让我对科研有了初步的探索，特别感谢朱灵挺学长以及夏天和李泊铮同学，带我走出了最迷茫的一段时光。还要感谢我的高中同学，那时纯粹的友谊一直延续至今。所有的你们共同构筑了我完整而充实的四年时光。

最后，要感谢我们组里的赵子楠、覃超、吴正楷、赵冠岚、何京城、孙寅瑞师兄和田烨、孙荣忆、王雪妮师姐，你们给了我生活上无微不至的照顾，让我感受到了和谐融洽的组内气氛，特别感谢姚泽宇师兄对我毕业设计中的理论的相关指导以及对我的论文提出的诸多修改建议。凡是过往，皆为序章。本科的终点也意味着未来科研学习的起点，希望我能永远保持那份热爱、那份勇气、那份坚持，在新的道路上继续书写自己的故事。

仿宋小四号或 12 磅，1.5 倍行距

此页单面打印

奇数页右侧页眉：论文题目

摘 要

考虑一个在线的假设检验问题：有一系列的可能是无穷批次的假设序列需要被决策，并且每一次的决策必须在下一批次的数据到来前做出。错误决策的比例需要在所有的拒绝时间点被控制。由于特殊的犯错限度以及未来数据的不可预知性，常规的大规模多重将设检验方法不再适用。进一步地，因为在检验过程中，总体错误决策额度容易耗竭，在线决策过程可能因此中止，无法对未来更多的假设进行决策。针对以上问题，这篇文章发展出了一套全新的在线虚假真实性指标规则（OPLIS）来控制在线的错误发现率（FDR）。该方法的关键是是使用了共形推断的框架来对决策过程中的正确拒绝量与错误拒绝量进行度量。OPLIS 方法使用了一种构造的数据集来帮助把握序列数据的结构性信息，同时使用一个镜像过程来学习拒绝的门槛。我们证明了 OPLIS 可以在有限样本的情况下控制在线的错误发现率。数值实验也证明了 OPLIS 方法的有效性与稳健性。

关键词：共形推断，在线错误发现率控制，镜像过程，序列化的[illegible]

摘要应说明研究目的、方法、结果和结论等，重点是结果和结论。不宜使用图、表、化学结构式、非公知公用的符号和术语。字数一般为 300-600 字以内，采用仿宋小四号或 12 磅，1.5 倍行距

“关键词”三个字：仿宋，小四号，加黑，左对齐；
关键词 3-8 个，仿宋，小四号，关键词之间可用顿号/逗号/分号隔开。

偶数页左侧页眉：浙江大学本科生毕业论文

此页单面打印

标题：Times New Roman 字体，三号，加粗，居中对齐

Abstract

Consider the online testing problem that a possibly infinite sequence of batches of hypotheses is tested before the next batch comes. The error rate is required to be controlled at all decision points. Conventional simultaneous testing rules are no longer applicable due to the more stringent error constraints and absence of future data. Moreover, the online decision-making process may come to a halt when the total error budget, or alpha-wealth, is exhausted. This work develops a new online pseudo local index of significance (OPLIS) rule for online false discovery rate (FDR) control. A key element in our proposal is a new alpha-investing algorithm that using the comformal inference framework to characterizes the gains and losses in sequential decision making. OPLIS uses a constructed dataset to capture time varying structures of the data stream and learns the decision threshold adaptively by a mirror process. We demonstrate that OPLIS offers finite-sample guarantees in online FDR control. Numerical results confirm the effectiveness and robustness of OPLIS.

Keywords: conformal inference, online FDR control, mirror process, s ntial Q value

关键词：Times New Roman 字体，小四号，关键词之间用逗号或分号隔开

英文摘要实词在 300 个左右，英文摘要应与中文摘要内容对应。正文用小四 Times New Roman 字体、1.5 倍行距

目录

第一部分　毕业论文

1　绪论 1
1.1 研究背景与目的 1
1.2 研究历史回顾 2
1.3 研究流程与方法 2
2　在线假设检验框架 3
2.1 问题框架 3
2.2 FDR 和在线假设检验方法 4
3　共形推断方法 6
3.1 模型自由的 FDR 控制方法 6
3.2 PLIS 方法 6
4　在线的 PLIS 方法 8
4.1 OPLIS 准则的描述 8
4.2 有效性的证明 11
4.3 对于“piggybacking”问题的改进 17
5　数值模拟结果 18
5.1 在线 FDR 与 MDR 的比较 18
5.2 OPLIS 方法关于信号强度的稳健性 22
5.3 OPLIS 方法关于信号稀疏度的稳健性 22
5.4 改进的 OPLIS vs 原始的 OPLIS 方法 25
6　后期目标 26
7　参考文献 28
本科生毕业论文（设计）任务书 29
本科生毕业论文（设计）考核 31

目录页使用罗马数字重新编页

第二部分　毕业论文开题报告

一、　文献综述 1
1　背景介绍 1
1.1 离线的假设检验与在线的假设检验 2
1.2 基于模型的统计推断和模型自由的统计推断 3
2　国内外研究现状 3
2.1 研究方向及进展 3
2.2 存在问题 5
3　研究展望 6
4　参考文献 7
二、　开题报告 9
1　问题提出的背景 9
1.1 背景介绍 9
1.2 本研究的意义和目的 10
2　项目的主要内容和技术路线 11
2.1 主要研究内容 11
2.2 技术路线 11
2.3 可行性分析 14
3　研究计划进度安排及预期目标 14
3.1 进度安排 14
3.2 预期目标 14
4　参考文献 17
三、　外文翻译 19
1　导言 19
2　问题的建立 22
3　广义的阿尔法投资框架（GAI） 23
四、　外文原文 25
毕业论文（设计）文献综述和开题报告考核 37

第一部分

毕业论文

1 绪论

一级标题，仿宋，三号，加粗

1.1 研究背景与目的

二级标题，仿宋，小三号，加粗

三级标题：1.1.1 四号，仿宋，加粗

在大数据时代，获取数据相比之前任何时期都要显得更加便利，但随之产生的一个问题是如何从海量数据和高维变量中获取我们所需的信息。这便是大规模假设检验所研究的重要课题。而在线的大规模假设检验问题是其中的一种特殊情况，它所考虑的场景是一种实时的监测过程，即决策者不能一下子获取所有的数据，也不清楚需要进行的决策的数量（很可能随着时间推移趋于无穷），但需要在每一个决策的时间点尽可能地保证决策的正确性。决策者所能依靠的仅仅是过去已经做出的决策和到当前时间点为止已经积累的数据。一旦决策者做出决定是否拒绝原假设，在未来他不再能对此决定进行修改。

在线的假设检验问题是大规模假设检验领域的前沿课题，因此它同样也在大规模假设检验的框架下，需要控制第一类错误与第二类错误。在实际的应用中人们往往更关心第一类错误（假阳性）的控制，因此在理论研究中对第一类错误有着多种的度量标准，比如家庭错误率（the familywise error rate ）和错误发现率（false discovery rate)。在实际的问题中往往需要控制错误发现率低于一个预先给定的数值 α。

正文：小四号或 12 磅仿宋，1.5 倍行距)

在线假设检验问题的应用非常广泛，其应用遍及医疗卫生、农业生产、交通物流、经济金融等领域。下面是两个简单的例子：

Example 1 在生物或医疗公司中，平台对比试验（*A/B Test*）作为一种流行的方法，经常用于检测药物是否对某种疾病产生作用。这种对比试验被多次执行，其中对疾病有作用的成分称为“信号”，而对疾病没有影响的成分称为“噪声”。决策者需要在每一次试验结束后判断出哪些数据源于“信号”，同时需要尽可能的保证正确性。

Example 2 在路面交通或是电子 *APP* 领域，实时的流量往往需要被监测。此时，类似于平台对比试验，流量发生变化的”异常点“为需要识别的”信号“，而流量正常情形为“噪声”。在这个例子中流量发生变化的“异常点”就是我们需要判别出的“信号”，而正常情形则是“噪声”。实时的异常点监测也是在线假设检验的重要问题。

在线的假设检验是大规模假设检验领域中比较前沿的课题。传统的大规模假设检验往往是离线的，这意味着决策者可以在做决策前预先知道需要检验的假设的数量以及用于决

我们选定了以下三种代表性的情况来反映数值模拟的效果：

- 强信号情形：我们确定了恒定的 $\pi_t = 0.2$ 以及 $\mu = 4$ 来模拟有较强信号时的情景。
- 弱信号情形：我们确定了恒定的 $\pi_t = 0.2$ 以及 $\mu = 2.5$ 来模拟有较弱信号时的情景。
- 变化稀疏度情景：我们确定恒定的 $\mu = 2.8$ 以及变化的 π_t 来模拟信号稀疏度变化的情景。其中 $\pi_t = 0.2$, for $t \in (1, 2000] \cup (4000, 6000] \cup (8000, 10000]$; $\pi_t = 0.3$, for $t \in (2000, 4000] \cup (6000, 8000]$

考虑到计算量，我们计算经验的在线错误发现率以及遗失发现率时使用错误发现比例和遗失发现比例的平均值，重复试验的量定为 100 次。为了展示在线的即时 FDR 和 MDR，我们每过一个批次就会展示对应的经验 FDR 和 MDR。我们使用 R 语言来实现数值模拟的结果，所有的结果可见图 5.1-图 5.3

在正文中要明示图号，如：“如图 3.1 所示”或“见图 3.1”。论文中不得出现复印的图、表、文字。

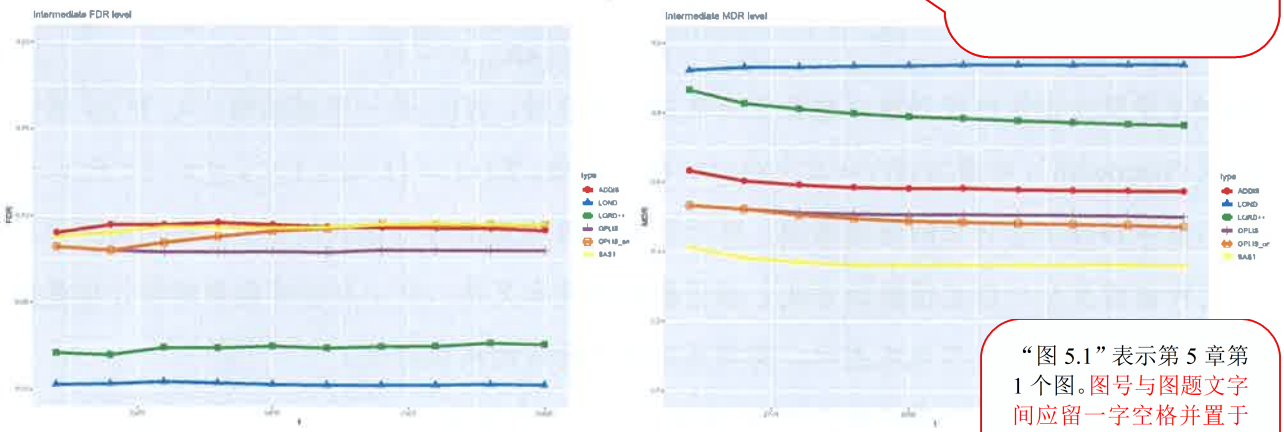


图 5.1 弱信号强度 $\mu = 2.5$ 时的 FDR 和 MDR

“图 5.1”表示第 5 章第 1 个图。图号与图题文字间应留一字空格并置于图正下方，图题用 5 号宋体加粗。

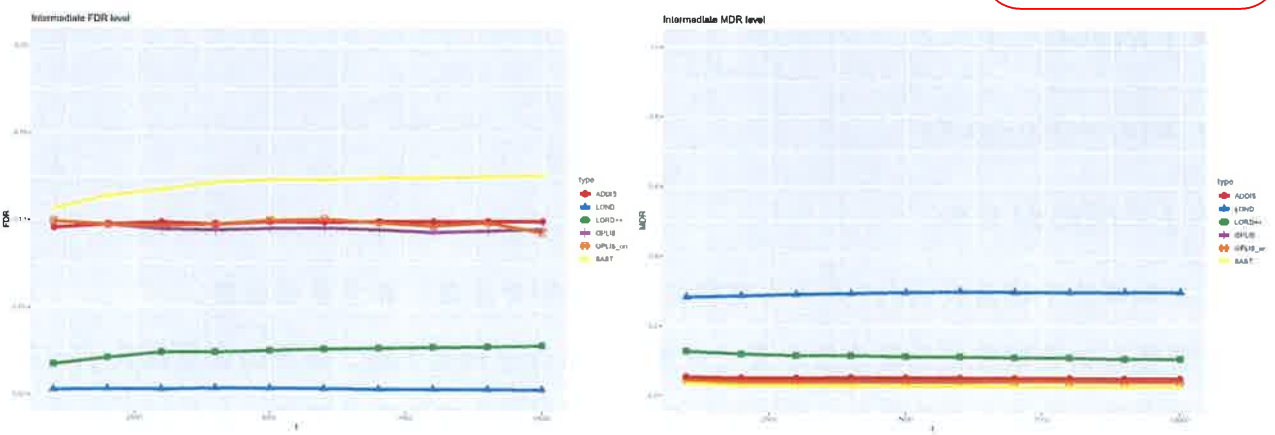


图 5.2 强信号强度 $\mu = 4$ 时的 FDR 和 MDR

的拒绝效率。这是因为我们的 OPLIS 及其改进方法也能够充分利用到数据间的结构信息。由于我们在模拟中设定的数据分布较为简单，在实际生活中，遇到较为复杂的数据，OPLIS 准则有着更好的适用性和准确性。

