

# 马尔科夫模型在医疗保险精算中的应用

专业：概率论与数理统计

硕士生：李瑞林

指导教师：方积乾教授

## 摘 要

为了更好地促进医疗保险事业的发展，我们从随机过程、统计和精算的角度来研究医疗保险。本文运用上述学科的有关知识，围绕医疗保险进行讨论，并给出了相应的精算模型。

第一章介绍马尔科夫模型在医疗保险中的应用背景。第二章首先介绍相关的移民—疾病—死亡过程，并基于此模型推导出各个状态期望人数的表达式，证明了在转移强度为常量的条件下，各个状态的期望人数是稳定的。第三章先介绍 Daniel 的有关多状态模型的保险精算定理，并将此定理由离散时间状态推广到连续时间状态。第四章基于马尔科夫随机过程描述重大疾病的发展过程及其相关保险因素，并针对重大疾病对人体损害大、费用高的特点，建立了重大疾病保险的多状态模型以及相关保险精算公式，研究如何设立该保险的保费问题；同时，在 Cordeiro 多状态模型的基础上，提出了长期健康保险的保险精算公式。第五章针对疾病不同状态所存在的风险叠加问题，提出了比例动态保费模型，并建立了相应破产概率和准备金的表达式。第六章基于所提出的重大疾病的多状态模型，给出了一个有关癌症的保险精算实例。

关键词：马尔科夫模型，多状态模型，医疗保险精算

# The application of Markov model in medical insurance actuary

Major: Probability and mathematical statistics

Name: Li Ruilin

Supervisor: Professor Fang Jiqian

## Abstract

We use the knowledge of stochastic process, statistics and actuary to explore some problems related to medical insurance. The backgrounds of applying Markov model in medical insurance are introduced in the first chapter, and in the second chapter, a migration—illness—death process is proposed and results in the formulas for the expectations of sub-populations in various states, which are stable if the transition intensities are constant. In chapter 3, the relationship between multi-state model and actuary is briefly discussed. We improve the Daniel's actuarial formulas about multi-states from discrete time to continue time. In chapter 4, the modeling for three types of medical insurance and related formulas is worked out based on Markov process. We get the actuarial formulas of long term health insurance on the fundamentals of Cordeiro's multi-states model of the insurance. And we set up the multi-states model for the critical illness insurance, and get the actuarial formulas about the insurance. In chapter 5, dynamic insurance, the premium and reserve

for dynamic insurance are discussed. In chapter 6, an example on cancer insurance is presented by using a multi-states Markov model.

**Key words:** Markov model, multi-states model, medical insurance actuary

## 第一章 引言

马尔科夫过程 (Markov Process) 是由俄国著名数学家马尔科夫 (1856~1922) 首先建立。建立之初, 它被用来描述和预测煤气分子在一个密闭容器中的状态。马尔科夫模型现已被广泛应用于科学领域, 特别是在医疗保险精算领域。

本文主要介绍马尔科夫模型如何在医疗保险精算中的应用。医疗保险经常被视为长期业务, 它表现为意外险或非寿险的一些特征, 不同于基本的人寿保险。疾病在某种程度上不像死亡那样是一个必然的问题, 被保险人需向保险人证明他们确实患有某种疾病, 并且这种疾病是属于承保范围的, 这样才能获得保险金。同时, 疾病不像死亡那样是一种最终的结果, 多数情况下被保险人都会康复, 并将会因疾病提出索赔。

人作为个体在健康、疾病、死亡三种基本状态之间转移的随机过程满足马尔科夫过程的规律, 即从一种状态转移到另外一种状态的概率只与现在的状态有关, 而与以前所处的状态无关。作为多状态模型的疾病—死亡模型, 它表现了“健康”、“疾病”和“死亡”之间可能的转换。由此, 只要在实际中选取足够数量的观察样本, 统计观察特定时间内每两个状态之间的转换次数和在每个状态中的总停留时间, 就可以得出转移强度的估计值, 从而估计转移概率, 并建立相应的方程来描述多状态模型。

在多状态模型中, 根据适当的利率计算相关的精算值, 其结果为在未来任意一年的保险给付的现值总和, 准备金和破产概率等。使用多状态模型还可以预测患者人数, 为新的保险产品开发和新旧保险产品的销售作一定的准备。

对于本文, 第二章介绍了如何建立简单的移民—疾病—死亡过程来预测住院患者人数, 这是本文主要工作之一。蒋庆琅<sup>[20]</sup>已经详细讨论了一般的移民—疾病—死亡过程, 在接下来的内容中, 使用蒋庆琅的模型来预测住院患者人数。上述的两个模型得出结论为: 在转移强度为常量的情况下, 患者人数是稳定的。在第三章讨论了多状态模型与保险精算的关系, 2004年 James W. Daniel<sup>[7]</sup>给出了离散时间状态的有关多状态模型的各种现金流的公式, 本文将其推广到连续时间状态。第四章中, 使用多状态模型来建立相应的保险产品模型, 依次讨论了长期护

理保险、长期健康保险和重大疾病保险。在 2004 年 Florian Helms 等人<sup>[8]</sup>已经详细地讨论了长期护理保险,本文引用该模型及其相关精算公式说明马尔科夫链如何在医疗保险精算中的应用; Isabel Maria Ferraz Cordeiro (2002)<sup>[11]</sup>讨论了长期健康保险的多状态模型,但没有给出相应的保险精算公式,本文在 Isabel Maria Ferraz Cordeiro 模型的基础上推导得到了相关的精算公式;本文的另外一个主要工作就是构造了重大疾病保险的多状态模型,并给出了相应的保险精算值。动态保费问题也是本文的主要工作,在第五章给出了动态保费的模型及公式,并根据 2005 年杜勇宏等人<sup>[29]</sup>的结果给出动态保费破产概率,准备金等。在第六章中,本文依据黄泓智等人<sup>[23]</sup>的相关癌症保险多状态模型并结合中国大陆的实际情况,给出了相关的癌症保险保费公式,并依据文献<sup>[26]</sup>中的相关数据给出一个一年期分年龄阶段的保费和理赔额表格。并在第七章中进行了讨论与总结。本文并在附录(附录 A: 马尔科夫模型,附录 B: 多状态模型)中给出了相关数学理论说明,附录 C 给出了相关的肿瘤病人的数据。

## 第二章 移民—疾病—死亡过程

在这一章里，将使用移民—疾病—死亡过程来描述患者人数，首先给出该过程的多状态模型，然后主要讨论各个状态的人口数，并推导相应的概率分布。

### 2.1 简单模型的建立

在本节中，以疾病—死亡过程为基础，并且允许不断有新的个体补充进去，通过建立简单的移民—疾病—死亡过程模型，来研究患者人数，这一小节是本文的主要工作。

对模型有三个必要的假设：

- 1) 假设有四个状态分别为健康、疾病(I)期、疾病(II)期和死亡(D)四个状态，四者关系如图2—1所示；
- 2)  $r_{ij}, i, j = 1, 2, 3, 4$  为各个状态之间的转移强度；
- 3) 并且假设不断有新的人员补充进入，该进入符合参数为  $\lambda$  的泊松流， $\lambda$  为常量。

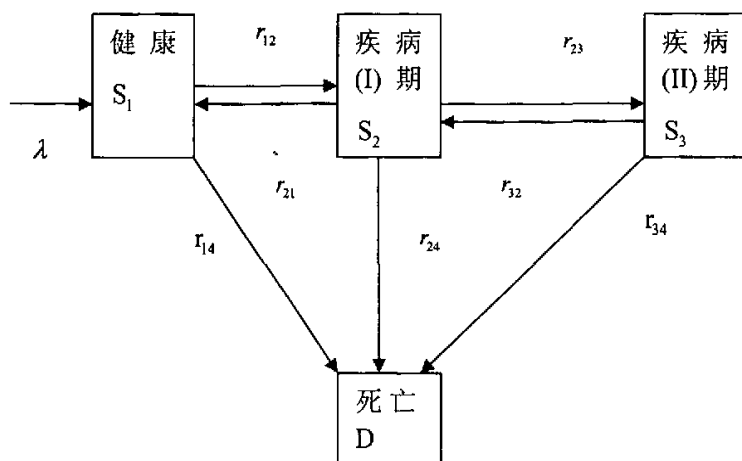


图2—1 简单的移民—疾病—死亡过程模型

记  $P_{n_1, n_2, n_3, n_4}(t)$  为在时刻  $t$ ， $S_1$  有  $n_1$  个个体， $S_2$  有  $n_2$  个个体， $S_3$  有  $n_3$  个个体， $S_4$  有  $n_4$  个个体的概率。则建立相应的方程有：

$$\begin{aligned}
 P_{n_1, n_2, n_3, n_4}(t + \Delta t) &= P_{n_1, n_2, n_3, n_4}(t)(1 - (n_1 r_{11} + n_2 r_{22} + n_3 r_{33} + \lambda)\Delta t) \\
 &+ (n_1 + 1)r_{14}P_{n_1+1, n_2, n_3, n_4-1}(t)\Delta t + (n_1 + 1)r_{12}P_{n_1+1, n_2-1, n_3, n_4}(t)\Delta t \\
 &+ (n_2 + 1)r_{23}P_{n_1, n_2+1, n_3-1, n_4}(t)\Delta t + (n_2 + 1)r_{24}P_{n_1, n_2+1, n_3, n_4-1}(t)\Delta t \\
 &+ (n_3 + 1)r_{34}P_{n_1, n_2, n_3+1, n_4-1}(t)\Delta t + (n_3 + 1)r_{32}P_{n_1, n_2-1, n_3+1, n_4}(t)\Delta t \\
 &+ \lambda P_{n_1-1, n_2, n_3, n_4-1}(t)\Delta t
 \end{aligned} \tag{2-1}$$

如果令  $\Delta t \rightarrow 0$ , 则有

$$\begin{aligned}
 \frac{dP_{n_1, n_2, n_3, n_4}(t)}{dt} &= P_{n_1, n_2, n_3, n_4}(t)(-(n_1 r_{11} + n_2 r_{22} + n_3 r_{33} + \lambda)) \\
 &+ (n_1 + 1)r_{14}P_{n_1+1, n_2, n_3, n_4-1}(t) + (n_1 + 1)r_{12}P_{n_1+1, n_2-1, n_3, n_4}(t) \\
 &+ (n_2 + 1)r_{23}P_{n_1, n_2+1, n_3-1, n_4}(t) + (n_2 + 1)r_{24}P_{n_1, n_2+1, n_3, n_4-1}(t) \\
 &+ (n_3 + 1)r_{34}P_{n_1, n_2, n_3+1, n_4-1}(t) + (n_3 + 1)r_{32}P_{n_1, n_2-1, n_3+1, n_4}(t) \\
 &+ \lambda P_{n_1-1, n_2, n_3, n_4-1}(t)\Delta t
 \end{aligned} \tag{2-2}$$

下面使用概率生成函数解上述微分方程 (2-2)。

令  $G(z_1, z_2, z_3, z_4; t)$  为联合概率生成函数, 即:

$$G(z_1, z_2, z_3, z_4; t) = \sum_{n_1, n_2, n_3, n_4=0}^{\infty} z_1^{n_1} z_2^{n_2} z_3^{n_3} z_4^{n_4} P_{n_1, n_2, n_3, n_4}(t) \tag{2-3}$$

从而,

$$\begin{aligned}
 \frac{dG}{dt} - (r_{11}z_1 + r_{14}z_4 + r_{12}z_2) \frac{dG}{dz_1} - (r_{21}z_1 + r_{22}z_2 + r_{23}z_3 + r_{24}z_4) \frac{dG}{dz_2} - (r_{33}z_3 + r_{34}z_4 + r_{32}z_2) \frac{dG}{dz_3} \\
 = G\lambda(z_1 - 1)
 \end{aligned} \tag{2-4}$$

根据 Lagrange 方法, 得辅助方程为

$$\frac{dt}{1} = \frac{-dz_1}{(r_{11}z_1 + r_{14}z_4 + r_{12}z_2)} = \frac{-dz_2}{(r_{21}z_1 + r_{22}z_2 + r_{23}z_3 + r_{24}z_4)} = \frac{-dz_3}{(r_{33}z_3 + r_{34}z_4 + r_{32}z_2)} = \frac{dG}{G\lambda(z_1 - 1)} \tag{2-5}$$

设  $z_4 = 1$ ,  $Z_i = z_i - 1$ , 我们有

$$\frac{dt}{1} = \frac{-dz_1}{(r_{11}z_1 + r_{12}z_2)} = \frac{-dz_2}{(r_{21}z_1 + r_{22}z_2 + r_{23}z_3)} = \frac{-dz_3}{(r_{33}z_3 + r_{32}z_2)} = \frac{dG}{G\lambda(z_1 - 1)} \quad (2-6)$$

其矩阵形式为:

$$\frac{\partial Z}{\partial t} = W\Delta Z \quad (2-7)$$

这里

$$W = \begin{pmatrix} r_{11} & r_{12} & 0 \\ r_{21} & r_{22} & r_{23} \\ 0 & r_{32} & r_{33} \end{pmatrix} \quad (2-8)$$

假设  $W$  是可逆的。

则式(2-7)有解, 且解记为  $K = \exp\{-Wt\}$ , 其中  $Z = \{Z_i\}$ 。

还有

$$\frac{1}{G} \frac{dG}{dt} = \sum_{i=1}^k \lambda_i Z_i = AZ \exp\{Wt\} K \quad (2-9)$$

其中  $A = (\lambda, 0, 0)$ ,  $\lambda$  为 Poisson 强度。

方程(2-9)有解  $G \exp\{-AW^{-1}Z\} = \text{constant}$ , 所以其通解为:

$$G = \exp\{AW^{-1}Z\} \Phi(\exp\{-Wt\}Z) \quad (2-10)$$

这里  $\Phi$  是关于  $W$ 、 $t$ 、 $Z$  的函数。

为了计算  $\Phi$ , 令  $t=0$ , 若假设每部分的初始大小由  $n_1 = v_1, n_2 = v_2, n_3 = v_3$  给出, 则可得到  $G$  的下面的形式:

$$G/G(z_1, z_2, z_3; 0) = z_1^{v_1} z_2^{v_2} z_3^{v_3} \quad (2-11)$$

所以  $z_1^{v_1} z_2^{v_2} z_3^{v_3} = \exp\{AW^{-1}Z\} \Phi(Z)$ 。

通过把方程 2-11 左边转化为关于  $Z$  的方程, 借助于矩阵表达形式, 我们有

$$\Phi(Z) = (1 + R_1 Z)^{v_1} (1 + R_2 Z)^{v_2} (1 + R_3 Z)^{v_3} \exp\{-AW^{-1}Z\} \quad (2-12)$$

其中  $R_i = \{\delta_y\}$ ,  $\delta_y$  为 Kronecker delta,  $\delta_y = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$ 。

所以,  $\Phi(\exp\{-Wt\}Z)$  的解为:



$$\Phi(\exp\{-Wt\}Z) = \prod_{i=1}^3 (1 + R_i \exp\{-Wt\}Z)^{v_i} \exp\{-AW^{-1} \exp\{-Wt\}Z\} \quad (2-13)$$

每一部分中个体个数的联合概率生成函数可由下式给出:

$$G = \prod_{i=1}^3 (1 + R_i \exp\{-Wt\}Z)^{v_i} \exp\{AW^{-1}\{I - \exp\{-Wt\}Z\}\} \quad (2-14)$$

对关于 G 的方程求导, 得

$$\frac{dG}{dZ} \Big|_{z=0} = AW^{-1}(I - \exp\{-Wt\}) + v \exp\{-Wt\} \quad (2-15)$$

若令  $z_1 = z_2 = z_3 = 1$ , 即  $Z_1 = Z_2 = Z_3 = 0$ , 则由(2-15)我们可以计算初始时刻的 G, 这里  $v=(v_1, v_2, v_3)$ 。

再对式(2-15)两边求导得到下面的方差-协方差形式:

$$\sum = \frac{\partial^2 G}{\partial t^2} \Big|_{z=0} + \frac{\partial G}{\partial t} \Big|_{z=0} - \left[ \frac{\partial G}{\partial t} \Big|_{z=0} \right]^2 \quad (2-16)$$

于是有以下的方差-协方差表达式:

$$\begin{aligned} \sum &= (AW^{-1}(I - \exp\{-Wt\}))^T (AW^{-1}(I - \exp\{-Wt\})) \\ &+ \sum_{i=1}^3 (R_i \exp\{-Wt\})^T (R_i \exp\{-Wt\}) v_i (v_i - 1) \\ &+ 2AW^{-1}(I - \exp\{-Wt\}) \sum_{i=1}^3 (R_i \exp\{-Wt\}) v_i + 2 \prod_{i=1}^3 (R_i \exp\{-Wt\}) v_i \\ &+ 2(R_i \exp\{-Wt\}) v_i (R_j \exp\{-Wt\}) v_j + 2 \prod_{i=1}^3 (R_i \exp\{-Wt\}) v_i + U[G] \\ &- [U[G]]^2 \end{aligned} \quad (2-17)$$

