
μ 子探测实验数据处理补充材料



地点: 物理楼 &220

更新时间: March 23, 2019

邮箱: zhaor25@mail2.sysu.edu.cn

版本: 1.0

目 录



1	数据处理背景简介	1
1.1	读取波形	1
1.2	未来更新计划	1
2	ElegantBook 设置说明	2
2.1	编译方式	2
2.2	选项设置	2
2.3	数学环境简介	3
2.4	可编辑的字段	3
2.5	参考文献	3
3	ElegantBook 写作示例	4
3.1	Lebesgue 积分	4
3.1.1	积分的定义	4
A	线性代数	8
A.1	矩阵的分块	8

第 1 章 数据处理背景简介



实验过程中我们给 PMT 设置合理的高压，设置合理的 ADC 采集参数，通过 ADC 模块采集数据并存储到一个 lvm 格式的文件。这个文件存储了大量（取决于你设置的事件数）的波形信息，我们需要计算每一个信号波形的积分面积，并算出此工作电压下 PMT 的增益大小。

由于我们要处理的波形数量比较多，需要进行简单的编程来计算和统计；python, MATLAB, C++ 等都可以满足需求¹。

1.1 读取波形

我们存储下来的 lvm 文件每行只有一个数字，代表一个电压值（单位 mV）；如果你在 labview 中设置采样长度是 1000，那么每 1000 行就组成一个波形；每一个点的时间间隔是 1ns。

1. 删除了自定义字体设置，改用 `ctex` 宏包用以支持中文；
2. 使用 `tcolorbox` 宏包改写了原先的定理类环境，目前定理环境等均可以跨页；
3. 重新命名了定理类环境的名称（`theorem`, `definition`, `proposition` 等）；
4. 重新命名了颜色的名字（`first`, `second`, `third`）；
5. 重新绘制 `ElegantLaTeX` 的 Logo；
6. 更新类定理环境，包括其交叉引用；
7. 特别修正附录相关内容。

1.2 未来更新计划

1. 重新设计 base 图标；
2. 重新设计一个好看的封面。
3. 将基础模板改为英文模板，添加选项使其支持中文。
4. 添加一个简化风格（`plain`）的颜色主题。

¹可以参考附录的 python, ROOT 程序

第 2 章 ElegantBook 设置说明



2.1 编译方式

本模板基于基础的 book 文类，所以 book 的选项对于本模板也是有效的。默认编码为 UTF-8，推荐使用 T_EX Live 编译。作者编写环境为 Win10(64bit) + T_EX Live 2018。由于使用的是 ctex 宏包，所以支持 pdf_latex 以及 xelatex 编译。










2.2 选项设置

本文特殊选项设置共有 2 类，分为主题颜色设置以及章标题显示风格设置。

第 1 类为主题颜色设置，内置 3 组颜色主题，分别为 green（默认）、cyan、blue，另外还有一个自定义的选项 nocolor。调用颜色主题 green 的方法为 `\documentclass[green]{elegantbook}` 或者 `\documentclass[color=green]{elegantbook}`。需要改变颜色的话请选择 nocolor 选项或者使用 `color=none`，然后在导言区定义 main、second、third 颜色，具体的方法如下：

```
\definecolor{main}{RGB}{70,70,70}    %定义 main 颜色值
\definecolor{second}{RGB}{115,45,2}   %定义 second 颜色值
\definecolor{third}{RGB}{0,80,80}     %定义 third 颜色值
```

表 2.1: ElegantBook 模板中的三套颜色主题

	green	cyan	blue	主要使用的环境
main				definition
second				theorem lemma corollary
third				proposition

第 2 类为章标题显示风格，包含 hang（默认）与 display 两种风格，区别在于章标题单行显示（hang）与双行显示（display），本说明使用了 hang。调用方式为 `\documentclass[hang]{elegantbook}` 或者 `\documentclass[titlestyle=hang]{elegantbook}`。

综合起来，同时调用三个选项使用 `\documentclass[color=X,titlestyle=Y]{elegantbook}`。其中 X 可以选择 green,cyan,blue,none；Z 可以选择 hang 或者 display。

