“十二五”国家重点图书出版规划项目

应用统计工程前沿丛书

风险模型：基于R的保险损失预测

孟生旺 著

清华大学出版社

内 容 简 介

保险是经营风险的行业，风险的评估和定价是保险公司最为核心的竞争力。本书以保险业为研究对象，讨论了相应的风险模型及其应用，主要包括损失概率、损失次数、损失金额和累积损失的分布模型以及它们的预测模型，同时还探讨了巨灾损失和相依风险的建模问题。在实证研究中，以 R 语言为计算工具，提供了详细的程序代码，方便读者再现完整的计算过程。

本书适合风险管理、保险与精算等相关专业的高年级学生、研究人员或从业人员参考。

图书在版编目(CIP)数据

风险模型：基于 R 的保险损失预测/孟生旺著. —北京：清华大学出版社，2017
(应用统计工程前沿丛书)
ISBN 978-7-302-48206-2

Ⅰ. ①风… Ⅱ. ①孟… Ⅲ. ①保险业－风险管理－研究 Ⅳ. ①F840.323

中国版本图书馆 CIP 数据核字(2017)第 209666 号

责任编辑：魏贺佳
封面设计：傅瑞学
责任校对：刘玉霞
责任印制：杨　艳

出版发行：清华大学出版社
网　　址：http://www.tup.com.cn，http://www.wqbook.com
地　　址：北京清华大学学研大厦 A 座　　**邮　　编**：100084
社 总 机：010-62770175　　**邮　　购**：010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
印 装 者：三河市金元印装有限公司
经　　销：全国新华书店
开　　本：170mm×230mm　　**印　　张**：27.75　　**字　　数**：513 千字
版　　次：2017 年 9 月第 1 版　　**印　　次**：2017 年 9 月第1次印刷
定　　价：89.00 元

产品编号：076146-01

“应用统计工程前沿丛书”
编　委　会

为中国的应用统计开拓奋进

（“应用统计工程前沿丛书”代序）

改革开放以来，我国统计事业取得了突飞猛进的发展。市场化、全球化和信息技术与网络经济的蓬勃发展，使统计在经济、社会、管理、医学、生物、农业、工程等领域中的应用迎来了又一春天。2011 年 2 月，国务院学位委员会第 28 次会议通过了新的《学位授予和人才培养学科目录(2011)》，将统计学上升为一级学科，这是国家对统计学科建设与发展的重大支持，它将全面推动统计学理论方法和应用研究的深入发展。

一

长期以来，我国统计学科主要在经济学、理学和医学等门类下发展，未来进一步发展，一级统计学科将成为一面旗帜。世界上先进国家的实践充分表明，统计广泛应用在各个学科，在信息网络技术与计算机强大能力的推动下，统计学科发展特别是统计的应用正展示出一种前所未有的时代特征，它将为创造新的人类文明、提升人类发展能力做出新的重要贡献。

新中国把中国从一盘散沙凝聚成高度集中的国家，推行计划经济发展模式。这一时期，统计直接为计划服务，为政府各级管理部门，为企事业单位的计划管理，为市场资源配置，为消费、投资的安排等提供全面系统的服务，因此在经济社会管理中发挥了重要作用。但是，由于权力至上的落后观念和体系机制呆滞，统计的科学性不被重视，统计数据搜集整理的简单化和主观操作造成了很多不良的后果。改革开放之后，市场的作用强化了统计的社会影响和地位，但是，惯性的从上向下的主观思维方式仍然没有彻底的改观，因此，统计的科学应用仍然需要依靠内生发展的强大魅力不断深入和扩大。

近年来，全球化进一步加速了经济结构的转型与效率的提高。事实上，一国的稳步可持续发展离不开扎实的基础。在当今的信息化网络化时代，信息基础设施及其运用效率成为基础的基础，伴随而来的是统计在搜集数据、整理数据、分析数据上发挥的重要基础性作用。电子金融、电子政务、电子商务、网上购物、微博等一系列以网络信息技术为支撑的经济社会活动创造了大数据的新时代，计算机科学、数据库技术、大数据统计分析成为新时代发展的耀眼之星，统计学理论方法在海量数据挖掘分析、高维分析和复杂系统模型分析，以及时空的统计图示图解分析等方面正显示出强劲发展的能量，应该讲现时期是统计应用最好的发展机遇，它将大大提高人类发展的创造力、生产力，造福社会、造福人类。