这里  $U[G]$  是  $3 \times 3$  的稀疏矩阵, 其每个主对角元素为患病人数的期望。还可以找到平衡态解。在平衡态下, 令  $t \rightarrow 0$ , 关于 G 的方程(2-14)变为:

$$G = \exp\{AW^{-1}Z\} \quad (2-18)$$

平衡态解的均值和方差由下式给出:

$$E[G] = AW^{-1}, \text{Var}[G] = G = AW^{-1} \quad (2-19)$$

从而可以得到平衡态的每一部分的均值和方差表达式:

$$\begin{aligned}
 E[S_1] = Var[S_1] &= \frac{\lambda(r_{22}r_{33} - r_{23}r_{32})}{r_{11}r_{22}r_{33} - r_{12}r_{21}r_{33} - r_{23}r_{32}r_{11}} \\
 E[S_2] = Var[S_2] &= \frac{\lambda r_{21}r_{33}}{r_{11}r_{22}r_{33} - r_{12}r_{21}r_{33} - r_{23}r_{32}r_{11}} \\
 E[S_3] = Var[S_3] &= \frac{\lambda r_{32}r_{21}}{r_{11}r_{22}r_{33} - r_{12}r_{21}r_{33} - r_{23}r_{32}r_{11}}
 \end{aligned} \tag{2-20}$$

根据式(2—20)，假设转移强度为常数时，我们可以得出各个状态之间的患者期望人数是稳定的。在此基础上，就可以使用往年的医疗卫生数据去构造医疗保险产品。

本文通过建立简单的移民—疾病—死亡模型来测算各个状态的患者期望人数，在接下来的一节中，将通过建立一般的移民—疾病—死亡模型来测算各个状态的患者期望人数。

## 2.2 一般的有关患者人数的模型

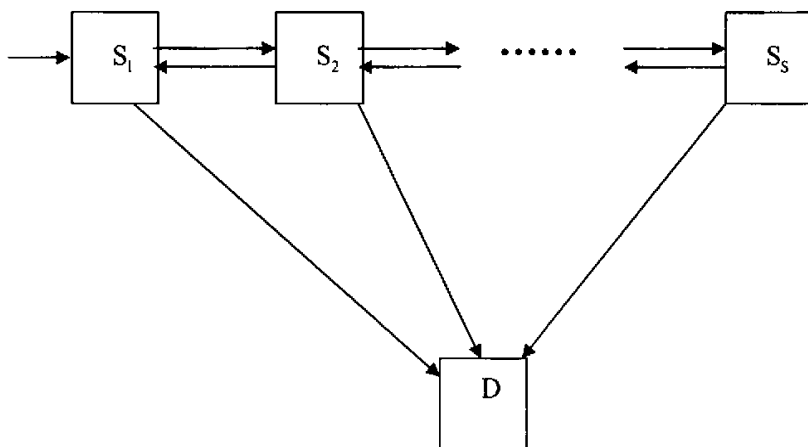


图2—2 一般的移民—疾病—死亡模型

在这一部分，通过建立一般的移民—疾病—死亡过程来讨论有关患者人数的问题，蒋庆琅<sup>[20]</sup>已经应用一般移民—疾病—死亡过程来讨论有关患者人数的问题，以下是有关基本假设和相应结论：

记 $S_1$ 是健康状态， $S_2, \dots, S_s$ 是疾病状态， $D$ 为死亡状态。用一个常数向量

$$i = \begin{pmatrix} i_1 \\ i_2 \\ \dots \\ i_s \end{pmatrix} \quad (2-21)$$

表示初始时刻0处于各状态的人口数，并用一个随机向量

$$X_t = \begin{pmatrix} X_{1t} \\ X_{2t} \\ \dots \\ X_{st} \end{pmatrix} \quad (2-22)$$

来表示时刻  $t$  处于各状态的人口数，例如， $X_{1t}$  是时刻  $t$  处于  $S_1$  的人口数。 $X_t$  的值用各个常数向量

$$j = \begin{pmatrix} j_1 \\ j_2 \\ \dots \\ j_s \end{pmatrix} \quad (2-23)$$

表示，其中  $j_\alpha$  是非负整数，我们感兴趣的是条件概率

$$P_y(0, t) = P_r(X_t = j | X_0 = i) \quad (2-24)$$

在本节所考虑模型中的概率（即式2—24）不仅依赖于初始人口数  $i$  和健康——死亡转移，而且还依赖于在区间  $(0, t)$  内发生的移民数。

对每个  $\tau, 0 \leq \tau < t$ ，假定在时间区间  $(\tau, \tau + \Delta)$  内处于状态  $S_\alpha$  的人口数的增长服从以下的瞬时概率：

$$\begin{aligned} \lambda_\alpha^*(\tau)\Delta + o(\Delta) &= P\{\text{处于状态 } S_\alpha \text{ 的人数由于移民在 } (\tau, \tau + \Delta) \text{ 内增加}\} \\ \nu_{\alpha\beta}^*(\tau)\Delta + o(\Delta) &= P\{\text{一个个体将在 } (\tau, \tau + \Delta) \text{ 内从 } S_\alpha \text{ 移至 } S_\beta, \beta \neq \alpha\} \\ \mu_\alpha^*(\tau)\Delta + o(\Delta) &= P\{\text{处于状态 } S_\alpha \text{ 的人数由于移民在 } (\tau, \tau + \Delta) \text{ 内减少}\} \end{aligned} \quad (2-25)$$

设

$$\nu_{\alpha\alpha}^*(\tau) = -[\sum_{\beta \neq \alpha} \nu_{\alpha\beta}^*(\tau) + \mu_\alpha^*(\tau)] \quad (2-26)$$

同时当没有外部人口迁入的时候，我们记：

$$\begin{aligned} \nu_{\alpha\beta}(\tau)\Delta + o(\Delta) &= P\{\text{一个个体将在 } (\tau, \tau + \Delta) \text{ 内从 } S_\alpha \text{ 移至 } S_\beta, \beta \neq \alpha\} \\ \mu_\alpha(\tau)\Delta + o(\Delta) &= P\{\text{处于状态 } S_\alpha \text{ 的人数由于移民在 } (\tau, \tau + \Delta) \text{ 内减少}\} \end{aligned} \quad (2-27)$$

在这里引入一个列向量

$$\delta_\alpha = \begin{pmatrix} \delta_{1\alpha} \\ \delta_{2\alpha} \\ \dots \\ \delta_{s\alpha} \end{pmatrix} \quad (2-28)$$

其中  $\delta_\alpha$  的元素是Kronecker delta,  $\delta_{\alpha\beta} = \begin{cases} 1 & \alpha = \beta \\ 0 & \alpha \neq \beta \end{cases}$ 。

由上, 我们可以得到  $P_{ij}(0, t) = P_r(X_t = j | X_0 = i)$  满足下面的微分方程组:

$$\begin{aligned} \frac{d}{dt} P_{ij}(0, t) = & -P_{ij}(0, t) \sum_{\alpha=1}^s [\lambda_\alpha^*(t) - \nu_{\alpha\alpha}^*(t)] + \sum_{\alpha=1}^s P_{i, i-\delta_\alpha}(0, t) \lambda_\alpha^*(t) \\ & + \sum_{\alpha=1}^s \sum_{\substack{\beta=1 \\ \alpha \neq \beta}}^s P_{i, j+\delta_\alpha-\delta_\beta}(0, t) \nu_{\alpha\alpha}^*(t) \\ & + \sum_{\alpha=1}^s P_{i, j+\delta_\alpha}(0, t) \mu_\alpha^*(t) \end{aligned} \quad (2-29)$$

当  $t=0$  时, 初始条件为:

$$P_{i,i}(0, 0) = 1, P_{i,j}(0, 0) = 0 \quad (j \neq i) \quad (2-30)$$

该差分方程的解法详见文献<sup>[20]</sup>。我们主要关心的是概率生成函数, 由文献<sup>[20]</sup>, 可得到下面的定理:

**定理 2.2.1** 向量  $X(t)$  的概率生成函数为

$$G_{X(t)}(\mathbf{u}; t) = \prod_{\alpha=1}^s [\pi_{\alpha 0}(0, t) + \pi_{\alpha 1}(0, t)u_1 + \dots + \pi_{\alpha s}(0, t)u_s]^{\alpha} \times \exp\left\{-\sum_{\alpha=1}^s (1-u_\beta)\theta_\beta(t)\right\} \quad (2-31)$$

其中  $\pi_{\alpha\beta}(0, t)$  为转移概率,  $\theta_\beta(t)$  是 Poisson 参数。

由式 (2-31) 得到:

期望

$$E[X_\beta(t)] = \prod_{\alpha=1}^s i_\alpha \pi_{\alpha\beta}(0, t) + \theta_\beta(t) \quad (2-32)$$

方差

$$V[X_{\beta}(t)] = \prod_{\alpha=1}^s i_{\alpha} \pi_{\alpha\beta}(0, t) [1 - \pi_{\alpha\beta}(0, t)] + \theta_{\beta}(t) \quad (2-33)$$

和协方差

$$\text{Cov}[X_{\beta}(t), X_{\gamma}(t)] = -\prod_{\alpha=1}^s i_{\alpha} \pi_{\alpha\beta}(0, t) \pi_{\alpha\gamma}(0, t) \quad (2-34)$$

$$\beta \neq \gamma; \beta, \gamma = 1, \dots, s$$

我们主要关心的是每个状态的期望人数，即式(2-32)，其中  $\pi_{\alpha\beta}(0, t)$  和  $\theta_{\beta}(t)$  与转移强度有关，如果转移强度为常量，则  $\pi_{\alpha\beta}(0, t)$  和  $\theta_{\beta}(t)$  也为常量，可以得到各个期望状态的人口数不随时间而急剧的变化，这个结果说明可以依据往年的疾病数据来构造医疗保险产品。

综上所述所做的工作可以证明在一定条件下，针对某些疾病，取转移强度为常量时，我们得到各个状态的期望人口数是稳定的。

## 第三章 多状态模型与保险精算现金流

在本章里，将从精算的角度来考虑多状态模型与现金流的关系，给出一些定理与公式。

### 3.1 状态之间的现金流与现值

在建立相关保险的多状态模型后，更感兴趣的是建立相关的现金流模型。比如在最基本的生存模型中，精算人员所关心的是在死亡状态应付的现金。

#### 3.1.1 有关各个状态转移产生的现金流

记  $C^{(i,j)}$  为有关从状态  $i$  到  $j$  的现金流 ( $i$  可以等于  $j$ )，同时假设  $C^{(i,j)}$  与时间无关。认为保险受益人获得赔偿是在被保险人伤残或死亡的当年年底。我们还有以下注记：如果  $t$  时刻个体在状态  $i$ ， $t+1$  时刻在状态  $j$ ，则记  ${}_{t+1}C^{(i,j)}$  为在时刻  $t+1$  从状态  $i$  转移到状态  $j$  的现金流，折现因子为  $v$ 。

#### 3.1.2 精算现值(Actuarial present values)

在这里使用三重积总计(triple-product summation)的方法来计算精算现值，如果在一定时间内，现金流发生，那么三重积所对应的要素是：现金流发生时刻的概率、现金流的总量和从现金流发生的时刻到现在时刻( $n$ )的折现。

在以上假设下进行精算现值的计算，不妨设现在位于状态  $S$ ，记现金流发生的时刻是  $n+k+1, k>0$ ，在该时刻现金流的总量为  ${}_{n+k+1}C^{(i,j)}$ ，从时刻  $n+k+1$  到现时(the present time)的折现是  ${}_{k+1}v_n$ ，从时刻  $n+k$  由状态  $i$  出发于时刻  $n+k+1$  到达状态  $j$  的概率是  $P_{ij}$ ，且  $P_{ij} = {}_kQ_n^{(i,i)}Q_{n+k}^{(i,j)}$ ，其中  $Q$  为转移强度矩阵。

定理 3.1 的离散时间状态是由 James W.Daniel<sup>[7]</sup> 给出的，我们在此基础上推广到了连续状态。

**定理 3.1:** 有关转移的现金流精算现值

记  ${}_{t+1}C^{(i,j)}$  为时刻  $t+1$  的现金流 (假定时刻  $t$  在状态  $i$ ，时刻  $t+1$  在状态  $j$ )。假设于时刻  $n$  处于状态  $S$ ，则在时刻  $n$  时现金流的精算现值

$$APV(C^{(i,j)}) = \sum_{k=0}^{\infty} [P_y] [{}_{n+k+1}C^{(i,j)}] [{}_{k+1}v_n] \quad (3-1)$$

若看作连续时间状态, 即  $v_n = v_n(t)$ , 则有

$$APV(C^{(i,j)}) = \sum_{k=0}^{\infty} \int_n^{n+k+1} [P_y] [{}_{n+k+1}C^{(i,j)}] [{}_{k+1}v_n(t)] dt \quad (3-2)$$

### 3.2 状态内的现金流与现值

假设对于状态  $i$  现金流于时刻  $l$  开始发生, 并假设到期给付年金是在每期期初付的。如果在时刻  $l$  处于状态  $i$ ,  ${}_l C^{(i)}$  为时刻  $l$  的现金流。假设于时刻  $n$  处于状态  $S$  中, 现金流发生在时刻  $n+k, k \geq 0$ , 此时在状态  $i$  的概率为  $P_i(P_i = {}_k Q_n^{(s,i)})$ 、此时现金流总量为  ${}_{n+k} C^{(i)}$ 、从时刻  $n+k$  到  $n$  的折现为  ${}_k v_n$ 。

定理 3.2 的离散时间状态也是由 James W. Daniel<sup>[7]</sup>给出的, 在此基础上将其推广到连续状态。

**定理 3.2:** 状态内的现金流精算现值

记  ${}_{l+1} C^{(i,j)}$  为时刻  $l+1$  的现金流 (假定时刻  $l$  在状态  $i$ , 时刻  $l+1$  在状态  $j$ ), 假定个体于时刻  $n$  处于状态  $S$ , 则此时的精算现值为

$$APV(C^{(i)}) = \sum_{k=0}^{\infty} [P_i] [{}_{n+k} C^{(i,j)}] [{}_k v_n] \quad (3-3)$$

若看作连续时间状态, 即  $v_n = v_n(t)$ , 则有

$$APV(C^{(i)}) = \sum_{k=0}^{\infty} \int_n^{n+k} [P_i] [{}_{n+k} C^{(i,j)}] [{}_k v_n(t)] dt \quad (3-4)$$

以上的公式就是与多状态模型有关的现金流的描述, 在以后的讨论中使用将这些公式, 并将其扩充到其他保险精算范围。

## 第四章 使用马尔科夫模型建立医疗保险模型

本章使用马尔科夫模型建立了医疗保险产品。在这里我们主要关心三种比较重要的医疗保险产品：长期护理保险(The long term care insurance)、长期健康险(Permanent Health Insurance)和重大疾病保险(Critical Illness Insurance)。在以下的部分，将分别对各个险种进行讨论，一般是首先使用马尔科夫模型去描述疾病过程以及相关保险因素，然后再利用相关精算公式来计算保险费用。

### 4.1 长期护理保险(The long term care insurance)

#### 4.1.1 长期护理保险的介绍

长期护理保险是指当被保险人非常衰弱以至于在没有其他人帮助的情况下不能照顾自己，甚至不能利用辅助设备时，给付保险金的一种保险。当一个人的自理能力出现困难时，往往需要医院或其他护理机构提供护理服务，而该项费用对一般家庭来说是相当高的，经常使许多家庭用尽一生的积蓄。当一个人年老、抵抗力下降、易患疾病的时候往往已经失去了收入的来源，如何使这个人及其配偶能够安度晚年呢？长期护理保险的就是提供这方面的保障，从保险的角度解决上诉问题，因此它对中老年人有着极其重要的意义。特别是当今中国正在进入老龄化社会，同时由于经济、医疗的极大的发展，人们的寿命都在提高，因此建立专门的长期健康保险是很重要的。在接下来的内容里，将介绍如何使用马尔科夫多状态模型去建立相关的长期护理保险数学模型。

2004年 Florian Helms<sup>[8]</sup>详细讨论了长期护理保险的多状态模型以及相关的保险精算值，下面的模型主要引自 Florian Helms<sup>[8]</sup>的论文，这里主要以这个模型来说明马尔科夫模型是如何应用到医疗保险领域的。

#### 4.1.2 长期护理保险的数学建模

在这里将围绕长期护理保险建立相关的马尔科夫多状态模型。从实际出发，假设有四个状态：存活(0)、家庭护理(1)、医院护理(2)和死亡(3)，四者的相关关系如图4—1所示，同时特别说明假设四个状态间不可逆转，且存活状态可以直接到医院护理状态，死亡状态(3)是吸收状态。



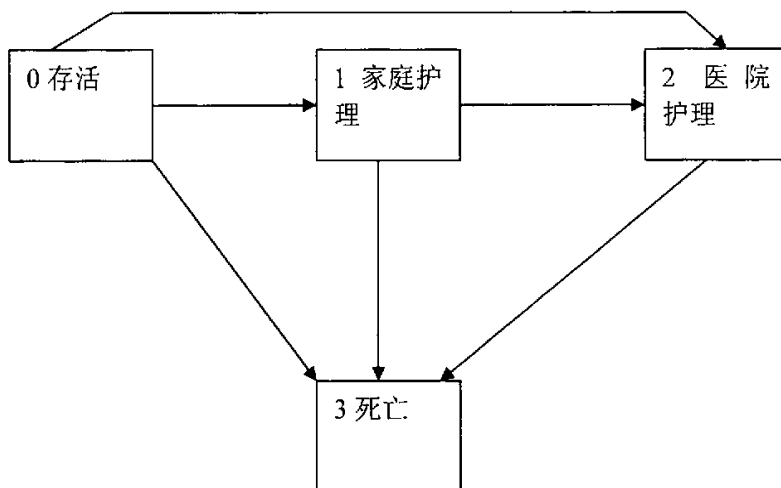


图 4—1 长期护理保险模型

### 4.1.3 有关模型的建立和推导

假定图 4—1 所示的个体生命史是符合齐次的马尔科夫链  $S(t)$ ，状态空间为  $S=0、1、2、3、$ ，记  $\mu_y$  为转移强度， $p_{ij}$  为转移概率，其中  $i, j=0,1,2,3$ 。

