李晓林 编著

精算学原理

JINGSUANXUE YUANLI

（第四卷）

复合生命状态模型

经济科学出版社

精 算 学 原 理

（第四卷）

复合生命状态模型

李晓林　编著

经 济 科 学 出 版 社

1999 年 · 北京

责任编辑：柳　敏
责任校对：孙　昉
版式设计：周国强
技术编辑：潘泽新

精算学原理（第四卷）复合生命状态模型

李晓林 编著

网址：www.esp.com.cn

电子邮件：esp@public2.east.net.cn

社址：北京海淀区万泉河路 66 号　邮编：100086

出版部电话：62630591　发行部电话：62568485

经济科学出版社出版、发行　新华书店经销

北京新华印刷厂印刷

河北省三河市三佳装订厂装订

850×1168 毫米　32 开 11.25 印张　290000 字

1999 年 10 月第一版　1999 年 10 月第一次印刷

印数：0001—3 000 册

ISBN 7－5058－1795－7/F·1275　定价：19.60 元

（图书出现印装问题，本社负责调换）

图书在版编目(CIP)数据

精算学原理　第四卷:复合生命状态模型/李晓林编著.
北京:经济科学出版社,1999.8
ISBN 7-5058-1795-7

Ⅰ.精…　Ⅱ.李…　Ⅲ.会计学:数学-应用-保险业　Ⅳ.F84

中国版本图书馆 CIP 数据核字(1999)第 27251 号

作者简介

李晓林，中央财经大学保险系执行主任，精算科学研究所副所长。多年来，负责中央财经大学的精算教育项目。已出版精算学、人身保险、社会保险等方面的著作，多次参与国家社科基金课题的研究工作，并多次出席国际精算代表大会及学术会议。

《精算学原理》简介

第四卷《复合生命状态模型》的主要内容是生命表技术与多生命状态和多生命个体的复合模型，包括单原因减员模型、死亡率的确定、原始死亡率数据的修匀、修匀检验的方法、复合状态模型、多原因减员模型、多原因减员表、联合生命、疾病保险、利润测算。

目　录

第一章
单原因减员模型及死亡率的确定

第一节　单原因减员模型

本节我们将探讨单原因减员模型的统计属性，该模型用于单一原因减员的样本总体，也就是说退出或离开样本总体的可能原因只有一个，我们将假定这个减员的原因就是死亡。

我们将探讨的模型有两个，一个是二项型模型，一个是泊松型模型。

在此，我们假定死亡率是不变的，因为如果我们研究一个很短的时间段里的某类人群总体，他们的死亡率确实是不变的。

一、二项型模型

（一）假设

二项型单原因减员模型是一个离散时间模型，在这个模型中，针对我们所探讨的减员事件（即死亡）的发生，我们做如下的假设：

① 人口总体最初的人数为 P 个，观测期为 1 年。

② 在这一年中任何一个人死亡的概率是 q。

③ 死亡的发生与否是独立的。

参数 q 叫做初始死亡率（initial rate mortality），在此，我们假设 q 对所有的个人都是相同的。

（二）死亡人数的分布

在上述假设的情况下，我们可以找出一个在某一时间段中一个特定人数死亡的概率计算公式。设随机变量 Θ 表示 1 年中死亡的人数，则死亡人数为 θ 的概率为：

$$P(\Theta = \theta) = \binom{P}{\theta} q^{\theta}(1-q)^{P-\theta} \tag{1.1}$$

其中，$\theta=0, 1, 2, \cdots, P$。上式说明，表示死亡人数的随机变量 Θ 服从参数为（P，q）的二项分布 binomial（P，q）。

证明：因为每个人是否死亡是独立的，P 人中某 θ 个人在一年内死亡、其余 $P-\theta$ 个人生存的概率是：

$$q^{\theta}(1-q)^{P-\theta}$$

而在 P 个人中选取 θ 个人的方式有：

$$\binom{P}{\theta} = \frac{P!}{\theta!(P-\theta)!}$$

种，所以，死亡人数为 θ 的概率是：

$$\binom{P}{\theta} q^{\theta}(1-q)^{P-\theta}$$

例 1.1　10 000 名学校儿童被选中参加一年的医疗研究，如果最初年死亡率是 0.00025，并且死亡的发生是独立的，计算两个或更多的被观测者在观测期间死亡的概率。