二

在发展非凡的年代，谁能插上翅膀自由翱翔，谁能潜下海底自由鱼跃，统计学科需当仁不让，测度方位、穿透迷雾、指引方向、科学决策，助国家繁荣昌盛，立世界之林，这是当今中国人民大学统计学科建设的基本认知和理念。中国人民大学统计学科成立于1950年，已有60多年的发展历程，为共和国建设培养了大批优秀人才。他们广泛分布在政府部门以及银行、保险、证券、数据调查与咨询等商业企业，发挥了骨干作用。几代人大统计学人的辉煌历程和奉献，铸就了中国人民大学应用统计的特色，其作为国家应用统计重点学科、教育部重点研究基地和国家统计局重点研究基地，在融入世界一流队列、开拓中国应用、培养高精尖应用统计人才、全方位支持国家建设和发展中，做出了重要的贡献。

今天，中国人民大学统计学科布局不仅深入经济社会发展领域和保险精算与金融风险管理领域，而且已经扩展到人文社会科学的许多领域，如法律、新闻、政治学、伦理学、教育学、心理学、文献计量等，展示出应用统计在量化人文社会科学研究中的重要作用。同时，我们也在生物、医学与公共健康领域开展了深入的统计交叉应用研究。建设扎实的概率论与数理统计基础，发展强大的应用统计是中国人民大学统计学院继往开来的基本目标。

三

为了系统总结和凝练中国人民大学在统计学各个领域的科研成果，引领和推动我国统计学学科建设，提高统计学在人文社会科学与自然科学各领域科学研究，以及在管理、决策支持等方面应用的科学化和普及水平，促进统计学及其交叉学科人才培养，我们组织编写了这套“应用统计工程前沿丛书”。丛书选题覆盖应用统计学的主要分支领域，如人文、社会、政治、经济、金融、管理、法律、教育、生物、卫生、网络、数据挖掘等，力求在科学性、应用性、创新性、前沿性和可读性上形成特色。

丛书针对各领域的实际问题，着重统计学方法、模型的创新、设计和应用。在应用领域的具体统计问题研究上，积极发展统计应用流程科学，强调应用背景描述清晰，基础问题明确，发挥对微观数据、大量数据归纳探索与挖掘的统计方法作用，发展标准化的统计思维方法，创建应用领域的重要统计模型，深入解决问题，推动应用领域适应信息社会的高速发展。我们首次提出应用统计工程一词。工程是将自然科学原理应用到工农业生产部门中去而形成的各学科的总称。“工程”是科学的某种应用，通过这一应用，使自然界的物质和能源的特性能够通过各种结构、机器、产品、系统和过程，以最短的时间和少而精的人力做出高效、可靠且对人类有用的东西。我们强调应用统计的工程性，也就是强调统计的实际应用价值、科学流程与先进的统计应用技术。

丛书要反映统计学科多个前沿领域的科研进展，反映信息化和网络化背景下在诸

多统计学应用领域产生的新的统计学问题及其方法和模型的发展，以及在人文社会科学各个领域的开创性应用研究。丛书选题覆盖了应用统计学的各主要分支学科和主要新兴应用领域，系统总结和凝练应用统计的专门技术方法，引领和推动我国大数据中的统计科学方法及其应用，提高网络信息统计处理与网络经济活动与经营活动的统计科学分析能力，提高统计学在企业经营管理、市场营销、科学决策，以及全面提升综合竞争力方面的作用，提高统计学在宏观经济产业政策、货币政策、收入分配政策等重大政策制定与效果分析，以及全面提升我国国际竞争力和国家软实力方面的作用。

