

中国精算师资格考试用书

精 算 模 型

Actuarial Models

主编 肖争艳

主审 孙佳美

图书在版编目 (CIP) 数据

精算模型 / 肖争艳主编. —北京: 中国财政经济出版社, 2010. 11

中国精算师资格考试用书

ISBN 978-7-5095-2554-8

I. ①精… II. ①肖… III. ①精算学—资格考核—自学参考资料
IV. ①F224.0

中国版本图书馆 CIP 数据核字 (2010) 第 200527 号

责任编辑: 陈 冰

责任校对: 张 凡

封面设计: 耕者设计

版式设计: 兰 波

URL: <http://www.cfeph.cn>

E-mail: cfeph@cfeph.cn

(版权所有 翻印必究)

社址: 北京市海淀区阜成路甲 28 号 邮政编码: 100142

发行处电话: 88190406 财经书店电话: 64033436

××印刷厂印刷 各地新华书店经销

787×1092 毫米 16 开 27 印张 652 000 字

2010 年 11 月第 1 版 2010 年 11 月北京第 1 次印刷

印数: 1— 定价: 58.00 元

ISBN 978-7-5095-2554-8 / F·2173

(图书出现印装问题, 本社负责调换)

本社质量投诉电话: 010-88190744

编写说明

BIAN XIE SHUO MING

精算科学的主要部分是构造和分析数学模型，这些模型刻画了保险赔付损失，以及资金流入和流出一个保险系统的过程。精算风险也可以用随机模型的方法进行表述，这些模型是对这些精算风险变量未来的概率分布及环境状况的假设。本教材的目的是向读者阐述精算建模的过程，即如何从实际数据出发建立一个合适的精算模型。

长期以来，我国精算课程和考试体系都包含了精算建模的内容，但是它们都分属于不同的课程体系。生存模型及估计是寿险精算的重要基础，用于确定身故、失效和伤残时间的概率模型，这部分内容被放在“生命表基础”课程中；理赔额和理赔次数一般模型是非寿险精算的重要基础，用来确定非寿险公司的赔付损失分布，这部分内容被放在“风险理论”课程中。虽然生存模型与理赔量和理赔数分布模型刻画的风险不同，但从统计方法上，生存模型的研究与理赔量和理赔数的一般模型并没有本质的差异。因此，国际精算师协会（IAA）在国际精算教育指南和国际认可精算师资格考试的培训大纲中，将这些内容放在一门课程（IAA6）中。为适应 IAA 考试体系的要求，中国精算师协会考试委员会将中国精算师资格考试课程“05 风险理论”和“06 生命表基础”整合成新课程体系中的“A3 精算模型”。作为这门课程的指定教材，本书试图将这些内容整合在一起，从精算建模的角度出发，以概率统计为研究工具，对保险经营中的损失风险和经营风险进行定量的刻画，建立精算模型并研究模型的性质。

为了保持与原考试课程体系的连贯性，本书是在中国精算师资格考试用书《风险理论》（吴岚、王燕主编，中国财政经济出版社 2006 版）和《生命表基础》（李晓林、孙佳美主编，中国财政经济出版社 2006 版）的基础上进行编写和修订。根据 IAA 课程大纲，本书增加了多状态生存模型、理赔额和理赔次数的分布、布朗运动与盈余过程、信度理论、Bootstrap 模拟和 MCMC 模拟等内容；保留了生存模型、生命表、短期个体风险模型、长期聚合风险模型、修匀理论、随机模拟等章节的基本内容，删除了人口统计和效用理论等内容。本书的最大特色是，重新编写了三章的内容来阐述在完整和非完整样本数据情况下生存函数、理赔额和理赔次数分布模型的估计和选择，并用两个案例来说明整个精算的建模过程。