3. 最后值得注意的是，我们的 OPLIS 及其改进方法是一种基于数据本身而具有一定模型自由度的方法。在与机器学习联系紧密的领域，我们往往不知道数据所服从的分布，因此我们可能得不到对应的 p 值，同时基于核密度的密度函数估计也可能有比较大的误差，这意味着传统的在线假设检验的方法可能都不再适用，而我们的 OPLIS 及其改进方法还有适用的空间，可以更好地适用现代化的数据分析与统计推断环境。

5.2 OPLIS 方法关于信号强度的稳健性

在这一节中我们将展示 OPLIS 方法对于不同信号强度所得到的决策结果的稳定性，我们希望 OPLIS 针对不同的信号强度 μ，都能有很好的在线 FDR 的控制。

我们分别选择了五个具有代表性的信号强度 $\mu = 2.5, 2.8, 3, 3.5, 4$，确定恒定的信号稀疏度 $\pi_t = 0.2$，以 1000 为时间间隔计算在线的 FDR 水平得到下面的五张表格（表 5.1）。

从表的结果中我们可以看到，在规定控制 FDR 水平在 $\alpha = 0.1$ 时，针对不同的信号强度，我们的 OPLIS 方法都能够给出较好的决策手段，使得各个拒绝时间段的 FDR 都很接近预先给定的水平。这表明我们的方法具有对信号的稳定性，针对不同的信号都能够给出较为稳健的决策。这意味着针对不同场景的信号识别或者异常监测问题，我们的方法都可以不加修正的给出有效的错误发现率控制。

5.3 OPLIS 方法关于信号稀疏度的稳健性

在这一节中我们将展示 OPLIS 方法对于不同信号稀疏度所得到的决策结果的稳定性，我们希望 OPLIS 针对不同的信号稀疏度 π，都能有很好的在线 FDR 的控制。

在正文中引用表中资料需标明表号

我们分别选择了五个具有代表性的信号强度 $\pi = 0.1, 0.2, 0.3, 0.4, 0.5$，确定性为 [illegible] 强度 $\mu_t = 2.8$，以 1000 为时间间隔计算在线的 FDR 水平得到下面的五张表格（表 5.2）。

根据表中的结果，我们可以看到我们的 OPLIS 准则在不同的信号稀疏度的情况下都能够很好的控制在线各时间点的 FDR，当信号较少时，我们的决策的 FDR 往往与给定的

“表 5.1”表示第 5 章的第 1 表。表格应有标题，与表号之间空 1 个汉字，表头置于表的上方居中，5 号，宋体，加粗。

表 5.1 OPLIS 准则关于不同 μ 得到的在线 FDR 水平

$\mu = 2.5$

t	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000
FDR	0.0820	0.0800	0.0790	0.0789	0.0794	0.0791	0.0804	0.0802	0.0803	0.0803

$\mu = 2.8$

t	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000
FDR	0.0830	0.0839	0.0842	0.0864	0.0871	0.0868	0.0852	0.0851	0.0846	0.0847

$\mu = 3$

t	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000
FDR	0.0911	0.0899	0.0873	0.0873	0.0858	0.0880	0.0889	0.0898	0.0902	0.0893

$\mu = 3.5$

t	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000
FDR	0.0931	0.0919	0.0902	0.0860	0.0886	0.0913	0.0909	0.0904	0.0916	0.0922

$\mu = 4$

t	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000
FDR	0.0993	0.0972	0.0947	0.0941	0.0951	0.0956	0.0946	0.0931	0.0939	0.0952

表中的内容和项目用 5 号宋体，单倍行间距。图表与下文空一行。在正文中引用要明示表号，表示方式与图号的表示方式相同。

公式：Times New Roman，小四号，倾斜，加粗，编号使用应分章编号，用（2-1），（2-2）…置后标注公式序列号，（2-1）表示第 2 章出现的第 1 个公式。

$$EQI=\sum_{i=1}^{n} Wj * rij \quad (2\text{-}1)$$

$$EHI=L_1 \times ESI + L_2 \times EQI \quad (2\text{-}2)$$

$$EE/EHI=\beta_0+\beta_1 PCG+\beta_2 RGP+\cdots+\beta_i X_i+$$
$$\cdots+\beta_9 ICWUR+\beta_{10} ECPG+\beta_{11} WCPG+\varepsilon_i \quad (2\text{-}3)$$

表 5.2 OPLIS 准则关于不同 π_t 得到的在线 FDR 水平

$\pi_t = 0.1$

t	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000
FDR	0.0966	0.0910	0.0934	0.0982	0.0983	0.0972	0.0971	0.0980	0.0983	0.0988

$\pi_t = 0.2$

t	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000
FDR	0.0830	0.0839	0.0842	0.0864	0.0871	0.0868	0.0852	0.0851	0.0846	0.0847

$\pi_t = 0.3$

t	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000
FDR	0.0882	0.0892	0.0897	0.0893	0.0890	0.0865	0.0862	0.0854	0.0856	0.0853

$\pi_t = 0.4$

t	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000
FDR	0.0820	0.0822	0.0814	0.0814	0.0827	0.0827	0.0833	0.0844	0.0850	0.0842

$\pi_t = 0.5$

t	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000
FDR	0.0676	0.0663	0.0668	0.0665	0.0671	0.0676	0.0669	0.0667	0.0667	0.0668

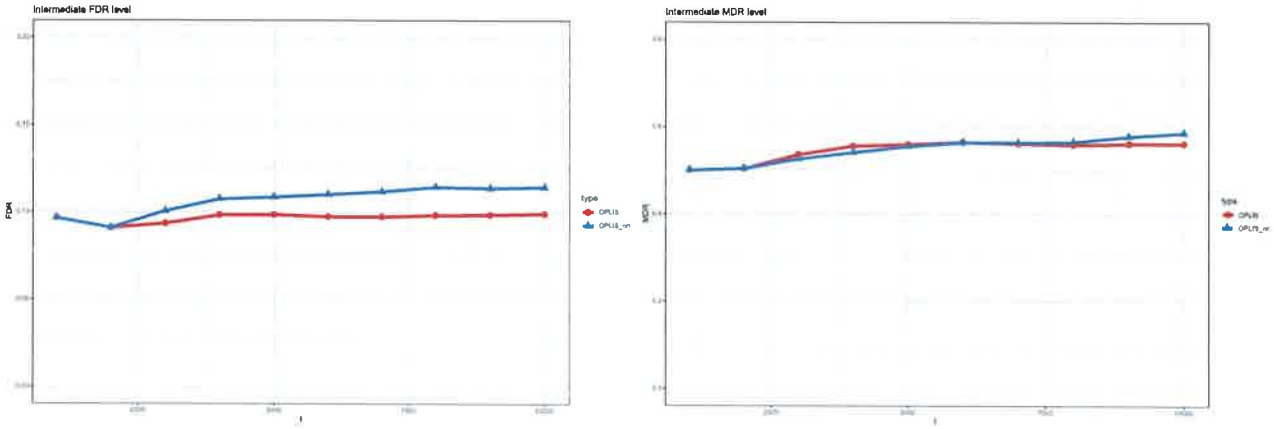


图 5.5 信号的稀疏度在 $\pi_t = 0.1$ 时的 FDR 与 MDR

绝任何假设，这导致了 FDR 一直无法因为后续的正确拒绝而得到下降。而改进后的 OPLIS 准则引入了离线的门槛，这使得得分函数较高的数据点不会被拒绝，以牺牲一些拒绝效率的代价抑制了“piggybacking”现象的发生，从模拟的结果中可以看到 FDR 可以始终被很好的控制。因此改进的 OPLIS 准则可以极大程度上减少“piggybacking”现象对决策的影响，保证了决策的有效性和稳健性。

6 后期目标

在本文中，我们提出了一种理想情形下的共形推断方法来控制在线的错误发生率，我们证明了这种方法能够有效的在有限样本的情况下控制任何时刻的 FDR，这为了已有的基于阿尔法投资的在线假设检验方法提供了一种全新的视角。我们的模型不需要依赖具体的参数模型，不需要对数据的分布有着很准确的估计，我们的基于“镜像过程”的共形 Q 值可以看成用于假设检验的阿尔法财富。它同样具有灵活性，当过去的错误拒绝数量较少时，可以有更大的限度拒绝新的假设。从这一点来看，我们的 OPLIS 准则可以看成是 SAST 准则的共形推断版本。而从另一个方面看，我们将 PLIS 准则推广到了可序列化进行的在线版本，进一步扩大了其的适用性。但由于我们目前考虑的模型比较的理想，我们可以在未来做出以下方面的推广：

首先，我们想要改进的是每次做出决策后能够知道过去的错误发现（错误拒绝）的个数这一个较为理想的模型设定，引入这个模型的原因是因为我暂时还不知道处理误差的累积的情况，一旦过去的决策错误累积起来将对未来的决策产生巨大的影响。因此我首先考

虑的是能结合过去假设的对错情况下给出对未来的最优的决策。而在这一问题得到解决之后，可以考虑对过去的错误拒绝情况做出保守估计，将这一估计量引入到我们共形 Q 值的构造中来代替这里已知的常数 a_n，为了保证错误拒绝的估计始终具有保守性，可能需要引入鞅论等工具进行证明。

其次，我们的 OPLIS 准则还是在共形推断的框架之下，这意味着为了确定精确的拒绝门槛，我们需要足够量的测试集。这也是为什么我们引入批次（batch）的原因。在未来我们会考虑进一步弱化这一条件，比如参考 FASI 方法[16]的思想，将测试集与足够的校准集结合起来构造共形 Q 值。我们最终希望能够将 OPLIS 方法拓展到一般的在线假设场景中。

最后我们希望我们的方法可以拓展到更为一般的数据中，我们希望可以开发出新的改进版本应对在线数据的分布的转移，或是带有先验信息的数据的选择性分类等问题。

7 参考文献

[1] FOST… …of Expected False Discoveries [J]. Jo… …gy, 2008, 70(2): 429-444.

[2] AHA… …ity Results and Application to Public… …tatistical Methodology, 2013, 76(4):…

[3] RAM… …lse discovery rate with decay-ing m… …lvances in Neural Information Proces… …dings.neurips.cc/paper_files/paper/2…

[4] BOWE… …ing for Online False Discovery Rate Control[J/OL]. Journal of the American Statistical Association, 2023, 118(541): 732-745. eprint: https://doi.org/10.1080/01621459.2021.1955688. https://doi.org/10.1080/01621459.2021.1955688. DOI: 10.1080/01621459.2021.1955688.

[5] ZHAO Z, SUN W. False Discovery Rate Control For Structured Multiple Testing: Asymmetric Rules And Conformal Q-values[Z]. 2023. arXiv: 2311.15322 [stat.ME].

[6] ZRNIC T, JIANG D, RAMDAS A, et al. The Power of Batching in Multiple Hypothesis Testing[C/OL]//CHIAPPA S, CALANDRA R. Proceedings of Machine Learning Research: Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics: vol. 108. PMLR, 2020: 3806-3815. https://proceedings.mlr.press/v108/zrnic20a.html.