2.3 数学环境简介

在我们这个模板中，定义了三大类环境

1. 定理类环境，包含标题和内容两部分。根据格式的不同分为 3 种
 - **definition** 环境，含有一个可选项，编号以章节为单位，颜色为main；
 - **theorem**、**lemma**、**corollary** 环境，颜色为主颜色second，编号均以章节为单位；
 - **proposition** 环境，含有一个可选项，编号以章节为单位，颜色为third。
2. 示例类环境，有 **example**、**exercise** 环境，自动编号，编号以章节为单位。
3. 证明类环境，有 **proof**、**note** 环境，特点是，有引导符或者结尾符，**note** 环境有引导符号，**proof** 环境有证明完毕标志。
4. 结论类环境，有 **conclusion**、**assumption**、**property**，**remark**、**solution** 环境，三者均以粗体的引导词为开头，和普通段落格式一致。

2.4 可编辑的字段

在模板中，可以编辑的字段分别为作者 \author、邮箱 \email、中文标题 \zhtitle、中文标题结尾 \zhend、英文标题 \entitle、英文标题结尾 \enend、名言 \myquote、版本号 \version。并且，可以根据自己的喜好把封面水印效果的 cover.pdf 替换掉，以及封面中用到的 logo.png。

2.5 参考文献

此模板使用了 BibTeX 来生成参考文献，默认使用的文献样式 aer 样式。参考文献示例：Chen et al. (2018) 使用了中国一个大型的 P2P 平台（人人贷）的数据来检验男性投资者和女性投资者在投资表现上是否有显著差异。你可以在谷歌学术，Mendeley，Endnote 中获得文献条目（bib item），然后把它们添加到 reference.bib 中。在文中引用的时候，引用它们的键值（bib key）即可。注意需要在编译的过程中添加 BibTeX 编译。



第 3 章 ElegantBook 写作示例



3.1 Lebesgue 积分

在前面各章做了必要的准备后,本章开始介绍新的积分。在 Lebesgue 测度理论的基础上建立了 Lebesgue 积分,其被积函数和积分域更一般,可以对有界函数和无界函数统一处理。正是由于 Lebesgue 积分的这些特点,使得 Lebesgue 积分比 Riemann 积分具有在更一般条件下的极限定理和累次积分交换积分顺序的定理,这使得 Lebesgue 积分不仅在理论上更完善,而且在计算上更灵活有效。

Lebesgue 积分有几种不同的定义方式。我们将采用逐步定义非负简单函数,非负可测函数和一般可测函数积分的方式。

由于现代数学的许多分支如概率论、泛函分析、调和分析等常常用到一般空间上的测度与积分理论,在本章最后一节将介绍一般的测度空间上的积分。

3.1.1 积分的定义

我们将通过三个步骤定义可测函数的积分。首先定义非负简单函数的积分。以下设 E 是 \mathcal{R}^n 中的可测集。

定义 3.1: 可积性

设 $f(x) = \sum_{i=1}^k a_i \chi_{A_i}(x)$ 是 E 上的非负简单函数,其中 $\{A_1, A_2, \dots, A_k\}$ 是 E 上的一个可测分割, a_1, a_2, \dots, a_k 是非负实数。定义 f 在 E 上的积分为

$$\int_E f dx = \sum_{i=1}^k a_i m(A_i). \quad (3.1)$$

一般情况下 $0 \leq \int_E f dx \leq \infty$ 。若 $\int_E f dx < \infty$, 则称 f 在 E 上可积。



一个自然的问题是, Lebesgue 积分与我们所熟悉的 Riemann 积分有什么联系和区别? 在 4.4 在我们将详细讨论 Riemann 积分与 Lebesgue 积分的关系。这里只看一个简单的例子。设 $D(x)$ 是区间 $[0, 1]$ 上的 Dirichlet 函数。即 $D(x) = \chi_{Q_0}(x)$, 其中 Q_0 表示 $[0, 1]$ 中的有理数的全体。根据非负简单函数积分的定义, $D(x)$ 在 $[0, 1]$ 上的 Lebesgue 积分为