虽然本书是精算师考试教材，但并不要求读者对保险系统已经具有很

好的知识背景。凡是在本书中首次出现保险术语的地方，我们都会给出定义。本书同时也是一本统计应用教材，既适合于具有中等概率统计知识的读者学习怎样运用统计学知识处理和研究保险业务的问题，也为已经掌握较多数理统计和随机过程知识的读者提供较深的理论内容，以便更好地掌握保险精算知识。

本书由中国人民大学统计学院肖争艳老师担任主编，负责全书统稿和编写第一、八、九、十、十二、十四章，并对第十一、十三章的初稿进行了修订；中央财经大学保险系郑苏晋老师负责编写第二至第四章，并对第五至第七章的初稿进行了修订；北京大学数学科学院的吴岚老师作为教材协调人，对全书的大纲和内容提出了方向性的指导意见；南开大学孙佳美老师作为教材主审，对全书进行了认真细致的审阅，并提出了许多宝贵的修改建议；李晓林、贾冬梅、钟颖、史森、郭程宁等作为试读人，为本书进行了认真的评审，为保证全书的出版质量提供了有力的支持；北京大学数学科学院的杨静平老师也对本书提出了一些参考意见。在本书的编写过程中，中国人民大学统计学院的一些研究生也参与了本书初稿编写和习题解答等工作，他们是：邵亚娣、李君、刘天营、王伟伟、左辰、鲍金辉、张逸铭、徐梦语、蒋安华和程夏莹等同学，同时还有许多读者、专家也提出了宝贵的意见和建议，在此一并表示衷心的感谢。

编者

2010年8月

目 录

第一章 绪论	(1)
§ 1.1 构建精算模型	(1)
§ 1.2 本书的结构	(4)

第一篇 基本风险模型

第二章 生存分析的基本函数及生存模型	(7)
§ 2.1 生存分析的基本函数	(7)
§ 2.2 参数生存模型举例	(11)
§ 2.3 条件随机变量的分布	(15)
§ 2.4 多元生存模型	(19)
习题	(26)
第三章 生命表	(29)
§ 3.1 生命表及其内容	(29)
§ 3.2 相邻整数年龄间的死亡分布	(35)
§ 3.3 选择—终极生命表	(40)
习题	(42)
第四章 理赔额和理赔次数的分布	(45)
§ 4.1 损失额分布	(45)
§ 4.2 理赔额分布	(50)
§ 4.3 理赔次数的分布	(56)
习题	(71)
第五章 短期个体风险模型	(74)
§ 5.1 引言	(74)

§ 5.2	个体保单的理赔分布	(75)
§ 5.3	总理赔额的分布——卷积法	(77)
§ 5.4	总理赔额的分布——矩母函数法	(81)
§ 5.5	总理赔额分布的正态近似	(83)
	习题	(88)
第六章	短期聚合风险模型	(90)
§ 6.1	引言	(90)
§ 6.2	理赔总量模型	(91)
§ 6.3	复合泊松模型	(95)
§ 6.4	聚合理赔量的近似模型	(105)
§ 6.5	个体风险模型与复合泊松模型的关系	(109)
	习题	(111)
第七章	破产模型	(114)
§ 7.1	盈余过程与破产概率	(114)
§ 7.2	总理赔过程	(118)
§ 7.3	连续时间终极破产概率的计算	(122)
§ 7.4	破产概率与调节系数	(128)
§ 7.5	离散时间破产模型	(133)
§ 7.6	最优再保险与调节系数	(137)
§ 7.7	布朗运动与盈余过程	(142)
	习题	(148)

第二篇 模型的估计和选择

第八章	经验模型	(150)
§ 8.1	数据类型	(150)
§ 8.2	完整数据情况下的经验分布函数估计	(154)
§ 8.3	非完整数据情况下的经验分布函数估计	(162)
§ 8.4	核密度估计	(172)
§ 8.5	大样本数据下的经验分布函数估计	(179)
	习题	(182)
第九章	参数模型的估计	(186)
§ 9.1	完整数据情况下参数的点估计	(186)