解：这是一个初始人口 $P=10\,000$ 的人口总体，其初始死亡率为常数 $q=0.00025$，假定死亡是独立发生的，所以，适用于二项型模型，死亡人数 Θ 服从参数为（10 000，0.00025）的二项分布。

根据二项分布公式

$$P(\Theta = 0) = (1-0.00025)^{10\,000} = 0.0821$$

$$P(\Theta = 1) = 10\,000 \times 0.00025(1 - 0.00025)^{9\,999}$$
$$= 0.2052$$

因此，两个或更多人死亡的概率是：

$$1 - 0.0821 - 0.2052 = 0.7127$$

（三）初始死亡率 q 的极大似然估计

在死亡率观测中，我们在观测期内死亡人数的观测值基础上估计 q。一个重要的估计方法是极大似然估计。

1. 初始死亡率 q 的极大似然估计

以一年中观测到的死亡人数为基础的初始死亡率 q 的极大似然估计是

$$\hat{q} = \frac{\Theta}{P} \tag{1.2}$$

其中，随机变量 Θ 表示死亡人数。

证明：如果初始死亡率的实际值是 q，那么观测中恰恰 θ 个人死亡的似然函数是：

$$L(q) = \binom{P}{\theta} q^{\theta}(1 - q)^{P-\theta}$$

我们通过求其对数的最大值来求它的最大值。

$$\log L(q) = \log\binom{P}{\theta} + \theta \log q + (P - \theta)\log(1 - q)$$

对 q 求偏导数，得

$$\frac{\partial}{\partial q}\log L(q) = \frac{\theta}{q} - \frac{P - \theta}{1 - q}$$

求使偏导数为 0 的 $\mathring{q}$，即

$$\theta(1 - \mathring{q}) = (P - \theta)\mathring{q}$$

即

$$\theta - \theta\mathring{q} = P\mathring{q} - \theta\mathring{q}$$

得

$$\mathring{q} = \frac{\theta}{P}$$

由于

$$\frac{\partial^2}{\partial^2 q}\log L(q) = -\frac{\theta}{q^2} - \frac{P-\theta}{(1-q)^2} < 0$$

所以，当 $q = \mathring{q} = \frac{\theta}{P}$ 时取得最大值。因此，$\hat{q} = \frac{\Theta}{P}$ 为 q 的极大似然估计，$\mathring{q} = \frac{\theta}{P}$ 是 q 的极大似然估计 $\hat{q}$ 的实现值。

我们能得到这个估计的均值和方差的公式，也能对它的分布作一个近似的描述。

2. 初始死亡率的极大似然估计的性质

① $\hat{q}$ 的均值等于 q 的实际值，即 $\hat{q}$ 是 q 的无偏估计。

② $\hat{q}$ 的方差是 $\frac{q(1-q)}{P}$（事实上，对 q 的无偏估计量，它是最小可能方差）。

③ $\hat{q}$ 渐近服从正态分布，即

$$\hat{q} \div N\left[q, \frac{q(1-q)}{P}\right] \tag{1.3}$$

证明：观测的死亡人数 θ 服从参数为（P，q）的二项分布，说明其均值为 Pq、方差为 Pq（$1-q$），所以

① $$E(\hat{q}) = E\left(\frac{\theta}{P}\right) = \frac{Pq}{P} = q$$

② $$\mathrm{Var}(\hat{q}) = \mathrm{Var}\left(\frac{\theta}{P}\right) = \frac{Pq(1-q)}{P^2} = \frac{q(1-q)}{P}$$

③ 因为 $\hat{q}$ 的均值为 q，方差为 $\frac{q(1-q)}{P}$，因此，我们在样本总体较大的情况下很容易地选择以 $\left[q, \frac{q(1-q)}{P}\right]$ 为参数的正态分布，即

$$\hat{q} \div N\left[q, \frac{q(1-q)}{P}\right]$$

这些性质说明 $\hat{q}$ 是一个很直观的估计，它的均值等于 q 的实际值，它距离实际值变化范围尽可能的小，通过正态近似我们可以计算其近似概率和置信区间。

在下一节中我们将研究怎样将模型一般化以探讨这一年中人口的其他变化。我们将用初始历险数（IETR，initial exposed to risk）E 来代替 P 。