本套丛书主要面向统计学及其交叉学科领域的科研人员、研究生和高年级本科生，以及在实际工作中需要应用统计学理论与方法的各领域专业人士。丛书在理论方法与应用领域深入结合研究上，强调增加关键点的细节内容，突出以统计知识为核心的应用领域的统计知识体系建设。丛书在内容上力求拥有清晰的逻辑结构；对方法、概念和统计问题的描述增加相关概念知识和应用背景及交叉学科知识运用的铺垫；同时给出相关参考文献或推荐阅读书目，以帮助有兴趣的读者进一步深入学习。奉献给相关专业的读者能读懂并能够学以致用的应用统计，这是本丛书追求的重要目标之一。

赵彦云　吕晓玲

2014 年 12 月

前言

保险是经营风险的行业，风险的评估和定价是保险公司最核心的竞争力。风险的内涵十分丰富，可以从不同的角度进行划分和归类。以保险风险为例，可以分为财产风险、人身风险、责任风险、信用风险等。本书所谓的风险，主要是指保险风险，或者更具体地说，是指保险损失的风险。保险损失具体表现为损失概率、损失次数和损失金额的大小，相应地，风险模型也就包括损失概率模型、损失次数模型、损失金额模型和累积损失模型。本书讨论的风险模型虽然以财产与责任保险业务为主要背景，但也可以扩展到信用风险评估和金融风险管理等领域，具有更加广泛的应用价值。

作者在中国人民大学统计学院为风险管理与精算专业的研究生讲授“风险模型”课程已有十余年，在此期间先后完成了包括国家社会科学基金重大项目、国家自然科学基金面上项目、教育部人文社会科学重点研究基地重大项目在内的十余项风险管理与精算方向的研究课题，取得了一定的研究成果。本书就是结合作者十余年的“风险模型”教学经验和部分课题的研究成果撰写而成。

全书共由十三章内容构成，主要介绍了风险模型的理论性质、数据拟合方法以及基于 R 的实际应用，适合风险管理、保险和精算等相关专业的研究生以及精算师、风险管理师等专业人士参考。

在写作过程中，注重内容的完整性、系统性和前沿性，强调理论模型在解决实际风险管理问题中的应用。为了方便读者重现有关实证分析的具体过程，提供了完整的 R 程序代码和数据集，可以通过书中提供的链接地址下载。

本书的部分内容是作者主持完成的下述科研项目的阶段性成果：国家社会科学基金重大项目“巨灾保险的精算统计模型及其应用研究”(16ZDA052)，教育部人文社会科学重点研究基地重大项目“基于大数据的精算统计模型与风险管理问题研究”(16JJD910001)。

对于本书可能存在的任何缺陷，作者负有不可推卸之责任，欢迎各位读者批评指正，以期再版时得以修正。今后如有补充或更新材料，将及时在作者的新浪博客上(http://blog.sina.com.cn/mengshw)发布。

孟生旺
中国人民大学统计学院教授，博士生导师
中国人民大学应用统计科学研究中心研究员
甘肃省“飞天学者”特聘计划兰州财经大学讲座教授

目 录

第1章　风险度量

保险是经营风险的行业，风险的评估和度量是保险公司最核心的竞争力。风险的内涵十分丰富，可以从不同的角度进行划分和归类，以保险风险为例，可以分为财产风险、人身风险、责任风险、信用风险等。本书所谓的风险，主要是指保险风险，或者更具体地说，是指保险损失的风险。

风险通常被定义为事件发生结果的不确定性。对于保险而言，风险是指保险损失的不确定性，具体表现为保险事故发生与否的不确定性，事故发生时间的不确定性，事故发生地点的不确定性，事故发生次数的不确定性，以及损失金额的不确定性。

随机变量是描述不确定性的常用工具，所以保险损失也可以用随机变量进行描述。为此，本章首先介绍描述随机变量的有关函数，包括分布函数、概率密度函数、生存函数、概率母函数、矩母函数和危险率函数，然后介绍一些常用的风险度量方法，包括VaR、TVaR和基于扭曲变换的风险度量方法。描述随机变量的这些函数都可以完整刻画损失的分布情况，而风险度量则是对这些函数的一种高度概括，它通过一个实值来反映风险的大小，可以更加容易地应用于实际的风险管理。

