

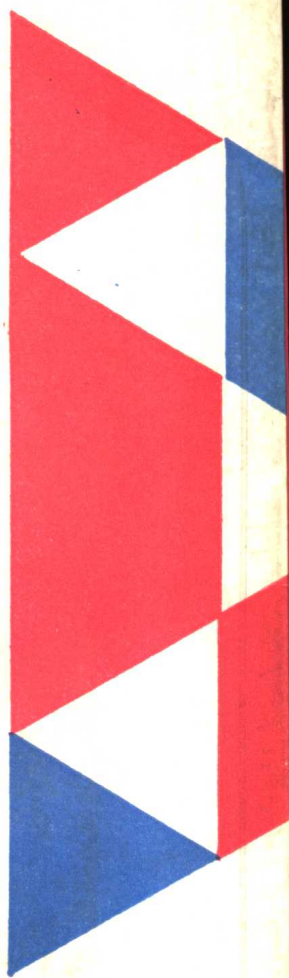
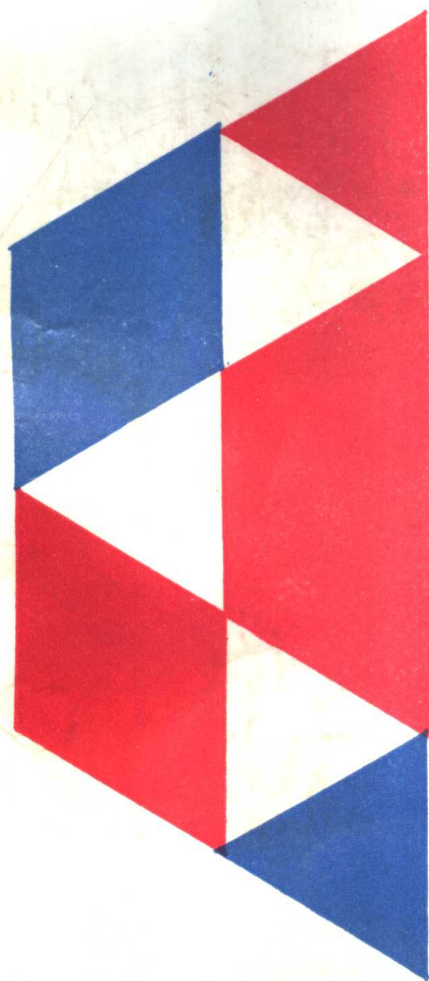


经济
数学
基础

概率统计



成世学 严颖 张诒兰等 编著



经济数学基础（三）

概 率 统 计

成世学 严 颖 张诒兰 等编著

中国人民大学出版社

(京) 新登字 156 号

经济数学基础 (三)

概 率 统 计

成世学 严 颖 张诒兰 等编著

出版者：中国人民大学出版社

发行者：中国人民大学出版社

(北京海淀路 39 号 邮码 100872)

印刷者：中国人民大学出版社印刷厂

经销者：新华书店总店北京发行所

开 本：850×1168 毫米 32 开

字 数：342 000

印 张：13.75

版 次：1994 年 1 月第 1 版

印 次：1994 年 1 月第 1 次印刷

册 数：1-31 000

书 号：ISBN7-309-01605-7/O·30

定 价：9.75 元

前 言

《概率统计》是作为中国人民大学财经与管理各专业本科生教材而写的，还可作为同类院校师生的教学参考书。本书着眼于管理科学的发展前景，为适应培养高水平管理人才的需要，除了注重概率统计的基础知识以外，还纳入了有着广泛应用的非参数统计方法、贝叶斯方法以及多元统计分析的部分内容，并为多元统计部分配备了教学软件“PUSE”（见书后说明书），以改变我国概率统计教学与计算机脱离的现状。在设备允许的情况下，可以尝试把概率统计教学与微机使用结合起来。在没有微机的情况下，本教材照样可以使用（计算量过大的部分可以只介绍方法而不进行计算）。

在讲述方面，概率论部分我们着力讲清重要概念；在统计部分，强调统计思想而往往省略数学推导。

凡标*的章、节、段可根据教学需要加以选用。

本书是由龚德恩任主编，张学贞、胡显估、严颖任副主编的《经济数学基础》系列教材中的第三分册。本分册由严颖主编，张治兰负责组织软件开发与调试。参加本册编写与软件开发的有：成世学（第1—4章）、袁荫棠（第5章）、严颖（第6—8章）、纪元庆（第9章及其软件）、赵晋（第10章及其软件）、张治兰（第11章及其软件以及PUSE说明书）、周邦珞（第12章及其软件）。全书由严颖负责统纂。

编 者

1991. 3.