§ 9.2	非完整数据情况下参数的点估计	(194)
§ 9.3	区间估计和方差	(202)
§ 9.4	多变量的参数模型	(207)
	习题	(221)
第十章	参数模型的检验和选择	(224)
§ 10.1	引言	(224)
§ 10.2	模型的直观选择	(225)
§ 10.3	分布的拟合优度检验	(230)
§ 10.4	最优模型的选择	(239)
	习题	(246)
第三篇 模型的调整和随机模拟		
第十一章	修匀理论	(250)
§ 11.1	修匀法概述	(250)
§ 11.2	表格数据修匀	(253)
§ 11.3	参数修匀	(264)
	习题	(273)
第十二章	信度理论	(278)
§ 12.1	引言	(278)
§ 12.2	有限波动信度	(279)
§ 12.3	贝叶斯信度	(286)
§ 12.4	最大精度信度模型	(296)
§ 12.5	经验贝叶斯信度参数估计	(303)
	习题	(315)
第十三章	随机模拟	(320)
§ 13.1	引言	(320)
§ 13.2	均匀分布随机数与伪随机数	(321)
§ 13.3	一般分布随机数	(324)
§ 13.4	模拟样本的容量	(336)
§ 13.5	Bootstrap 模拟	(338)
§ 13.6	MCMC 模拟	(344)
§ 13.7	精算建模中的随机模拟实例	(351)

----->

习题	(354)
第十四章 案例分析	(357)
§ 14.1 引言	(357)
§ 14.2 退休人员的死亡时间和养老金	(357)
§ 14.3 再保险定价案例分析	(363)
附 录	(384)
附录一 中国人寿保险业经验生命表	(384)
附录二 常用概率分布及其性质	(392)
附录三 部分习题解答	(399)
附录四 名词索引	(409)
参考文献	(414)
特别鸣谢	(418)

第一章 绪 论

学习目标

- 了解精算建模的一般过程
- 了解参数模型与经验模型的优缺点
- 了解本书的基本结构

§ 1.1 构建精算模型

所谓模型，就是对现实的一种数学简化。对任何给定问题的研究，都可以用模型化的方法来解决。精算中许多问题的解决都需要借助于模型。例如，在寿险中，通过建立生存模型，对人口的死亡规律进行分析来预测被保险人未来的赔付；在非寿险中，精算师通过估计被保险人的索赔次数和索赔额的分布来进行费率厘定、准备金计提、再保险安排等一系列精算问题。因此，北美精算协会在公开发表的《精算学基本原理》中指出：“精算风险可以用随机模型的方法进行表述，这些模型是对这些精算风险变量未来的概率分布以及未来的环境状况的假设”^①。这里的精算风险变量一般指：风险是否发生、发生的时间和损失量——索赔事件的发生机率、如果索赔事件发生其发生的时间以及围绕索赔的所有成本。

精算模型的构建有两种方法：经验法和参数模型法。经验法就是不对模型做任何分布假设，直接使用经验数据建模。当统计数据特别充足而且完整时，经验分布趋近于真实分布。但在通常情况下，我们所获得的样本数据是有限的，尤其是关于高额赔付的数据更为有限。有时得到的数据还是不完整的，有可能被截断或被删失。这些都会导致经验法存在偏差。下面两个例子将阐明这一点。

【例 1-1】 某团体人寿保险合同由不同年龄和不同受益水平的 500 个雇员组成。在过去的 5 年中，已有 8 名雇员身故并共计得到 45 万元。由于该计划的身故赔付与雇员的工资水平挂钩，所以需要将赔付进行通货膨胀调整。假设下一年通货膨胀率是 10%，试根据以上信息对该合同下一年的

^① 这句话来自 Society of Actuaries Committee on Actuarial Principles, “principles of actuarial science”, Transactions of the Society of Actuary, 1992, 第 571 页的原理 3.1。

预期身故赔付进行经验估计。

解：5年内年平均赔付额为9万元，考虑到通货膨胀因素，预计下一年预期身故赔付为9.9万元。当然这个估计的缺陷在于，过去5年的经验不一定完全能够反映这个合同在未来一年的情况，因为在如此短（5~6年）的时间内身故赔付的表现可能会有很大波动。

