中国精算师资格考试用书

生命表基础

主编 李晓林 孙佳美

中国财政经济出版社

中国精算师资格考试用书

生命表基础

修订版主编 李晓林 孙佳美

原书主编 周江雄 刘建华 黎颖芳

审稿 王晓军 李勇权

中国财政经济出版社

图书在版编目（CIP）数据

生命表基础/李晓林、孙佳美主编．—北京：中国财政经济出版社，2006.11

中国精算师资格考试用书

ISBN 7－5005－9385－6

Ⅰ．生…　Ⅱ．①李…②孙…　Ⅲ．寿命表－理论－资格考核－自学参考资料

Ⅳ．R195.3

中国版本图书馆 CIP 数据核字（2006）第 113629 号

中国财政经济出版社 出版

URL：http：//www.cfeph.cn

E－mail：cfeph @ cfeph.cn

社址：北京市海淀区阜成路甲 28 号　邮政编码：100036

发行处电话：88190406　财经书店电话：64033436

北京财经印刷厂印刷　各地新华书店经销

787×1092 毫米　16 开　19 印张　431 000 字

2006 年 11 月第 1 版　2006 年 11 月北京第 1 次印刷

印数：1—5 060　定价：38.00 元

ISBN 7－5005－9385－6/F·8142

（图书出现印装问题，本社负责调换）

中国精算师资格考试用书
编审委员会

总　序

1997年，由中国人民银行保险司牵头，开始筹划中国精算师资格考试体系的构建与考试教材的编写。当时中国寿险市场刚开始发展，市场只销售普通型产品，精算制度建设刚刚起步，在设计考试体系和考试内容时主要参考了北美、英国的资格考试体系。经过两年多的努力，2000年，中国精算师资格考试用书《利息理论》、《寿险精算数学》、《生命表的构造理论》、《风险理论与非寿险精算》和《寿险精算实务》陆续出版。1999年，中国保险监督管理委员会组织了中国首批精算师资格认证考试，其中有43名具有精算理论和实务背景的考生通过考试，获得了中国精算师资格。2000年12月，中国保监会组织了基于中国精算师资格考试体系的首次考试，中国精算师职业建设开始进入一个新的历史时期。

自2000年至今的6年中，中国精算师资格考试取得了长足的发展。到目前为止，已经设立了中国准精算师层次的全部9门考试和中国精算师层次的6门考试，在全国建立了15个考试中心，有13人通过考试获得中国精算师资格（共有56名中国精算师），269人通过考试获得中国准精算师资格。在中国精算师职业发展的同时，中国的精算实践也取得了快速发展，以1999年发布《人身保险精算规定》为开端，在寿险业的共同努力下，中国保监会逐步建立了包括精算规定、精算责任人制度、精算报告、内含价值报告、生命表在内的较为完整的精算制度体系。这其中，在中国从事精算工作的精算人员起到了非常重要的作用。

随着中国保险业的发展以及精算师工作领域的不断扩展，参加中国精算师资格考试的人数不断增加，而第一版的考试用书在使用过程中逐渐发现了一些问题。因此，2005年开始启动修订计划。本次修订的教材包括《利息理论》、《寿险精算数学》、《风险理论》（将原书的风险理论部分独立出来）、《生命表基础》（原书名《生命表的构造理论》）和《寿险精算实务》。考试用书修订的宗旨是在不进行大的调整的基础上，对原考试用书进行完善，包括结构、内容、语句等，目的是给考生提供更为规范的考试用书。与此同时，中国保监会、中国保险行业协会精算工作委员会开始启动“中国精算师资格考试体系”研究课题，对考试体系、结构、内容进行深入研究，以中国精算理论与

实践为核心，结合国际精算理论与实践的发展，构建更为科学、全面的资格体系，这将是一项长期的系统工程。随着精算师工作领域的不断扩大，在精算师资格体系的建设过程中逐步突出精算在不同领域的应用，使精算师职业不断壮大。

本次修订得到了瑞士再保险公司和中国平安人寿保险股份有限公司的支持，在此表示感谢。

我们希望更多的有志之士投身于中国精算事业，也希望中国精算师职业的专业品质不断提高，为中国保险业、金融业以及中国社会保障的发展贡献力量。

中国精算师资格考试用书编审委员会

2006 年 7 月

前 言

本书是在周江雄、刘建华、黎颖芳编著的《生命表的构造理论》的基础上修订而成的。在过去几年的使用中，逐渐发现原书中存在的一些问题，本书在对原书不进行大的调整的基础上，对原书进行完善，希望能为考生提供较为规范的考试用书。同时，为了与中国精算师资格考试体系的课程名称保持一致，将原书名改为《生命表基础》。