[7] EFRON B, TIBSHIRANI R, STOREY J D, et al. Empirical Bayes Analysis of a Microarray Experiment [J/OL]. Journal of the American Statistical Association, 2001, 96: 1151-1160. https://api.semanticscholar.org/CorpusID:9076863.

[8] JAVANMARD A, MONTANARI A. Online rules for control of false discovery rate and false discovery exceedance[J/OL]. The Annals of Statistics, 2018, 46(2): 526-554. https://doi.org/10.1214/17-BAOS1559. DOI: 10.1214/17-BAOS1559.

[9] … …m for Online Control … Learning Research: …R, 2018: 4286-4294.

[10] … …ntrol with conserva-… Advances in Neural …eedings.neurips.cc/p…

[11] … Multiple Haystacks … eprint: https://doi.o… …0.1198/jasa.2009.tm08415.

[12] VOVK V, GAMMERMAN A, SHAFER G. Algorithmic Learning in a Random World[C/OL]//. 2005. https://api.semanticscholar.org/CorpusID:118783209.

[13] BATES S, CANDÈS E, LEI L, et al. Testing for outliers with conformal p-values[J/OL]. The Annals of Statistics, 2023, 51(1): 149-178. https://doi.org/10.1214/22-BAOS2244. DOI: 10.1214/22-BAOS2244.

[14] … R F, CANDÈS E J, SAMWORTH R J. Robust inference with knockoffs[J/OL]. The Annals of … …ttps://doi.org/10.1214/19-BAOS1852. DOI: 10.1214/19-BAOS1852.

[15] … …olling the false discovery rate via knockoffs[J/OL]. Annals of Statis-… …i.semanticscholar.org/CorpusID:8467059.

[16] … al. A Burden Shared is a Burden Halved: A Fairness-Adjusted Ap-… …l Report, 2023. Preprint: 2110.05720 (stat.ME).

请根据 浙江大学本科生毕业论文（设计） 编写规则（2018）要求，对参考文献引用格式进行重点检查。文献查阅 8 篇以上（外文不少于 3 篇）

用“[J]”等在文章标题后标识各参考文献的类型。专著“[M]”，论文集“[C]”，期刊文章“[J]”，学位论文“[D]”，报告“[R]”，标准“[S]”；网上资料，如引用的是数据库，用“[DB/OL]”表示；如引用的是电子文献，用“[EB/OL]”表示。如直接引用网页，用“[Z]”表示。

参考文献序号采用“[1]、[2]、[3]…”编排

此页单面打印

本科生毕业论文（设计）任务书

一、题目：在线错误发现率控制的共形推断方

题目的内容与格式必须与封面题目完全一致！

二、指导教师对毕业论文（设计）的进度安排及任务要求：

（1）11 月 12 日——11 月 28 日：导师下达任务书，对进度、文献和开题提出要求。

（2）11 月 29 日——1 月 9 日：学生确认任务书，对确定的课题搜集相关文献资料，了解问题的背景、应用、研究历史与现状。从中确定论文最终题目。

（3）1 月 10 日——3 月 3 日：对确定的题目进一步展开学习，包括所必需的基础知识及近几年涉及此问题的文章。初步撰写并完成开题报告、文献综述，并提交导师审核。

（4）3 月 3 日——3 月 15 日：组织开题，每位学生准备 10 分钟左右的答辩。

（5）3 月 16 日——3 月 20 日：根据答辩小组意见进行修改，将定稿的开题报告、文献综述、外文翻译稿上传至教务系统。

（6）3 月 21 日——3 月 28 日：做中期检查报

（7）3 月 29 日——4 月 29 日：完成论文初稿。进 修改并最终完成，向导师提交论文终稿。

指导教师前往本科教学管理服务平台“毕业论文”—“任务书上传”—“查看”。复制至此模版中，适当编辑调整至一页，打印后由指导教师签字。

（8）4 月 30 日——5 月 12 日：导师评阅 表”及符合规范格式要求的送审论文。

（9）5 月 12 日——5 月 15 日：毕业论文专家评阅。

（10）5 月 15 日——5 月 18 日：评阅结果有修改意见的，根据评阅意见对论文进行修改。

（11）5 月 19 日——5 月 31 日：组织毕业论文答辩。提交最终版毕业论文，并将论文上传至教务系统。起讫日期 2023 年 11 月 1 日 至 2024 年 6 月 2 日

2026年10月10日 至 2027年5月28日

师（签名） 职称 教授

三、系或研究所审核意见：

同意该计划！

签名处留空白

负责人（签名）

2023 年 11 月 1 日

2026年11月11日

此页单面打印

本科生毕业论文（设计）考核

一、指导教师对毕业论文（设计）的评语：

大范围多重假设检验方法为[illegible]高维的数据中迅速准确的找到重要变量提供了一个高效有力的工具[illegible]态数据，动态数[illegible]题。

评语不少于 200 个字，建议评语分四段，分别对应文献综述、外文翻译、开题报告和论文正文四块。例如“文献综述材料是否准备充分？观点阐述是否明确？是否具有代表性？是否具有自己的评价？外文翻译是否与论文方向有关？翻译是否准确？开题报告思路是否清晰？框架布局是否合理？论文时间安排是否合理？论文研究内容是否具有理论和现实意义？研究方法是否切实可行？论文创新点在哪里？论文质量与理论水平如何？论文写作是否规范？”等等。

针对大数据[illegible]大数据流中的序列结构信息，提出了高效的错误率动态控制策略，显著提高了已有在线分析工具的功效。作者严格证明了所提出方法可以控制在线的错误发现率，并通过数值实验验证了方法的有效性与稳健性。论文内容充实，有一定深度，反映出作者的扎实基础和出色的科研能力。论文设计结构合理，逻辑性强，论述层次清晰结论合理。达到了任务书规定的要求。

指导教师（签名）

2024 年 5 月 22 日

早于答辩日期

二、答辩小组对毕业论文（设计）的答辩评语及总评成绩：

该同学在答辩期间阐述了选题的主要动机和目的，展示了基本方法和和主要结果，答辩时陈述准确，思路清晰，能较好地回答答辩老师所提出的主要问题。答辩委员会认为该论文选题恰当，所获结果具有较好的实际应用参考价值，结构完整清楚，写作规范，图表制作规范。文献综述和引用符合要求。答辩小组[illegible]

答辩评语不少于 200 个字，对论文选题、开题报告、文献综述、外文翻译和论文研究内容与方法、创新点、论文质量与理论水平、论文写作规范以及学生答辩表现和质量等方面分别加以评价。

成绩比例	文献综述（10%）	开题报告（15%）	外文翻译（5%）	毕业论文质量及答辩（70%）	总评成绩
分值	10	14	5	67	96

负责人（签名）

2024 年 5 月 24 日

第二部分

毕业论文开题报告

浙 江 大 學

本 科 生 毕 业 论 文

文献综述和开题报告



编排方法同论文封面的排版

姓名与学号	顾润哲 3200103625
指导教师	孙文光
年级与专业	2020级统计学
所在学院	数学科学学院

题目的内容与格式必须与封面题目完全一致！

此页单面打印

一、题目：在线错误发现率控制的共形推断方法

二、指导教师对文献综述、开题报告、外文翻译的具体要求：

文献综述要求：根据阅读的国内[illegible]文献综述报告一要求根据题展开，文献综述内容要切题，包括

指导教师对文献综述和开题报告的具体内容要求各不少于 150 个字。

（1）简述与在线假设检验[illegible]展，掌握当前该领域的研究前沿，分析你毕业论文研究的内容与这些论文的差异和相关。

（2）分析 SAST 和 comformal Q-value 的论文所采用的研究方法，简述这些方法的优劣和你的思考。

（3）简述各参考文献的创新性、存在的问题或未能解决的问题。

（4）要求翻译其中的一篇外文文献，结构完整，语句通顺。

开题报告要求：

（1）分析 SAST 准则和 comformal inference 的统计方法的意义（包括理论意义和实际意义，并分析课题与本专业的关系）

（2）根据文献综述分析课题的研究背景（即要解决什么问题？这些问题在其他文献中有没有讨论过？本文所讨论问题的角度与已有参考文献中所涉及的问题的差异，课题的主要创新点是什么？

（3）选题的可行性分析（一般从数据的可获得性、研究技术的可行性等方面去说明）。

（4）主要研究内容（这部分要展开写，主要包括[illegible]容内容、实证分析内容和调研内容等）。

（5）根据研究的内容写出具体的实施计划[illegible]

（6）明确论文最后预期结果。

指导教师前往本科教学管理服务平台“毕业论文”—“任务书上传”—“查看”。复制至此模版中，适当编辑调整至一页，打印后由指导教师签字。

指导教师（签名）

2023 年 11 月 1 日

目录

一、 文献综述 1
1 背景介绍 1
1.1 离线的假设检验与在线的假设检验 2
1.2 基于模型的统计推断和模型自由的统计推断 3
2 国内外研究现状 3
2.1 研究方向及进展 3
2.2 存在问题 5
3 研究展望 6
4 参考文献 7
二、 开题报告 9
1 问题提出的背景 9
1.1 背景介绍 9
1.2 本研究的意义和目的 10
2 项目的主要内容和技术路线 11
2.1 主要研究内容 11
2.2 技术路线 11
2.3 可行性分析 14
3 研究计划进度安排及预期目标 14
3.1 进度安排 14
3.2 预期目标 14
4 参考文献 17
三、 外文翻译 19
1 导言 19
2 问题的建立 22
3 广义的阿尔法投资框架（GAI） 23
四、 外文原文 25

毕业论文（设计）文献综述和开题报告考核 37

文献综述，字数 3000 字以上。

一、文献综述

1 背景介绍

在线假设检验问题是统计推断领域的重要研究问题，它是大规模假设检验问题中的重要分支。在如今这样一个大数据的时代，大规模假设检验可以说是无处不在，它在医疗卫生、农业生产、交通物流、经济金融等领域都有着极其重要的应用。在在线假设检验问题的基本设定中，假设流以在线的形式一个个地出现，在每一步的决策时，决策者必须在不知道假设的个数（很有可能是无穷）和未来数据的情况下决定是否拒绝当前的原假设，他所能依靠的仅仅是过去已经做出的决策和到目前为止已经积累的数据。一旦决策者做出决定是否拒绝原假设，在未来他不再能对此决定进行修改。

在线假设检验同样在大规模假设检验的框架下需要对决策的误差进行控制，因此也需要控制第一类错误与第二类错误。在实际的应用中人们往往更关心第一类错误（假阳性）的控制，因此在理论研究中对第一类错误有着不同的度量，比如家庭错误率（the familywise error rate ）和错误发现率（false discovery rate）。在实际的问题中往往需要控制错误发现率在一个预先给定的数值 α 以下。在线假设检验问题的应用非常广泛，比如在医药领域会进行大量的对比试验（A/B Tests）来确定药物中对疾病产生作用的成分，这在数学上就是一种在线假设检验的问题。另外在社会科学、交通、金融等领域存在着实时的异常检测问题，运用在线错误发现率控制的方法能有效对异常进行监测，从而减小对应的损失。

在线错误发现率控制的基本问题如下：$\mathbb{T}$ 代表了进行假设检验的时域，$\theta_t = 0/1$ 代表了假设 H_t 是原假设/不是原假设。信号的稀疏度可能随着时间的推移而发生变化由 $\{\pi_t \equiv P(\theta_t = 1) : t \in \mathbb{T}\}$ 表示。因此我们的观察可以别看成一个分层模型：

$$\theta_t \sim \text{Bernoulli}(\pi_t), \quad X_t \mid \theta_t \sim (1-\theta_t) F_0 + \theta_t F_{1t}.$$

其中 X_t 代表 t 时刻的数据观测值，F_0 代表原假设的数据分布，它是已知的且不随时间发生变化，F_{1t} 代表备择假设的数据分布，它可能随着时间变化。$\boldsymbol{X}^t = (X_i : i \in \mathbb{T}; i \le t)$ 代表到 t 时刻收集的统计量（如 p 值或 z 值）的集合，考虑一个在线的决策规则 $\boldsymbol{\delta} = \{\delta_t(\boldsymbol{X}^t) : t \in \mathbb{T}\} \in \{0,1\}^{\mathbb{T}}$，其中 $\delta_t(\boldsymbol{X}^t)$ 代表了一个实时的决策过程，这意味着 δ_t 值仅