1.1 描述随机变量的函数

对于保险而言，损失随机变量一般是非负的，可以分为连续型变量（如损失金额）和离散型变量（如损失次数）两大类。当然，也存在一些混合型损失随机变量，如保单的累积损失，一方面在零点有一个较高的概率堆积，另一方面在大于零的部分又是连续的。无论是哪种类型的损失随机变量，都可以用一个函数进行描述。本节主要介绍刻画损失随机变量的常用函数，如分布函数、概率密度函数、生存函数、概率母函数、矩母函数、危险率函数，这些函数是建立风险模型的基本工具。

1.1.1 分布函数

令 X 表示损失随机变量，则其分布函数定义为

$$F(x) = \Pr(X \leqslant x)$$

上式表明，损失随机变量 X 的分布函数就是 X 小于或等于 x 的概率。

【例 1-1】 随机变量 X 的取值范围为(10,30,40,70,90)，取每个值的概率均为1/5，求 X 的分布函数。

【解】 根据分布函数的定义，随机变量 X 的分布函数如下：

$$F(10) = \Pr(X \leqslant 10) = 1/5 = 0.2$$

$$F(30) = \Pr(X \leqslant 30) = 2/5 = 0.4$$

$$F(40) = \Pr(X \leqslant 40) = 3/5 = 0.6$$

$$F(70) = \Pr(X \leqslant 70) = 4/5 = 0.8$$
$$F(90) = \Pr(X \leqslant 90) = 5/5 = 1$$

绘制分布函数的 R 程序代码如下，其中函数 ecdf()表示**经验累积分布函数**(empirical cumulative distribution function)。输出结果如图 1-1 所示，其中横轴表示随机变量 x 的取值，纵轴表示分布函数 $F(x)$的取值。

```
#随机变量的取值
x = c(10, 30, 40, 70, 90)

#绘制分布函数
plot(ecdf(x), main = '', ylab = 'F(x)')
```

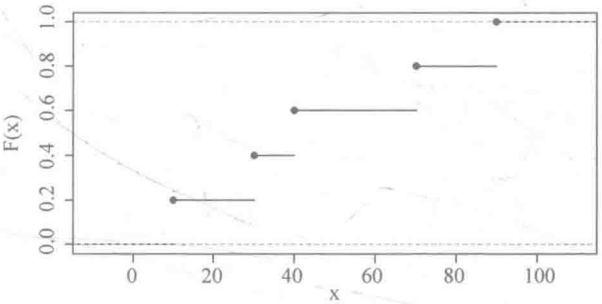

图 1-1　随机变量的分布函数*

【例 1-2】 假设损失服从伽马分布，形状参数为 shape＝2，比率参数为 rate＝1/3，绘制其分布函数，并计算损失小于 10 的概率和损失大于 4 的概率。

【解】 在 R 程序中，可以用函数 pgamma()计算伽马分布的分布函数。在该函数中，第一个参数默认为形状参数 shape，第二个参数默认为比率参数 rate，相应地，伽马分布的均值等于形状参数除以比率参数。

绘制分布函数的 R 程序代码如下，输出结果如图 1-2 所示，其中横轴表示随机变量 x 的取值，纵轴表示分布函数 $F(x)$的取值。

```
curve(pgamma(x, 2, 1/3), lwd = 2, xlim = c(0, 20), ylab = 'F(x)')
```

伽马分布的第二个参数也可以表示为尺度参数 scale，此时在程序代码中必须明确声明 scale＝1/rate。譬如，在本例的程序代码中，如果使用 scale 参数，则伽马分布函数应该表示为 pgamma(x, shape＝2, scale＝3)。计算损失概率的程序代

* 本图为上述代码运行后直接输出的结果。本书中为了使程序代码简洁明了、重点突出，均省略了对图片、字体等进行修饰的代码，后文类似情况不再赘述。