目 录

* 第一章 引论	1
§ 1. 1 概率论与数理统计发展简史	1
§ 1. 2 数理统计学的基本内容与应用	12
第二章 数据的初步处理	21
§ 2. 1 引言	21
§ 2. 2 频率分布与累积频率分布	24
§ 2. 3 数字特征	36
习题二	43
第三章 随机事件与概率	46
§ 3. 1 随机事件	46
§ 3. 2 随机事件的概率	53
* § 3. 3 概率论的公理化结构	65
§ 3. 4 条件概率、乘法公式与全概公式	69
§ 3. 5 事件的独立性与贝努里概型	76
习题三	84
第四章 随机变量及其分布	89
§ 4. 1 引言	89
§ 4. 2 离散型随机变量及其分布	91
§ 4. 3 连续型随机变量及其分布	97
§ 4. 4 随机变量的数字特征	113
§ 4. 5 大数定律及中心极限定理	128
习题四	135

第五章	随机向量	139
§ 5. 1	多元分布	139
§ 5. 2	边缘分布与条件分布	144
§ 5. 3	随机变量的独立性	157
§ 5. 4	随机向量的数字特征	160
	习题五	167
第六章	参数估计	171
§ 6. 1	引言	171
§ 6. 2	矩估计法	177
§ 6. 3	极大似然估计法	178
* § 6. 4	贝叶斯估计法	183
	习题六	186
第七章	假设检验与区间估计	188
§ 7. 1	假设检验的概念	188
§ 7. 2	抽样分布	193
§ 7. 3	一个正态总体的假设检验	197
§ 7. 4	两个正态总体的假设检验	202
§ 7. 5	正态总体参数的区间估计	208
	习题七	211
* 第八章	非参数假设检验	214
§ 8. 1	χ^2 拟合优度检验	214
§ 8. 2	独立性的 χ^2 检验	219
§ 8. 3	柯尔莫哥洛夫-斯米尔诺夫检验	221
§ 8. 4	秩和检验	225
§ 8. 5	配对观测数据的检验问题	230
	习题八	235
第九章	方差分析	238
§ 9. 1	单因素方差分析	238

* § 9. 2	双因素方差分析	246
	习题九	256
第十章	回归分析	263
§ 10. 1	一元线性回归	264
* § 10. 2	多元回归分析	277
* § 10. 3	逐步回归分析	289
	习题十	308
* 第十一章	判别分析	311
§ 11. 1	距离判别	312
§ 11. 2	费希尔判别	317
§ 11. 3	贝叶斯判别	321
§ 11. 4	逐步判别	329
	习题十一	340
* 第十二章	相关分析	344
§ 12. 1	主成分分析	344
§ 12. 2	典型相关分析	355
	习题十二	366
附录	PUSE 软件说明书	373
	习题答案	395
	附表 1—7	416
	参考文献	431

* 第一章 引 论

概率论和数理统计都是应用数学的重要分支，但就它们二者之间的关系来说，通常认为概率论是数理统计的基础，数理统计则是概率论的一种应用。本书侧重于数理统计部分，致力于介绍有实际应用价值的一些统计推断方法。概率论部分则为这些统计上的应用提供必要的基本知识。在接触本学科的基本理论与方法之前，了解其发展简史、特点以及应用情况，有助于对本学科的深入领会与实际运用。这就是本章要讲述的内容。

§ 1. 1 概率论与数理统计发展简史

一、概率论发展简史

我们将概率论的发展历史大致上划分为以下五个阶段。

1. 概率论发展的原始阶段

这一阶段很难确定它发展的起端，有些学者把它的发展追溯到人类最早期的文明史。在史前的圣堂和神殿中曾经发现过一些用马的腿骨制成的小立方体，这表明在早期的宗教仪式中可能使用过骰子。至 15 世纪，在意大利兴起了航运保险事业，这说明那时人们至少已有了概率论的初步知识。

2. 概率论发展的早期阶段

这一阶段大体上从 17 世纪中叶开始，至 18 世纪初期结束。主要的代表人物是法国的巴斯加、费尔马和荷兰的惠更斯。突出的事件是概率论在赌博问题中的应用。1654 年，一些热衷于赌博