看来，更合理的方法是建立一个模型。依例1-1，应建立一个生命表，而要构造这样的表需要积累很多个体的经验，500个人的经验是不够的。有了这张生命表，不仅可以估计下一年的预期赔付，还可以度量我们所作的估计本身的风险。

【例1-2】 考虑一个公司团体牙医保险计划。目前保单规定，每次事故的免赔额为50元，即只对一次医疗费用超过50元的保单赔付超出的部分。为了减少保险公司的平均赔付成本，有3种修改方案。第一种方案认为应该取消免赔额，这样员工就会经常去看牙，从而减少高昂的医疗费用；第二种方案认为应该提高免赔额到100元，以降低赔付成本；第三种方案认为应该限制对高额损失的赔付，建议保持免赔额50元不变，但每次最高理赔额不超过2000元。作为精算师，你认为哪种方案比较合理？为了研究方便，假设你已经随机抽取了10个赔付数据：141、16、46、40、351、259、317、1511、107、567。

解：在免赔额为50元的条件下，每次赔付的平均值为335.5元。如果免赔额提高到100元，上述保单的赔付额数据将变为： $141-50=91$ ， $351-50=301$ ， \dots ， $567-50=517$ 元，其中赔付额低于50元的保单将为0。于是平均理赔额为 $(91+301+\dots+517)/7=2903/7=414.7$ 元。保险公司的成本将减少 $(3355-2903)/3355=13.47\%$ 。

当免赔额为0，上述理赔额分别为： $141+50=191$ ， $16+50=66$ ， \dots ， $567+50=617$ 元。平均理赔额为 $3855/10=385.5$ 元。保险公司的成本将提高 $(3855-3355)/3355=14.90\%$ 。

实际上，当取消免赔额时，经验法得到的结果是没有意义的，因为我们使用的数据并不是真正来自原始样本。如果免赔额为0，则任何有医疗费用的保单都可以获得赔付，也都应该有可能被随机抽取到。但是由于最初的保单规定免赔额为50元，因此医疗费用低于50元的保单损失不可能被保险公司所记录，更不可能被随机抽取。这样会损失了大量的原始数据，造成估计结果的偏差。

对于第三种方案，经验法无法衡量这个修改对平均赔付额的影响，因为样本数据中没有理赔额超过2000元的数据。

经验法的上述缺陷可以通过建立参数模型来解决。在例1-2中，假设每张保单的原始医疗费用服从对数正态分布 $LN(\mu, \sigma)$ ，利用极大似然法得

到参数估计值为 $\hat{\mu}=5.262$, $\hat{\sigma}=1.112$ 。经计算得到免赔额为 50 元时, 每张保单的平均赔付额为 308.88 元。当免赔额提高到 100 元时, 平均赔付额为 268.93 元, 平均赔付额将减少 14.88%。当取消免赔额时, 则每张保单的平均赔付额等于对数正态分布的期望值 356.49 元。由于医疗费用大于 50 元的概率为 0.8876, 若取消免赔额, 理赔次数将会增加 $0.11569/0.88776=12.66\%$ 。当每张保单的最高赔付额不超过 2 000 元时, 每张保单的平均赔付额为 287.22 元, 平均赔付额将减少 7.1%。

相比较而言, 参数模型法对分析问题更加有用。首先, 理论分布中有丰富的应用性质(如中心极限定理、独立同分布泊松随机变量的可加性), 这些性质有助于对实际问题进行分析; 其次, 参数模型法更加简单, 完全可以由少数几个参数概括, 如泊松分布、指数分布只有一个参数, 正态分布、对数正态分布、伽玛分布、帕累托分布和负二项式分布也仅有两个参数; 最后, 模型法不仅可以给出各种相关量的点估计值, 还可以估计置信区间, 进行误差分析。