定义 4.1.1 (转移概率) 设  $S(t)$  为时间齐次的马尔科夫过程， $0 \leq t \leq \mu$  且  $i, j \in S$ ，

$$p_{ij}(t, u) := p(S(u) = j | S(t) = i) \quad (4-1)$$

我们可有转移强度的表达

$$\mu_{ij} := \lim_{dt \rightarrow 0} \frac{p_{ij}(t, t+dt) - p_{ij}(t, t)}{dt} \quad (4-2)$$

进一步可得一年期的转移概率为：

$$p_{ij}(t, u) := p_{ij}(t, t+1) \quad (4-3)$$

同时由 Chapman-Kolmogorov 方程，可得：

$$p_{ij}(t, u) = \sum_{k \in S} p_{ik}(t, t+1) p_{kj}(t+1, u) \quad (4-4)$$

并使用差分方程可得：

$$\begin{aligned}
 p_{00}(z, t) &= \exp\left(-\int_z^t [\mu_{01}(u) + \mu_{02}(u) + \mu_{03}(u)] du\right) \\
 p_{11}(z, t) &= \exp\left(-\int_z^t [\mu_{12}(u) + \mu_{13}(u)] du\right) \\
 p_{22}(z, t) &= \exp\left(-\int_z^t [\mu_{23}(u)] du\right) \\
 p_{23}(z, t) &= 1 - p_{22}(z, t) \\
 p_{12}(z, t) &= \int_z^t p_{11}(z, u) \mu_{12}(u) p_{22}(u, t) du \\
 p_{13}(z, t) &= 1 - p_{11}(z, t) - p_{12}(z, t) \\
 p_{01}(z, t) &= \int_z^t p_{00}(z, u) \mu_{01}(u) p_{11}(u, t) du \\
 p_{02}(z, t) &= \int_z^t [p_{00}(z, u) \mu_{02}(u) + p_{01}(z, u) \mu_{12}(u) p_{22}(u, t)] du \\
 p_{03}(z, t) &= 1 - p_{00}(z, t) - p_{01}(z, t) - p_{02}(z, t)
 \end{aligned} \tag{4-5}$$

从保险精算的角度，设：

- 如果  $S(t)=i$ ，时刻  $t$  的连续保费的费率为  $p_i(t)$ ；
- 如果  $S(t)=j$ ，时刻  $t$  的连续年金的费率为  $b_j(t)$ ；
- 在时刻  $t$ ，整付(a lump sum)  $C_j(t)$  是与从状态  $i$  转移到状态  $j$  有关的；
- $v = \exp\{-\delta\}$ ， $\delta$  为利息力。

考虑时刻  $t$ ， $S(u)=j$ ，我们令

$$Y_t^{P_j}(u, u+du) = v^{u-t} I_{\{S(u)=j\}} P_j(u) du \tag{4-6}$$

则得到的保费的随机现值为：

$$Y_t^{P_j}(u_1, u_2) = \int_{u_1}^{u_2} v^{u-t} I_{\{S(u)=j\}} P_j(u) du, t \leq u_1 \leq u_2 \tag{4-7}$$

在时刻  $u$ ， $S(u)=j$ ，我们令

$$Y_t^{b_j}(u, u+du) = v^{u-t} I_{\{S(u)=j\}} b_j(u) du \tag{4-8}$$

得到年金受益的随机现值为：

$$Y_t^{b_j}(u_1, u_2) = \int_{u_1}^{u_2} v^{u-t} I_{\{S(u)=j\}} b_j(u) du, t \leq u_1 \leq u_2 \quad (4-9)$$

考虑整付  $C_{jk}(u)$  在时刻  $u$  后付, 且在时刻  $u$  从状态  $j$  到状态  $k$  时发生, 我们有该整付的现值为:

$$Y_t^{C_{jk}}(u) = v^{u-t} I_{\{S(u)=j, S(u)=k\}} C_{jk}(u) \quad (4-10)$$

精算值是现值的期望。假定被保的风险在时刻  $t$  处于状态  $i$ , 精算值则由随机现值的条件期望给出如下:

$$\text{趸缴费用: } E[Y_t(u) | S(t) = i] \quad (4-11)$$

$$\text{年金: } E[Y_t(u, u + du) | S(t) = i] \quad (4-12)$$

则由(4-11), (4-12) 得到:

在时刻  $u$ , 状态  $j$  下, 连续保费精算值为:

$$E[Y_t^{P_j}(u, u + du) | S(t) = i] = v^{u-t} P_j(u) du \quad (4-13)$$

进一步, 当  $u_1 \leq u_2$  时, 有:

$$E[Y_t^{P_j}(u_1, u_2) | S(t) = i] = \int_{u_1}^{u_2} v^{u-t} p_j(u) du \quad (4-14)$$

在区间  $[t, n]$ , 我们有聚合的保费函数  $Pre_j(t, n)$ :

$$Pre_j(t, n) = \int_t^n v^{u-t} \sum_{k \in S} p_{jk}(t, u) p_j(u) du \quad (4-15)$$

在时刻  $u$ , 状态  $S(u) = j$ , 设连续年金的费率为  $b_j(t)$ , 则该连续年金受益的精算值为:

$$E[Y_t^{b_j}(u, u + du) | S(t) = i] = v^{u-t} p_j(u) b_j(u) du \quad (4-16)$$

当  $u_1 \leq u_2$  时, 定义:

$$E[Y_t^{b_j}(u_1, u_2) | S(t) = i] = \int_{u_1}^{u_2} v^{u-t} p_j(u) b_j(u) du \quad (4-17)$$

针对整付  $C_{jk}(u)$  在时刻  $u$  时应付的精算值, 如果在时刻  $u$ , 从状态  $j$  到状态  $k$ , 则有:

$$E[Y_t^{C_{jk}}(u) | S(t) = i] = v^{u-t} p_{ij}(u) \mu_{jk}(u) C_{jk}(u) du \quad (4-18)$$

当  $u_1 \leq u_2$  时, 有:

$$E[Y_t^{C_{jk}}(u) | S(t) = i] = \int_{u_1}^{u_2} v^{u-t} p_{ij}(u) \mu_{jk}(u) C_{jk}(u) du \quad (4-19)$$

由以上的讨论, 可以得出总的收益(the cumulative benefit)函数  $B_i(t, n)$ :

$$B_i(t, n) = \int_t^n v^{u-t} \sum_{k \in S} p_{ij}(u) b_j(u) du + \int_t^n v^{u-t} \sum_{k \in S} \sum_{k \neq j} p_{ij}(u) \mu_{jk}(u) C_{jk}(u) du \quad (4-20)$$

## 4.2 长期健康险(Permanent Health Insurance)

在国内, 长期健康险一般指保险期限为一年以上(不含一年)的健康险业务, 该保险的显著特点是时间跨度比较长, 有 20 年期的还有 50 年期的等不同时间的跨度。由于保险时期过长也给对有关风险的评估带来了挑战, 同时长期健康险中还设有延迟期限, 该时期往往比伤残的时期长, 其目的是观察被保险人的健康状况, 以有利于核保理赔, 可见只有对长期健康险的被保险人进行详细的研究才能更好规避风险, 进一步为被保险人服务。本节使用马尔科夫模型来对该问题建立随机过程模型, 研究长期健康险的相关问题。Isabel Maria Ferraz Cordeiro<sup>[1]</sup>详细讨论了长期健康保险的多状态模型, 本文在其模型的基础上进行推导得到了有关保险精算公式。

### 4.2.1 模型的假设与简介

根据长期健康保险的特点, 即在很长的一个时期内个体将受到多种疾病的侵害, 我们给出各个状态之间的关系如图 4—2 所示, 其中各类疾病之间无直接的联系。

1) 假设长期健康保险的模型有  $n+2$  个状态: 健康状态(记为  $H$ ), 死亡状态(记为  $D$ )为吸收状态, 一类疾病状态(记为  $S_1$ ), 二类疾病状态(记为  $S_2$ ), ...,  $n$  类疾病状态(记为  $S_n$ ), 同时记  $S_0 = H, S_{n+1} = D$ ;

2) 设  $x$  为被保险人达到某种状态的年龄,  $z$  为该被保险人在疾病状态停留的

时间；设  $\sigma(i)(x)$  是被保险人在年龄  $x$  从健康状态转移到  $i$  类疾病状态的转移强度； $\rho(i)_{x,z}$  是从疾病状态转移到健康状态的转移强度； $\nu(i)_{x,z}$  是从疾病状态转移到健康状态的转移强度； $\mu_x$  为从健康状态转移到死亡状态的转移强度； $\sigma(i)(x), \rho(i)_{x,z}, \nu(i)_{x,z}$  不仅与  $x, z$  有关，还与它们所联系的状态有关。

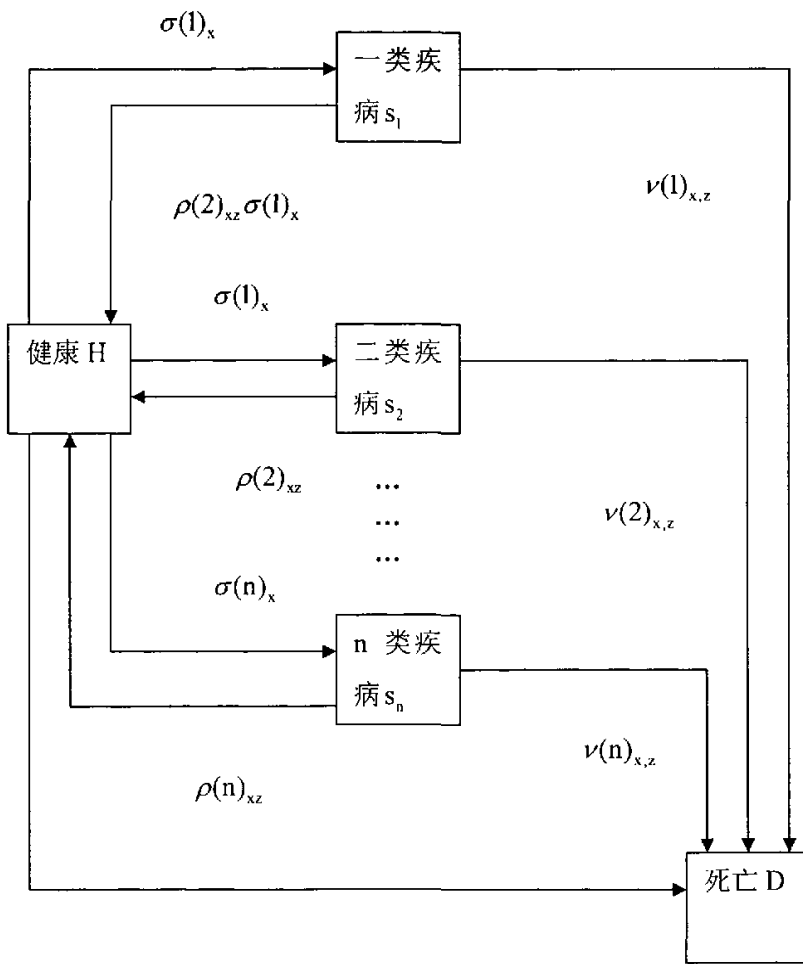


图 4—2 长期健康保险模型

### 4.2.2 多状态模型描述

这里认为所研究的情况可用一个连续时间的多状态随机过程模型来描述,记该多状态的随机过程模型为:

$$\{Y(x), Z(x)\}, x > 0 \quad (4-21)$$

$Y(x)$ 表示被保险人在年龄  $x$  时所处的状态,  $Y(x) \in \{H, S_1, \dots, S_n, D\}$ ,

$Z(x)$ 表示被保险人在年龄  $x$  于状态  $Y(x)$ 所逗留的时间,

$Z(x) \in [0, \infty), Z(x) = \max\{t: t \leq x \text{ and } Y(x-h) = Y(x) \forall 0 \leq h \leq t\}$ 。

同时  $\{Y(x), Z(x)\}, x > 0$  具有马尔科夫性质, 年龄  $X$  后的过程只与年龄  $x$  时的  $Y(x)$ 、 $Z(x)$  有关, 而与年龄  $x$  以前的  $Y(x)$ 、 $Z(x)$  无关, 这也是符合实际情况的, 一位被保险人的下一个状况只与现在所处的状况有关, 而不受以前各个状态的影响。

同时在这个模型里给出一些假设和相应的记号:

$$1) {}_tP_{x,z}^{jk} = P[Y(x+t) = k | Y(x) = j, Z(x) = z], j, k = H, S_1, S_2, \dots, S_n, D; t, x, z \geq 0 \quad (4-22)$$

假设  $Y(x) = H$ , 年龄  $x$  以后的过程并不依赖  $Z(x)$ , 因此可以认为

${}_tP_{x,z}^{HH}, {}_tP_{x,z}^{HS_i} (i=1, 2, \dots, n), {}_tP_{x,z}^{HD}$  是独立于  $Z$  的, 记这些符号为

${}_tP_x^{HH}, {}_tP_x^{HS_i} (i=1, 2, \dots, n), {}_tP_x^{HD}$ ; 同时还假设

$${}_tP_{x,z}^{Dk} = 0, k = H, S_1, S_2, \dots, S_n, {}_tP_{x,z}^{DD} = 1 \quad (4-23)$$

$$2) {}_tP_x^{\overline{HH}} = p\{Y(x+t) = H, Z(x+t) \geq t | Y(x) = H\} \quad (4-24)$$

式(4-24)表示: 若被保险人在年龄  $x$  是处于健康状态, 其在年龄为  $x+t$  还处于健康状态的概率。这里要指出的是这个概率是不同于  ${}_tP_x^{HH}$ 。

$$3) {}_tP_{x,z}^{\overline{S_i S_i}} = p\{Y(x+t) = S_i, Z(x+t) = x+t | Y(x) = S_i, Z(x) = z\} \quad (4-25)$$

式(4-25)表示: 若现在年龄为  $x$  的某被保险人处于  $i$  类疾病状态且逗留期间为  $z$ , 在年龄为  $x+t$  时仍在该状态的概率。若  $z=0$ , 记其概率为  ${}_tP_x^{\overline{S_i S_i}}$ 。

$$4) {}_{\omega,t}P_x^{HS_i} = p\{Y(x+t) = S_i, Z(x+t) \leq \omega | Y(x) = H\}, i=1, 2, \dots, n \quad (4-26)$$

式(4—26)表示:一个年龄为  $x$ , 处于健康状态(H)的被保险人,在其年龄为  $x+t$  处于  $i$  类疾病状态,且逗留期不小于  $\omega$  ( $\omega > 0$ )的概率。当  $\omega \geq t$  时,  ${}_{\omega,t}P_x^{HS_i} = {}_tP_x^{HS_i}$ ;  ${}_{\omega+d\omega,t}P_x^{HS_i} - {}_{\omega,t}P_x^{HS_i}, 0 < d\omega \leq t - \omega$  的含义是一个被保险人现在年龄是  $x$ 、处于健康状态(H),他在年龄  $x+t$  处于  $i$  类疾病状态,且逗留时间为  $[\omega, \omega + d\omega]$  的概率。

假设以上的概率函数都是  $x, z, t, \omega$  的连续函数。

在此模型中,  $n+2$  个状态的转移强度定义如下:

$$\begin{aligned} \sigma(i)_x &= \lim_{t \rightarrow 0^+} \frac{{}_tP_x^{HS_i}}{t}, \mu_x = \lim_{t \rightarrow 0^+} \frac{{}_tP_x^{HD}}{t}, \rho(i)_{x,z} = \lim_{t \rightarrow 0^+} \frac{{}_tP_{x,z}^{S_iH}}{t} \\ \nu(i)_{x,z} &= \lim_{t \rightarrow 0^+} \frac{{}_tP_{x,z}^{S_iD}}{t}, i = 1, 2, \dots, n \end{aligned} \quad (4-27)$$

假设式(4—27)定义的极限都存在且它们是  $x$  或  $(x, z)$  的连续函数。同时假设这些取值于  $x$  或者  $(x, z)$  的转移强度是有界的。

由以上的假设和定义,我们在区间  $(t, t+dt)$  上有:

$$\begin{aligned} p[\text{在}(t, t+dt)\text{两个或多于两个状态}] &= o(dt), \\ p[Y(x+t+dt) = S_i | Y(x+t) = H] &= \sigma(i)_{x,t} dt + o(dt), \\ p[Y(x+t+dt) = D | Y(x+t) = H] &= \mu_{x,t} dt + o(dt), \\ p[Y(x+t+dt) = H | Y(x+t) = S_i, Z(x+t) = z] &= \rho(i)_{x,t,z} dt + o(dt), \\ p[Y(x+t+dt) = D | Y(x+t) = S_i, Z(x+t) = z] &= \nu(i)_{x,t,z} dt + o(dt), \\ i = 1, 2, \dots, n, \text{且} \lim_{dt \rightarrow 0^+} \frac{o(dt)}{t} &= 0 \end{aligned} \quad (4-28)$$

在这个模型里,还可以定义当被保险人在年龄  $x$  时处于健康状态(H),未来在年龄  $x+t$  处于疾病状态(不是特指哪类疾病)的概率,即:

$$p[Y(x+t) = s_1 \text{ 或 } Y(x+t) = s_2 \text{ 或 } \dots \text{ 或 } Y(x+t) = s_n | Y(x) = H] = \sum_{i=1}^n {}_tP_x^{HS_i} \quad (4-29)$$