的人向巴斯加尔提出了还不能归入那时数学范畴的划分赌本的问题，这引起了巴斯加尔的兴趣。他把解法写信告诉了费尔马，从而二人采用通信的方式展开了讨论。现在这些信件被公认为是概率论中最早有详尽文字记载的文献。基于这一原因，人们通常把这看成是概率论发展的起点。在费尔马和巴斯加尔的讨论中，已提炼出了像概率和数学期望这样一些概率论中最重要的概念。稍后，另一位年轻的荷兰人惠更斯也独立地讨论了类似的赌博问题，并于1655年发表了概率论中的第一篇论文《赌博中的计算》。这一阶段，由于科学发展水平不高，所以充满机遇性的赌博便成为那时发展概率论的直观背景。解决概率论中问题时用到的数学工具也仅是初等的排列和组合算法。这里我们需要强调一下，尽管概率论的早期应用主要限于赌博，但在讨论赌博问题中发展起来的 思想却对于认识随机现象有着普遍的意义。上面提及的几位数学家在当时就已觉察到了这点。惠更斯在《赌博中的计算》一文中写到：“我相信如果读者仔细地研究这些材料，当可发现你所处理的不仅仅是赌博，其中实际上包含了既深刻又有趣的一种新理论的基础。”概率论以后发展的历史确实验证了惠更斯的预言。

3. 概率论发展的中期阶段

这一阶段从18世纪初期开始，一直延续到19世纪中叶。这一阶段中，概率论的应用已不局限于赌博，而是扩展到观察误差理论、射击学、人口统计学等众多领域。解析的方法也被引入到概率论的研究之中。这一时期的代表人物是以拉普拉斯为首的一批法国数学家，其中比较杰出的还有贝努里、德莫哇佛尔和普阿松等人。此外，英国的贝叶斯和德国著名数学家高斯对于概率论发展作出的贡献也不容忽视。

贝努里的遗作《猜测的艺术》于1713年发表，这一事件被认为是概率论这一发展阶段的开端。现在称之为贝努里大数定律的

概率论中的第一个极限定理便是首次在此书中刊出的。

拉普拉斯领导了概率论在这一发展阶段中的潮流。他一生发表了不少概率论方面的论文，其中影响最为深远的是于1812年发表的集大成作品《分析概率论》。在这一著作中，他不仅总结了自己的研究成果，而且还系统地整理了概率论至那时为止已取得的主要进展。他还利用等可能性事件的概念明确地给出了概率的古典定义，从而建立了古典概率论基础的理论框架。直到20世纪初，关于概率论的教科书还一直沿用他所给出的概率论的古典基础，由此足见他这部著作对后世所产生的影响。拉普拉斯不仅研究一般的理论问题，而且热衷于概率论的应用。他成功地运用概率论去估计观察误差，在人口统计学的研究中也有所建树。

贝叶斯的重要论文《论机遇理论中一个问题的解决》是他逝世后两年，即1763年发表的。这篇论文中包含了归纳推理的一种思想，在当时并未引起人们足够的重视，现在已被发展成为关于统计推断的一种系统的理论和方法。

高斯系统地建立了观察的误差理论，推导了随机误差的分布规律，即正态分布律，这是概率和统计中最重要的分布律之一。他在研究观察数据的随机误差时所提出的最小二乘法，也在以后统计学中发展起来的回归分析中找到了重要的应用。

这一阶段，概率论在理论中取得的另一进展是，自从贝努里开始，德莫娃佛尔、拉普拉斯和普阿松等人都研究了概率论中古典的极限定理。这些定理既有助于深刻理解概率论的基础，而且诱发了以后更为深入的工作，使得极限定理成为概率论研究中久盛不衰的课题。

综上所述，这一阶段中法国古典概率学派处于概率论发展的中心地位，在他们的努力下，古典概率论得到了全面的发展，在应用方面也取得了长足的进步。但是这种发展势头到了19世纪后半半个世纪却在法国出现了停滞，这主要归因于古典概率基础的局

限性。这样概率论发展的中心移出了法国，开始了一个新的发展时期。

4. 概率论发展的近期阶段

这一阶段经历的时间大体上可认为是 19 世纪后半世纪以及 20 世纪的前 40 年。这一阶段中，概率论的应用范围不仅继续得以扩展，而且也更为深入。概率论应用的一个重要方面是统计学。以皮尔逊和费希尔为首的英国统计学派领导了统计学这一时期的发展潮流。现代数理统计学中几乎所有重要的统计理论和统计方法都是他们在 19 世纪末和 20 世纪初这一段时间内建立起来的，统计学从而成为应用数学的一个独立的分支。这方面的背景材料我们将在稍后详细地介绍。