1.2 基于模型的统计推断和模型自由的统计推断

在线假设检验的另一大特点是数据的结构可能很复杂，可能并不会服从我们已知的某些分布比如正态分布。而在传统的统计推断中往往需要假设数据服从某种参数分布来构造对应的统计量，而这在在线的假设检验中可能应用比较有限。因此近年来随着机器学习领域的不断发展，也有不少研究直接从数据本身出发借鉴机器学习的思想构造对应统计量进行推断，此类方法有着更为广阔应用场景，结果往往也更稳定，是统计推断领域的前沿方向。能否通过模型自由的方式进行在线的假设检验是一个很有意思的问题。

2 国内外研究现状

2.1 研究方向及进展

在线错误发现率控制方法最早是基于所谓的“阿尔法投资（alpha-investing）”[1]以及它的一般化（GAI）[2]。任何 GAI 准则都会有个犯错误的限度成为“α 财富”分配给每一个假设。这可以将犯错误的限度看成一种“投资”，如果此假设正好被拒绝，则“α 财富”得到回收，可以看成是投资的回报。阿尔法投资方法背后的直觉是随着假设的增加，当错误发生比例（FDP）的分子增大时，同样也允许分母增大即更多的假设被拒绝来保证错误发现率始终能被控制。

通常来说，$R_t = 0/1$ 是指 t 时刻根据任何假设检验准则的做出的决定，其中 1 代表拒绝原假设而 0 代表不拒绝。而一个 GAI 框架会产生一列每个假设的拒绝限度 $(\alpha_1, \alpha_2, \ldots)$，基于这些限度可以得到一系列的决定的集合 $(R_1, R_2, \ldots)$。需要注意的是 α_t 的产生必须仅仅基于 $R_1, R_2, \ldots, R_{t-1}$。在每一个决策的时间点 t，总体的 α 财富值 $W(t)$ 会减少 $\phi(t)$，如果此假设 H_t 被拒绝即 $R_t = 1$, 则总体的 α 财富值会增加 $\varphi(t)$。换句话说 $\phi(t)$ 是用来投资给新的假设的成本而 $\varphi(t)$ 是当这个假设被拒绝后得到的收益。因此最初的 α 财富值为 $W(0) = w_0$，同时遵照下面的更新原则：

$$W(t) = W(t-1) - \phi(t) + R_t\varphi(t).$$

由于总体的财富值 $W(t)$ 必须保持非负，因此有 $\phi(t) \leq W(t-1)$, 除此之外 $\alpha_t, \phi(t), \varphi(t)$ 的选取有一定的自由度但需要满足错误发现率的控制。

另一大研究方向是模型自由的假设检验准则，这在离线的大规模假设检验领域有了一定的发展。共形推理的框架[9]仅仅需要可交换性的假设，就可以为一类灵活的现成机器学习算法提供有限样本不确定性保证。机器学习和错误发现率控制的重要联系显示了共形分数比如共形的 p 值可以通过复杂机器学习算法构造[10]。

（zhao 和 Sun）[11]提出了全新的基于共形 Q 值的 PLIS 准则。在共形推断中，校准数据集（calibration data）$\mathbf{Y}$ 扮演了重要的角色，它有多种方式生成，比如可以从已知的原假设的分布 F_0 中产生。假设多重假设检验问题中数据来自未知的假设状态 $\boldsymbol{\Theta}=(\theta_i: i\in\mathcal{G})$ 构成了一个图（graph）$\mathcal{G}$，$m=|\mathcal{G}|$ 为假设的数量，测试数据和校准数据被并排放置并记为 $\{(X_i,Y_i), i\in\mathcal{G}\}$. 方法的第一步是构造基本数据集 $\mathbf{W}=(W_i\equiv h(X_i,Y_i): i\in\mathcal{G})$，其中

$$h(x,y)=\begin{cases} x & \text{if } |x|\geq|y| \\ y & \text{其他} \end{cases}$$

接下来生成通过用 W_i 分别代替 X_i 和 Y_i 来生成两组平行的数据集 $\left\{\tilde{\mathbf{X}}_i: i\in\mathcal{G}\right\}$ 和 $\left\{\tilde{\mathbf{Y}}_i: i\in\mathcal{G}\right\}$. 然后计算虚假的分数函数 $s_i^X\equiv\mathbb{P}_{\mathcal{M},\mathcal{A}}\left(\theta_i=0\mid\tilde{\mathbf{X}}_i\right)$ 和 $s_i^Y\equiv\mathbb{P}_{\mathcal{M},\mathcal{A}}\left(\theta_i=0\mid\tilde{\mathbf{Y}}_i\right)$。其中，$\mathcal{M}$ 和 $\mathcal{A}$ 代表了工作模型和对应的算法，它们可以使用机器学习工具的简化表示。我们记 $\mathbf{S}_X=\left\{s_i^X: i\in\mathcal{G}\right\}$ 和 $\mathbf{S}_Y=\left\{s_i^Y: i\in\mathcal{G}\right\}$ 为测试集和校准集的分数函数构成的集合。我们只选择小的 s_i^X 从目标拒绝集合 $\mathcal{G}^r=\left\{i\in\mathcal{G}: s_i^X<s_i^Y\right\}$ 取出。与之对应的是镜像过程集合 $\mathcal{G}^c=\left\{i\in\mathcal{G}: s_i^Y<s_i^X\right\}$. 由此可以计算共形的 Q 值：

$$Q(t)=\frac{1+\sum_{j\in\mathcal{G}^c}\mathbb{I}\left\{s_j^Y\leq t\right\}}{\left(\sum_{j\in\mathcal{G}^r}\mathbb{I}\left\{s_j^X\leq t\right\}\right)\vee 1},\quad t>0$$

根据这个共形 Q 值我们可以确定拒绝的门槛：

$$\tau=\sup\left\{t\in\mathbf{S}_X\cup\mathbf{S}_Y: Q(t)\leq\alpha\right\}$$

根据 PLIS 准则，我们拒绝 $H_{0,i}$ 如果 $\delta_i\equiv\mathbb{I}(i\in\mathcal{G}^r, s_i^X\leq\tau)=1$. 在可交换性的假设下，PLIS 准则能够控制错误发现率在预先给定的值 α 以下。

2.2 存在问题

在传统的在线错误发现率控制方法如 GAI 框架中，普遍存在的问题是由于想避免“alpha-death”（拒绝到某一时刻就用完了 α 财富，无法再进行拒绝），每个时间点的拒绝门

槛往往比较高，这导致拒绝假设显得比较得保守，拒绝假设的效率比较低。在早期的GAI框架中还可能出现“piggybacking”的问题，即 α 财富经过一段时间的积累可能在某个时刻拒绝一个统计量（p值）比较极端的假设从而导致后续的 α 财富变的很小，从而无法拒绝后续本可以拒绝的假设。

2021年提出的SAST准则可以很好地避免“alpha-death”和“piggybacking”的问题，拒绝假设的效率也比之前的方法有显著的提升。但存在的一个问题是它的数据驱动的准则只能证明在大样本的情况下能够控制错误发现率（FDR），无法保证在有限样本的情况下控制错误发现率。

而最近比较前沿的模型自由的假设检验准则能够在有限样本的情况下控制错误发现率（FDR），但是目前还只是在离线的情况下进行研究。在在线情况下数据整体的可交换性可能不能得到满足，因此还需要对共形推断等方法进行修改与创新。

3 研究展望

本课题的目标是引入共形推断的思想，基于共形的Q值（conformal Q-value）来建立与之前方法等效的阿尔法投资框架，通过基于数据而模型自由的方法在有限样本的情况下控制每一个时间点的在线错误发现率。于此同时，希望尽可能地提高在线假设检验的效率，避免过于保守的决策。为了简单起见，本课题研究的情况数据的稀疏度和分布不随时间的发展而发生变化。在未来可以做更复杂的拓展。

4 参考文献

[1] FOSTER D P, STINE R A. α-Investing: a Procedure for Sequential Control of Expected False Discoveries [J]. Journal of the Royal Statistical Society Series B: Statistical Methodology, 2008, 70(2): 429-444.

[2] AHARONI E, ROSSET S. Generalized α-investing: Definitions, Optimality Results and Application to Public Databases[J]. Journal of the Royal Statistical Society Series B: Statistical Methodology, 2013, 76(4): 771-794.

[3] JAVANMARD A, MONTANARI A. Online rules for control of false discovery rate and false discovery exceedance[J/OL]. The Annals of Statistics, 2018, 46(2): 526-554. https://doi.org/10.1214/17-BAOS1559. DOI: 10.1214/17-BAOS1559.

[4] RAMDAS A, YANG F, WAINWRIGHT M J, et al. Online control of the false discovery rate with decaying memory[C/OL]//GUYON I, LUXBURG U V, BENGIO S, et al. Advances in Neural Information Processing Systems: vol. 30. Curran Associates, Inc., 2017. https://proceedings.neurips.cc/paper_files/paper/2017/file/7f018eb7b301a66658931cb8a93fd6e8-BPaper.pdf.

[5] RAMDAS A, ZRNIC T, WAINWRIGHT M, et al. SAFFRON: an Adaptive Algorithm for Online Control of the False Discovery Rate[C/OL]//DY J, KRAUSE A. Proceedings of Machine Learning Research: Proceedings of the 35th International Conference on Machine Learning: vol. 80. PMLR, 2018: 4286-4294. https://proceedings.mlr.press/v80/ramdas18a.html.

[6] TIAN J, RAMDAS A. ADDIS: an adaptive discarding algorithm for online FDR control with conservative nulls[C/OL]//WALLACH H, LAROCHELLE H, BEYGELZIMER A, et al. Advances in Neural Information Processing Systems: vol. 32. Curran Associates, Inc., 2019. https://proceedings.neurips.cc/paper_files/paper/2019/file/1d6408264d31d453d556c60fe7d0459e-BPaper.pdf.

[7] BOWEN GANG W S, WANG W. Structure-Adaptive Sequential Testing for Online False Discovery Rate Control[J/OL]. Journal of the American Statistical Association, 2023, 118(541): 732-745. eprint: https://doi.org/10.1080/01621459.2021.1955688. https://doi.org/10.1080/01621459.2021.1955688. DOI: 10.1080/01621459.2021.1955688.

[8] CAI T T, SUN W. Simultaneous Testing of Grouped Hypotheses: Finding Needles in Multiple Haystacks [J/OL]. Journal of the American Statistical Association, 2009, 104(488): 1467-1481. eprint: https://doi.org/10.1198/jasa.2009.tm08415. https://doi.org/10.1198/jasa.2009.tm08415. DOI: 10.1198/jasa.2009.tm08415.

[9] VOVK V, GAMMERMAN A, SHAFER G. Algorithmic Learning in a Random World[C/OL]//. 2005. https://api.semanticscholar.org/CorpusID:118783209.

[10] BATES S, CANDÈS E, LEI L, et al. Testing for outliers with conformal p-values[J/OL]. The Annals of Statistics, 2023, 51(1): 149-178. https://doi.org/10.1214/22-BAOS2244. DOI: 10.1214/22-BAOS2244.

[11] ZHAO Z, SUN W. False Discovery Rate Control For Structured Multiple Testing: Asymmetric Rules And Conformal Q-values[Z]. 2023. arXiv: 2311.15322 [stat.ME].