图 1-1 是建立参数模型流程的示意图^①, 这个流程由以下六个阶段组成:

第一阶段, 根据分析人员对现有数据的性质和形式的先验认知和经验, 初步选择一个或多个模型。例如, 在研究死亡率时, 所选的模型也许会包含以下这些协变量: 年龄、性别、保单生效期限、保单类型、健康方面的信息和生活方式等。在研究保险的损失量大小时, 会对统计分布类型有一些自然的选择(例如: 对数正态、伽玛、威布尔)。

第二阶段, 基于观测数据进行模型的校准。在研究死亡率的情形, 数据可能是某寿险保单群体的信息; 在研究财产保险的索赔时, 数据可能是某财产保险群体的实际赔付数据。

第三阶段, 确定所拟合模型的有效性是否充分考虑了数据中的所有信息。这里可以采用各种诊断检验。例如一些著名的统计检验: 卡方拟合优度检验、Kolmogorov-Smirnov 检验, 或者按照事物的本质进行的定性检验。检验方法的选择直接依赖于建模的目的。在保险有关的研究中, 经常要求最终的模型能够从整体上复制出实际经验数据所表现的损失, 保险实务中也常称之为“模型的无偏性”。

第四阶段, 要有适当的机会考虑其他可选模型。特别是, 如果在第三阶段揭示出前面的模型不适用, 这个步骤将特别有价值。在这个阶段也许会考虑不止一个的有效模型。

^① 引自 Klugman S. A., H. H. Panjer., G. E. Willmot 著:《损失模型: 从数据到决策》, 吴岚译, 人民邮电出版社 2009 年版, 第 1~2 页。

第五阶段，将所有第一至第四阶段考虑的有效模型按照一定的准则进行比较选择。这时可以利用前面获得的一些检验结果，也可以考虑其他的准则。一旦某个模型胜出，则淘汰的那些模型可用于敏感性分析。

第六阶段，最终要保证被选出的模型适用于未来的应用。也许要对参数进行适当的调整，以反映从观测数据时期到未来模型使用时期之间的通货膨胀变化。

当新的数据产生或者环境有所变化时，需要重复进行以上六个步骤以改进模型。

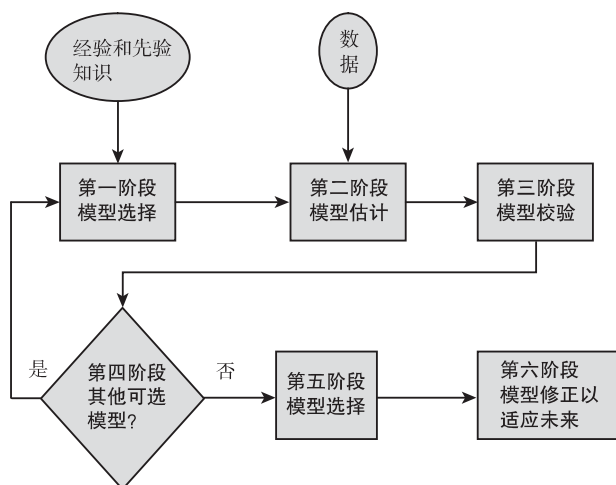


图 1-1 建模的流程

在模型选择中，要注意对模型的拟合程度和简单程度进行平衡，对模型的优劣给予客观的评价。**简洁性**是指模型中待估计的未知参数个数尽可能少，**准确性**是指模型对实际数据的拟合的误差尽可能小。另一个需要注意的问题是，对损失分布估计应考虑获得该分布的具体精算要求。在不同的精算目的下，对损失分布估计的精度要求不一样，代价也不一样。例如，如果是为了制定费率，则对损失分布的中间部分的分布情况要求较高；如果是为了考虑再保险自留额的问题，则对损失分布的尾部要作更细致的估计。

§ 1.2 本书的结构

本书将向读者展示整个精算建模过程。我们假定读者已经掌握了高等数学和概率统计的基础知识。我们首先要学习寿险和非寿险中的基本风险模型并介绍如何使用这些模型，然后再讨论模型参数的估计以及如何选择模型。本书结构如下：