概率论在这一阶段中令人瞩目的另一重要应用是在统计物理学方面。对此作出贡献的代表人物是奥地利物理学家玻茨曼和美国物理学家吉布斯。他们首先将概率论和统计学的论证方法贯穿到分子物理学的研究中。鉴于大量分子的存在以及它们热运动的极其紊乱的性质，使得概率论和统计学在这一领域中的应用获得了极大的成功。概率论在这样一种严肃科学的应用中取得了成功，而且这还是由物理学家实现的，就不得不使人们对概率论刮目相看了。概率论在物理学中的应用至今仍是概率论研究中相当活跃领域之一。

这一阶段在概率论的理论研究工作中取得突破性成就的，当首推俄罗斯和前苏联概率论学派。他们中间的代表人物是契比雪夫、马尔科夫、伯恩斯坦、欣钦和柯尔莫哥洛夫等人。这一阶段中他们对于概率论作出的主要贡献可以归纳为两个方面。其一是继续在由贝努里和拉普拉斯开创的古典极限定理的研究方向上进行探索，但他们已把所研究的内容大大地推广了。

前苏联概率论学派作出的另一重大贡献是彻底改造了概率论的古典基础。随着时间的推移，概率论古典基础的局限性就越来越

越暴露出来了。它极大地限制了概率论在物理、生物、统计等领域中的进一步应用。这样到了 20 世纪初,对于概率论的古典基础进行改造已是一种客观上的要求。这时的数学界正风行公理化的潮流,由罗巴契夫斯基在研究非欧几何时所建立起来的公理化方法在那时已渗透到数学的各个分支,于是一些卓越的数学家自然会想到以公理化手段来重建概率论的逻辑基础。概率论公理化的先行者是伯恩斯坦,他在 1917 年发表了第一篇关于概率论公理基础的随笔。概率论公理化的最终完成是由柯尔莫哥洛夫实现的。他在 1933 年发表的《概率论的基本概念》一书中,系统地表述了现在已被广为接受的概率论公理系统,第一次把概率论建立在严密的逻辑基础之上。从此概率论确立了它作为一门严格数学分支的地位,它的发展也开始了一个全新的阶段。

5. 概率论发展的现期阶段

这一阶段可认为是从第二次世界大战结束后直至现在的一段时期。在这一段时间内,概率论发展的步伐极大地加速了,其理论研究工作的内容得以大幅度地拓广,深刻程度也远非古典概率论所能比拟。概率论在应用方面也取得了辉煌的成就,许多应用概率论的分支相继形成。

二、数理统计学发展简史

我们分三个阶段来介绍统计学的发展历史。

1. 统计学发展的初级阶段

我们把 19 世纪中叶以前都划归为这一阶段。统计学在这一阶段的发展过程中反映出以下二种变化趋势:

- (1) 由纯粹收集数据转变到借助数据进行简单的推断;
- (2) 由最初和概率论互相分离到逐渐结合。

先谈第一种变化。统计学在英语中称为 Statistics。Statistics 源出于拉丁文中的“Status”一词,它和英语中的“State”(国家)同

义。所以在西方，统计最早指的就是收集国情资料。若从这个意义上理解统计学的发展，它的历史就源远流长了，和概率论发展的情形一样，几乎也可追溯到人类的早期文明史。拿我国来说，早在公元前 2 000 多年前的周朝就保存有收征钱粮和朝廷消费支出的大量记录，也保存有人口和征兵的记录。但类似这种纯粹属于收集数据的统计工作，和我们现在对于统计学的理解相去甚远。这种情况一直延续到 17 世纪中叶才有所变化。

西方学术界普遍认为统计学的研究事实上是由英国人格兰特开始的，并称他为人口统计学之父。人口统计学是从事人类总体的统计研究的。格兰特是一个商人，他在经营服饰及缝纫用品的同时，潜心研究死亡记录，统计了不同死因的百分比。结果他发现死于不同原因，如自杀，意外事故和某种疾病的百分比，不仅极有规律，而且逐年之间几乎保持不变。格兰特通过统计分析还发现，男婴的出生要比女婴多，但由于男人因职业上公害（工业废气和废水对健康带来的损害）、疾病和战争致死的较多，所以到了结婚的年龄，男人和女人的数目近乎相等。这些观察结果都总结在他于 1662 年发表的《基于死亡记录的自然和政治方面的观察》一书中。由于他在人口统计学上作出的贡献，他成为于 1660 年成立的英国最古老的科学协会——皇家学会的特约会员。格兰特的上述著作为以后死亡率表的制定铺平了道路，也对统计学上一些先驱者的工作产生了影响，其中包括预言哈雷彗星回归的天文学家哈雷。在 1700 年，哈雷发表了他关于死亡率表和寿命估计的研究成果。这些成果对于人寿保险事业当然是很有用的，这使人们感到，统计推断作为一门科学而言，不仅是可信的，而且是值得花费时间和精力。