考虑  $\sum_{i=1}^n \sigma(i)_x = \lim_{t \rightarrow 0^+} \frac{\sum_{i=1}^n {}_tP_x^{HS_i}}{t}$ , 可认为  $\sum_{i=1}^n \sigma(i)_x$  是全局的疾病强度而不是针对

某种特指的疾病状态。

同时假定不同类别的疾病状态的疾病强度是不同的。

概率  ${}_t P_x^{\overline{HH}}$ 、 ${}_t P_{x,z}^{\overline{S_i S_i}}$ 、 ${}_t P_x^{HH}$ 、 ${}_{\omega,t} P_x^{HS_i}$  和  ${}_t P_x^{HD}$  ( $i=1,2,\dots,n$ ) 是该模型的基本概率, 这是研究长期健康保险的基础。

### 4.2.3 基本概率的公式表达

利用以上的假设以及马尔科夫过程的性质, 给出以下方程, 并且可以看作 Kolmogorov 向前方程

$$\begin{aligned} \frac{\partial}{\partial t} {}_t P_x^{\overline{HH}} &= -{}_t P_x^{\overline{HH}} \left( \sum_{i=1}^n \sigma(i)_{x+t} + \mu_{x+t} \right) \\ \frac{\partial}{\partial t} {}_t P_{x,z}^{\overline{S_i S_i}} &= -{}_t P_{x,z}^{\overline{S_i S_i}} \left( \rho(i)_{x+t,z+t} + \nu(i)_{x+t,z+t} \right), i=1,2,\dots,n \\ \frac{\partial}{\partial t} {}_t P_x^{HH} &= -{}_t P_x^{HH} \left( \sum_{i=1}^n \sigma(i)_{x+t} + \mu_{x+t} \right) + \sum_{i=1}^n \left( \int_0^t {}_u P_x^{HH} \sigma(i)_{x+u} {}_{t-u} P_{x+u}^{\overline{S_i S_i}} \rho(i)_{x+t,t-u} du \right) \\ \frac{\partial}{\partial \omega} {}_{\omega,t} P_x^{HS_i} &= \begin{cases} {}_{t-\omega} P_x^{HH} \sigma(i)_{x+t-\omega} {}_{\omega} P_{x+t-\omega}^{\overline{S_i S_i}} & 0 \leq \omega < t \\ 0 & \omega \geq t \end{cases} \\ \frac{\partial}{\partial t} {}_t P_x^{HD} &= {}_t P_x^{HH} \mu_{x+t} + \sum_{i=1}^n \left( \int_0^t {}_u P_x^{HH} \sigma(i)_{x+u} {}_{t-u} P_{x+u}^{\overline{S_i S_i}} \nu(i)_{x+t,t-u} du \right) \end{aligned} \quad (4-30)$$

对以上方程进行积分求解可得:

$$\begin{aligned} {}_t P_x^{\overline{HH}} &= \exp \left\{ - \int_0^t \left( \sum_{i=1}^n \sigma(i)_{x+s} + \mu_{x+s} \right) ds \right\} \\ {}_t P_{x,z}^{\overline{S_i S_i}} &= \exp \left\{ - \int_0^t \left( \rho(i)_{x+s,z+s} + \nu(i)_{x+s,z+s} \right) ds \right\} \\ {}_{\omega,t} P_x^{HS_i} &= \int_{\max\{0,t-\omega\}}^t {}_s P_x^{HH} \sigma(i)_{x+s} {}_{t-s} P_{x+s}^{\overline{S_i S_i}} ds \\ {}_t P_x^{HD} &= \int_0^t {}_s P_x^{HH} \mu_{x+s} + \sum_{i=1}^n \left( \int_0^s {}_u P_x^{HH} \sigma(i)_{x+u} \int_0^{t-s} {}_r P_{x+u}^{\overline{S_i S_i}} \nu(i)_{x+s+r,t-r} du \right) \end{aligned} \quad (4-31)$$

针对  ${}_t P_x^{HD}$  不像其他四个容易求解的问题, 在实际应用中可采用数值计算来解决, 其中  $d$  为该段时间内延缓期。

### 4.2.4 理赔平均时间

假设被保险人所持的保单到 65 岁或者死亡时失效。记  $T_y^i$  为第  $i$  类疾病理赔发生, 被保险人于年龄  $y$  患该类疾病, 则相应的该类疾病持续时间为:



$$E(T_y^r) = \frac{\int_t^{65-y} {}_tP_y^{\overline{S,S}} \rho(i)_{y+t,t} dt}{\int_t^{65-y} {}_tP_y^{\overline{S,S}} \rho(i)_{y+t,t} dt} - d \quad (4-32)$$

可以应用上面的公式，在获得平均医疗费用的前提下，估计患病所用的费用。

#### 4.2.5 现值和准备金

考虑时刻  $t$ ,  $S(u) = i$  ( $S_0 = H, S_{n+1} = D$  见模型假设), 由

$$Y_t^P(u, u + du) = v^{u-t} I_{\{S(u)=i\}} P_i(u) du \quad (4-33)$$

可得保费的随机现值为:

$$\begin{aligned} Y_t^P(u_1, u_2) &= \int_{u_1}^{u_2} v^{u-t} I_{\{S(u)=i\}} P_i(u) du \\ &= \int_{u_1}^{u_2} v^{u-t} \left[ I_{\{S(u)=i\}} ({}_tP_x^{HH}) + I_{\{S(u)=i\}} ({}_tP_x^{\overline{HH}}) + I_{\{S(u)=i\}} \left( \sum_z {}_tP_{x,z}^{S,S} \right) \right. \\ &\quad \left. + \sum_z I_{\{S(u)=i\}} ({}_tP_{x,z}^{\overline{S,S}}) + \sum_z I_{\{S(u)=i\}} ({}_tP_x^{HS_i}) + I_{\{S(u)=i\}} ({}_tP_x^{HD}) \right] du, \quad (4-34) \\ &t \leq u_1 \leq u_2 \end{aligned}$$

时刻  $u$ ,  $S(u) = i$ , 由

$$Y_t^b(u, u + du) = v^{u-t} I_{\{S(u)=i\}} b_i(u) du \quad (4-35)$$

可得年金受益的随机现值为:

$$Y_t^b(u_1, u_2) = \int_{u_1}^{u_2} v^{u-t} I_{\{S(u)=i\}} b_i(u) du, t \leq u_1 \leq u_2 \quad (4-36)$$

考虑整付  $C_{ik}(u)$ , 在时刻  $u$  后付, 且在时刻  $u$  从状态  $i$  到状态  $k$  发生, 则该整付的现值为

$$Y_t^{C_{ik}}(u) = v^{u-t} I_{\{S(u)=i, S(u)=k\}} C_{ik}(u) \quad (4-37)$$

在时刻  $u$ , 状态  $j$  下, 连续保费精算值为:

$$E[Y_t^P(u, u + du) | S(t) = i] = v^{u-t} P_j(u) du \quad (4-38)$$

进一步, 当  $u_1 \leq u_2$  时, 有

$$E\left[Y_t^P(u_1, u_2) | S(t) = i\right] = \int_{u_1}^{u_2} v^{u-t} p_j(u) du \quad (4-39)$$

在区间  $[t, n]$ , 有聚合的保费函数  $Pr e_i(t, n)$ :

$$\Pr e_i(t, n) = \int_t^n v^{u-t} \left[ \begin{array}{l} I_{\{S(u)=i\}}({}_t P_x^{HH}) + I_{\{S(u)=i\}}({}_t P_x^{\overline{HH}}) + \sum_i I_{\{S(u)=i\}}({}_t P_{x,z}^{S_i}) + \\ \sum_i I_{\{S(u)=i\}}({}_t P_{x,z}^{\overline{S_i}}) + \sum_i I_{\{S(u)=i\}}({}_{\omega,t} P_x^{HS_i}) + I_{\{S(u)=i\}}({}_t P_x^{HD}) \end{array} \right] du \quad (4-40)$$

在时刻  $u$ , 状态  $S(u) = j$ , 以费率  $b_j(t)$  的连续年金收益的精算值为:

$$E \left[ Y^{b_j}_t(u, u+du) | S(t) = i \right] = v^{u-t} \left[ \begin{array}{l} I_{\{S(u)=i\}}({}_t P_x^{HH}) + I_{\{S(u)=i\}}({}_t P_x^{\overline{HH}}) \\ + \sum_i I_{\{S(u)=i\}}({}_t P_{x,z}^{S_i}) + \sum_i I_{\{S(u)=i\}}({}_t P_{x,z}^{\overline{S_i}}) \\ + \sum_i I_{\{S(u)=i\}}({}_{\omega,t} P_x^{HS_i}) + I_{\{S(u)=i\}}({}_t P_x^{HD}) \end{array} \right] du \quad (4-41)$$

当  $u_1 \leq u_2$  时, 定义

$$E \left[ Y^{b_j}_t(u_1, u_2) | S(t) = i \right] = \int_{u_1}^{u_2} v^{u-t} \left[ \begin{array}{l} I_{\{S(u)=i\}}({}_t P_x^{HH}) + I_{\{S(u)=i\}}({}_t P_x^{\overline{HH}}) \\ + \sum_i I_{\{S(u)=i\}}({}_t P_{x,z}^{S_i}) + \sum_i I_{\{S(u)=i\}}({}_t P_{x,z}^{\overline{S_i}}) \\ + \sum_i I_{\{S(u)=i\}}({}_{\omega,t} P_x^{HS_i}) + I_{\{S(u)=i\}}({}_t P_x^{HD}) \end{array} \right] b_j(u) du \quad (4-42)$$

针对整付  $C_{jk}(u)$  在时刻  $u$  时应付的精算值, 如果是在时刻  $u$ , 从状态  $j$  到状态  $k$ , 则有:

$$E \left[ Y_t^{C_{jk}}(u) | S(t) = i \right] = v^{u-t} \left[ \begin{array}{l} I_{\{S(u)=i\}}({}_t P_x^{HH}) + I_{\{S(u)=i\}}({}_t P_x^{\overline{HH}}) \\ + \sum_i I_{\{S(u)=i\}}({}_t P_{x,z}^{S_i}) + \sum_i I_{\{S(u)=i\}}({}_t P_{x,z}^{\overline{S_i}}) \\ + \sum_i I_{\{S(u)=i\}}({}_{\omega,t} P_x^{HS_i}) + I_{\{S(u)=i\}}({}_t P_x^{HD}) \end{array} \right] \mu_{jk}(u) C_{jk}(u) du \quad (4-43)$$

当  $u_1 \leq u_2$  时, 定义

$$\begin{aligned}
 E[Y_t^{C,jk}(u) | S(t) = i] &= \int_{u_1}^{u_2} v^{u-t} \\
 &\left[ \begin{aligned}
 &I_{\{S(u)=i\}}({}_t P_x^{HH}) + I_{\{S(u)=i\}}({}_t P_x^{\overline{HH}}) \\
 &+ \sum_i I_{\{S(u)=i\}}({}_t P_{x,z}^{S,S}) + \sum_i I_{\{S(u)=i\}}({}_t P_{x,z}^{\overline{S,S}}) \\
 &+ \sum_i I_{\{S(u)=i\}}({}_{\omega,t} P_x^{HS}) + I_{\{S(u)=i\}}({}_t P_x^{HD})
 \end{aligned} \right] \\
 &\mu_{jk}(u) C_{jk}(u) du
 \end{aligned} \quad (4-44)$$

从以上的讨论，我们可以得出总的收益(the cumulative benefit)函数  $B_1(t, n)$ ，

$$\begin{aligned}
 B_1(t, n) &= \int_t^n v^{u-t} \left[ \begin{aligned}
 &I_{\{S(u)=i\}}({}_t P_x^{HH}) + I_{\{S(u)=i\}}({}_t P_x^{\overline{HH}}) \\
 &+ \sum_i I_{\{S(u)=i\}}({}_t P_{x,z}^{S,S}) + \sum_i I_{\{S(u)=i\}}({}_t P_{x,z}^{\overline{S,S}}) \\
 &+ \sum_i I_{\{S(u)=i\}}({}_{\omega,t} P_x^{HS}) + I_{\{S(u)=i\}}({}_t P_x^{HD})
 \end{aligned} \right] b_j(u) du \\
 &+ \int_t^n v^{u-t} \left[ \begin{aligned}
 &I_{\{S(u)=i\}}({}_t P_x^{HH}) + I_{\{S(u)=i\}}({}_t P_x^{\overline{HH}}) \\
 &+ \sum_i I_{\{S(u)=i\}}({}_t P_{x,z}^{S,S}) + \sum_i I_{\{S(u)=i\}}({}_t P_{x,z}^{\overline{S,S}}) \\
 &+ \sum_i I_{\{S(u)=i\}}({}_{\omega,t} P_x^{HS}) + I_{\{S(u)=i\}}({}_t P_x^{HD})
 \end{aligned} \right] \mu_{jk}(u) C_{jk}(u) du
 \end{aligned} \quad (4-45)$$

### 4.3 重大疾病保险(Critical Illness Insurance)

随着人们保险意识的不断提高，重大疾病险也逐渐被越来越多的人了解和接受。为自己和家人购买一份大病保险，已成了多数人家家庭理财规划中的首选。重大疾病保险的保险事件是指合同所约定的重大疾病，这类疾病一般是对人体损伤比较严重、费用较高的疾病，可能有多种。同时，为了控制风险，保险公司一般对所保障的重大疾病进行明确、清晰的定义，因此各保险公司对重大疾病的定义可能有一定的差别。而一般寿险的保险事件仅仅是身故，比较单一，所以重大疾病保险相对复杂。

重大疾病保险是针对一些重大疾病而开发的，它的特点就是到一定阶段后是不可逆转的，故此我们给出以下多状态模型。假设有四个状态：健康、轻病、重病和死亡四个状态，它们之间的关系如图 4—3 所示，其特点就是健康状态与轻

病状态之间可以互相转换，健康状态亦可直接到重病状态，这种假设是比较符合实际的；同时假设轻病状态与重病状态之间的转移只有从轻病状态到重病状态，但不可逆转。一般更有死亡状态是吸收状态。

本小节是本文主要创新点之一，提出了重大疾病保险模型，并给出了相应精算值。

### 4.3.1 使用马尔科夫模型建立疾病保险模型

把投保人群划分为四个状态：健康状态，轻病状态，重病状态和死亡状态。四个状态的关系如图 4—3 所示，并且一旦进入重病状态将不可转归，死亡状态是吸收状态。并设  $\mu_{ij}$  为状态之间的转移强度 ( $i, j = 1, 2, 3, 4, i \neq j$ )。

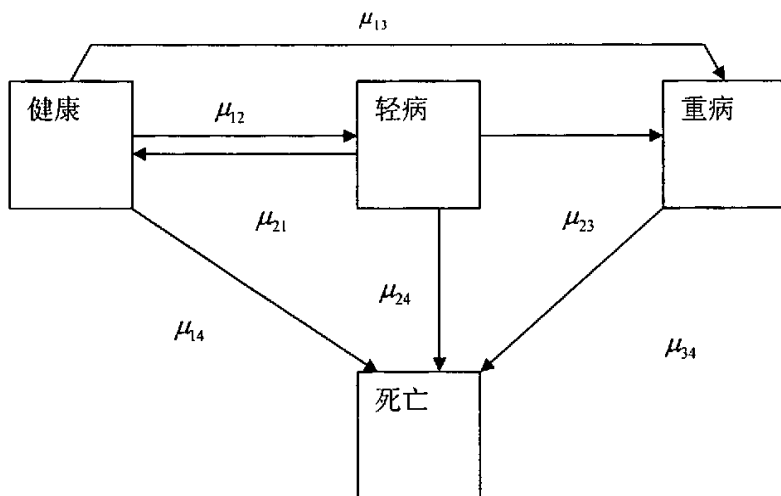


图 4—3 重大疾病保险模型

令  $S(t)$  为  $t$  时刻个体所处的状态，同时设  $S = \{S(t), t \geq 0\}$  是齐时马尔科夫过程 (time-homogeneous Markov process)， $S(0)=1$ ，其转移概率表示为：

$$p_{hi}(s) = p_{hi}(t, t+s; F_t) = p[S(t+s) = i | S(t) = h], t \geq 0 \quad (4-46)$$

同时设  $Q$  为转移强度矩阵，则有 Kolmogorov 后向方程

$$P'(s) = QP(s) \quad (4-47)$$

其中  $p(s) = (p_{ij})_{4 \times 4}$ ,  $\frac{dp(s)}{ds} = \left(\frac{dp_{ij}(s)}{ds}\right)_{4 \times 4}$ 。

图 4—3 所示的各个状态之间的转移强度矩阵为:

$$Q = \begin{pmatrix} -(\mu_{12} + \mu_{13} + \mu_{14}) & \mu_{12} & \mu_{13} & \mu_{14} \\ \mu_{21} & -(\mu_{21} + \mu_{23} + \mu_{24}) & \mu_{23} & \mu_{24} \\ 0 & 0 & -\mu_{34} & \mu_{34} \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad (4-48)$$