二、泊松型（固定人口）单原因减员模型

下面我们探讨泊松型单原因减员模型。泊松型模型是连续型时间模型，有两种形式，一种是样本总体的人口总数保持不变的形式，也就是在观测期内死亡的人口能够立即得到补充，这是最简单的形式，我们称之为固定人口形式，我们将首先探讨这种形式；另一种是变动人口形式，我们将在前一种形式之后探讨它。

泊松型模型是一个连续的时间模型，它与瞬间死亡力（instantaneous forces of mortality）与中心死亡率（central mortality rate）（即通常称为 m 类比率）有关。

在此，我们假定死亡率是不变的。事实上，如果我们研究一个很短的时间段里的一个同类的人口总体，事实的确如此（注意这意味着中心死亡率和死亡力是相同的）。

（一）泊松型模型（固定人口）单原因减员模型的假设

在这种最简单的形式中，我们做如下的假设：

① 在 T 年中人数总是包括 P 个人，人口总是保持不变，死亡的人立刻被替换。

② 在任何短时间间隔中 $(t,\ t+h)$：

第一，单个死亡发生的概率是 μhP。

即 P［在 $(t,\ t+h)$ 期间 1 人死亡］$=\mu hP+o\ (h)^{*}$

第二，与仅发生一个死亡的事件相比，发生一个以上死亡的可以忽略，即

$$P[\text{在}(t, t+h)\text{期间 1 人以上死亡}] = o(h)$$

③ 在各个不同的时间间隔中死亡的发生是独立的。

参数 μ 在上一卷中称为死亡力，在本卷中有时称为灾害率。本节中，我们假设 μ 对所有的人是相同的。

（二）死亡人数的概率分布

为了计算在某一给定期间的特定死亡人数发生的概率，我们可以采用这些假设来导出一个公式。

对于给定期间的死亡人数 θ，T 年期间恰恰发生 θ 人死亡的概率是：

$$\frac{(\mu PT)^{\theta} e^{-\mu PT}}{\theta!} \tag{1.4}$$

即代表死亡人数 θ 的随机变量服从（μPT）泊松分布。

证明：我们可以把这段时间分成 n 个长度为 h 的短时间段，在每个间隔中，单个死亡发生的概率近似于 μhP，没有死亡发生的概率近似于 $1-\mu hP$。

因为我们假设在各时间间隔中这些事件的发生与否是独立的，我们可以认为这种情况是由 n 个独立的试验组成的，死亡的概率是常量 μhP，所以死亡总数服从（n，μhP）二项分布。

如果让 $h \to 0$，近似值变得精确，二项分布接近泊松分布，这一泊松分布的均值是 $n \times \mu hP = \mu PT$（因为 $nh = T$），所以死亡数服从（μPT）泊松分布，概率如公式所示。

例 1.2　某大公司总是保持 5 000 名年轻工人的员工规模，离开的立刻补上，计算 6 个月内不超过 3 人死亡的概率，假设所有工人的死亡率为 0.0008。

解：固定人数 $P = 5\,000$，死亡率为常数 $\mu = 0.0008$，如果假设死亡是独立的，则泊松分布适用，6 个月死亡人数服从泊松分布，其均值为 $0.0008 \times 5\,000 \times \frac{6}{12} = 2$，于是