从格兰特和哈雷的工作开始，统计学的发展已具有这样一种特色：针对某些特定的问题进行观察试验，收集资料，且已不满足于数据的简单统计，有时还能越出收集数据的范围，进行某种

简单的推断。这比人类早期纯属例行公务式地收集数据的活动已有不少长进。但必须指出，这样一种推断只是基于一种朴素直观的想法，还没有坚实的理论作为其后盾。出现这种情况也是很自然的，因为作为现代数理统计基础的概率论的发展，在那时也才刚刚起步。这样，直至贝努里《猜测的艺术》一书的发表，尽管概率论和统计学均已被人们承认，但它们的发展基本上是互相平行的，其内容也互相分离。这一情况在随后的一个半世纪中才有所变化：概率论的思考方法开始逐步进入统计学的研究之中。这就是我们要谈的统计学在这一阶段中发生的第二种变化。

前面已提及，从18世纪初开始至19世纪中叶是古典概率论全面发展的时期。在这一时期内，由格兰特开创的人口统计学研究成为概率论的重要应用领域之一。古典概率论的奠基人拉普拉斯曾对此产生过浓厚的兴趣。他在根据一次讲演写成的《概率的哲学探讨》一文中，讨论了许多人口统计学方面的具体问题。他运用概率论中的概念叙述了死亡率表的制定方法，引入了平均寿命的概念。特别地，他还考虑了如何根据死亡率表去计算人口总数的问题，以及当某种死亡原因消除后如何确定平均寿命的问题。这些工作促进了概率论在人口统计学中的应用，对于统计学的发展也有很积极的意义。此外，我们还须指出，概率论在这一时期的发展过程中也为日后统计学的发展积累了许多重要的、必不可少的概念与方法。其中有作为极限分布形式和随机误差分布形式出现的正态分布，它是现代统计学中最重要，也是使用得最为广泛的一种概率分布，这一分布的性质已为高斯在这一时期内所详尽研究。高斯在研究观察数据随机误差时所发展起来的最小二乘法，也是现代统计学中回归分析的重要基础之一。

这样我们看到，至19世纪中叶，统计学已从纯粹地收集数据过渡到简单的统计推断，概率论的研究方法也已渗透到统计学的研究之中。但总的说来，统计学的应用范围到那时为止还相当狭

隘，主要仍集中在人口统计学的领域之中，而且只限于解决具体的问题，还不能将问题归结成具有一般意义的统计模型。另外，统计学的理论水平也不高。重要的统计思想和方法在那时还未见端倪。统计学基本上附属于概率论，还未形成一个独立的应用数学分支。造成这一局面的原因，除了由于受到概率论发展水平的牵制，更重要的恐怕还是来自应用方面的要求不是那么迫切，以至于对统计学的发展未能形成一股强大的推动力。这一情况到了19世纪后半期，特别是本世纪初有了很大的改观，于是统计学的发展开始了一个全新的时期。

2. 现代数理统计学的形成阶段

这一阶段自19世纪中叶始，至本世纪40年代止。和概率论在这一段时间内的发展情形一样，统计学也取得了突破性的进展。其推动力如前所述，是来自应用方面的迫切要求。其发展中心则是在英国，它代表了当时科学与生产力发展的最高水平。领导统计学在这一阶段发展的是以费希尔、皮尔逊为首的英国统计学派。通过他们的努力，不仅大大拓广了统计学的应用范围，而且确立了现代数理统计学中几乎所有重要的统计理论与统计方法。这样，统计学在经历了和概率论互相分离到逐渐结合的阶段以后，又最终从概率论中分离出来，成为应用数学的一个独立分支。以下我们介绍一些对统计学的现代发展作出过重要贡献的英国统计学家，借此可了解统计学在这一阶段的发展概况。