则有相关表达式:

$$\begin{aligned} \frac{dp_{11}(s)}{ds} &= -(\mu_{12} + \mu_{13} + \mu_{14})p_{11}(s) + \mu_{12}p_{21}(s) \\ \frac{dp_{12}(s)}{ds} &= -(\mu_{12} + \mu_{13} + \mu_{14})p_{12}(s) + \mu_{12}p_{22}(s) \\ \frac{dp_{13}(s)}{ds} &= -(\mu_{12} + \mu_{13} + \mu_{14})p_{13}(s) + \mu_{12}p_{23}(s) + \mu_{13}p_{33}(s) \\ \frac{dp_{14}(s)}{ds} &= -(\mu_{12} + \mu_{13} + \mu_{14})p_{14}(s) + \mu_{12}p_{24}(s) + \mu_{13}p_{34}(s) \\ \frac{dp_{21}(s)}{ds} &= \mu_{21}p_{11}(s) - (\mu_{21} + \mu_{23} + \mu_{24})p_{21}(s) \\ \frac{dp_{22}(s)}{ds} &= \mu_{21}p_{12}(s) - (\mu_{21} + \mu_{23} + \mu_{24})p_{22}(s) \\ \frac{dp_{23}(s)}{ds} &= \mu_{21}p_{13}(s) - (\mu_{21} + \mu_{23} + \mu_{24})p_{23}(s) + \mu_{23}p_{33}(s) \\ \frac{dp_{24}(s)}{ds} &= \mu_{21}p_{14}(s) - (\mu_{21} + \mu_{23} + \mu_{24})p_{24}(s) + \mu_{23}p_{34}(s) \\ \frac{dp_{33}(s)}{ds} &= -\mu_{34}p_{33}(s) \\ \frac{dp_{34}(s)}{ds} &= -\mu_{34}p_{34}(s) \end{aligned} \quad (4-49)$$

$Q = ADA^{-1}$ , 其中  $D = \text{diag}(d_1, d_2, \dots, d_n)$ , A 的第  $i$  列是伴随矩阵  $d_i$  的右特征向量,

$$P(s) = A \text{diag}(e^{d_1 s}, e^{d_2 s}, \dots, e^{d_n s}) A^{-1} \quad (4-50)$$

$$P_{ij}(s) = \sum_{l=1}^n a_{il} c_{lj} e^{d_l s} \quad (4-51)$$

其中  $a_{il}$  是 A 的元素,  $c_{lj}$  是  $C = A^{-1}$  的元素。因为不易得到 Q 的特征值, 故还利用相关符号来表示, 在数值解时, 可以得到相关特征值。

### 4.3.2 精算计算

在  $t$  时刻, 状态  $h$  下, 设  $\delta$  为年利息力,  $p_i(t)$  为每年的被保险人的保费的连续费率,  $b_i(t)$  为连续年金的费率,  $c_{ij}(t)$  为从状态  $i$  到状态  $j$  所花费的医疗费率,  $i=1, 2, 3; j=1,2,3,4$ 。

由时间齐次马尔科夫过程的性质和重大疾病保险的多状态模型, 该保险保费的精算现值为:

$$Y_t^p = \sum_{i=1}^3 \int_0^{\infty} e^{-\delta s} p_{hj}(s) p_i(s) ds + \sum_{i=1}^3 \sum_{\substack{j=1 \\ i \neq j}}^4 e^{-\delta s} p_{hi}(s) \mu_{ij} c_{ij} ds \quad (4-52)$$

连续年金的精算现值为:

$$Y_t^b = \sum_{i=1}^3 \int_0^{\infty} e^{-\delta s} p_{hj}(s) b_i(s) ds \quad (4-53)$$

则可直接给出总的受益函数的精算现值

$$B_t = \sum_{i=1}^3 \int_0^{\infty} e^{-\delta s} p_{hj}(s) p_i(s) ds + \sum_{i=1}^3 \int_0^{\infty} e^{-\delta s} p_{hj}(s) b_i(s) ds + \sum_{i=1}^3 \sum_{\substack{j=1 \\ i \neq j}}^4 e^{-\delta s} p_{hi}(s) \mu_{ij} c_{ij} ds \quad (4-54)$$

这里只须计算从状态 1 到 3 的和, 因为  $c_{4i} = 0, i=1,2,3,4$ 。

以上公式就是在疾病发展的多状态模型基础计算得出的, 在实际中可根据具体情况来计算相应的保费。

## 4.4 总结与讨论

综上, 使用马尔科夫多状态模型的方法研究了三个有关医疗保险的问题, 并给出了相应的模型以及有关的保费问题。这里主要运用多状态模型详细的描述了疾病保险过程, 然后使用数学、统计和精算等工具来计算保费和相关精算受益。但现实生活中的情况是很复杂的, 多状态模型并没有概括所有的发病原由, 对于其他情形有待进一步研究, 对于疾病状态的划分, 应和医学专家一起合作, 并应用相关统计知识进行分类, 确定转移强度和转移概率。

## 第五章 比例动态保费模型

首先要说明是为什么要建立一个动态的保费模型。因为我们的被保人群是很复杂的，既有健康人，又有患者，同时患者中有重病患者，又有比较轻微的患者，疾病状态又分轻重缓急，又有不同的发展阶段。为了更好地去收取保费，又体现公平，同时控制风险，我们给出一个动态的保费模型使得患者和保险机构都获得较大的利益。

一般的方法是根据实际情况和数据并结合相关知识将疾病过程分解为各个状态得出各个状态的分布，即子分布（见图 5—1），然后再拟合混合的分布（见图 5—2）。由图 5—1 可以看出子分布有交叉的地方，也就是说不同风险有叠加的地方，即不同的疾病风险相互叠加。那么对处于不同疾病风险状态的人员怎么收取保费呢？本文提出一个比例动态保费模型，就是给各个风险分布一个比例，然后再计算保费。下面给出相应的数学叙述和证明。

比例动态模型保费模型是该文的创新点之一。

### 5.1 比例动态保费模型

这里主要使用保费均衡原则和纯保费模型来计算动态保费。下面以疾病保险的过程为例进行说明。首先介绍一下纯保费的公式，记纯保费为  $I_t$ ，则

$$I_t = \int xf(x)dx, \text{ 其中 } f(x) \text{ 为相应的损失分布。}$$

#### 5.1.1 一般比例动态保费

为简单起见，首先从两个不同的损失分布来考虑此问题。

一般动态保费模型是按照一定的比例去收取保费的，假设有某种疾病分为疾病状态(1)和疾病状态(2)，且分别对应不同的疾病损失模型  $f_1(x)$  和  $f_2(x)$ ，有一般动态保费模型

$$I_t(A) = \int_{-\infty}^{\infty} xL_1(t)f_1(x)dx + \int_{-\infty}^{\infty} xL_2(t)f_2(x)dx \quad (5-1)$$

$$L_1(t) = \frac{l_1(t)}{l_1(t) + l_2(t)}, L_2 = \frac{l_2(t)}{l_1(t) + l_2(t)} \quad (5-2)$$

其中的  $f_1(x)$ ,  $f_2(x)$ , 式(5-2)具体表示见图 5-1。

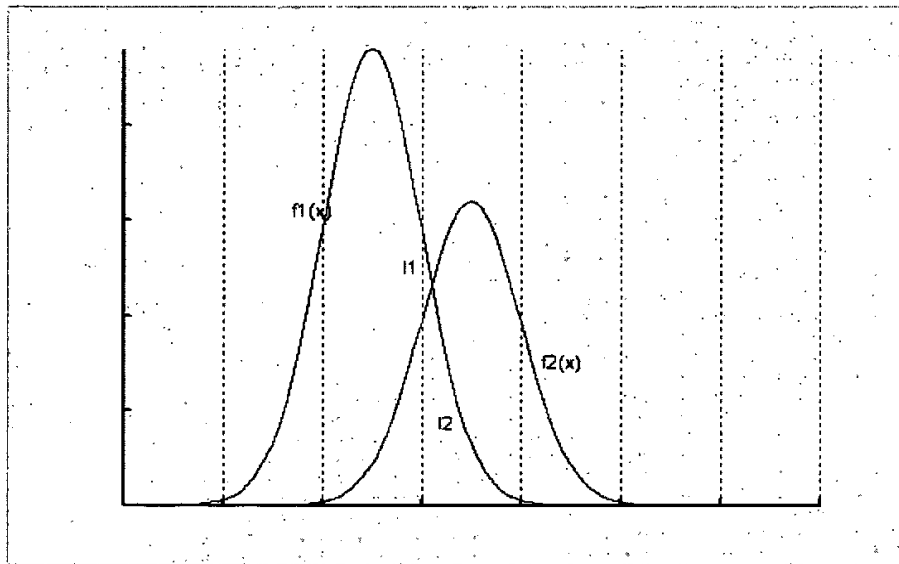


图 5-1 两个不同分布的示意图

下面证明这个按比例收取的纯保费动态模型与由(1)和(2)的混合分布建立的保费模型是等价的。

令  $f(x)$  是这个混合分布, 则有

$$f(x) = L_1(t)f_1(x) + L_2(t)f_2(x), L_1(t) = \frac{l_1(t)}{l_1(t) + l_2(t)}, L_2(t) = \frac{l_2(t)}{l_1(t) + l_2(t)} \quad (5-3)$$

混合分布的保费模型:

$$I_t(B) = \int_{-\infty}^{\infty} x f(x) dx \quad (5-4)$$

下证  $I_t(A) = I_t(B)$ 。

证明:

$$\begin{aligned} I_t(B) &= \int_{-\infty}^{\infty} x f(x) dx = \int_{-\infty}^{\infty} x (L_1(t)f_1(x) + L_2(t)f_2(x)) dx = \\ &= \int_{-\infty}^{\infty} x L_1(t)f_1(x) dx + \int_{-\infty}^{\infty} x L_2(t)f_2(x) dx = I_t(A) \end{aligned}$$



证毕。

由此可以看出建立的这个动态保费模型是可行的。若将其推广到多个分布的情形，可得出一个较为复杂的动态保费模型：

$$I_t = \int_{-\infty}^{\infty} xL_1(t)f_1(x)dx + \int_{-\infty}^{\infty} xL_2(t)f_2(x)dx + \dots + \int_{-\infty}^{\infty} xL_n(t)f_n(x)dx \quad (5-5)$$

其中， $L_i(t) = \frac{l_i(t)}{\sum_{j=1}^n l_j(t)}$ 。同理可以证明，它与相应的基于混合分布的保费模型是

等价的。

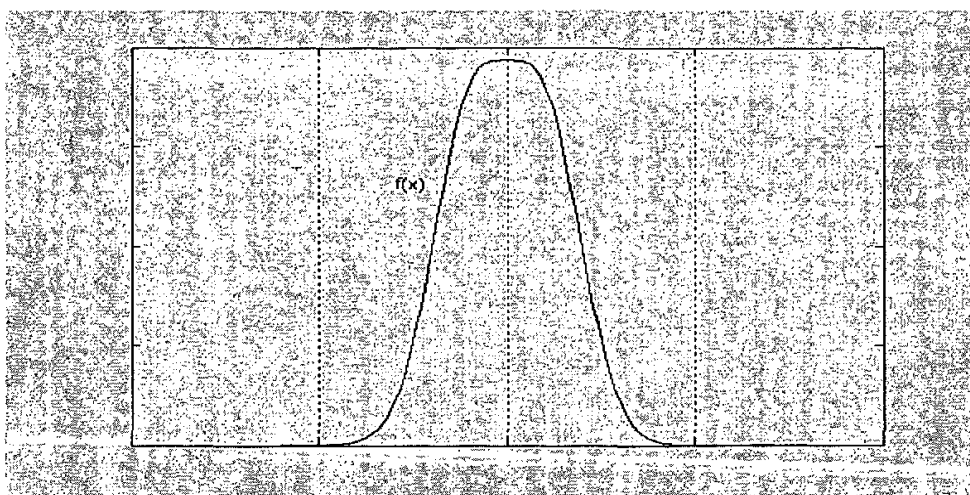


图 5—2 混合分布示意图

### 5.1.2 权重比例动态保费模型

以上给出的一般比例动态保费模型，对于各个分布并没有给予一定的权重，即默认权重为 1，下面将给予各个损失分布权重，使比例动态模型更贴近实际。

先给出简单的权重动态模型，仍记  $I(t)$  为动态保费，分别赋予分布  $f_1(x)$ 、

$f_2(x)$  权重  $p, q$ , 且  $p+q=1$ 。

$$\begin{aligned} f(x) &= L_1(t)pf_1(x) + L_2(t)qf_2(x), \\ \text{记} \quad L_1(t) &= \frac{l_1(t)}{l_1(t)+l_2(t)}, L_2(t) = \frac{l_2(t)}{l_1(t)+l_2(t)} \end{aligned} \quad (5-6)$$

则动态保费为:

$$I_t(A) = \int_{-\infty}^{\infty} xL_1(t)pf_1(x)dx + \int_{-\infty}^{\infty} xL_2(t)qf_2(x)dx \quad (5-7)$$

对应于混合分布  $f(x)$  的动态保费  $I_t(B) = \int_{-\infty}^{\infty} xf(x)dx$ , 下证

$$I_t(A) = I_t(B).$$

证明:

$$\begin{aligned} I_t(B) &= \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{\infty} x(L_1(t)pf_1(x) + L_2(t)qf_2(x))dx = \\ &= \int_{-\infty}^{\infty} xL_1(t)pf_1(x)dx + \int_{-\infty}^{\infty} xL_2(t)qf_2(x)dx = I_t(A) \end{aligned}$$

证毕。

由此, 可以看出在给各个子分布加上权重之后, 混合分布的保费和各个子分布对应保费之和仍然相等。

若将其推广到多个分布的情形, 可得出一个较为复杂的动态保费模型, 即:

$$I_t = \int_{-\infty}^{\infty} xL_1(t)p_1f_1(x)dx + \int_{-\infty}^{\infty} xL_2(t)p_2f_2(x)dx + \cdots + \int_{-\infty}^{\infty} xL_n(t)p_nf_n(x)dx \quad (5-8)$$

其中,  $p_1+p_2+\cdots+p_n=1$ ,  $L_i(t) = \frac{l_i(t)}{\sum_{i=1}^n l_i(t)}$ 。同理可以证明, 它与相应的基于混

合分布的保费模型是等价的。

## 5.2 破产概率和准备金

这里将讨论比例动态保费模型的破产概率及其准备金的问题。杜勇宏等人<sup>[29]</sup>

已经讨论了具有两类索赔的风险过程的破产概率，本文将其结论引用如下：

### 5.2.1 复合二元风险模型

#### 1) 复合二元风险模型

假设保险人承保多种风险，从而有多个类别的索赔。为简明起见，只考虑两种风险的情况。与古典情形类似定义盈余过程为  $U(t) = u + ct - \sum_{i=1}^{N_1(t)} X_i - \sum_{j=1}^{N_2(t)} Y_j$ ，式中  $u$  为初始准备金， $c > 0$  为单位时间的保费收入， $N_1(t)$  和  $N_2(t)$  是两类索赔的计数过程， $X_i, Y_j, N_1(t), N_2(t)$  相互独立， $X_i$  的共同分布为  $F(x)$ ， $Y_j$  的共同分布为  $G(x)$ ， $N_1(t), N_2(t)$  均为 Erlang(2) 过程，它们的跳跃时间间隔分别为  $L_1^2$  和  $L_2^2$ ，均服从 Erlang(2) 分布，为了能够得出关于破产概率的一些明确结论，把 Erlang(2) 分布表示为两个独立的指数分布的和，即假设  $L_1^2 = T_{11} + T_{12}$ 、 $L_2^2 = \tau_{11} + \tau_{12}$ ，其中  $T_{11}$ ， $T_{12}$ ， $T_{21}$ ， $T_{22}$ ……为具有均值  $1/\lambda_1$  的服从指数分布的独立随机变量列， $\tau_{11}$ ， $\tau_{12}$ ， $\tau_{21}$ ， $\tau_{22}$ ……为具有均值  $1/\lambda_2$  的服从指数分布的独立随机变量列。

根据均值原理，保费收入应大于期望索赔，因此有  $c > (\lambda_1 \mu_1 + \lambda_2 \mu_2) / 2$ ，式中  $\mu_1$ 、 $\mu_2$  为  $F(x)$  和  $G(x)$  的均值。

#### 2) 几种破产概率及其微分方程组

令  $\Phi(u)$  表示生存概率，即  $\Phi(u) = p(U_t \geq 0, t \geq 0)$ 。此时， $\Psi(u) = 1 - \Phi(u)$  为破产概率，为了研究  $\Phi(u)$  满足的方程，需引入如下辅助的生存概率及破产概率：

a: 改变基本模型的  $L_1^2 = \tau_{11} + \tau_{12}$  为  $L_1^2 = \tau_{12}$ ，记相应的生存概率为  $\Phi_1(u)$ ；

b: 改变基本模型的  $L_1^2 = T_{11} + T_{12}$  为  $L_1^2 = T_{12}$ ，记相应的生存概率为  $\Phi_2(u)$ ；

c: 改变基本模型的  $L_1^2 = T_{11} + T_{12}$  为  $L_1^2 = T_{12}$ ，同时改变  $L_2^2 = \tau_{11} + \tau_{12}$  为  $L_2^2 = \tau_{12}$  记相应的生存概率为  $\Phi_3(u)$ 。

先考虑基本模型，令  $W = \min\{T_{11}, T_{12}\}$ ，在时刻  $W$  之前没有索赔发生，可以利用指数分布的无记忆性。如果  $W = \tau_{11}$ ，则此时从  $W$  开始，模型变为改变后的模型 a；如果  $W = T_{11}$ ，则此时从  $W$  开始，模型变为改变后的模型 b。简单的计算可知，

$$p(W = \tau_{11}) = \lambda_2 / (\lambda_1 + \lambda_2) \quad (5-9)$$

$$p(W = T_{11}) = \lambda_1 / (\lambda_1 + \lambda_2) \quad (5-10)$$

$$p(|W > x | W = \tau_{11}) = p(|W > x | W = T_{11}) = e^{-(\lambda_1 + \lambda_2)x} \quad (5-11)$$