1. 生存模型和生命表。死亡风险是寿险精算中的最主要的风险。在第二章中将介绍寿险精算中生存模型的基本概念及形式，对生存函数进行研究，进而对死亡率和条件死亡率的概念及公式进行阐述。生命表是表达生存模型的一种最常见的形式，在第三章中将给出生命表的传统形式和标准精算符号，以之为基础，推导死亡年龄随机变量的概率密度函数，给出风险暴露数的概念和计算公式，并对非整数年龄的死亡概率进行阐述。

2. 单张保单的理赔额和理赔次数分布。保单理赔发生的时间、发生次数和每次理赔的金额是非寿险中最主要的风险来源。在第四章中我们将介绍财产险中单张保单的理赔额和理赔次数分布。一般情况下，保险合同不可能对标的损失提供全额的赔偿。因此我们首先说明理赔额和损失额的区别与联系，其次介绍常见的保险责任：免赔额、保单限额和比例分保的基本概念，并研究带有这些保险责任后理赔额和理赔次数的分布。最后介绍几种常见的理赔次数分布的性质，并讨论免赔额对理赔次数分布的影响。

3. 短期多张保单的总理赔额的分布。在对一些保单组合、某个业务线或者某个公司进行损失建模时，我们将关心赔付的整体情况。第五章研究个体风险模型，即保单组合中保单数已知，假设每张保单至多只发生一次理赔，且相互独立。第六章研究聚合风险模型，是既考虑赔付次数，又考虑每次赔付的金额总理赔额模型。

4. 破产概率。保险公司的经营过程是动态的。第七章将不仅考虑保险人长期的赔付情况，还将引入保费收入、投资收入和费用支出以及其他可能影响现金流的因素来研究保险人盈余的变化规律。我们将分别建立连续时间和离散时间盈余过程的概念和破产概率的计算公式及性质。

5. 经验模型。有时人们需要利用数据的经验分布，这也许是因为数据的规模足够大或者是因为需要很好的表现数据本身的特点。第八章将讨论这些内容，包括对完整个体数据直接计算的简单情形、对截断或删除数据的调整、对大数据集的适当修正，特别是对生存模型的研究。

6. 参数模型的估计和选择。构建精算模型的关键在于选择一个合适的分布模型并估计参数。第九章对不同数据类型下参数估计方法进行了详细阐述，讨论参数的区间估计和极大似然估计的方差，介绍含伴随变量的参数模型。第十章将考虑模型选择问题，包括被选模型与实际数据函数图形上的直观比较和筛选直观的选择、用统计学方法对模型分布函数进行拟合优度检验、评分法选择最优模型。

7. 估计的调整。在此，还需要对结果作进一步的调整。本书中介绍了两种调整方法，首先是修匀调整。有时我们得到的估计值不一定满足先前的观点。例如死亡率，我们认为它应该是关于年龄的连续变化、递增的光滑曲线。而在大多数情形下，这个序列的每个元素是彼此独立得到的，它

们不一定满足光滑性和递增性，所以需要采用修匀方法进行必要的调整。第十一章讨论了表格数据修匀和参数修匀的几种方法。其次是信度调整。有时我们得到的几个估计都是基于很小的观测量得到的，数据的可信度不高，这时可以考虑添加一些相关的先验信息来提高估计的精度。第十二章的信度理论考虑了如何结合先验信息进行适当调整的机制和方法。

8. 随机模拟。在很难得到解析结果时，随机模拟（利用随机数）方法也许可以提供一些答案。第十三章讨论各种随机变量随机数的产生方法，还简单介绍了目前流行的 *bootstrap* 法和 *MCMC* 模拟法的技术。