$$\int_0^1 D(x) dx = \int_0^1 \chi_{Q_0}(x) dx = m(Q_0) = 0 \quad (3.2)$$

即 $D(x)$ 在 $[0, 1]$ 上是 Lebesgue 可积的并且积分值为零。但 $D(x)$ 在 $[0, 1]$ 上不是 Riemann 可积的。

有界变差函数是与单调函数有密切联系的一类函数。有界变差函数可以表示为两个单调递增函数之差。与单调函数一样，有界变差函数几乎处处可导。与单调函数不同，有界变差函数类对线性运算是封闭的，它们构成一线空间。练习题 3.1 是一个性质的证明。

练习 3.1: 设 $f \in L(\mathcal{R}^1)$, g 是 \mathcal{R}^1 上的有界可测函数。证明函数

$$I(t) = \int_{\mathcal{R}^1} f(x+t)g(x)dx \quad t \in \mathcal{R}^1 \quad (3.3)$$

是 \mathcal{R}^1 上的连续函数。

定理 3.1: Fubini 定理

(1) 若 $f(x, y)$ 是 $\mathcal{R}^p \times \mathcal{R}^q$ 上的非负可测函数，则对几乎处处的 $x \in \mathcal{R}^p$, $f(x, y)$ 作为 y 的函数是 \mathcal{R}^q 上的非负可测函数， $g(x) = \int_{\mathcal{R}^q} f(x, y)dy$ 是 \mathcal{R}^p 上的非负可测函数。并且

$$\int_{\mathcal{R}^p \times \mathcal{R}^q} f(x, y)dx dy = \int_{\mathcal{R}^p} \left(\int_{\mathcal{R}^q} f(x, y)dy \right) dx. \quad (3.4)$$

(2) 若 $f(x, y)$ 是 $\mathcal{R}^p \times \mathcal{R}^q$ 上的可积函数，则对几乎处处的 $x \in \mathcal{R}^p$, $f(x, y)$ 作为 y 的函数是 \mathcal{R}^q 上的可积函数，并且 $g(x) = \int_{\mathcal{R}^q} f(x, y)dy$ 是 \mathcal{R}^p 上的可积函数。而且 3.4 成立。



注意: 在本模板中，引理 (lemma)，推论 (corollary) 的样式和定理 3.1 的样式一致，包括颜色，仅仅只有计数器的设置不一样。

我们说一个实变或者复变量的实值或者复值函数是在区间上平方可积的，如果其绝对值的平方在该区间上的积分是有限的。所有在勒贝格积分意义下平方可积的可测函数构成一个希尔伯特空间，也就是所谓的 L^2 空间，几乎处处相等的函数归为同一等价类。形式上， L^2 是平方可积函数的空间和几乎处处为 0 的函数空间的商空间。

命题 3.1: 最优性原理

如果 u^* 在 $[s, T]$ 上为最优解，则 u^* 在 $[s, T]$ 任意子区间都是最优解，假设区间为 $[t_0, t_1]$ 的最优解为 u^* ，则 $u(t_0) = u^*(t_0)$ ，即初始条件必须还是在 u^* 上。

我们知道最小二乘法可以用来处理一组数据，可以从一组测定的数据中寻求变量之间的依赖关系，这种函数关系称为经验公式。本课题将介绍最小二乘法的精确定义及如何寻求点与点之间近似成线性关系时的经验公式。假定实验测得变量之间的 n 个数据，则在平面上，可以得到 n 个点，这种图形称为“散点图”，从图中可以粗略看出这些点大致散落在某直线近旁，我们认为其近似为一线性函数，下面介绍求解步骤。



考虑函数 $y = a + bx$, 其中 a 和 b 是待定常数。如果离散点完全的在一直线上, 可以认为变量之间的关系为一元函数。但一般说来, 这些点不可能在同一直线上。但是它只能用直线来描述时, 计算值与实际值会产生偏差。当然要求偏差越小越好, 但由于偏差可正可负, 因此不能认为总偏差时, 拟合函数很好地反映了变量之间的关系, 但是因为此时每个偏差的绝对值可能很大。为了改进这一缺陷, 就考虑用平均值来代替。但是由于绝对值不易作解析运算, 因此, 进一步用残差平方和函数来度量总偏差。偏差的平方和最小可以保证每个偏差都不会很大。于是问题归结为确定拟合函数中的常数和使残差平方和函数最小。