高尔顿毕业于剑桥大学的数学系。当达尔文的《物种起源》于1859年发表后，他的兴趣转向人类学。他对人类的特征进行了详尽的研究。在身高的研究中，他对确定儿子的身高与其父母平均身高之间的关系发生兴趣。通过他的观察与分析，于1885年发表了《身高遗传中向平凡的回归》一文。高尔顿认为，如果父母的身材很高（或很矮），他们儿子的身材一般说来也较高（或较矮），但不会和他们的父母一般高矮。从上述意义上去理解“向平凡回

归”的含意，可理解成儿子的身高有恢复到平均高度的趋势，而不是越来越走向极端（越来越高或越来越矮）。随后高尔顿的朋友皮尔逊对此作了更为详尽的研究，通过对上千个家庭成员身高数据的分析，他将儿子身高 y 和父亲身高 x 之间的关系大致地用下述方程来表示

$$y = 0.516x + 33.73 \quad (\text{以英寸为单位})$$

由高尔顿和皮尔逊在研究上述问题过程中所建立起来的方法，以后又得到了进一步的推广。这部分内容在现代数理统计学中称为回归分析，被广泛地用于分析诸变量之间的关系，而且允许自变量不止一个。

科赛特曾获得牛津大学的数学和化学学位。自 1899 年起一直在都柏林（爱尔兰共和国首都）的一家啤酒酿造公司——盖尼斯公司供职。鉴于酿造啤酒时只能进行小规模的试验，使他感到有必要发展一种方法去解释从小样本中获得的信息。为此他于 1908 年在有影响的杂志 *Biometrika* 上发表了一篇文章，该文给出了现在被称为样本 t 统计量的密度函数。由于盖尼斯公司不希望它的竞争对手知道他们利用统计方法去改进啤酒的发酵过程，所以不允许它的雇员们发表任何材料。这样，科赛特只得以笔名“*Student*”（学生）发表上述那篇文章。因此现在也通常把 t 统计量的分布称为 *Student* 分布。科赛特的研究成果是数理统计学发展史上的一件大事，被公认为是小样本理论发展的先导。

从高尔顿和科赛特开创性的统计研究工作中我们已看到，统计学中新思想和新理论的提出往往都密切地联系着科学与生产的实践活动。从英国另一位统计学家费希尔的研究工作中，我们也能看到这一特点。

费希尔是英国统计学派中最有影响的代表人物，在现代数理统计学的发展中占有特别显赫的地位。他于 1912 年毕业于剑桥大学，接受过数学与理论物理的正规训练。他曾教过一段时间书，不

久便辞去这一工作，于1919年到罗萨姆斯泰德试验站供职。该试验站主要是从事农业上的研究的。日复一日地收集与解释农业试验数据的实践活动，使他突然感到有必要考虑这样的问题：响应变量（如作物产量）的差异经常有被试验环境中无法控制的非齐次性因素（不同的土质，不同的灌溉量等）掩埋的危险。他迅速地发现传统的技巧已不适于处理这样的问题，从而他决定另辟蹊径。通过几年的努力，他成功地找到了全新的统计方法，其中涉及收集数据的原理，分析试验数据的数学方法等。这部分内容现在称为试验设计，其核心便是方差分析。费希尔的这些成果后来大都总结在他的名著《试验设计》中，除此以外，费希尔还另外发表了一些重要的统计著作，如《统计的数学基础》、《统计估计理论》、《研究人员用的统计方法》等，几乎对统计学中所有重要的统计思想、统计方法和统计理论的形成都作出了重要的贡献，是本世纪推动数理统计发展的当之无愧的领路人。

在统计学发展中有多方面建树的另一位英国统计学家是前面已提到过的皮尔逊。他也毕业于剑桥大学，学习过物理、哲学和法律。他和费希尔一起被公认为是20世纪数理统计发展的奠基人。除了在创立回归分析方面有突出的成就外，他还对统计估计和假设检验理论的建立作出过贡献。特别是他在1900年发表的一篇论文中推导了 χ^2 统计量的极限分布，利用这一结果他开辟了拟合优度检验的研究方向，是非参数统计方法的先导，因此也被认为是数理统计发展中的一件大事。皮尔逊在统计发展史上的另一贡献是他和韦尔登一起创办了统计学中第一份有权威性的杂志 *Biometrika*（《生物统计学》），并从1901年起担任该杂志的主编，直至他去世（1936年）。数理统计学中许多早期奠基性的工作都首先发表在这一杂志上，*Biometrika* 也因此 in 统计学界享有盛名。

综上所述，到了本世纪40年代，现代数理统计学已形成了自己完整的体系。第二次世界大战刚结束不久，瑞典统计学家克拉