$$P(\text{没有死亡}) = e^{-2} = 0.1353$$

$$P(1\text{ 人死亡}) = 2e^{-2} = 0.2707$$

$$P(2\text{ 人死亡}) = \frac{2^2}{2!\,e^{-2}} = 0.2707$$

不超过 3 人死亡的概率是：

$$0.1353 + 0.2707 + 0.2707 = 0.6767$$

（三）灾害率的极大似然估计

我们可以用极大似然估计法，根据观测期观测到的死亡数估计 μ。

1. 灾害率的极大似然估计

基于 T 年中的死亡人数观测值，灾害率 μ 的极大似然估计为：

$$\hat{\mu} = \frac{\Theta}{PT} \tag{1.5}$$

证明：如果灾害率的实际值是 μ，θ 人死亡的似然函数是：

$$L(\mu) = \frac{(\mu PT)^{\theta} e^{-\mu PT}}{\theta!} \tag{1.6}$$

取对数

$$\log L(\mu) = \theta(\log\mu + \log PT) - \mu PT - \log\theta!$$

对 q 求偏导数

$$\frac{\partial}{\partial\mu}\log L(\mu) = \frac{\theta}{\mu} - PT$$

求使偏导数为 0 的 $\mathring{\mu}$，得

$$\mathring{\mu} = \frac{\theta}{PT}$$

由于

$$\frac{\partial^2}{\partial^2\mu}\log L(\mu) = -\frac{\theta}{\mu^2} < 0$$

所以，当 $\mu = \mathring{\mu} = \dfrac{\theta}{PT}$ 时取得最大值。因此，$\hat{\mu} = \dfrac{\Theta}{PT}$ 为灾害率 μ

的极大似然估计，$\mathring{\mu}=\dfrac{\theta}{PT}$是$\mu$的极大似然估计$\hat{\mu}$的实现值。

我们可以得到该估计的均值和方差的确切公式，也可以对它的分布给出近似地描述。

2. 灾害率的极大似然估计的性质

① $\hat{\mu}$的均值等于μ的实际值，即$\hat{\mu}$是μ的无偏估计。

② $\hat{\mu}$的方差是$\dfrac{\mu}{PT}$（事实上，对μ的无偏估计量，它是最小可能方差）。

③ $\hat{\mu}$渐近服从正态分布，即

$$\hat{\mu} \dot{\sim} N\left(\mu,\frac{\mu}{PT}\right) \tag{1.7}$$

证明：观测的死亡人数θ服从参数为（μPT）的泊松分布，其均值为μPT、方差为μPT，所以

① $$E(\hat{\mu}) = E(\theta/PT) = \frac{\mu PT}{PT} = \mu$$

② $$\mathrm{Var}(\hat{\mu}) = \mathrm{Var}(\theta/PT) = \frac{\mu PT}{PT^2} = \frac{\mu}{PT}$$

③ 因为$\hat{\mu}$的均值为μ，方差为$\dfrac{q\ (1-q)}{P}$，因此，我们在样本总体较大的情况下很容易地选择以$\left[\mu,\ \dfrac{\mu}{PT}\right]$为参数的正态分布为其近似分布，即

$$\hat{\mu} \dot{\sim} N\left(\mu,\frac{\mu}{PT}\right)$$

这些性质说明$\hat{\mu}$是一个很直观的估计，它的均值等于μ的实际值，它距离实际值变化范围尽可能的小，通过正态近似我们可以计算其近似概率和置信区间。

在下一节中我们将研究怎样将模型一般化以探讨这一年中人口的其他变化。我们将用中心历险数（CETR，central exposed to risk）E^c来代替PT。

三、泊松型（变动人口）单原因减员模型

下面，我们探讨一种更普遍的泊松型模型，在这种模型中，观测期内死亡的个人将不被补充，允许个人在观测期内加入或离开被观测群体。

（一）泊松型（变动人口）单原因减员模型的假设

在泊松型（变动人口）单原因减员模型中，我们做如下的假设：

① 在时刻 t 时由 P_t 人组成的人口总体，在 T 年期限内被追踪观测。

② 在任何（t，$t+h$）间隔时间内：

第一，发生 1 人死亡的概率是 μhP_t，即

P［在（t，$t+h$）期间内恰好发生 1 人死亡］

$$=\mu hP+o\ (h)$$

第二，与 1 人死亡相比，发生多于 1 人死亡的概率可忽略不计，即

P[在($t, t+h$)期间内发生 1 人死亡以上]$=o(h)$

③ 不同时间间隔里死亡的发生与否是独立的。

我们再次假定死亡率 μ 对所有人是相同的。

（二）历险数（ETR）

历险数（ETR）是为了衡量我们所观测的样本总体的人口规模而定义的。对于我们在前面探讨的固定人口模型，由于自始至终人口总数不发生变化，也就是说，人口规模一直为初始人口水平，我们把这种情况下的初始人口规模用初始历险数（initial exposed to risk）来描述，其缩写为 IETR，记为 E，我们定义其值为初始人口总数 P。为了使固定人口模型一般化，我们需要引入中心历险数这一概念。中心历险数（central exposed to risk）的缩写为 CETR，记为 E^c，它是衡量整个观测期间平均人口规模