因此, 以  $W$  为求生存概率的条件, 并区分  $W = T_{11}$  和  $W = \tau_{11}$  两种情况, 可以得出:

$$\Phi(u) = \frac{\lambda_2}{\lambda_1 + \lambda_2} \int_0^{\infty} (\lambda_1 + \lambda_2) e^{-(\lambda_1 + \lambda_2)t} \Phi_1(u + ct) dt + \frac{\lambda_1}{\lambda_1 + \lambda_2} \int_0^{\infty} (\lambda_1 + \lambda_2) e^{-(\lambda_1 + \lambda_2)t} \Phi_2(u + ct) dt \quad (5-12)$$

利用类似的方法, 从其他几种辅助模型出发, 并注意到是否有索赔发生, 可以得到

$$\Phi_1(u) = \int_0^{\infty} \pi_1(t) \int_0^{u+ct} \Phi(u + ct - x) dG(x) dt + \int_0^{\infty} \pi_2(t) \Phi_3(u + ct) dt \quad (5-13)$$

$$\Phi_2(u) = \int_0^{\infty} \pi_1(t) \Phi_3(u + ct) dt + \int_0^{\infty} \pi_2(t) \int_0^{u+ct} \Phi(u + ct - x) dF(x) dt \quad (5-14)$$

$$\Phi_3(u) = \int_0^{\infty} \pi_1(t) \int_0^{u+ct} \Phi_2(u + ct - x) dG(x) dt + \int_0^{\infty} \pi_2(t) \int_0^{u+ct} \Phi_1(u + ct - x) dF(x) dt \quad (5-15)$$

**定理 5.1:** 生存概率  $\Phi(u)$ ,  $\Phi_1(u)$ ,  $\Phi_2(u)$ ,  $\Phi_3(u)$ , 满足如下微积分方程组:

$$c\Phi'(u) = -\lambda_2\Phi_1(u) - \lambda_1\Phi_2(u) + (\lambda_1 + \lambda_2)\Phi(u) \quad (5-16)$$

$$c\Phi_1'(u) = -\lambda_2 \int_0^u \Phi(u-x) dG(x) - \lambda_1\Phi_3(u) + (\lambda_1 + \lambda_2)\Phi_1(u) \quad (5-17)$$

$$c\Phi_2'(u) = -\lambda_2\Phi_3(u) - \lambda_1 \int_0^u \Phi(u-x) dF(x) + (\lambda_1 + \lambda_2)\Phi_2(u) \quad (5-18)$$

$$c\Phi_3'(u) = -\lambda_2 \int_0^u \Phi_2(u-x) dG(x) - \lambda_1 \int_0^u \Phi_1(u-x) dF(x) + (\lambda_1 + \lambda_2)\Phi_3(u) \quad (5-19)$$

该方程可以通过线性微分方程组的求解公式得到。由以上的方程, 可以结合实际问题的就可以解得相关的破产概率。下面将讨论准备金的问题。

**定义 5.1:** 设聚合理赔过程  $S(t)$  为复合泊松过程, 泊松参数为  $\lambda$ , 个别理赔额为  $C$ , 其矩母函数为  $M_C(r)$ , 方程:

$$\lambda + rc = \lambda M_C(r) \quad (5-20)$$

的非零正解记作  $R$ , 称为调节系数, 其中  $c$  为保险费率可记

$$c = (1 + \theta)\lambda p_1 \quad (5-21)$$

即保费在纯保费  $\lambda p_1$  的基础上按附加费率  $\theta$  收取, 则上述公式为:

$$1 + (1 + \theta) p_1 r = M_C(r) \quad (5-22)$$

其中  $p_1 = E(C)$ 。

定理 5.2: 由破产概率定理

$$\psi(u) = \frac{e^{-Ru}}{E[e^{-RU(T)} | T < \infty]} \quad (5-23)$$

可知调节系数  $R$  是一个非常小的数, 近似可得:

$$\psi(u) < e^{-Ru} \quad (5-24)$$

其中  $u$  是准备金, 而  $\psi(u)$  是准备金在不同水平的破产概率, 则可利用已知数据做准备金与破产概率的探索研究。

由以上的定理和定义, 当得到调节系数和破产概率在基础上就可以得到相应的准备金。

在以上的内容里, 主要讨论了有关动态保费在公式、破产概率和准备金等问题, 为其实际应用打下了坚实的理论基础。

## 第六章 癌症保险

### 6.1 癌症保险模型和公式

本节以上海杨浦区的恶性肿瘤的调查数据为例,介绍如何使用马尔科夫链在实际中建模求解保费的。在这里所用的方法公式引自文献<sup>[23]</sup>。为计算方便,将整个疾病的状态空间划分如下:健康(H)、癌症(C)和死亡(D)三个状态。第 $t$ 年 $x$ 岁末时未来 $n$ 年的疾病过程见图6—1。…

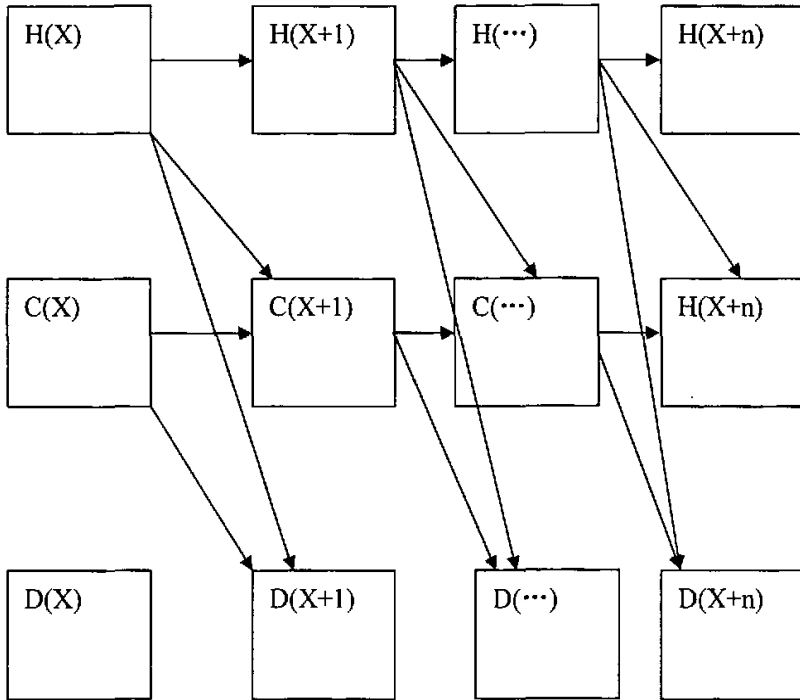


图6—1 癌症保险模型

在癌症的疾病过程中健康状态在下一年有健康、患病及死亡三种可能状态(出于计算的方便,与第五章中的重大疾病保险模型相比,我们省略一个疾病状

态), 而患病在下一年有患病以及死亡两种可能状态, 其中  $x$  岁的人在下一年度各状态间的转移概率可以通过转移概率矩阵表示如式(6—1):

$$\begin{bmatrix} (1-m_x)(1-I_x) & (1-m_x)I_x & m_x \\ 0 & 1-m_x^c & m_x^c \\ 0 & 0 & 1 \end{bmatrix} \quad (6-1)$$

多年度各状态转移概率矩阵可以通过转移概率矩阵相乘的方式来计算, 在死亡率递减的假设下多年度各状态转移概率的计算可以由下式表示:

$${}_n P_x(t) = {}_{n-1} P_x(t) * \begin{bmatrix} (1-m_{x+n-1})(t+n-1)(1-I_{x+n-1}) & (1-m_{x+n-1})(t+n-1)I_{x+n-1} & m_{x+n-1}(t+n-1) \\ 0 & 1-m_{x+n-1}^c(t+n-1) & m_{x+n-1}^c(t+n-1) \\ 0 & 0 & 1 \end{bmatrix} \quad (6-2)$$

其中,  ${}_n P_x(t)$  表示  $t$  年  $x$  岁末时未来  $n$  年转移概率矩阵;

$m_{x+n-1}(t+n-1)$  表示  $t$  年  $x$  岁末的一般人  $n-1$  年后死亡率;

$I_{x+n-1}$  表示  $x+n-1$  岁癌症发生率;

$m_{x+n-1}^c(t+n-1)$  表示  $t$  年  $x$  岁末的癌症患者  $n-1$  年后死亡率。

每位被保险人的癌症保险费用可以依据精算中保费均衡原则计算如下:

$$pre_x = \frac{\sum_{j=1}^n w_x(t+j-1) * {}_n P_x(t)_{(1,2)} * MTE_{x+j-1} * v_1^{-0.5}}{1 + \sum_{j=1}^n (1-m_x(t+j-1)) * v_2'} + \frac{\sum_{j=0}^n (1-w_x)(t+j-1) * {}_n P_x(t)_{(2,2)} * MTE_{x+j-1} * v_1^{-0.5}}{1 + \sum_{j=1}^n (1-m_x(t+j-1)) * v_2'} \quad (6-3)$$

其中,

$pre_x$  表示以保费均衡原则下所计算  $x$  岁的人在癌症门诊部分的平衡年交保费;

${}_n P_x(t)_{(i,j)}$  表示  $t$  年  $x$  岁末时未来  $n$  年转移概率矩阵中位于  $i$  行  $j$  列的元素;

$w_x(t+j) = \frac{I_x(t+j) - s_x(t+j)}{I_x(t+j)}$  表示第  $t+j$  年  $x$  岁健康人口占总人口的比例;

$MTE_{x+j}$  表示  $x+j$  岁每人年平均癌症门诊医疗金额;

$$v_1 = \frac{1+r_m}{1+r_i}, \quad r_i \text{ 为通货膨胀率, } r_m \text{ 为医疗物资水平增长率;}$$

$$v_2 = \frac{1+r_s}{1+r_i}, \quad r_s \text{ 为工资水平增长率。}$$

## 6.2 计算一年期癌症保险保费

下面给出一年期的癌症保险保费, 该保费的计算是混合型的, 没有分男女性别来计算。首先说明数据来源, 数据来自《人口 疾病 保险》<sup>[26]</sup>中上海杨浦区恶性肿瘤调查分析的数据, 具体数据见附录 C, 这里使用的寿命表是中国人寿保险业经验生命表(1990-1993)(混合表)。根据均衡保费原则给出纯保费。

参数的假设: 不妨假设  $r_i = 2\%$ ,  $r_m = 3\%$ ,  $r_s = 4\%$ , 则  $v_1 = \frac{1+r_m}{1+r_i} = 1.0098$ ,

$$v_2 = \frac{1+r_s}{1+r_i} = 1.0196。$$

以 40 岁为例, 取  $n=1$ ,  $x=40$ ,  $j=1$ ,  $v_1 = 1.0098$ ,  $v_2 = 1.0196$ ,  $w_{40}(t) = 0.998471$ ,  $w_{40}(t-1) = 0.999194$ ,  $MTE_{40} = 22901.99$ ,  $MTE_{40-1} = 29743.435$ ,  ${}_1P_{40}(t)_{(1,2)} = 0.00155255565$ ,  ${}_1P_{40}(t)_{(2,2)} = 0.6083$ ,  $m_{40}(t) = 0.3197$ , 代入(6—3)可以得出 40 岁时所应交的一年期保费为 55.33652 元, 在癌症发生时即可获得 22901.99 元的医疗补助。

以 60 岁为例, 同理可计算出一年期的癌症保险有关费用, 投保人交纳 222.9514 元的保费, 在疾病发生时即可一次获得 25607.46 元的医疗补助。分年龄阶段来计算年趸缴保费和保险理赔额(见表 6—1)(为简单起见把金额四舍五入到个位数)。

以天安保险公司面向大众的抗癌安康保险为例, 该险种每份年保费为 90 元, 保额 1 万元, 如被保险人购买保险 180 天后不幸患上癌症, 且经确诊为初次患癌症, 便可按保险条款约定获得保险公司的赔偿。此外, 天安保险还专门为女性度身打造了抗癌的女性安康保险。女性安康险也是一年期短期健康险, 但与抗癌安康险相比, 此险种保费更低, 女性只需缴纳 25 块钱的年保费, 如在购买保险后



180天不幸初次患上癌症，就可按保险条款约定获得保险公司赔偿。我们可以看到，我们的结果在被保险人在分年龄阶段所要交纳保费和实际还是很接近的。

表 6—1 一年期的癌症保险的保费表格(元)

年龄阶段	年趸缴保费	保险理赔额
30~34	26	33952
35~39	44	29743
40~44	55	22902
45~49	70	21286
50~54	120	26245
55~59	140	21148
60~64	223	25607
65~69	240	24236
70~74	448	31280
75~79	559	41588
80~	364	49947

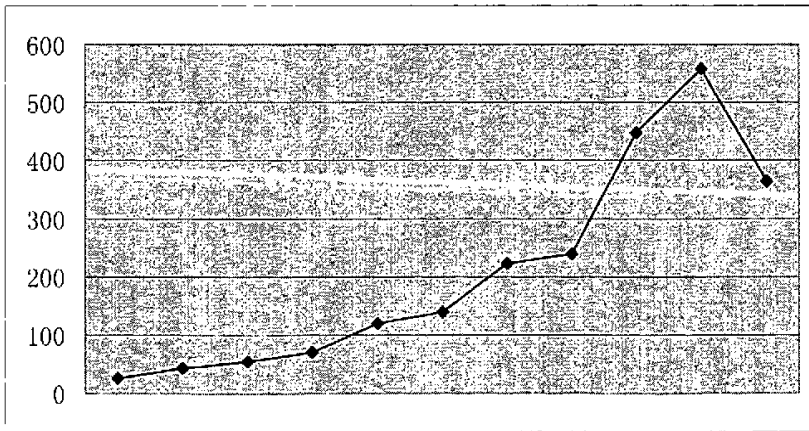


图 6—1 癌症保险保费折线示意图（横坐标表示年龄，纵坐标为金额）

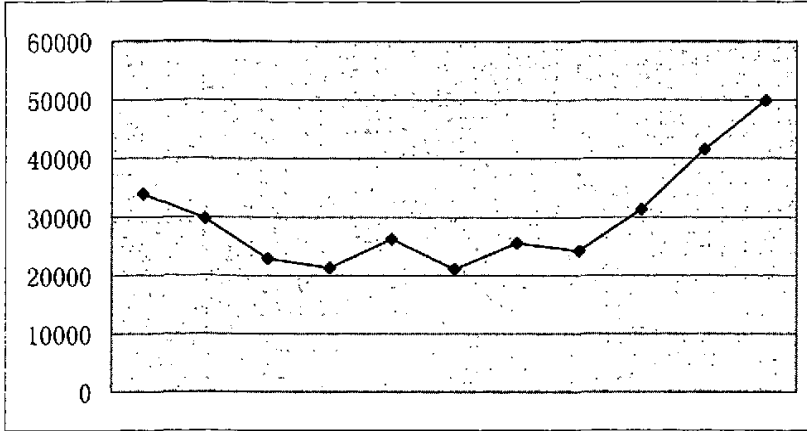


图 6—2 癌症保险理赔额折线示意图（横坐标表示年龄，纵坐标为金额）

## 第七章 结论与讨论

本文主要讨论多状态模型在医疗保险中的应用，将多状态模型引入保险精算中，使精算人员更好的估计疾病以及被保险人的风险，同时更好的进行保险产品的定价，我们并给出了几个相应的保险产品的具体的模型以及保费，但很遗憾由于时间和数据的关系没能针对中国大陆市场进行细致的保费和其他费用的测算，以后将更进一步改进模型并将其引入实际应用层次。

第一章简要的介绍了马尔科夫模型在医疗保险中的应用的情况和相关背景。

在第二章中，给出了移民—疾病—死亡过程，并用该过程描述了各个状态的期望人数，得到一个相关结论：即在转移强度为常量的情况下，各个状态的期望人数是稳定的。这个结论给使用往年的数据建立相关的医疗保险模型提供了理论依据。因为如果每年的患者人数波动很大的话，将无法参照往年的数据来进行数学建模处理有关数据。在这里，主要是以数学的语言进行描述和证明。

在第三章我们讨论了多状态模型与保险精算的关系，2004年 James W.Daniel<sup>[7]</sup>给出了离散时间状态的有关多状态模型的各种现金流的公式，本文将推广到连续时间状态。

在第四章中，讨论了如何使用多状态的马尔科夫模型去描述相关保险产品，重点讨论了三个保险：长期健康保险、长期护理保险和重大疾病保险，给出相关的多状态模型和有关精算值。同时应该指出的是虽然长期健康保险，重大疾病保险，长期护理保险都应属于医疗保险，但它们之间还有细微的差别，因此以上所给出的模型是有所不同的，还应该结合实际与其它专家更深一层的讨论状态的划分以及转移强度，转移概率的确定，保证保险机构的收益以及被保险人的权益等。

第五章中我们讨论一个有关多风险情况下的动态保费模型，我们并在这里讨论了相关的破产概率和准备金问题。在这里建立动态保费的主要目的就是收取更公平科学的保费，保护被保险人的权益。

在第六章中，我们应用有关数据得到了一个一年期的癌症保险，给出了一个有关年龄分段的保费，但我们的数据不足，未能进行很好的分析建立相应更好的

保费、准备金和利润模型。

在以上的讨论中，我们有许多工作还没有做好，如：怎么更好地得到转移概率，有待更好的研究。有研究人员已经提出一些比较好的方法如广义线性模型法、贝叶斯方法等。还有，我们的数据搜集工作没有做好，所得数据不足。这些都要一一加以注意，建立完备的统计数据库、方法库等等，以备研究之用。