开题报告，字数 3500 字以上。

二、开题报告

1 问题提出的背景

1.1 背景介绍

在大数据时代，获取数据相比之前任何时期都要显得更加便利，但随之产生的一个问题是如何从海量数据和高维变量中获取我们所需的信息，这便是大规模假设检验所研究的重要课题。而在线的大规模假设检验问题是其中的一种特殊情况，它所考虑的场景是一种实时的监测过程，即决策者不能一下子获取所有的数据也不清楚需要进行的决策的数量（很可能随着时间推移趋于无穷），但需要在每一个决策的时间点尽可能地保证决策的精确性。决策者所能依靠的仅仅是过去已经做出的决策和到当前时间点为止已经积累的数据。一旦决策者做出决定是否拒绝原假设，在未来他不再能对此决定进行修改。

在线的假设检验问题是大规模假设检验领域的前沿课题，因此它同样也在大规模假设检验的框架下，需要控制第一类错误与第二类错误。在实际的应用中人们往往更关心第一类错误（假阳性）的控制，因此在理论研究中对第一类错误有着多种的度量标准，比如家庭错误率（the familywise error rate ）和错误发现率（false discovery rate）。在实际的问题中往往需要控制错误发现率在一个预先给定的数值 α 以下。

在线假设检验问题的应用非常广泛，其应用遍及医疗卫生、农业生产、交通物流、经济金融等领域。下面是两个简单的例子：

Example 3 在生物或医疗公司中，存在着一种非常常见的方法来检测药物中对某种疾病产生作用的成分。它叫做平台对比试验（*A/B Tests*）。这种对比试验会进行非常多次，其中对疾病有作用的成分称为“信号”而对疾病没有影响的成分称为“噪声”，而决策者需要在每一次试验结束后判断出哪些数据源于“信号”，同时需要尽可能的保证精确度。

Example 4 在路面交通或是电子 *APP* 领域，都需要对流量进行实时的监测。在这个例子中流量发生变化的“异常点”就是我们需要判别出的“信号”，而正常情形则是“噪声”。实时的异常点监测也是在线假设检验的重要问题。

1.2 本研究的意义和目的

在线的假设检验问题在大规模假设检验的框架下，同样需要对我们决策所犯的错误进行约束。为此我们在这里引入相关记号进行说明。在线错误发现率控制的基本问题如下：

$\mathbb{T}$ 代表了进行假设检验的时域，$\theta_t = 0/1$ 代表了假设 H_t 是原假设/不是原假设。信号的稀疏度可能随着时间的推移而发生变化由 $\{\pi_t \equiv P(\theta_t = 1) : t \in \mathbb{T}\}$ 表示。因此我们的观察可以别看成一个分层模型：

$$\theta_t \sim \text{Bernoulli}(\pi_t), \quad X_t \mid \theta_t \sim (1-\theta_t) F_0 + \theta_t F_{1t}.$$

其中 X_t 代表 t 时刻的数据观测值，F_0 代表原假设的数据分布，它是已知的且不随时间发生变化，F_{1t} 代表备择假设的数据分布，它可能随着时间变化。$\boldsymbol{X}^t = (X_i : i \in \mathbb{T}; i \le t)$ 代表到 t 时刻收集的统计量（如 p 值或 z 值）的集合，考虑一个在线的决策规则 $\boldsymbol{\delta} = \{\delta_t(\boldsymbol{X}^t) : t \in \mathbb{T}\} \in \{0,1\}^{\mathbb{T}}$，其中 $\delta_t(\boldsymbol{X}^t)$ 代表了一个实时的决策过程，这意味着 δ_t 值仅仅依赖于 t 时刻已知的信息。若 $\delta_t = 1$ 意味着 H_t 被拒绝 $\delta_t = 0$ 则表示接受原假设。记 $\boldsymbol{\delta}^t = \{\delta_i(\boldsymbol{X}^i) : i \in \mathbb{T}; i \le t\}$ 为到 t 时刻的决策集合。而 t 时刻错误发生的比例（FDP）的定义如下

$$\text{FDP}^t(\boldsymbol{\delta}^t) = \frac{\sum_{i \le t; i \in \mathbb{T}} (1-\theta_i)\delta_i}{\left(\sum_{i \le t; i \in \mathbb{T}} \delta_i\right) \vee 1}.$$

FDP 代表了拒绝的假设中错误拒绝的比例，其中 $x \vee y = max(x, y)$. 但是在这里，错误拒绝的假设的个数我们是未知的，FDP 并不可以被直接地观察到，所以我们尝试去控制 FDP 的期望即错误发现率（false discovery rate），而 t 时刻的错误发现率的定义如下：

$$\text{FDR}^t(\boldsymbol{\delta}^t) = \mathbb{E}\{\text{FDP}^t(\boldsymbol{\delta}^t)\}$$

在过去离线的大规模假设检验的问题中，我们能不受限制的观察到所有的数据，并且只用控制整体的 FDR 就可以，但在在线的场景下，我们的目标是寻找一种寻找一种决策规则 $\boldsymbol{\delta} = \{\delta_t(\boldsymbol{X}^t) : t \in \mathbb{T}\} \in \{0,1\}^{\mathbb{T}}$ 使得对任意的 $t \in \mathbb{T}$ 都有 $FDR^t \le \alpha$. 为了比较不同决策规则的拒绝效率（power），我们也有对应的平均效率（AP）和遗漏发现率（MDR），它们是大规模假设检验中第二类错误的度量。

$$\text{AP}^t(\boldsymbol{\delta}^t) = \frac{\mathbb{E}\left(\sum_{i \le t; i \in \mathbb{T}} \theta_i \delta_i\right)}{\mathbb{E}\left(\sum_{i \le t; i \in \mathbb{T}} \theta_i\right)}; \quad \text{MDR}^t(\boldsymbol{\delta}^t) = 1 - \text{AP}^t(\boldsymbol{\delta}^t).$$

我们希望本研究的方法在严格控制FDR的同时能够尽可能地提高拒绝的平均效率，从而不会导致过于保守的决策。我们还希望得到基于数据本身而非数据服从的参数模型的方法来控制FDR，因为在机器学习等工具发展的今天，这种方法有着更为广阔的应用前景。在线的错误发现率控制的方法在信号检测、异常监测等领域有着重要的研究意义。

2 项目的主要内容和技术路线

2.1 主要研究内容

以下的内容我都会统一使用上文提到的记号。

主要研究的是控制在线错误发现率（FDR）的共形推断方法。首先学习前辈们在在线假设领域所提出的控制错误发现率的方法，理解他们的思想，并分析他们的方法的优势与劣势。于此同时，根据孙老师的指示，学习共形推断的相关方法，理解他们是如何基于数据本身而有一定的模型自由度的构造出检验方法来控制FDR。接下来，尝试将共形推断中的共形Q值方法改造出适合在线假设检验的版本，建立一套在线共形推断框架，并与之前的方法进行比较。最后，希望证明提出的在线共形Q值方法能够在所有决策时间点控制FDR。

2.2 技术路线

2.2.1 基于模型的在线错误发现率控制方法

目前提出的控制在线错误发现率的方法均是基于模型的。2008年Foster和Stine提出了“阿尔法投资(alpha-investing)”的框架[1]。这一框架又在2014年得到了一般化，即GAI框架[2]。一个GAI框架会产生一列每个假设的拒绝限度$(\alpha_1, \alpha_2, \dots)$，基于这些限度可以得到一系列的决定的集合$(R_1, R_2, \dots)$。需要注意的是$\alpha_t$的产生必须仅仅基于$R_1, R_2, \dots, R_{t-1}$。在每一个决策的时间点$t$，总体的$\alpha$财富值$W(t)$会减少$\phi(t)$，如果此假设$H_t$被拒绝即$R_t = 1$, 则总体的$\alpha$财富值会增加$\varphi(t)$。换句话说$\phi(t)$是用来投资给新的假设的成本而$\varphi(t)$是当这个假设被拒绝后得到的收益。因此最初的$\alpha$财富值为$W(0) = w_0$，同时遵照下

4 参考文献

[1] FOSTER D P, STINE R A. α-Investing: a Procedure for Sequential Control of Expected False Discoveries [J]. Journal of the Royal Statistical Society Series B: Statistical Methodology, 2008, 70(2): 429-444.

[2] AHARONI E, ROSSET S. Generalized α-investing: Definitions, Optimality Results and Application to Public Databases[J]. Journal of the Royal Statistical Society Series B: Statistical Methodology, 2013, 76(4): 771-794.

[3] JAVANMARD A, MONTANARI A. Online rules for control of false discovery rate and false discovery exceedance[J/OL]. The Annals of Statistics, 2018, 46(2): 526-554. https://doi.org/10.1214/17-BAOS1559. DOI: 10.1214/17-BAOS1559.

[4] RAMDAS A, YANG F, WAINWRIGHT M J, et al. Online control of the false discovery rate with decaying memory[C/OL]//GUYON I, LUXBURG U V, BENGIO S, et al. Advances in Neural Information Processing Systems: vol. 30. Curran Associates, Inc., 2017. https://proceedings.neurips.cc/paper_files/paper/2017/file/7f018eb7b301a66658931cb8a93fd6e8-BPaper.pdf.

[5] RAMDAS A, ZRNIC T, WAINWRIGHT M, et al. SAFFRON: an Adaptive Algorithm for Online Control of the False Discovery Rate[C/OL]//DY J, KRAUSE A. Proceedings of Machine Learning Research: Proceedings of the 35th International Conference on Machine Learning: vol. 80. PMLR, 2018: 4286-4294. https://proceedings.mlr.press/v80/ramdas18a.html.

[6] TIAN J, RAMDAS A. ADDIS: an adaptive discarding algorithm for online FDR control with conservative nulls[C/OL]//WALLACH H, LAROCHELLE H, BEYGELZIMER A, et al. Advances in Neural Information Processing Systems: vol. 32. Curran Associates, Inc., 2019. https://proceedings.neurips.cc/paper_files/paper/2019/file/1d6408264d31d453d556c60fe7d0459e-BPaper.pdf.

[7] BOWEN GANG W S, WANG W. Structure - Adaptive Sequential Testing for Online False Discovery Rate Control[J/OL]. Journal of the American Statistical Association, 2023, 118(541): 732-745. eprint: https://doi.org/10.1080/01621459.2021.1955688. https://doi.org/10.1080/01621459.2021.1955688. DOI: 10.1080/01621459.2021.1955688.

[8] CAI T T, SUN W. Simultaneous Testing of Grouped Hypotheses: Finding Needles in Multiple Haystacks [J/OL]. Journal of the American Statistical Association, 2009, 104(488): 1467-1481. eprint: https://doi.org/10.1198/jasa.2009.tm08415. https://doi.org/10.1198/jasa.2009.tm08415. DOI: 10.1198/jasa.2009.tm08415.

[9] VOVK V, GAMMERMAN A, SHAFER G. Algorithmic Learning in a Random World[C/OL]//. 2005. https://api.semanticscholar.org/CorpusID:118783209.

[10] BATES S, CANDÈS E, LEI L, et al. Testing for outliers with conformal p-values[J/OL]. The Annals of Statistics, 2023, 51(1): 149-178. https://doi.org/10.1214/22-BAOS2244. DOI: 10.1214/22-BAOS2244.

[11] ZHAO Z, SUN W. False Discovery Rate Control For Structured Multiple Testing: Asymmetric Rules And Conformal Q-values[Z]. 2023. arXiv: 2311.15322 [stat.ME].

[12] ZRNIC T, JIANG D, RAMDAS A, et al. The Power of Batching in Multiple Hypothesis Testing[C/OL]//CHIAPPA S, CALANDRA R. Proceedings of Machine Learning Research: Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics: vol. 108. PMLR, 2020: 3806-3815. https://proceedings.mlr.press/v108/zrnic20a.html.

