

统计学

[在线教程](#) | [案例与参考资料](#) | [模拟试卷](#) | [教学录像](#) | [相关网站](#) | [在线题库](#)



案例与参考资料

● 案例与参考资料

两则趣味贝叶斯统计案例

- 第一章 绪论
 - 曾五一：关于经济统计
 - 曾五一：关于经济统计
 - 陈希孺：关于我国统计
 - 贾洪卫：计算机与统计
 - 曾五一：关于进一步
- 第二章 统计资料的搜集
 - 朱建平：《大学生统计
 - 李露华：建立国家统计
 - 郑京平：中国国家统计
 - 李强：加强普查与统计
 - 汤以伦：必须坚持统计
- 第三章 统计描述
 - 徐明生：关于平均统计
 - 朱建平：山西省晋中
- 第四章 抽样分布与参数估计
 - 黄长全：用一例典型
 - 黄长全：趣味小案例
 - 国家统计局：2005年
 - 周晓燕：高学历青年
- 第五章 假设检验
 - 朱平辉：能从高尔文
 - 朱平辉：我国经济统计
 - 朱平辉：统计检验中
- 第六章 方差分析
 - 朱平辉：“地域”与
 - 朱平辉：统计大家
 - 朱建平：正交试验设计
- 第七章 相关与回归分析
 - 袁卫：“回归”一词

趣味案例一

在1787到1788年，也就是纽约州带头鼓动通过新的美国宪法期间，美国第四任总统詹姆斯·麦迪逊和开国政治家亚历山大·汉密尔顿都写了许多文章支持通过宪法。但这些都是以署名“联邦主义者”匿名发表的。19世纪初，麦迪逊和汉密尔顿两人开始确认各自的著作，但其中有12篇一直颇具争议，就像笼罩着一层面纱，无法以真面目示人。那么如何用统计方法进行分析 and 识别呢？

类似于《红楼梦》某些章节作者考证中的曹雪芹和高鹗之争，我们当然可以用多元统计分析中的聚类分析进行论证，但在当时，聚类方法还没有得到很好的发展。而且麦迪逊和汉密尔顿在已有著作中的平均句长几乎完全相同，这使得这一能反映写作风格特征的数据失效了。直到1964年，美国统计学家莫斯特勒和华莱士转而从用词习惯上来找出这两位作者的有区别性的风格特征，运用贝叶斯定理判定了《联邦主义论文集》中这些署名有争议的文章的作者。他们找出了几百个无“特定内容”的英文单词，如“if”、“while”、“because”、“over”、“upon”、“whilst”、“as”、“and”等。这些单词在句子里只有语法上的意义，本身并没有什么特定的含义，其使用主要取决于作者的语言习惯。对这两位作者的其他已有确定性著作进行统计分析，结果发现大约有30个虚词的使用频率是不同的。例如，汉密尔顿在他已有的18篇文章中，有14篇使用了“enough”一词；而麦迪逊在他的14篇文章中根本未使用“enough”。汉密尔顿喜欢用“while”，而麦迪逊总是用“whilst”。麦迪逊使用“upon”这个词的频率是每千字平均0.23次，而汉密尔顿对这个词的使用频率非常高，每千字平均3.24次。在12篇署名有争议的文章里，有11篇根本没有用“upon”这个词，而在剩下的那篇文章中，平均每千字出现1.1次。

需要解决的问题是，这些文章用词的分布形态是来自于麦迪逊相群的概率分布