## 参考文献

- [1] Maria, I. and Cordeiro, F., **A multiple state model for the analysis of permanent health insurance claims by cause of disability**, Insurance: Mathematics and Economics 2002, 30: 167-186.
- [2] Hesselager, O. and Norberg, R., **On probability distributions of present values in life insurance**, Insurance: Mathematics and Economics 1996, 18:35-42.
- [3] Pitacco, E., **Multistate models for long-term care insurance and related indexing problems**, Applied stochastic models in business and industry, 1999,15:429-441.
- [4] Gutierrez, M.C. and Macdonald, A.S., **Huntington's disease, Critical Illness Insurance and Life Insurance**, Scandinavian Actuarial Journal, 2004, 279-313.
- [5] Macdonald, A.S., Waters, H.R. and Wekwete, C.T., **A model for coronary heart disease and stroke, with applications to critical illness insurance underwriting I: The model**, North American Actuarial Journal, 2005,9: 1,41-56.
- [6] Macdonald, A.S., Waters, H.R. and Wekwete, C.T., **A model for coronary heart disease and stroke, with applications to critical illness insurance underwriting II: Applications**, North American Actuarial Journal, 2005,9:1, 13-40.
- [7] Daniel, W.J., **Multi-state transition models with actuarial applications**, Casualty Actuarial Society and Society of Actuaries 2004
- [8] Helms, F., Czado, C. and Gschlobl, S., **Calculation of LTC Premiums based on direct estimates of transition probabilities**, Discussion Paper 393 beim SFB 386 "Diskrete Strukturen", Technische Universität München, 2004.
- [9] Steven, H., Zoltan, B., Rickayzen, B. D., **Multiple state models, simulation and insurer insolvency**, Actuarial Research Paper 2001 no. 136.
- [10] Ocon, P.R. and Ruizcastro, J.E., **A Multiple-Absorbent Markov Process in Survival Studies: Application to Breast Cancer**, Biometrical Journal 2003, 45:7, 783-797

- [11] Cole, F.B., Bonetti, M., Zaslavsky, M.A. and Gelber, D.R., **A multistate Markov chain model for longitudinal, categorical quality-of-life data subject to non-ignorable missing ness**, *Statistic. Med.* 2005; 24:2317–2334
- [12] Lagazio, C., Biggeri, A. and Dreassi, E., **Age–period–cohort models and disease mapping**, *Environmetrics* 2003; 14: 475–490.
- [13] Sama1, W., Owusu, A.S., Felger, I., Vounatsou, P. and Smith,T., **An immigration–death model to estimate the duration of Malaria infection when detectability of the parasite is imperfect**, *Statist. Med.* 2005; 24:3269–3288.
- [14] Hougaard, P., **Multi-state Models: A Review**, *lifetime Data Analysis*,1999 5,239-264.
- [15] Keiding, N., Klein, P.J. and Horowitz, M.M., **Multi-state models and outcomepre diction in bonemarrow transplantation**, *Statist. Med.* 2001, 20:1871–1885.
- [16] Capasso, V., Salani, C., **Stochastic birth-and-growth processes modeling crystallization of polymers with spatially heterogeneous parameters**, *Nonlinear Analysis: Real World Applications* 2000,1: 485-498
- [17] Taylor, J.G., McClean, I.S., and Millard, H.P., **Using a continuous-time Markov model with poisson arrivals to describe the movements of geriatric patients**, *Appl. Stochastic Models Data Anal.*, 1998,14: 165-174
- [18] 柳青, 方积乾, 胡孟璇 **鼻咽癌疾病自然史的随机模型研究** *中国卫生统计* 1997年第14卷第4期。
- [19] 王心旺 ,方积乾 **糖尿病保险费精算模型研究** *广州医学院学报* 2004,32(3): 4-7.
- [20] 蒋庆琅 著 方积乾译 **随机过程原理与生命科学** 上海翻译科学出版社, 上海 1987。
- [21] 方积乾 著 **医学统计学与电脑实验** 上海科学技术出版社 上海。
- [23] 黄泓智, 刘明昌, 余清祥 **台湾地区重大伤病医疗费用推估** *人口学刊* 2004.12 第29期。
- [24] 王梓坤 杨向群 著 **生灭过程与马尔科夫链** 科学出版社, 北京。

- [25] 钱敏平 龚光鲁 著 随机过程论 北京大学出版社 北京。
- [26] 尚汉冀 著 人口 疾病 保险 复旦大学出版社 上海。
- [27] 谢志刚 韩天雄 著 风险理论与非寿险精算 南开大学出版社, 2000。
- [28] (荷兰)R.卡尔斯 等著 现代精算风险理论 科学出版社北京 2005。
- [29] 杜勇宏, 兰德新, 阮淑萍, 具有两类索赔的风险过程的破产概率, 华中科技大学学报(自然版) 2005, 8。
- [30] 周江雄, 刘建华, 黎颖芳 著, 生命表在构造理论, 南开大学出版社, 天津, 2001。

## 附录 A: 马尔科夫 (Markov) 模型

### 一 马尔科夫 (Markov) 过程

我们设:  $(X, B)$  为完备的可测空间,  $(\Omega, F, P)$  为完备概率空间  
 $\zeta_t, t \in I \subset \square$  为  $(\Omega, F, P)$  上取值于  $(X, B)$  的随机过程。

定义:

$$\begin{aligned} F_{\zeta_t} &:= \delta\{\zeta_s, s \in I, s \leq t\}, \\ F_{\zeta_t} &:= \delta\{\zeta_s, s \in I, s \geq t\}, \\ F_{[s,t]} &:= \delta\{\zeta_u, u \in I, u \in [s,t]\}, \\ F_{\zeta_t} &:= \delta\{\zeta_t\}, \end{aligned} \tag{A-1}$$

定义 A.1 若对任意  $A \in F_{\zeta_t}, B \in F_{\zeta_t}$  有

$$P(AB | F_{\zeta_t}) = P(A | F_{\zeta_t})P(B | F_{\zeta_t}) \tag{A-2}$$

则  $\zeta_t$  称为马尔科夫过程。

假设马尔科夫过程  $\{X_n, n \in T\}$  的参数集  $T$  是离散时间集合, 即  
 $T = \{0, 1, 2, \dots\}$ , 相应  $X_n$  可能取值的全体组成的状态空间是离散的状态集  
 $I = \{i_1, i_2, i_3, \dots\}$ 。

定义 A.2 设有随机过程  $\{X_n, n \in T\}$ , 若对任意的整数  $n \in T$  和任意  
 $i_0, i_1, \dots, i_{n+1} \in I$ , 条件概率满足:

$$\begin{aligned} P\{X_{n+1} = i_{n+1} | X_0 = i_0, X_1 = i_1, \dots, X_n = i_n\} \\ = P\{X_{n+1} = i_{n+1} | X_n = i_n\} \end{aligned} \tag{A-3}$$

则称  $\{X_n, n \in T\}$  为马尔科夫链。

考虑取非负整数的连续时间随机过程  $\{X(t), t \geq 0\}$ 。

定义 A.3 随机过程  $\{X(t), t \geq 0\}$  称为连续时间的马尔科夫链, 如果:



对一切  $s, t \geq 0$ , 及非负整数  $i, j, x(u), 0 \leq u < s$ , 有

$$\begin{aligned} P\{X(t+s) = j \mid X(s) = i, X(u) = x(u), 0 \leq u < s\} \\ = P\{X(t+s) = j \mid X(s) = i\} \end{aligned} \quad (\text{A-4})$$

记  $P\{X(t+s) = j \mid X(s) = i\} = P_{ij}(s, t)$ , 它表示系统在  $s$  时刻处于状态  $i$ , 经过时间  $t$  后转移到状态  $j$  的转移概率。

**定理 A.1** (科尔莫哥洛夫向后方程) 假设  $\sum_{k \neq i} q_{ik} = q_{ii}$ , 则对一切  $i, j$  及  $t \geq 0$ , 有

$$P'_{ij}(t) = \sum_{k \neq i} q_{ik}(t) p_{kj}(t) - q_{ii}(t) p_{ij}(t) \quad (\text{A-5})$$

**定理 A.2** (科尔莫哥洛夫向前方程) 在适当的正则条件下

$$P'_{ij}(t) = \sum_{k \neq j} q_{jk}(t) p_{ik}(t) - p_{ij}(t) q_{jj}(t) \quad (\text{A-6})$$

## 二 马尔科夫 (Markov) 模型与疾病

### 1 简单马尔科夫模型

在不可逆的进行性三状态模型中,  $P(t)$  表示在时刻  $t$  时处在状态  $S_p$  的概率, 即为临床前的患病率,  $\omega(t)\Delta t$  是  $S_0 \rightarrow S_p$  在  $(t, t + \Delta t)$  转移概率,  $I(t)\Delta t$  是  $S_p \rightarrow S_c$  在  $(t, t + \Delta t)$  转移概率,  $I(t)$  为疾病的点发病率,  $q(t)$  是在状态  $S_p$  的驻留时间的概率密度函数, 则  $Q(t) = \int_0^t q(x) dx$ , 令  $[0, T]$  表示随访区间,  $t_0 < t_1 < t_2 \cdots t_n$ , 第  $i$  个时间间隔为  $(t_{i-1}, t_i)$ ,  $\Delta i = t_i - t_{i-1}, i = 1, 2, \cdots, n, t_0 = 0, t_n = T$ 。假定  $t_0$  时进入  $S_p$  状态, 记为 0 代, 个体在第  $i$  个区间进入称为第  $i$  代, 则:  $p_i(t)$  表示个体在第  $i$  代在时间  $t_i$  进入  $S_p$  状态的概率,  $q_i(t)$  表示第  $i$  代个体向前发展的概率密度函数。

$$p_i(t) = \int_{t_{i-1}}^{t_i} \omega(x) Q(t_i - x) dx \quad (\text{A-7})$$

$$q_i(t) = \frac{\int_{t_{i-1}}^{t_i} \omega(x)q(t_i - x)dx}{\int_{t_{i-1}}^{t_i} \omega(x)Q(t_i - x)dx} \quad (\text{A-8})$$

当  $\omega(t)$  不依赖于时间  $t$  时, 即  $\omega(t) = \omega$  常数, 这个模型为静态模型,  $P = mI, I = \omega$ , 即为 Zelen & Feinleib 提出的模型。Day & walter 在不可逆转的进行性疾病模型下, 用  $I(t)$  表示年龄时临床疾病的发病率,  $f(y)$  是临床前状态驻留时间的密度函数,  $J(t)$  表示临床前状态的发病率。假定  $f(y)$  与年龄  $t$  独立, 则

$$I(t) = \int_0^t J(s)f(t-s)ds \quad (\text{A-9})$$

如果  $J(s)$  为常数  $J$ , 第  $n$  次随访时,

$$P_n = J \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \min(y - (t_n - t_i), t_i - t_{i-1}) f(y) dy \quad (\text{A-10})$$

Brookmeyer & Day 在可逆性多状态模型中, 临床前状态分为两阶段, 设  $X$  表示在临床前的第一阶段所花的时间,  $Y$  表示临床前的第二阶段所花的时间,  $X$ 、 $Y$  均为随即变量,  $X$  与  $Y$  的联合密度函数  $f(x,y)$ , 累积分布函数为  $F(x,y)$ 。令  $T = X + Y$ , 表示临床前的驻留时间, 则其分布函数为

$$F_T(t) = \int_0^t \int_0^{t-s} f(s-y,y) dy ds \quad (\text{A-11})$$

简单随机模型从概率论的角度出发, 假定随机变量的密度函数, 一般假定它服从指数分布, 也可假定其它分布。这方法只能估计临床前期持续时间的分布规律, 并没有讨论各状态之间的转归。

## 2 时间齐次的 MARKOV 模型 (Time-homogeneous Markov Models)

在生物统计、人口学和保险业中, MARKOV 模型是非常有用的模型。因为它不仅可以估计临床前期的持续时间的分布规律, 而且可以分别计算各阶段的持续时间分布, 也可以根据在给定时刻处的状态预测其后的各种可能转归的统计规律。

给定 Markov 过程  $\{X(t), t \geq 0\}$ , 如果对一切  $t > 0$ , 一切  $i, j \in E, E = \{0, 1, 2, \dots\}$ , 转

移概率  $P(X(s+t) = j | X(s) = i)$  与  $s$  无关, 那么称  $\{X(t), t \geq 0\}$  为时间齐次的马尔科夫过程。

假定多状态模型  $(X(t); t > 0)$  是马尔可夫过程,  $X(t)$  取值于有限状态空间  $(1, 2, \dots, k)$ ,  $k$  为吸收态, 则状态概率矩阵

$$\begin{aligned}
 P(t) &= (p_{ij}(t))_{k \times k}, Q(t) = (q_{ij}(t))_{k \times k} \\
 p_{ij}(s, t) &= P(X(t) = j | X(s) = i) \\
 i &= 1, 2, \dots, k-1, j = 1, 2, \dots, k, s < t \\
 q_{ij}(t) &= \lim_{\Delta t \rightarrow 0} \frac{P(X(t+\Delta t) = j | X(t) = i) - p_{ij}(t)}{\Delta t}, i \neq j
 \end{aligned} \tag{A-12}$$

当  $q_{ij}(t) = q_{ij}$  与当时间独立,  $(X(t); t > 0)$  是时间齐次的马尔可夫过程, 则转移强度矩阵

$$Q(t) = Q \tag{A-13}$$

$$\frac{dP(t)}{dt} = P(t)Q, P(0) = I \tag{A-14}$$

$I$  是单位矩阵, 则转移矩阵为

$$P(t) = \exp(Qt) = \sum_{s=0}^{\infty} \frac{Q^s t^s}{s!} \tag{A-15}$$

这个数列是收敛的。如果  $Q$  有明显的特征值:  $d_1, d_2, \dots, d_n$ , 我们有

$$\begin{aligned}
 Q &= A D A^{-1} \\
 D &= \text{diag}(d_1, d_2, \dots, d_n) \\
 P(t) &= A \text{diag}(e^{d_1 t}, e^{d_2 t}, \dots, e^{d_n t}) A^{-1} \\
 P_{hi}(t) &= \sum_{l=1}^n a_{hl} c_{li} e^{d_l t}
 \end{aligned} \tag{A-16}$$

其中  $D = \text{diag}(d_1, d_2, \dots, d_n)$ ,  $A$  的第  $i$  列是伴随  $d_i$  的右特征向量,

其中  $a_{hl}$  是  $A$  的元素,  $c_{li}$  是  $C = A^{-1}$  的元素。

### 3 非齐次马尔可夫过程

由于疾病在发展过程中转移强度不可能是一个恒定的常数, 可能是随时间变化的同时也受协变量的影响, Aalen, Mau & Steinke, Andersen, Keiding 等许多研

究者集中于转移强度,提出了非齐次马尔可夫过程和非齐次马尔可夫链拟合了宫颈癌隐期的二阶段模型。其中转移强度有二种形式:

COX 乘法强度模型为:

$$q_{ij}(t) = a_{ij0} e^{\beta X}, i, j = 1, 2, \dots, k \quad (\text{A—18})$$

加法强度模型为:

$$q_{ij}(t) = a_{ij0} \beta X, i, j = 1, 2, \dots, k \quad (\text{A—19})$$

如果不考虑协变量,  $q_{ij}(t)$  是时间  $t$  的函数:

$$\begin{aligned} q_{ij}(t) &= a_{ij0} e^{\beta t} \\ q_{ij}(t) &= a_{ij0} t^{\beta} \\ q_{ij}(t) &= a_{ij} + \beta t, i, j = 1, 2, \dots, k \end{aligned} \quad (\text{A—20})$$

在这种情况下, 往往用来对时间齐次的马尔可夫模型进行诊断, 是否是时间齐次。则转移概率矩阵  $P(t)$ , 转移强度矩阵  $Q(t)$  有关系为

$$\frac{dP(t)}{dt} = P(t)Q(t) \quad (\text{A—21})$$

则 
$$P(t) = e^{\int Q(t) dt} \quad (\text{A—22})$$

#### 4 半马尔科夫过程模型 (Semi-Markov Process Model)

在疾病进展过程中, 从某一状态向下一状态转移时, 在这一状态所停留的时间是非常重要的, 一般马尔可夫模型只与当前状态有关, 而半马尔可夫模型, 转移概率不仅考虑当前状态而且考虑在这个状态停留的时间。

考虑一具有状态  $0, 1, 2, \dots$  的随机过程, 满足以下条件: 每当它进入状态  $i, i \geq 0$  时,

- 下一个进入的状态是  $j$  的概率为  $p_{ij}, i, j \geq 0$ ;
- 在下一个进入的状态是  $j$  的条件下, 直到发生从  $i$  到  $j$  转移为止的时间有分布  $F_{ij}$ 。

若以  $Z(t)$  记时刻  $t$  的状态, 则  $\{Z(t), t \geq 0\}$  为半马尔科夫过程。

Lagkos, somner & Zelen 提出非参数来分析多状态半马尔可夫模型。Joseph,

voelkel & crouley 对马尔可夫过程进行非参数统计推断。转移概率矩阵

$$\begin{aligned}
 P &= (p_{ij})_{k \times k} \\
 P_{ij} &\geq 0, i, j = 1, 2, \dots, k \\
 \sum_{j=1}^k p_{ij} &= 1, i = 1, 2, \dots, k
 \end{aligned}
 \tag{A-23}$$