外文翻译，译文 3000 字以上。

三、外文翻译

出处：Advances in Neural Information Processing Systems 30(NeurIPS 2017)

作者：Ramdas, Aaditya and Yang, Fanny and Wainwright, Martin J and Jordan, Michael I

题目：Online control of the false discovery rate with decaying memory

摘要：在在线的多重假设检验问题中，对应于不同原假设的 p 值被一个接一个的观察到，并且是否拒绝当前原假设的决定必须立刻做出，然后再观察下一时刻的 p 值。Alpha 投资算法控制错误发现率（FDR），由 Foster 和 Stine 提出，已经被推广并应用于多个场景，包括科学领域的质量保护的数据库以及互联网商业中的多重对比试验（A/B Tests）和多臂老虎机试验。这篇论文从四个方面提升了一般化的 alpha 投资框架（GAI）：（1）我们展示了一致提升整个单调类（GAI）准则的拒绝效率的方法通过给予每个假设更多的 alpha 财富值，这是一个双赢的方式来解决由 Javanmard 和 Montanari 提出的困境。（2）我们展示了如何利用先验的权重来显示哪些假设很有可能不是噪声。（3）我们允许对错误发现的不同权重的惩罚来显示假设重要性程度的不同。（4）我们定义了一种新的度量叫做衰减记忆错误发现率（mem-FDR），它对于真实的时间序列的场景可能更有意义，并且能够减轻我们所熟知的“piggybacking”和“alpha-death”问题。我们的 GAI++ 合并了这四种推广，并且在衰减与权重都设置为一致时会退化为早期更有效率的算法的变体。最后，我们还描述了一种基于估计的错误发现比例的规则来导出新的在线 FDR 的简单方法。

1 导言

Tukey 的开创性专著首次认识到了多重比较的问题。简单地说，给定要测试的多个假设的集合，目标是区分原假设和备择假设，并对不同类型的错误进行适当的控制。对于每个假设，我们可以得到一个 p 值，我们以此来判断哪部分的假设应该被拒绝，有效的宣称被拒绝的假设不是噪声。被拒绝的假设被称为发现，其中真正无效的子集——因此被错误地拒绝——被称为虚假发现。在这项工作中，我们使用错误发现率（FDR）来度量方法的性能，定义为错误发现占全体发现的期望比例。具体来说，我们需要所有的准则必须保证

FDR 在一个预先给定的常数 α 以下。

传统意义上的多重假设检验实际上是离线的，意思是一个要检验 N 个假设的算法在一个时刻接受一整个批次的 p 值 $\{P_1, \ldots, P_N\}$。在这个问题的在线版本中，我们不知道我们提前测试了多少假设；相反，一个可能无限的 p 值序列一个接一个地出现，并且必须在接收到下一个 p 值之前做出关于拒绝原假设的决定。至少有两个动机来考虑在线环境。

M1 我们可能从一开始就有整批 p 值可供使用，但我们仍然可以选择按特定顺序逐一处理 p 值。事实上，如果可以使用先验知识来确保非噪声通常在排序中更早出现，那么精心设计的在线过程可能会比离线算法（在没有先验知识的情况下运行）（如经典 Benjamini Hochberg 算法）产生更多的发现，当使用同样的 FDR 控制准则时。这一动机是 Foster 和 Stine 的原始在线多重假设检验论文的基础。

M2 我们可以真正逐一进行一系列测试，其中下一个原假设的选择和测试的水平可能取决于之前测试的结果。在这个动机下的应用包括希望为下面的场景提供一个实时的保证（i）互联网公司随着时间的推移进行一系列的 A/B 测试（ii）制药公司使用多臂老虎机规则进行一系列临床试验，或（iii）质量保护的数据库提供保证，在这些数据库中，不同的研究团队随着时间的流逝对相同数据测试不同的假设。

这篇论文包含的算法应用于以上两种场景，尤其适用于 M2 场景。

让我们首先重申，在在线环境中测试一系列假设时，即使所有 p 值都是独立的，也需要进行更正。如果每个假设 i 都是独立于在它之前或之后进行的测试总数进行测试的，那么我们无法控制随着时间的推移所做出的错误发现的数量。事实上，如果我们在检验中对每个 P_i 使用 $1\{P_i \leq \alpha\}$ 的形式，那么尽管每一个假设的第一类错误能被 α 控制，整体的发现的 FDR 控制往往不能保证。举个例子，在全局假设下，即所有的假设都是真的噪声（原假设），只要检验的数量 N 足够的大，并且原假设的 p 值服从均匀分布，这个方法会以很高的概率做出一次发现，但在这个场景下，所有的发现均是错误，因此 FDR 会以很高的概率等于 1。

一个很自然的替代方式是把多重性进行 Bonferroni 校正。如果决策者知道要拒绝的假设的数量 N，那么决策准则 $1\{P_i \leq \alpha/N\}$ 对每个 $i \in \{1, \ldots, N\}$ 控制了有一个错误发现的概率——这是一个被称为家庭错误率（FWER）的度量——在 α 的水平下。这一准则可

四、外文原文

插入外文原文图片，外框黑色。

Online control of the false discovery rate with decaying memory

Aaditya Ramdas **Fanny Yang** **Martin J. Wainwright** **Michael I. Jordan**
University of California, Berkeley
{aramdas, fanny-yang, wainwrig, jordan} @berkeley.edu

Abstract

In the online multiple testing problem, p-values corresponding to different null hypotheses are observed one by one, and the decision of whether or not to reject the current hypothesis must be made immediately, after which the next p-value is observed. Alpha-investing algorithms to control the false discovery rate (FDR), formulated by Foster and Stine, have been generalized and applied to many settings, including quality-preserving databases in science and multiple A/B or multi-armed bandit tests for internet commerce. This paper improves the class of generalized alpha-investing algorithms (GAI) in four ways: (a) we show how to uniformly improve the power of the entire class of monotone GAI procedures by awarding more alpha-wealth for each rejection, giving a win-win resolution to a recent dilemma raised by Javanmard and Montanari, (b) we demonstrate how to incorporate prior weights to indicate domain knowledge of which hypotheses are likely to be non-null, (c) we allow for differing penalties for false discoveries to indicate that some hypotheses may be more important than others, (d) we define a new quantity called the decaying memory false discovery rate (mem-FDR) that may be more meaningful for truly temporal applications, and which alleviates problems that we describe and refer to as "piggybacking" and "alpha-death." Our GAI++ algorithms incorporate all four generalizations simultaneously, and reduce to more powerful variants of earlier algorithms when the weights and decay are all set to unity. Finally, we also describe a simple method to derive new online FDR rules based on an estimated false discovery proportion.

1 Introduction

The problem of multiple comparisons was first recognized in the seminal monograph by Tukey [12]: simply stated, given a collection of multiple hypotheses to be tested, the goal is to distinguish between the nulls and non-nulls, with suitable control on different types of error. We are given access to one p-value for each hypothesis, which we use to decide which subset of hypotheses to reject, effectively proclaiming the rejected hypothesis as being non-null. The rejected hypotheses are called *discoveries*, and the subset of these that were truly null—and hence mistakenly rejected—are called *false discoveries*. In this work, we measure a method's performance using the *false discovery rate* (FDR) [2], defined as the expected ratio of false discoveries to total discoveries. Specifically, we require that any procedure must guarantee that the FDR is bounded by a pre-specified constant α.

The traditional form of multiple testing is *offline* in nature, meaning that an algorithm testing N hypotheses receives the entire batch of p-values $\{P_1, \ldots, P_N\}$ at one time instant. In the *online* version of the problem, we do not know how many hypotheses we are testing in advance; instead, a possibly infinite sequence of p-values appear one by one, and a decision about rejecting the null must be made before the next p-value is received. There are at least two different motivating justifications for considering the online setting:

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

M1. We may have the entire batch of p-values available at our disposal from the outset, but we may nevertheless choose to process the p-values one by one in a particular order. Indeed, if one can use prior knowledge to ensure that non-nulls typically appear earlier in the ordering, then carefully designed online procedures could result in more discoveries than offline algorithms (that operate without prior knowledge) such as the classical Benjamini-Hochberg algorithm [2], while having the same guarantee on FDR control. This motivation underlies one of the original online multiple testing paper, namely that of Foster and Stine [5].

M2. We may genuinely conduct a sequence of tests one by one, where both the choice of the next null hypothesis and the level at which it is tested may depend on the results of the previous tests. Motivating applications include the desire to provide anytime guarantees for (i) internet companies running a sequence of A/B tests over time [9], (ii) pharmaceutical companies conducting a sequence of clinical trials using multi-armed bandits [13], or (iii) quality-preserving databases in which different research teams test different hypotheses on the same data over time [1].

The algorithms developed in this paper apply to both settings, with emphasis on motivation M2.

Let us first reiterate the need for corrections when testing a sequence of hypotheses in the online setting, even when all the p-values are independent. If each hypothesis i is tested independently of the total number of tests either performed before it or to be performed after it, then we have no control over the number of false discoveries made over time. Indeed, if our test for every P_i takes the form $\mathbf{1}\{P_i \leq \alpha\}$ for some fixed α, then, while the type 1 error for any individual test is bounded by α, the set of discoveries could have arbitrarily poor FDR control. For example, under the "global null" where every hypothesis is truly null, as long as the number of tests N is large and the null p-values are uniform, this method will make at least one rejection with high probability (w.h.p.), and since in this setting every discovery is a false discovery, w.h.p. the FDR will equal one.

A natural alternative that takes multiplicity into account is the Bonferroni correction. If one knew the total number N of tests to be performed, the decision rule $\mathbf{1}\{P_i \leq \alpha/N\}$ for each $i \in \{1, \ldots, N\}$ controls the probability of even a single false discovery—a quantity known as the familywise error rate or FWER—at level α, as can be seen by applying the union bound. The natural extension of this solution to having an unknown and potentially infinite number of tests is called *alpha-spending*. Specifically, we choose any sequence of constants $\{\alpha_i\}_{i \in \mathbb{N}}$ such that $\sum_i \alpha_i \leq \alpha$, and on receiving P_i, our decision is simply $\mathbf{1}\{P_i \leq \alpha_i\}$. However, such methods typically make very few discoveries—meaning that they have very low power—when the number of tests is large, because they must divide their error budget of α, also called alpha-wealth, among a large number of tests.

Since the FDR is less stringent than FWER, procedures that guarantee FDR control are generally more powerful, and often far more powerful, than those controlling FWER. This fact has led to the wide adoption of FDR as a de-facto standard for offline multiple testing (note, e.g., that the Benjamini-Hochberg paper [2] currently has over 40,000 citations).

Foster and Stine [5] designed the first online alpha-investing procedures that use and earn alpha-wealth in order to control a modified definition of FDR. Aharoni and Rosset [1] further extended this to a class of generalized alpha-investing (GAI) methods, but once more for the modifed FDR. It was only recently that Javanmard and Montanari [9] demonstrated that monotone GAI algorithms, appropriately parameterized, can control the (unmodified) FDR for independent p-values. It is this last work that our paper directly improves upon and generalizes; however, as we summarize below, many of our modifications and generalizations are immediately applicable to all previous algorithms.

Contributions and outline. Instead of presenting the most general and improved algorithms immediately, we choose to present results in a bottom-up fashion, introducing one new concept at a time so as to lighten the symbolic load on the reader. For this purpose, we set up the problem formally in Section 2. Our contributions are organized as follows:

1. **Power.** In Section 3, we introduce the generalized alpha-investing (GAI) procedures, and demonstrate how to uniformly improve the power of monotone GAI procedures that control FDR for independent p-values, resulting in a win-win resolution to a dilemma posed by Javanmard and Montanari [9]. This improvement is achieved by a somewhat subtle modification that allows the algorithm to reward more alpha-wealth at every rejection but the first. We refer to our algorithms as *improved generalized alpha-investing* (GAI++) procedures, and provide intuition for why they work through a general super-uniformity lemma (see Lemma 1 in Section 3.2). We

Existing online multiple testing algorithms control some variant of the FDR over time, as we now define. At any time T, let $R(T) = \sum_{t=1}^{T} R_t$ be the total number of rejections/discoveries made by the algorithm so far, and let $V(T) = \sum_{t \in \mathcal{H}^0} R_t$ be the number of false rejections/discoveries. Then, the false discovery proportion and rate are defined as

$$\text{FDP}(T) := \frac{V(T)}{R(T)} \text{ and } \text{FDR}(T) = \mathbb{E}\left[\frac{V(T)}{R(T)}\right],$$

where we use the dotted-fraction notation corresponds to the shorthand $\frac{a}{\dot{b}} = \frac{a}{b \vee 1}$. Two variants of the FDR studied in earlier online FDR works [5, 8] are the *marginal* FDR given by $\text{mFDR}_\eta(T) = \frac{\mathbb{E}[V(T)]}{\mathbb{E}[R(T)]+\eta}$, with a special case being $\text{mFDR}(T) = \frac{\mathbb{E}[V(T)]}{\mathbb{E}[R(T) \vee 1]}$, and the *smoothed* FDR, given by $\text{sFDR}_\eta(T) = \mathbb{E}\left[\frac{V(T)}{R(T)+\eta}\right]$. In Appendix A, we summarize a variety of algorithms and dependence assumptions considered in previous work.

3 Generalized alpha-investing (GAI) rules

The generalized class of alpha-investing rules [1] essentially covers most rules that have been proposed thus far, and includes a wide range of algorithms with different behaviors. In this section, we present a uniform improvement to monotone GAI algorithms for FDR control under independence.