《生命表基础》是寿险精算的重要基础，主要内容包括生存模型、人口统计和修匀法三个方面。为了保持教材一定的连续性，本次修订未对原书结构做重大调整。此次修订，将原书第七章和第八章合并为一章——“死亡和生育测度”；原书第十一、十二和十四章合并为一章——“表格数据修匀”；原书第十五章改为附录，并增加了中国人寿保险业经验生命表（2000—2003）。此外，对原书各章节的内容作了部分调整，也对原书正文和习题中的一些细节问题作了修正。

修订工作由李晓林、孙佳美负责，参与者包括李元生、祝伟、辛士波、温贤烃、寇业富、陈辉等同志。本书聘请了王晓军老师和李勇权老师审稿，同时还有许多读者、专家提出了宝贵的意见和建议，在此表示衷心的感谢。

由于水平所限，书中难免存在错误和不足，敬请读者指正。

编 者

2006 年 8 月

目 录

第一篇 生存模型及其估计

第二篇 人口统计

第三篇 修 匀 法

第一篇　生存模型及其估计

第一章

生存模型的概念及生存模型数学

本章主要内容： 本章以概率论为基础介绍生存模型的基本概念及形式，着重对生存模型 $S(t)=Pr(T>t)$进行研究，给出描述生存模型的参数分布形式，包括均匀分布、指数分布、Gompertz 分布、Makeham 分布和 Weibull 分布，进而对死亡率和条件死亡率的概念及公式进行详细阐述，在此基础上，给出中心死亡率的定义及计算公式。

本章主要词汇： 生存模型　截尾分布　死亡率　条件死亡率　中心死亡率　平均余命

§1.1　生存模型简介

1.1.1　生存模型的基本概念

生存模型是一类特殊随机变量的概率分布。

我们考察一台在高温实验室运行的空调设备，假设这台设备从时刻 $t=0$ 开始运行，连续运转直至报废(或失效)，用 T 表示该设备从时刻 $t=0$ 开始直至报废或失效的时间，显然，T 是一个随机变量，我们所关心的是该设备在任意时刻 $t(t\geqslant 0)$仍然正常运行的概率 $Pr(T>t)$，它是 t 的函数，记为

$$S(t)=Pr(T>t) \tag{1.1.1}$$

显然有

(1) $T\geqslant 0$;

(2) $S(0)=1$;

(3) $S(t)$是 t 的非增函数，且 $\lim\limits_{t\to+\infty} S(t)=0$。

我们称随机变量 T 为设备从 $t=0$ 开始的"未来寿命"。一般地，总是从某一时刻开始记录某种设备、某类动物、某类人群的"未来寿命"。我们将这一起始时刻记为 $t=0$，在 $t=0$ 时发生的事件称为初始事件。值得注意的是并没有考虑设备在 $t=0$ 之前已使用的时间。下面再讨论一个类似的例子。

考察一群注射了某种致癌物的实验动物(如老鼠)的生存状态。注射在时刻 $t=0$ 进行，观察这些动物注射后的生存时间 T。此时概率 $Pr(T>t)$仅与时间 t 有关，而与 $t=0$ 之前已存活的时间和其他因素关系不大。

上述两个实例都是在给定条件下，不考虑初始时刻 $t=0$ 之前已有年限(或年龄)等因素对生存概率的影响。某些情形下,确实存在这样的生存模型。例如,对已确诊患有某种重大疾病的病人而言,一旦确诊,他的生存分布仅依赖于时间 t，而与病人确诊时的年龄无关。此时,用 $S(t)$表示这样的生存模型。

1.1.2 精算生存模型

1.1.1 节中的两个例子，是机械设备的可靠性问题和临床医学统计学的问题，所研究对象的确切年龄是无足轻重的，甚至可以不知道。然而用于人寿保险和养老金计划的精算生存模型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 $Pr(T>t)$有很大的不同。

考虑一个关于 x 岁(通常 x 取整数)的投保群体的生存模型。在签发保单的时刻 $t=0$，我们探讨生存概率 $S(t)=Pr(T>t)$。

显然，$x=25$ 与 $x=55$ 所对应的函数 $S(t)$是不一样的。为了刻画 $Pr(T>t)$与 x 的关系，精算师通常将其记为

$$S(t;\ x)=Pr(T>t) \tag{1.1.2}$$

上例中的年龄 x 称为伴随变量，而从年龄 x 开始，未来的寿命 t 称为主要变量，这时，精算生存模型为式(1.1.2)。

值得注意的是，年龄不是惟一对生存函数 $S(t)$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，年龄为 x 岁的男性(m)吸烟者(s)的生存模型为 $S(t;\ x,\ m,\ s)$。更一般的生存模型为 $S(t;\ x_1,\ x_2,\ \cdots,\ x_m)$，其中 x_1，x_2，$\cdots$，x_m 是对生存有影响的 m 个伴随变量。这样的模型也称为选择模型。