9. 案例分析。在第十四章中，我们给出了两个案例来说明精算建模的流程。

第一篇 基本风险模型

第二章 生存分析的基本函数及生存模型

学习目标

- 了解对一元和多元生存模型进行分析的基本函数：生存函数、概率密度函数、危险率函数、剩余寿命均值以及剩余寿命中位数
- 了解五类参数生存模型：均匀分布、指数分布、Gompertz 分布、Makeham 分布以及韦伯分布
- 熟悉生存分析基本函数的概念及其相互关系，熟悉这些基本函数对应的精算符号
- 掌握五类参数生存模型的假设及结果，并能熟练运用相应的精算符号进行演算

§ 2.1 生存分析的基本函数

简单地说，生存分析就是对特定事件发生的时间进行分析和推断。由于研究领域不同，这一特定事件可以是生物体的死亡、疾病的出现、设备的失效或者债券的违约。这些事件发生的时间受随机因素的影响，是一个随机变量，通常我们称其为生存时间随机变量，用 T 表示，其含义是个体从初始时刻开始直至死亡、发生疾病、失效或者违约的时间，以下为了简便起见，我们称 T 是个体从初始时刻到死亡的时间。

T 的分布特征可以通过以下四个函数来描述：（1）反映个体存活时间超过时间 t 的概率的生存函数；（2）反映给定年龄的个体在下一瞬间死亡概率的危险率函数；（3）反映无条件瞬间死亡概率的概率密度函数；（4）反映平均死亡时间的剩余寿命均值。

这四个函数与另一个常用函数——累积危险率函数一起，刻画了随机变量 T 不同的特征。

2.1.1 生存函数

生存时间随机变量 T 表示个体从初始时刻开始的“未来寿命”，通常我

们总是从某一时刻开始记录某种生物体、某种设备的“未来寿命”，我们将这一起始时刻记为 $t=0$ ，在 $t=0$ 发生的事件称为**初始事件**。

描述生存时间统计特征的基本函数是生存函数，它反映被观察个体在任意时刻 $t(t \geq 0)$ 仍然生存的概率，我们将其定义为：

$$S(t) = P(T > t) \quad (2.1.1)$$

显然有：① $T \geq 0$ ；② $S(0) = 1$ ；③ $S(t)$ 是 t 的非增函数，且 $\lim_{t \rightarrow +\infty} S(t) = 0$ 。

生存函数有时与初始时刻个体的年龄无关，例如对已确诊患有某种重大疾病的病人而言，一旦确诊，其生存概率仅依赖于时间 t ，而与病人确诊时的年龄无关。但在人寿保险与养老金计划的生存模型中，必须考虑所研究对象的确切年龄（自然年龄），这是因为不同年龄群体的生存概率 $P(T > t)$ 有很大的不同。

考虑一个 x 岁（通常 x 取整数，称其为**选择年龄**）的投保群体的生存函数，我们把保单签发的时刻作为初始时刻 $t = 0$ ，讨论生存函数 $S(t)$ 。

显然， $x = 25$ 与 $x = 45$ 所对应的函数 $S(t)$ 不一样。为了刻画 $P(T > t)$ 对 x 的依赖关系，我们通常将其记为

$$P(T > t) = S(t; x) \quad (2.1.2)$$

式 (2.1.2) 中的选择年龄 x 称为**伴随变量**，未来的寿命 t 称为**主要变量**。这时，生存函数为式 (2.1.2)。

年龄不是唯一对生存函数 $S(t)$ 有影响的伴随变量，性别、吸烟与否等因素对未来寿命都有影响。如果考虑这些因素，那么对年龄为 x 岁的男性 (m) 吸烟者 (s) 来说，其生存函数则为 $S(t; x, m, s)$ ，其中 x, m, s 均为伴随变量。更一般的生存函数为 $S(t; x_1, x_2, \dots, x_m)$ 。其中， x_1, x_2, \dots, x_m 是对生存有影响的 m 个伴随变量，我们称这样的生存函数为**选择生存函数**。

