

清华大学  
高能物理中心

《高能物理中的数据分析》  
阅读笔记和习题解答

学习小组

二〇一七年夏

目 录

序言 .....	III
第 1 章 基本概念 .....	1
1.1 练习 .....	1
第 2 章 参数估计 .....	2
2.1 练习 .....	2
第 3 章 假设检验 .....	3
3.1 练习 .....	3
第 4 章 区间估计 .....	4
4.1 练习 .....	4
第 5 章 分类 .....	5
5.1 练习 .....	5
第 6 章 解谱法 .....	6
6.1 练习 .....	6
第 7 章 约束拟合 .....	7
7.1 练习 .....	7
第 8 章 如何处理系统不确定度 .....	8
8.1 练习 .....	8
第 9 章 理论不确定度 .....	9
9.1 练习 .....	9
第 10 章 高能物理领域常用的统计方法 .....	10
10.1 练习 .....	10
第 11 章 分析实例 .....	11
11.1 练习 .....	11

## 序言

在当今时代，随着科学研究对精确度、准确度要求的逐渐提高，统计推断在科研中正扮演着日益重要的角色。事实上，许多研究结果的获取，可以并且只能借助于复杂精妙的统计学手段。这一现象在高能物理实验领域当中体现的尤为明显，在我们的数据分析工作中，统计推断的知识无处不在。

近些年来，新的统计手段和以其为基础的复杂的软件工具如雨后春笋般地出现，这一方面为高能物理实验的数据分析工作提供了极大的便利，另一方面，其背后所隐含的复杂精巧的统计学原理则对实验人员的知识水平提出了更高的要求，这也是本书被撰写的目的所在——即为高能物理的从业人员提供统计推断方面的知识基础。本书面向的对象包括从学生到教授各个知识层面的研究人员，旨在对其高能物理数据分析上的工作提供全面而有具有可操作性的建议。为此，全书分为十二个章节，每个章节具有一个明确的主题，并由该方面的专家进行撰写，具体说明如下：

第一章里，我们将为读者介绍统计分析最为基本的概念，例如概率密度函数及其性质，常见的分布类型（高斯分布、泊松分布等），以及“概率”这一概念的两种不同含义——频率论述和贝叶斯论述。

接下来，我们将利用三个章节的篇幅，讲解如何利用统计学知识，从实验数据中提取各种信息：“参数估计”主要介绍如何通过拟合实验数据，得到模型参数的最优估计值，例如信号强度的估计等等；“假设检验”主要介绍如何评估某一假设正确与否，例如评估某一数据集究竟应该由已知本底的统计涨落来解释，还是其中真正包含了我们所关心的物理信号；“区间估计”则主要讨论如何给出参数的置信区间，例如信号强度的上限等。

在此之后，我们将对一些常见的话题进行更加深入的讨论。“分类”一章将会介绍数种不同的方法，以更好地进行事例区分，例如利用多变量分析的手段从本底样本中进行信号提取。这些方法往往能够非常有效地提高测量的灵敏程度，在某些情况下，能够借此观察到传统的事例区分方法下无法观测到的信号。“解谱法”一章中，我们将讨论在估计了各种系统偏差之后，如何对数据进行修正使之恢复原貌。这种方法在对微分分布的测量中具有较多的应用。“约束拟合”中，我们将说明如何在测量中引入物理定律所决定的约束条件，用来提高测量精度，或是估计未知的参量。

系统误差的估计往往是数据分析的最后一步，也往往是最关键的步骤之一。因

此，在接下来的两个章节中，我们将着重介绍系统误差这一在其他教材中往往被忽视的话题。“如何处理不确定度”一章将介绍如何估计各种来源的系统误差及其对最终结果的影响，以及如何避免这一影响。“理论不确定度”介绍了以强相互作用为代表的各种理论误差的处理方式。

行百里路者半九十。本书的最后三个章节，将为读者介绍这些统计学知识在实际的高能物理学中的具体应用。第十章，“高能物理领域常用的统计方法”，将为读者介绍多种在实践中常用的方法，如样本成分估计中常用的样板矩阵（卧槽这里怎么这么奇怪）法，以及用于估计分析过程导致的系统偏差的综合测试法。“分析实例”一章则主要列举了包含新粒子的寻找和新粒子性质测量在内的两个实际的物理分析，并以此为依托综述全书。最后一章“天文学中的应用”，将带领读者一探数据分析技巧在天文学中的巧妙应用。

在撰写各个章节的过程中，我们都尽可能地插入实例，以保证内容的实际、具体。通过完成每个章节后的习题，读者可以对本书中的材料有更加准确、深刻的理解。习题的提示和答案，以及一些必要的软件，都可以在出版社提供的网站上找到。此外，也欢迎读者们向本书的作者提供与本书相关的各类反馈。详情请访问 [www.wiley.com](http://www.wiley.com)，屠龙宝刀，点击就送。

在此，我们要对在成书过程中提供帮助的团体和个人表示由衷的感谢。首先，要感谢每个章节的撰写者。其次，要感谢在编辑过程中参与讨论的大量同事们，但由于人数众多，篇幅所限，并不能一一列举：感谢 **Katarina Brock** 对图片的调整和排版；感谢来自 **Wiley** 的 **Konrad Kielsing** 对本书的排版；感谢同样来自 **Wiley** 的 **Vera Palmer** 和 **Ulrike Werner** 在成书过程中的持续帮助。另外，还要感谢 **Tatsuya Nakada** 对我们使用他的习题材料的许可。

最后，真诚地感谢我们的朋友、伴侣和家人。没有你们的支持，这本书也将无法问世。

## 主要符号对照表

缩写

全称

## 第 1 章 基本概念

### 1.1 练习

## 第 2 章 参数估计

### 2.1 练习

## 第 3 章 假设检验

### 3.1 练习



## 第 4 章 区间估计

### 4.1 练习

## 第 5 章 分类

### 5.1 练习

## 第 6 章 解谱法

### 6.1 练习

## 第 7 章 约束拟合

### 7.1 练习

## 第 8 章 如何处理系统不确定度

### 8.1 练习

## 第 9 章 理论不确定度

### 9.1 练习

## 第 10 章 高能物理领域常用的统计方法

### 10.1 练习

## 第 11 章 分析实例

### 11.1 练习