下面考察一个特殊情况，即在初始事件发生的时刻 $t=0$ 时，$x=0$。根据式(1.1.2)，生存模型为 $S(t;\ 0)$，通常也可简写为 $S(t)$。我们注意到时刻 t 时，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，用 x 代替 t，此时的生存模型为 $S(x)$，对应的随机变量用 X 表示，它表示新生婴儿的死亡年龄(或称为未来寿命随机变量)，且 $S(x)=Pr(X>x)$。这里，$S(x)$是新生婴儿的生存模型。

1.1.3 生存模型的形式

目前我们只给出函数 $S(t)$或 $S(x)$的概念，下面将介绍生存模型的形式。当生存概率 $S(t)$表示为 t 的某种数学函数形式时，称 $S(t)$为参数生存模型。例如，如果随机变量 T

服从指数分布，则 $S(t)=e^{-\lambda t}$，其中 λ 为参数。一般地，单参数分布(如均匀分布或指数分布)不能很好地描述精算生存模型，一般用两参数的 Gompertz 分布或 Weibull 分布，甚至三参数的 Makeham 分布拟合 $S(t)$更好。

在实际应用中，常见的是用表格描述的生存模型。但是表格只给出了 $x=0$，1，…，时 $S(x)$的值，而当 x 不是整数时，就得不到 $S(x)$的值。为此，在相邻整数间，通常假设 $S(x)$为某种形式的表达式，诸如线性假设，常值死力假设，Balducci 假设等，当这些假设运用于整个模型时，对所有 $x\geqslant 0$ 的值，$S(x)$均可求出。用表格描述生存模型已有一百多年的历史，它就是我们熟知的生命表(或死亡表)。

生存模型的另一种形式是含有伴随变量的生存模型 $S(t;\ x_1,\ x_2,\ \cdots,\ x_m)$。在以后的讨论中，我们经常把某种假设附加于含伴随变量的生存模型，由于这些假设中通常含有待估参数，因此，这样的生存模型又称为含伴随变量的参数模型。

如前所述，我们把生存模型定义为 $S(t)$[或 $S(x)$]，$S(t)=Pr(T>t)$，它是随机变量 T 的概率分布。在以后的章节，将深入地讨论上述生存模型的性质。当然，我们可以建立一个对 $S(t)$的估计，通常表示为 $\hat{S}(t)$。我们将依样本数据的性质、研究的目标和选择各种分布假设等不同情况，用各种不同的方法来估计 $S(t)$。这是本书的主要工作。

1.1.4 研究方法

一般而言，精算师和人口统计学家研究较大的样本空间，我们称之为横向研究。在这一研究过程中，首先确定一个研究团体——对其生存状态进行研究的独立人群，这一研究团体可以是某个城市或国家的人口，或者是某家人寿保险公司的保单持有者，也可以是参与养老金计划的成员等。还要选取一个观察期，在这期限开始的时候，已有许多观察对象是研究团体的成员，并且从一开始就接受观察，另一些观察对象将在观察期内任意时刻加入该研究团体，同时也可能有一些研究对象在观察期内退出研究。在观察期内进入和退出的行为称为迁移。通过适当的分类和数据处理，尤其是根据某一指定的程序对所观察的死亡人数进行分类，可以估计出每个年龄的死亡概率。这些概率值就构成了一个过渡性的表格生存模型。

与大样本的横向研究相反，大多数临床研究通常先选定一个研究团体,然后对研究团体中的每个人都研究到死亡时为止,这种研究称为纵向研究。当生存的时间较短时，通常采用这种方法。在许多情况下，临床研究将会出现队列研究问题。在对列研究中,首先确定一个队列,这个队列或初始群体可以是注射过致癌物的老鼠；或者是正在进行某种特殊治疗的人群；也可以是等待检验的灯泡样品。在所有情况下，对每一研究的个体均从初始时刻 $t=0$ 开始观察。初始群体确定后，对群体中的每个现象进行观察直至全部死亡，然后记录死亡的确切时间。显然，进行此项研究，观察者必须具备不让任何研究个体在死亡之前消失的这种控制能力，因此，这种研究有时称为“可控数据研究”。

在一些情况下，为了避免研究周期过长，纵向研究可以在研究个体全部死亡前终止。此时对应的数据称为不完全数据。通常有两种方法终止研究。一种是研究在某个预定日期前终止,此时称数据在该日期被截尾；另一种是研究在观察到预计的死亡数时终止，此时称数据被删截。

以上是对生存模型研究的两种思路，我们将会在以后的章节详细地加以阐述。

§1.2 T 的分布函数

在上一节中，已经定义了生存模型 $S(t)=Pr(T>t)$，式中的随机变量 T 表示失效时间(或死亡时间)，由随机变量 T 确定的函数 $S(t)$ 也称为生存分布函数，且有 $S(0)=1$，$S(+\infty)=0$。