特殊情况下，如果在初始时刻 $x = 0$ ，即个体为新生婴儿，那么根据式 (2.1.2)，生存函数为 $S(t; 0)$ ，也可简写为 $S(t)$ 。我们注意到，在时刻 t ，被观察者的自然年龄也刚好是 t 岁，在不致引起混淆的情况下，可用 x 代替 t ，此时新生儿的生存函数就是 $S(x)$ ，对应的随机变量 X 表示新生儿的死亡年龄（或称为**未来生命随机变量**）， $S(x) = P(X > x)$ 。

当 T 为连续型随机变量时，生存函数与累积分布函数互补，即 $S(t) = 1 - F(t)$ ，这里 $F(t) = P(T \leq t)$ ，同时，生存函数也是概率密度函数 $f(t)$ 的积分，即 $S(t) = P(T > t) = \int_t^{\infty} f(y) dy$ ，因此，

$$f(t) = - \frac{dS(t)}{dt} \quad (2.1.3)$$

$S(t)$ 的图形叫做**生存曲线**，陡峭的生存曲线表示较低的生存概率或较短的生存时间，平缓的生存曲线表示较高的生存率或较长的生存时间。

【例 2-1】 生存时间随机变量服从指数分布，其生存函数为 $S(t) = e^{-\frac{1}{\theta}t}$, $t \geq 0, \theta > 0$, 图 2-1 为指数生存曲线。

当 T 为离散型随机变量时，假设其概率分布函数 $p(t_j) = P(T = t_j)$, $j = 1, 2, \dots$, 其中 $t_1 < t_2 < \dots$, 那么 T 的生存函数为：

$$\begin{aligned} S(t) &= P(T > t) \\ &= \sum_{t_j > t} p(t_j) \end{aligned} \quad (2.1.4)$$

【例 2-2】 假设生存时间 T 服从离散均匀分布，概率分布函数为：

$$p(t_j) = P(T = j) = \frac{1}{3},$$

$$j = 1, 2, 3$$

其生存函数为：

$$S(t) = P(T > t) = \sum_{t_j > t} p(t_j) = \begin{cases} 1, & 0 \leq t < 1 \\ 2/3, & 1 \leq t < 2 \\ 1/3, & 2 \leq t < 3 \\ 0, & t \geq 3 \end{cases}$$

其生存曲线如图 2-2 所示。

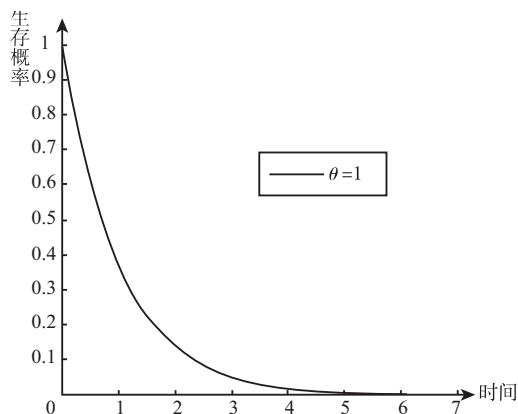


图 2-1 指数生存曲线

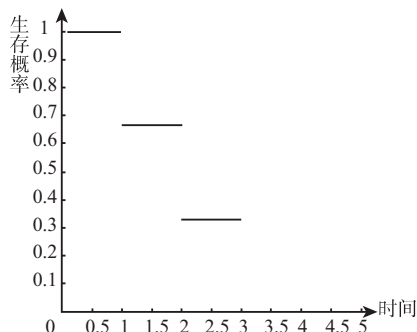


图 2-2 离散型随机寿命的生存曲线

2.1.2 危险率函数

危险率函数是生存分析中的另一个基本函数，它描述被观察个体在某时刻存活的条件下，在以后的单位时间内死亡的（条件）概率。危险率函数也称为条件瞬时死亡率、死亡密度。在人口学中，它也被称为死亡力，在可靠性研究中，也称为条件失效率。

危险率函数的定义为：

$$h(t) = \frac{f(t)}{S(t)} \quad (2.1.5)$$