Any algorithm of the GAI type begins with an *alpha-wealth* of $W(0) = W_0 > 0$, and keeps track of the wealth $W(t)$ available after t steps. At any time t, a part of this alpha-wealth is used to test the t-th hypothesis at level α_t, and the wealth is immediately decreased by an amount ϕ_t. If the t-th hypothesis is rejected, that is if $R_t := \mathbf{1}\{P_t \le \alpha_t\} = 1$, then we award extra wealth equaling an amount ψ_t. Recalling the definition $\mathcal{F}^t := \sigma(R_1, \ldots, R_t)$, we require that $\alpha_t, \phi_t, \psi_t \in \mathcal{F}^{t-1}$, meaning they are predictable, and $W(t) \in \mathcal{F}^t$, with the explicit update $W(t) := W(t-1) - \phi_t + R_t\psi_t$. The parameters W_0 and the sequences α_t, ϕ_t, ψ_t are all user-defined. They must be chosen so that the total wealth $W(t)$ is always non-negative, and hence that $\phi_t \le W(t-1)$ If the wealth ever equals zero, the procedure is not allowed to reject any more hypotheses since it has to choose α_t equal to zero from then on. The only real restriction for α_t, ϕ_t, ψ_t arises from the goal to control FDR. This condition takes a natural form—whenever a rejection takes place, we cannot be allowed to award an arbitrary amount of wealth. Formally, for some user-defined constant B_0, we must have

$$\psi_t \le \min\{\phi_t + B_0, \frac{\phi_t}{\alpha_t} + B_0 - 1\}. \tag{4}$$

Many GAI rules are not monotone (cf. equation (3)), meaning that α_t is not always a coordinatewise nondecreasing function of $R_1, \ldots, R_{t-1}$, as mentioned in the last column of Table 2 (Appendix A). Table 1 has some examples, where $\tau_k := \min_{s \in \mathbb{N}} \mathbf{1}\{\sum_{t=1}^{s} R_t = k\}$ is the time of the k-th rejection.

Name	Parameters	Level α_t	Penalty ϕ_t	Reward ψ_t
[5] Alpha-investing (AI)	—	$\frac{\phi_t}{1+\phi_t}$	$\le W(t-1)$	$\phi_t + B_0$
[1] Alpha-spending with rewards	$\kappa \le 1, c$	$cW(t-1)$	$\kappa W(t-1)$	satisfy (4)
[9] LORD'17	$\sum_{t=1}^{\infty} \gamma_t = 1$	ϕ_t	$\gamma_t W_0 + B_0 \sum_{j:\tau_j < t} \gamma_{t-\tau_j}$	$B_0 = \alpha - W_0$

Table 1: Examples of GAI rules.

3.1 Improved monotone GAI rules (GAI++) under independence

In their initial work on GAI rules, Aharoni and Rosset [1] did not incorporate an explicit parameter B_0; rather, they proved that choosing $W_0 = B_0 = \alpha$ suffices for mFDR_1 control. In subsequent work, Javanmard and Montanari [9] introduced the parameter B_0 and proved that for monotone GAI rules, the same choice $W_0 = B_0 = \alpha$ suffices for sFDR_1 control, whereas the choice $B_0 = \alpha - W_0$ suffices for FDR control, with both results holding under independence. In fact, their monotone GAI rules with $B_0 = \alpha - W_0$ are the only known methods that control FDR. This state of affairs leads to the following dilemma raised in their paper [9]:

> A natural question is whether, in practice, we should choose W_0, B_0 as to guarantee FDR control (and hence set $B_0 = \alpha - W_0 \ll \alpha$) or instead be satisfied with mFDR or sFDR control, which allow for $B_0 = \alpha$ and hence potentially larger statistical power.

Acknowledgments

We thank A. Javanmard, R. F. Barber, K. Johnson, E. Katsevich, W. Fithian and L. Lei for related discussions, and A. Javanmard for sharing code to reproduce experiments in Javanmard and Montanari [9]. This material is based upon work supported in part by the Army Research Office under grant number W911NF-17-1-0304, and National Science Foundation grant NSF-DMS-1612948.

References

[1] Ehud Aharoni and Saharon Rosset. Generalized α-investing: definitions, optimality results and application to public databases. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(4):771–794, 2014.

[2] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, 57(1): 289–300, 1995.

[3] Yoav Benjamini and Yosef Hochberg. Multiple hypotheses testing with weights. *Scandinavian Journal of Statistics*, 24(3):407–418, 1997.

[4] Gilles Blanchard and Etienne Roquain. Two simple sufficient conditions for fdr control. *Electronic journal of Statistics*, 2:963–992, 2008.

[5] Dean P. Foster and Robert A. Stine. α-investing: a procedure for sequential control of expected false discoveries. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(2):429–444, 2008.

[6] Christopher R Genovese, Kathryn Roeder, and Larry Wasserman. False discovery control with p-value weighting. *Biometrika*, 93(3):509–524, 2006.

[7] Philipp Heesen and Arnold Janssen. Dynamic adaptive multiple tests with finite sample fdr control. *arXiv preprint arXiv:1410.6296*, 2014.

[8] Adel Javanmard and Andrea Montanari. On online control of false discovery rate. *arXiv preprint arXiv:1502.06197*, 2015.

[9] Adel Javanmard and Andrea Montanari. Online rules for control of false discovery rate and false discovery exceedance. *The Annals of statistics*, 2017.

[10] Ang Li and Rina Foygel Barber. Multiple testing with the structure adaptive benjamini-hochberg algorithm. *arXiv preprint arXiv:1606.07926*, 2016.

[11] Aaditya Ramdas, Rina Foygel Barber, Martin J. Wainwright, and Michael I. Jordan. A unified treatment of multiple testing with prior knowledge. *arXiv preprint arXiv:1703.06222*, 2017.

[12] John Tukey. *The Problem of Multiple Comparisons: Introduction and Parts A, B, and C*. Princeton University, 1953.

[13] Fanny Yang, Aaditya Ramdas, Kevin Jamieson, and Martin J. Wainwright. A framework for Multi-A(rmed)/B(andit) testing with online FDR control. *Advances in Neural Information Processing Systems*, 2017.

此页单面打印

毕业论文（设计）文献综述和开题报告考核

对文献综述、外文翻译和开题报告评语及成绩评定：

文献综述主要介绍了在线假设检验问题的一系列广泛应用，研究历史，已有的方法及存在问题。文献综述切题准确，阅读的参考文献较丰富，系统的讨论了国际前沿最新发展，显示了作者对该领域的良好理解。开题报告中阐述了大数据流在线分析的研究思路，基本框架，研究流程与方法，并给出进度安排和预期目标。外文翻译资料与毕业论文选题密切相关，译文语句通顺、流畅，字数达到要求。

评语不少于 200 个字，对论文文献综述、开题报告和外文翻译以及学生答辩表现和质量等方面分别加以评价

成绩比例	文献综述 占（10%）	开题报告 占（15%）	外文翻译 占（5%）
分值	10	14	5

非指导老师

开题报告答辩小组负责人（签名）

2024 年 3 月 1 日

此页单面打印

浙江大学本科生毕业论文（设计）开题答辩记录表

学院：数学科学学院　　毕业届别：2024届

学生姓名	顾润哲	学号	3200103625	专业	统计学
毕业论文（设计）题目	在线错误发现率控制的共形推断方法				
指导教师姓名	孙文光	职 称	教授	所在单位	浙江大学数据科学研究中心
答辩时间	2024年3月1日		答辩地点	行政楼1417	
答辩组成员（签名）					

本科生毕业论文（设计）答辩记录：（要点：主答辩陈述和回答问题等方面具体加以记录与评价）

答辩组评审老师签字

Q. online 之前的

在已有的 online 的文献中，我们认为反复修改的决策是不合理的。在这里的研究中，我希望沿用之前的设定

Q. online 和 offline 条件的区别是什么？

在 offline 的情况下，我们知道需要拒绝的假设个数，我们可以根据 [illegible] 的统计量（p值）然后进行排序来决定拒绝的门槛

但在 online 条件下，我们不知道需要假设的个数，我们连排序统计量

答辩记录不可过于简单，记录人需签字，时间为实际答辩日期

记录人（签名）：彭鑫霖

2024 年 3 月 1 日

非指导老师

答辩小组负责人（签名）：

2024 年 3 月 1 日

实际答辩日期

数学科学学院本科毕业论文（设计）中期检查表

学号	3200103625	姓名	顾润哲	班级	统计2001
指导教师姓名	孙文光			职称	教授
论文题目	在线错误发现率控制的共形推断方法				

目前已完成任务

主要内容：（毕业论文进展情况，字数不少于800字）

已完成的任务主要分理论研究与实践操作两个方面展开，其中在理论研究方面，我首先完成了阅读孙老师发给我的各类参考文献的任务，在阅读的基础上完成了对过往文献的梳理，明确了在线假设检验基本设定和应用场景以及相关的用来衡量假设检验中第一类错误的错误发现率的度量。搞清楚了在线多重假设检验与一般的离线的多重假设检验的区别。在明确了问题的基本设定后，我接下来完成了对过往在线错误发现率控制方法的学习，对已经有了的阿尔法投资的方法如LORD、SAFFRON 和 SAST 等方法，我明白了他们控制错误发现率的思路，明确了他们各自的创新点，也发现了他们的缺点，即只能在大样本的前提下渐进地控制错误发现率，而无法，无法保证在有限样本的情况下控制错误发现率。在此基础上我以有限样本下控制错误发现率作为研究的出发点，来开始我的方法的设计。

为了实现错误发现率在有限样本下的控制，我开始借鉴目前统计推断领域比较前沿的共形推断方法。这种方法让数据不依赖于某个特定的参数模型，而是基于数据本身。它通过将一部分数据作为校准集来使辅助进行测试集数据的推断，在已有的文献中已经证明了共形推断是个能在有限样本的情况下控制错误发现率的方法，需要的前提仅仅是数据具有某种可交换性。但是对于这类方法的讨论仅仅停留于离线的多重假设检验，我试图将其改造成一个序列化的假设准则，使得在每个时间点都能够控制错误发现率。在共形推断的框架下，我们需要大量的数据来使得拒绝假设的门槛变得精准，为了解决这个问题，我引入了在线批次假设的设定，这样能够保证每个决策的时间点，我们都有足够的数据来决定拒绝门槛。我构建了拒绝假设的序列化准则的雏形，下一步需要确定一个关键参数来保证错误发现率控制的证明。

在实践操作方面，我进行了一些比较有效的数值模拟实验。我首先学习了 R 语言生成特定的模拟数据流的方式，并主动调用或是编写了过去的在线多重假设检验方法的程序，在自己生成的数据上进行了反复的实验，比较他们在这些数据上表现出的错误发现比例的大小。并在次基础上，编写了自己构建的序列化的共形推断方法的雏形的代码，后续只需要在此基础上进行修改即可。

是否符合任务书要求的进度（导师填写）	是

<table>
<tr><td>尚需完成的任务</td><td colspan="3">1. 需要完成错误发现率控制的证明。
2. 需要完成自己构建的方法的数值模拟。</td></tr>
<tr><td rowspan="2">存在的问题和解决办法</td><td>存在问题</td><td colspan="2">在错误发现率控制的证明中需要确定某个可迭代的参数来代替原来离线准则中的常数。这依赖于鞅论等工具。</td></tr>
<tr><td>采取的办法</td><td colspan="2">学习鞅的理论，并结合过去的论文进行大胆尝试。</td></tr>
<tr><td colspan="2">指导教师签字</td><td>日期</td><td>2024.3.26</td></tr>
<tr><td>检查小组意见</td><td colspan="3">抽查评分：☑优 ☐良 ☐合格 ☐不合格
意见和建议：
检查老师签字：</td></tr>
</table>

请正反面打印

论文中期检查小组负责人签字

浙江大学本科生毕业论文（设计）专家评阅意见

毕业论文（设计）题目	在线错误发现率的共形推断方法				
学生姓名	顾润哲	学 号	3200103625	年级	大四
所在学院	数学科学学院		专业	统计学	
指导教师姓名	孙文光	职 称	教授	所在单位	浙江大学数据科学研究中心