在概率论中定义的分布函数

$$F(t)=Pr(T\leqslant t) \tag{1.2.1}$$

实际上是累积分布函数，显然有

$$F(t)=1-S(t) \tag{1.2.2}$$

并且 $F(0)=0$，$F(+\infty)=1$。

但由于讨论生存函数的需要，我们着重讨论 $S(t)$。对于连续型随机变量 T，概率密度函数为

$$f(t)=\frac{d}{dt}F(t)=-\frac{d}{dt}S(t) \quad (t\geqslant 0) \tag{1.2.3}$$

从而有

$$F(t)=\int_0^t f(y)\,dy \tag{1.2.4}$$

$$S(t)=\int_t^{+\infty} f(y)\,dy \tag{1.2.5}$$

显然有

$$\int_0^{+\infty} f(y)\,dy=1 \tag{1.2.6}$$

概率密度函数 $f(t)$ 表示起始时刻 $t=0$ 时生存的实体在时刻 t 失效(或死亡)的密度，或者称 $f(t)$ 是在时刻 t 死亡的无条件密度。

在生存到时刻 t 的条件下，在时刻 t 处的瞬间死亡密度称为时刻 t 处的危险率(或称为危险率函数)，记为 $\lambda(t)$。显然，$\lambda(t)$ 是在生存到时刻 t 的条件下的死亡密度，从而有

$$\lambda(t)=\frac{f(t)}{S(t)} \tag{1.2.7}$$

值得注意的是，式(1.2.7)和式(1.2.3)分别为随机变量 T(死亡时间)的危险率函数和概率密度函数，它们都是在时刻 t 失效密度的瞬时度量。它们的区别是，$\lambda(t)$ 是以生存到时刻 t 为条件的，而 $f(t)$ 是无条件的(只是给定条件 $t=0$ 时生存)。

由于 $f(t)=-\frac{d}{dt}S(t)$，所以

$$\lambda(t)=-\frac{\frac{d}{dt}S(t)}{S(t)}=-\frac{d}{dt}\ln S(t) \tag{1.2.8}$$

等式两边积分得

$$\int_0^t \lambda(y)dy = -\ln S(t) \tag{1.2.9}$$

或

$$S(t) = \exp[-\int_0^t \lambda(y)dy] \tag{1.2.10}$$

累积危险函数定义为

$$\Lambda(t) = \int_0^t \lambda(y)dy = -\ln S(t) \tag{1.2.11}$$

则

$$S(t) = e^{-\Lambda(t)} \tag{1.2.12}$$

在$(0, +\infty)$上 $f(t)$可积的条件下，连续随机变量 T 的一阶矩为

$$E(T) = \int_0^{+\infty} t \cdot f(t)dt \tag{1.2.13}$$

分部积分可得

$$E(T) = \int_0^{+\infty} S(t)dt \tag{1.2.14}$$

同样可得 T 的二阶矩为

$$E(T^2) = \int_0^{+\infty} t^2 \cdot f(t)dt \tag{1.2.15}$$

只要上述积分存在。

由此 T 的方差为

$$Var(T) = E(T^2) - [E(T)]^2 \tag{1.2.16}$$

如果 $Pr(T > y) = Pr(T \leqslant y) = \frac{1}{2}$，则称 y 为随机变量 T 的中位数。

显然，若 y 为未来寿命随机变量 T 的中位数，则有

$$S(y) = F(y) = \frac{1}{2} \tag{1.2.17}$$

以上我们考察了随机变量 T 和 T 的数字特征以及它们之间的关系。实际上，在用生存分布函数 $S(x)(x \geqslant 0)$表示的精算生存模型中，这些特征以及它们之间的关系依然存在，只是使用的符号有所不同罢了，如危险率也称为死力，用 μ_x 表示，而不是用 $\lambda(x)$，即

$$\mu_x = -\frac{\frac{d}{dx}S(x)}{S(x)} = -\frac{d}{dx}\ln S(x) \tag{1.2.18}$$

习惯上用 $\overset{\circ}{e}_0$ 表示随机变量 X 的一阶矩，即

$$\overset{\circ}{e}_0 = E(X) = \int_0^{+\infty} x \cdot f(x)dx \tag{1.2.19}$$

因为 $\overset{\circ}{e}_0$ 是 X 的无条件期望，因此也称为出生婴儿未来寿命的完全期望。

对于选择生存模型 $S(t;\ x)$，t 为随机变量的值，x 为选择年龄，那么 T 的期望值 $E(T;\ x)$给出了 x 岁的人群的未来预期寿命(或寿命期望)，用 $\overset{\circ}{e}_{[x]}$表示，它的危险率函数用 $\mu_{[x]+t}$表示，并且