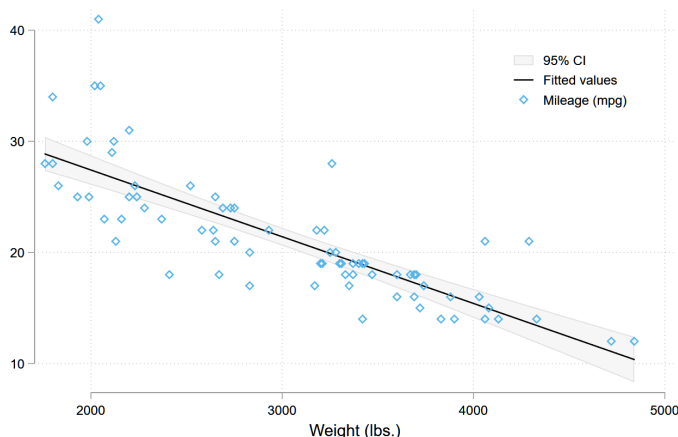


图 3.1: MPG 和 Weight 的关系图

以最简单的一元线性模型来解释最小二乘法。什么是一元线性模型呢? 监督学习中, 如果预测的变量是离散的, 我们称其为分类 (如决策树, 支持向量机等), 如果预测的变量是连续的, 我们称其为回归。回归分析中, 如果只包括一个自变量和一个因变量, 且二者的关系可用一条直线近似表示, 这种回归分析称为一元线性回归分析。如果回归分析中包括两个或两个以上的自变量, 且因变量和自变量之间是线性关系, 则称为多元线性回归分析。对于二维空间线性是一条直线; 对于三维空间线性是一个平面, 对于多维空间线性是一个超平面。

性质: 柯西列的性质

1. $\{x_k\}$ 是柯西列, 则其子列 $\{x_k^i\}$ 也是柯西列。
2. $x_k \in \mathcal{R}^n$, $\rho(x, y)$ 是欧几里得空间, 则柯西列是收敛的, (\mathcal{R}^n, ρ) 空间是完备的。

结论: 回归分析 (regression analysis) 是确定两种或两种以上变量间相互依赖的定量关系的一种统计分析方法。运用十分广泛, 回归分析按照涉及的变量的多少, 分为一元回归和多元回归分析; 按照因变量的多少, 可分为简单回归分析和多重回归分析; 按照自变量和因变量之间的关系类型, 可分为线性回归分析和非线性回归分析。如果在回归分析中, 只包括一个自变量和一个因变量, 且二者的关系可用一条直线近似表示, 这种回归分析称为一元线性回归分析。如果回归分析中包括两个或两个以上的自变量, 且自变量之间存在线性相关, 则称为多重线性回归分析。



参考文献



- Chen, Jia, Jiajun Jiang, and Yu jane Liu**, “Financial Literacy and Gender Difference in Loan Performance,” *Journal of Empirical Finance*, 2018, 48 (71673007), 307–320.
- Einav, Liran and Amy Finkelstein Mark R. Cullen**, “Estimating Welfare in Insurance Markets Using Variation in Prices,” *Quarterly Journal of Economics*, 2010, CXXV (August), 877–921.
- Havrylchyk, Olena and Marianne Verdier**, “The Financial Intermediation Role of the P2P Lending Platforms,” *Comparative Economic Studies*, 2018, 60 (1), 115–130.

附录 线性代数



A.1 矩阵的分块

$$\begin{vmatrix} A_1 & 0 \\ C & A_2 \end{vmatrix} = |A_1| \times |A_2| \quad (\text{A.1})$$

由矩阵 A 的若干行、若干列的交叉位置元素按照原来顺序排成的矩阵称为 A 的一个子矩阵。