本科生毕业论文（设计）评阅意见：（对论文选题、文献综述、外文翻译、研究内容与方法、创新点、论文质量与理论水平、论文写作规范与文风和修改建议等方面加以评阅）

该论文选题具有一定的现实意义和学术价值。选题紧扣当前统计学界的热点问题，具有较高的研究价值。学生在选题过程中展示了对研究领域的深刻理解。

论文中外文翻译部分准确地传达了原文的意思，翻译流畅、用词准确，体现了较高的外语水平。翻译过程中对专业术语的处理恰当，没有明显的错误。

该论文研究方法选择得当并与研究目标紧密结合，并采用了多种研究方法，从而确保了研究结果的可靠性和有效性。

总体来说，论文质量与理论水平论文质量较高，理论水平较为扎实。

修改建议：进一步丰富文献综述，仔细检查论文的格式和排版，确保没有错漏之处。

同意答辩	√	同意修改后答辩		未达到答辩要求	
评阅人签名	宋逸凡 非指导教师	评阅人职称	研究员	评阅人单位	浙江大学数据科学研究中心

注：请评阅专家经综合评价后，在相应栏内打“√”。

此页单面打印

浙江大学本科生毕业论文（设计）现场答辩记录表

学院：数学科学学院　　毕业届别：2024

学生姓名	顾润哲	学号	3200103625	专业	统计学
毕业论文（设计）题目	在线错误发现率控制的共形推断方法				
指导教师姓名	孙文光	职 称	教授	所在单位	浙江大学数据科学研究中心
答辩时间	2024年5月24日		答辩地点	行政楼1312	
答辩组成员（签名）	林俊宏				

本科生毕业论文（设计）答辩记录：（要求在答辩陈述和回答问题等方面具体加以记录与评价）

答辩组评审老师签字

讲述了在线假设检[illegible]的共形推断方法，并能够控制错误发现率。

Q. 在问题设定中假定数据点服从 Efron's two group model 是否必要？

这是假设检验问题中最为简单的假设便于模拟实现。但实际的共形方法并不需要~~共形~~这一数据模型的条件，可以适用于 HMM 等数据条件。

[illegible]？如果估不准怎么办？

[illegible]的情况，我们的方法还能控制错[illegible]方法的效率，但不会影响有效性。

答辩记录不可过于简单，记录人需签字，时间为实际答辩日期

记录人（签名）：陈希

2024 年 5 月 24 日

非指导教师　答辩小组负责人（签名）：

2024 年 5 月 24 日

实际答辩日期

大学生论文检测系统

文本复制检测报告单(简洁)

№:ADBD2024R_2024031814543420240514151519475172857230　　检测时间:2024-05-14 15:15:19

篇名: **顾润哲毕业论文**
作者: 顾润哲
指导教师:
检测机构: 浙江大学
文件名: 通过API上传
检测系统: 大学生论文检测系统
检测类型: 大学生论文
检测范围: 中国学术期刊网络出版总库
中国博士学位论文全文数据库/中国优秀硕士学位论文全文数据库
中国重要会议论文全文数据库
中国重要报纸全文数据库
中国专利全文数据库
图书资源
优先出版文献库
大学生论文联合比对库
互联网资源(包含贴吧等论坛资源)
英文数据库(涵盖期刊、博硕、会议的英文数据以及德国Springer、英国Taylor&Francis 期刊数据库等)
港澳台学术文献库
互联网文档资源
源代码库
CNKI大成编客-原创作品库
时间范围: 1900-01-01至2024-05-13

检测结果

去除本人文献复制比: 5%　　跨语言检测结果: 0%
去除引用文献复制比: 5%　　总文字复制比: 5%
单篇最大文字复制比: 2.7% (肺结节的识别研究)

重复字数:	[1622]	总段落数:	[3]
总字数:	[32632]	疑似段落数:	[2]
单篇最大重复字数:	[892]	前部重合字数:	[459]
疑似段落最大重合字数:	[1163]	后部重合字数:	[1163]
疑似段落最小重合字数:	[459]		



指标: 疑似剽窃观点　✓ 疑似剽窃文字表述　疑似整体剽窃　过度引用

相似表格: 0　　相似公式: 没有公式　　疑似文字的图片: 0

4.5%(459)	4.5%(459)	顾润哲毕业论文_第1部分 (总10112字)
0%(0)	0%(0)	顾润哲毕业论文_第2部分 (总12620字)
11.7%(1163)	11.7%(1163)	顾润哲毕业论文_第3部分 (总9900字)

指导教师审查结果
指导教师:
审阅结果:
审阅意见: 指导老师未填

指导教师在此对查重情况签字确认

1. 顾润哲毕业论文_第1部分　　总字数：10112

相似文献列表

去除本人文献复制比：4.5%(459)　去除引用文献复制比：4.5%(459)　文字复制比：4.5%(459)　疑似剽窃观点：(0)

1	大规模同步推断中的稀疏性估计方法	3.7%（372）
	覃超 - 《大学生论文联合比对库》- 2023-05-21	是否引证：否
2	大规模同步推断中的稀疏性估计方法	3.7%（372）
	覃超 - 《大学生论文联合比对库》- 2023-05-25	是否引证：否
3	关于流体力学方程组的Liouville型问题研究	0.9%（89）
	石宇轩 - 《大学生论文联合比对库》- 2023-05-23	是否引证：否

2. 顾润哲毕业论文_第2部分　　总字数：12620

相似文献列表

去除本人文献复制比：0%(0)　去除引用文献复制比：0%(0)　文字复制比：0%(0)　疑似剽窃观点：(0)

3. 顾润哲毕业论文_第3部分　　总字数：9900

相似文献列表

去除本人文献复制比：11.7%(1163)　去除引用文献复制比：11.7%(1163)　文字复制比：11.7%(1163)　疑似剽窃观点：(0)

1	肺结节的识别研究	9.0%（892）
	孔睿至 - 《大学生论文联合比对库》- 2018-05-23	是否引证：否
2	随机梯度算法用于学习理论	8.6%（854）
	沈悦 - 《大学生论文联合比对库》- 2019-05-30	是否引证：否
3	波动率模型下KOSPI200期权定价实证研究	8.6%（854）
	尹琇映 - 《大学生论文联合比对库》- 2022-05-16	是否引证：否
4	BAB因子在中国股票市场的实证分析	8.6%（854）
	朱成浩 - 《大学生论文联合比对库》- 2022-05-16	是否引证：否
5	粉丝杂质检验的计算机视觉方法	8.6%（854）
	许乐乐 - 《大学生论文联合比对库》- 2022-05-16	是否引证：否
6	基于广义重心坐标的平面图案设计	8.6%（853）
	庄桁 - 《大学生论文联合比对库》- 2020-05-26	是否引证：否
7	一类投影算子的能量谱改进	8.6%（852）
	龚学平 - 《大学生论文联合比对库》- 2019-05-16	是否引证：否
8	医学影像精准分析中的配准问题研究	8.5%（841）
	佚名 - 《大学生论文联合比对库》- 2021-05-18	是否引证：否
9	Lubin Tate形式群与局部类域论	8.4%（827）
	陈琦元 - 《大学生论文联合比对库》- 2021-05-23	是否引证：否
10	两相动力学自由界面问题的适定性研究	8.3%（826）
	唐宁 - 《大学生论文联合比对库》- 2022-05-16	是否引证：否
11	浙大数学系1932级同学研究	7.9%（779）
	赵浩江 - 《大学生论文联合比对库》- 2022-05-16	是否引证：否
12	全纯分歧覆盖的研究	6.0%（598）
	许煜辉 - 《大学生论文联合比对库》- 2022-05-16	是否引证：否
13	大规模同步推断中的稀疏性估计方法	5.1%（509）
	覃超 - 《大学生论文联合比对库》- 2023-05-21	是否引证：否
14	大规模同步推断中的稀疏性估计方法	5.1%（509）
	覃超 - 《大学生论文联合比对库》- 2023-05-25	是否引证：否

15	圆纹曲线曲面造型	2.4%（237）
	李瑞祥 - 《大学生论文联合比对库》- 2022-05-16	是否引证：否
16	3160101779_马存_知识图谱在中小企业信贷风控中的应用	2.4%（237）
	马存 - 《大学生论文联合比对库》- 2021-01-15	是否引证：否
17	3160101779_马存_知识图谱在中小企业信贷风控中的应用	2.4%（237）
	马存 - 《大学生论文联合比对库》- 2020-05-28	是否引证：否
18	3140102558_冯烨_基于深度学习的机器阅读理解系统	2.4%（237）
	冯烨 - 《大学生论文联合比对库》- 2018-06-01	是否引证：否

说明：1. 总文字复制比：被检测论文总重合字数在总字数中所占的比例

2. 去除引用文献复制比：去除系统识别为引用的文献后，计算出来的重合字数在总字数中所占的比例

3. 去除本人文献复制比：去除作者本人文献后，计算出来的重合字数在总字数中所占的比例

4. 单篇最大文字复制比：被检测文献与所有相似文献比对后，重合字数占总字数的比例最大的那一篇文献的文字复制比

5. 复制比：按照“四舍五入”规则，保留1位小数

6. 指标是由系统根据《学术论文不端行为的界定标准》自动生成的

7. 红色文字表示文字复制部分；绿色文字表示引用部分(包括系统自动识别为引用的部分)；棕灰色文字表示系统依据作者姓名识别的本人其他文献部分

8. 本报告单仅对您所选择的比对时间范围、资源范围内的检测结果负责

amlc@cnki.net
https://check.cnki.net/

知网大学生论文检测系统

此页单面打印

浙江大学数学科学学院本科生毕业论文（设计）提交前检查表

论文题目	在线错误发现受控制的共形推断方法					
姓名	顾润哲	学号	3200103625	导师	孙文光	
检查内容	检查重点				学生自查（打√或×）	导师自查（打√或×）
一、教师课题申报	1、是否一学生一题(本届学生不重题，与往届学生也不重题)				√	√
	2、是否有项目（课题）的详细介绍和具体要求（字数一般不少于 200 字）				√	√
	3、课题是否经院系审定批准				√	√
二、教师任务下达	1、任务书是否填写完整、详细，对毕业论文各环节时间结点有明确要求				√	√
	2、任务书是否有指导老师签名、学院审核签名				√	√
	3、文献综述、开题报告要求是否填写详细（字数一般不少于各 150 字），签名和时间是否正确（时间是否早于开题时间）				√	√
三、文献综述、开题报告、外文文献翻译	1、文献查阅是否 8 篇以上（外文不少于 3 篇），文献综述是否切题，内容完整，字数 3000 以上				√	√
	2、开题报告是否符合要求，内容完整，字数 3500 以上				√	√
	3、文献引用是否规范，正文中引用到的文献是否都列在参考文献中，参考文献中的文献是否在论文中引用，文献是否依照数学论文规范按作者字母次序排列，文献作者名、论文名、期刊名、期刊号、出版年是否符合规范				√	√
	4 外文翻译的译文是否 3000 字以上				√	√
	5、译文是否规范，是否标注标题、原作者姓名，外文翻译是否提供原文及原文的出处（如发表的期刊、论著的章节等）				√	√
四、开题答辩与考核	1、答辩考核评语是否详细完整（字数一般 200 字以上），成绩评定是否与评语相符，答辩考核时间是否正确				√	√
	2、答辩考核小组组长（不能是指导老师）是否签名				√	√
五、毕业论文（设计）	1、毕业论文（设计）格式是否规范；				√	√
	2、毕业论文（设计）研究内容是否与开题报告一致				√	√
	3、前人成果是否明确标注，成果的所有人、作者引用是否规范，没有抄袭现象				√	√
六、毕业论文（设计）考核	1、指导老师考核评语是否详细、完整（字数一般 200 字以上），是否签名，签名时间是否正确（早于答辩时间）				√	√
	2、是否有专家评阅意见，评阅意见是否详细、完整，是否在“同意答辩”等处打 √ 确定，专家（不能是指导老师）是否签名，签名时间是否正确（早于答辩时间）				√	√
	3、答辩记录是否完整、规范，是否详细记录了答辩过程，包括问题及回答情况。是否有答辩委员会名单，是否有记录人、答辩考核小组签名，签名时间是否正确				√	√
	4、答辩考核小组的考核评语是否详细、完整（字数一般 200 字以上），成绩评定是否与评语相符				√	√
	5、答辩考核小组组长（不能是导师）是否签名				√	√

学生签名：顾润哲　　日期：2024年5月27日

导师签名：孙文光　　日期：2024年5月27日

学院复核时间：2024.5.28

签字，时间为答辩后