设在状态  $i$  到状态  $j$  所等待的时间为  $\tau_{ij}$ ,  $\tau_{ij}$  的密度函数为  $f_{ij}(\tau)$ , 则区间转移概率矩阵  $\Phi(t) = (\varphi_{ij}(t))$ 。Howard 给出  $\varphi_{ij}(t)$  的积分方程

$$\begin{aligned}
 \varphi_{ij}(t) &= \delta_{ij} \sum_{i=1}^k p_{ij} (1 - \int_0^t f_{ij}(\tau) d\tau) + \sum_{i=1}^k p_{ij} \int_0^t f_{ij}(\tau) \varphi_{ij}(t-\tau) d\tau \tag{A-24} \\
 i, j &= 1, 2, \dots, k, t > 0, \delta_{ij} = \begin{cases} 1, & i=j \\ 0, & i \neq j \end{cases}
 \end{aligned}$$

$\tau_{ij}$  密度函数可以根据实际资料假定它服从指数分布或伽玛分布等。

## 5 拟合优度检验

在针对疾病的多状态模型的研究中, 对其进行拟合优度检验是必要的, 否则就可能出现错误的结论, 对马尔科夫模型拟合问题, 在某种程度上是通过比较观察的转移频数  $n_{ijk}$  和预期转移频数  $e_{ijk} = n_{i \cdot k} \hat{p}_{ij}(\omega_k)$ , 其中  $k$  是观察次数,  $\omega_k$  是两次观察时间间隔即  $\omega_k = t_k - t_{k-1}$ ,  $n_{i \cdot k} = \sum_j n_{ijk}$ , 这样就可通过似然统计量或渐近 Pearson  $\chi^2$  统计量来检验马尔科夫模型的拟合优度, 如果  $p_{ij}(\omega_k) \neq 0$ , 则似然统计量

$$\Lambda = 2 \sum_{k=1}^m \sum_j^k n_{ijk} \log \left( \frac{n_{ijk}}{e_{ijk}} \right) \tag{A-25}$$

当  $m$  固定时,  $n \rightarrow \infty$ , 近似  $\chi^2$  分布, 其自由度为  $mp(p-1)-b$ ,  $b$  为参数个数, 渐近 Pearson 卡方统计量为

$$\chi^2 = \sum_{k=1}^m \sum_j^k \frac{(n_{ijk} - e_{ijk})^2}{e_{ijk}} \tag{A-26}$$

后者是一种常用且容易理解的拟合优度检验方法。

把马尔科夫模型引入疾病的研究, 可以更好的认识疾病发展的过程, 转归的

规律，认识影响疾病进展的因素，从而对疾病更好进行预防和控制，并对医疗保险有一定的发展作用。我们对疾病进行建模，可以使我们更清楚在各个疾病状态的风险所在，使我们建立更细致的保费模型，更好地控制风险因素。

## 附录 B: 多状态模型简介

### 一 基本概念

多状态模型是描述纵向生存资料的最常见模型,多状态模型定义为一个随机过程模型,随机过程取几个离散状态之一。医学研究中,状态有健康、疾病、疾病综合症及死亡(dead)。状态的改变称为转移或事件,如疾病的发生、疾病综合症的发生和死亡(death)。状态分暂态(transient state)和(absorbing state),暂态指从此状态可发生状态的改变,转移到其他状态,而吸收态指从他状态转移到此状态后,不能再发生状态的转移。可直观地图示状态结构(state structure),明确状态及哪些状态间的转移是可能的。一个完整的统计模型应明确状态结构及每一个可能的转移所对应的危险函数形式。

多状态模型可由其转移强度或转移概率描述。转移强度描述状态间的瞬时转移危险,即在 $t$ 时刻前处于状态 $h$ 的个体,将在很短的时间区间内转移到状态 $j$ 的概率:

$$\alpha_{hj}(t) = \lim_{\Delta t \rightarrow 0} \frac{p_{hj}(t, t + \Delta t)}{\Delta t} \quad (\text{B—1})$$

$$\alpha_{hh} = -\sum_{h \neq j} \alpha_{hj} \quad (\text{B—2})$$

转移概率给出在时刻 $s$ 处在状态 $h$ 的个体,将在后来的时刻 $t$ 处在状态 $j$ 的可能性

$$p_{hj}(s, t) = \text{prob}(X(t) = j | X(s) = h) \quad (\text{B—3})$$

转移概率是进行长期预测的关键。 $k \times k$  Markov链转移概率矩阵用 $P(s, t)$ 表示,其中 $h$ 行 $j$ 列元素对应于 $p_{hj}(s, t)$ 。重要的Chapman—Kolmogorov等式为

$$p(s, t) = p(s, u)p(u, t), s < u < t \quad (\text{B—4})$$

据此可根据适当的时间和状态进行分割。

## 二 标准状态结构

多状态模型有以下五种标准状态结构：

### 1 两状态模型



图 B—1 两状态模型

最简单的多状态模型是生存资料的两状态模型见图B—1，其中有一个暂态“0:生存”和一个吸收态“1:死亡”，即对所有时间 $t$ ， $\alpha_{10}(t)=0$ 。只有一个可能的转移：从状态 0 到状态 1，相应的转移强度由危险率函数给出，即每个时间单位从状态 0 到状态 1 的瞬时概率。

$$\alpha_{10}(t) = \lim_{\Delta t \rightarrow 0} \frac{P(T \leq t + \Delta t | T \geq t)}{\Delta t} \quad (\text{B—5})$$

$T$ 的分布由概率分布函数  $F(t) = P(T \leq t)$  或  $S(t) = 1 - F(t) = P(T > t)$  确定。

$$S(t) = \exp\left(-\int_0^t \alpha(\mu) d\mu\right) \quad (\text{B—6})$$

$S(t)$ 和  $F(t)$  对应于时刻  $t$  处于状态 0 或 1 的概率。 $P_{00}(0, t)$  即传统的生存函数  $S(t)$  的 Kaplan-Meier 估计。若假定每个个体时刻 0 时处于状态 0，则  $F(t)$  是时间区间 0 到  $t$  从状态 0 到状态 1 的转移概率。

### 2、竞争风险模型



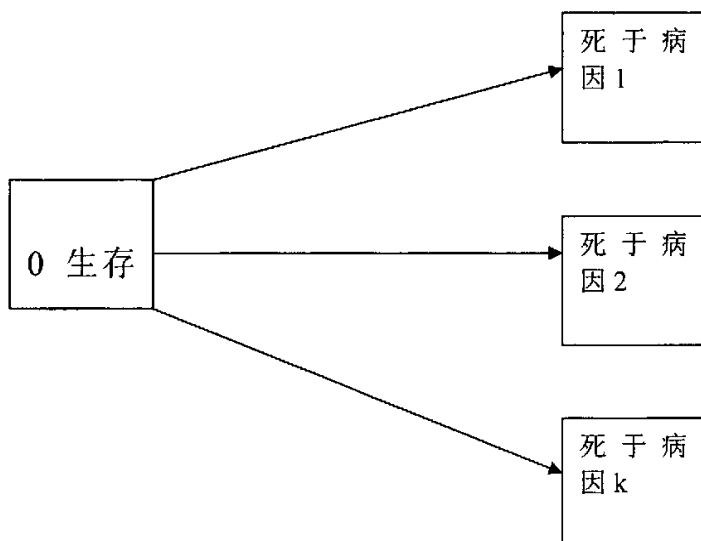


图 B—2 竞争风险模型

该模型从状态0 到状态 $h$  的转移强度表示为 $\alpha_{0h}(t)$ , 描述了死于病因 $h$  的瞬间危险, 即 $\alpha_{0h}(t)$  是个体在 $t$  时刻前尚存活, 将在很小的时间区间 $[t, t + \Delta t]$  死于病因 $h$  的可能性, 称为死因别危险函数。其中

$$\alpha_{0h}(t) = \lim_{\Delta t \rightarrow 0} \frac{p(\text{dead from cause } h \text{ by } t + \Delta t)}{\Delta t} \quad (\text{B—7})$$

对 $h = 1, 2, \dots, k$ , 我们记 $P_{0h}(s, t)$  为个体在时刻 $s$  处在状态0 (即存活), 将在后来时间 $t$  处于状态 $h$  (即死于病因 $h$ ) 的概率, 这些转移概率常称为累积发生函数(cumulative incidence function, CIF) 或边际死亡概率、粗死亡概率(marginal or crude failure probability)。其中

$$P_{0h}(0, t) = p(x \leq t, D = h) = \int_0^t s(u-) \alpha_{0h}(u) du \quad h = 1, 2, \dots, k \quad (\text{B—8})$$

令 $P_{00}(s, t)$  表示个体在时间 $s$  存活 (即状态0) 在将来某时间 $t$  仍存活的概率, 则

$$P_{00}(0, t) = S(t) = p(x > t) = \exp\left(-\int_0^t \sum_{h=1}^k \alpha_{0h}(u) du\right) \quad h = 1, 2, \dots, k \quad (\text{B—9})$$

即为边际生存概率。

### 3 疾病—死亡模型或残疾模型

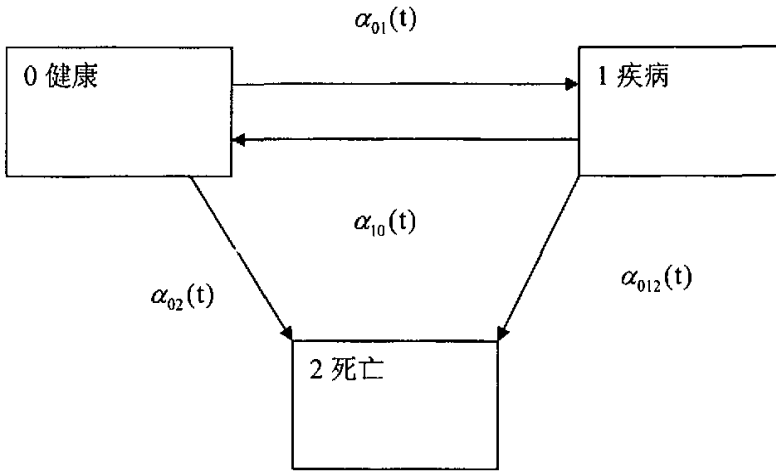


图 B—3 疾病—死亡模型或残疾模型

此模型适用于研究慢性病及死亡在同质人群中的发生规律,是非常重要的多状态模型之一。疾病—死亡模型有两种形式,一种为可逆模型(图B—3),即个体可在两个状态如健康和疾病间发生转移。另外一种为单方向模型,即个体只能从健康转向疾病,此模型统计分析较容易。 $\alpha_{01}(t)$ 是在时刻 $t$ 前尚健康个体,将在一较小时间区间 $[t, t + \Delta t]$ 患病的可能性大小,而 $\alpha_{02}(t)$ 和 $\alpha_{012}(t)$ 分别表示在时刻 $t$ 前健康或患病的个体将在一较小时间区间 $[t, t + \Delta t]$ 死亡的概率。对在时刻 $s$ 的健康个体(即状态0),我们用 $P_{01}(s, t)$ 表示其在后来时间 $t$ 患病(即状态1)的概率,而 $P_{00}(s, t)$ 是其仍在健康状态(状态0)的概率。类似地,对时刻 $s$ 患病(即状态1)的个体,令 $P_{11}(s, t)$ 表示其在时刻 $t$ 仍患病(即状态1)的概率,则

$$P_{00}(s, t) = \exp\left(-\int_s^t (\alpha_{02}(u) + \alpha_{01}(u)) du\right) \quad (\text{B—10})$$

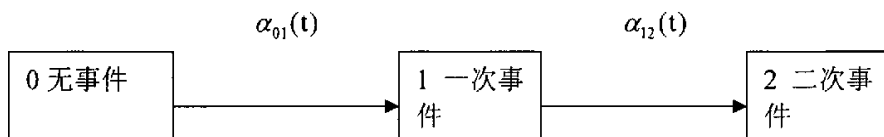
$$P_{01}(s, t) = \int_s^t P_{00}(s, u) \alpha_{01}(u) P_{11}(u, t) du \quad (\text{B—11})$$

$$P_{11}(s, t) = \exp\left(-\int_s^t \alpha_{12}(u) du\right) \quad (\text{B—12})$$

$\alpha_{12}(t)$ 除依赖于时间 $t$ 外,有时还依赖于进入状态1的时间长短的过程为Markov过程(semi-Markov process);如 $\alpha_{12}(t)$ 不依赖于时间的过程为Markov过程。

#### 4 重复事件或复发事件模型

这一模型研究兴趣在于某一给定事件的重复发生,如住院、生育、感染等(图B—4)。生育模型是一个标准的重复事件模型。应用模型时,感兴趣的通常是时间区间 $[t, t + \Delta t]$ 上事件的期望发生数。



图B—4 重复事件或复发事件模型

#### 5 双变量模型

该模型是一种针对双变量失效时间资料的多状态模型,例如描述孪生子的生存情况(图B—5)。状态结构的图示很重要,状态结构不是唯一的,通过恰当地确定状态和转移,可简化概率计算,把一些非Markov模型转化为Markov模型。

### 三 多状态模型分析

#### 1, 常数和分段常数转移强度

最简单的模型是转移强度为常数的情形,记作 $\alpha'_{ij}(t) = \alpha'_{ij}$ ,称为同质Markov过程。该过程是一个传统的概率模型,已经得到广泛应用。模型假定所有的转移危险在时间上是常数,当然转移危险的真值可不同,依赖于转移发生在哪两个状

态之间。模型可引入固定协变量,但不能引入时协变量。分段常数强度假定

$$\alpha_{hj}^i(t) = \alpha_{hj}^{i(l)} \quad t_{l-1}^{hj} < t \leq t_l^{hj}, \quad t_0 = 0 \quad (\text{B—13})$$

此模型应用很广,特别在经济学、流行病学、社会学和人口学的大型研究中。常数和分段常数Markov 模型的转移概率是转移强度的显式函数,允许直接“代入”进行最大似然估计及利用delta 法估计标准误。

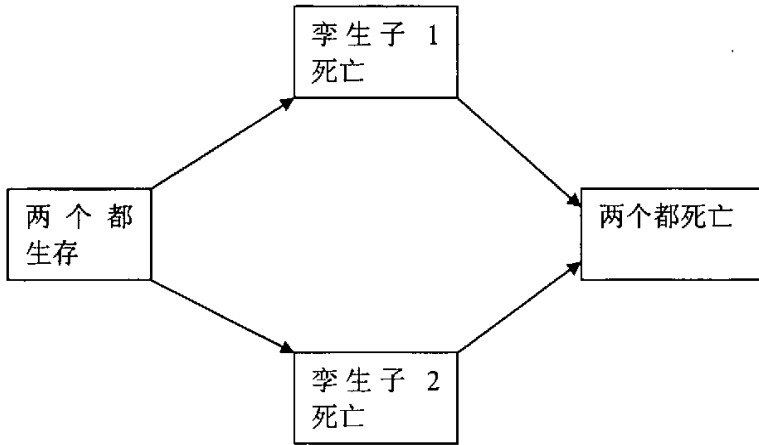


图 B—5 双变量模型

## 2. 自由变化(非参数)转移强度

假定所有个体的转移强度相等,但允许它们随时间自由变化。

$$\alpha_{hj}^i(t) = \alpha_{hj} \quad (\text{B—14})$$

## 3. Markov 回归模型

生存分析中使用最多的回归模型具有乘法结构,其中假定基准转移强度对所有个体均相同。

(1) Andersen-Gill Cox Markov 模型(AG - CoxMarkov 模型)

若假定每个可能转移各自的基准危险和回归系数,则通过拟合每个转移各自的Cox 模型可估计模型参数。

$$\alpha_{hj}^i(t) = \alpha_{hj_0}(t) \exp(\beta'_{hj} z_j) \quad (\text{B—15})$$

### (2) Klein-Keiding Cox Markov 模型( KK-Cox Markov 模型)

假定某些基准转移强度成比例或假定某些协变量对某些转移具有相同效应, 可得到较节省的模型。该模型是AG-Cox Markov 模型的特例, 假定所有转移到相同事件的危险成比例。

### (3) Lin-Ying 可加危险模型和Aalen 可加危险模型(additive risk model)

可加模型可估计依协变量变化的转移强度, Aalen等(2001) 阐明了基于非参数可加模型, 如何估计依协变量变化的转移概率。比例危险模型是非局部模型可能在一个较长时间段依时间变化的协变量效应平均化, 而可加模型是一个局部模型, 可迅速捕捉协变量的这种变化。

### (4) 半Markov 模型( semi-Markov model ) 及Markov 扩展模型(Markov extension model)

实际应用中会出现各种时间依赖, 如转移强度依赖于其他时间起点如进入当前状态的时间而非  $t = 0$ , 从而偏离Markov 模型, 只要各转移强度依赖于仅仅一个时间点, 可通过汇总各转移强度的独立子模型得到多状态过程模型, 如常数、分段常数、非参数或半参数模型。另外, 更一般的模型即Markov 扩展模型假定转移除依赖距起点的时间外, 还依赖于当前状态的停留时间。

## 四 多状态模型的应用

多状态模型是一种极为方便的方法, 可对几乎任何纵向资料建模, 特别是涉及到几种不同的事件而且概率与事件相联系。构造多状态模型的目的是对疾病过程有更深入细致的了解, 估计将来某一时间患者处在某状态的可能性大小。估计转移概率后可通过合成这些概率预测患者的最终预后。患者的预后是依赖给定时间病史的动态整体, 多状态模型可帮助我们基于观察到的病史估计一系列预测概率, 这些病史包括患者进入研究时的信息(固定协变量) 和中间事件发生的时间信息。利用这些信息我们可以很好的进行计算保费, 预测患者人数, 死亡率等等。

应用多状态模型时, 应选择解决当前问题最恰当的状态结构。从概率的角度, Markov 模型较简单, 特别是计算转移概率, 但某些情况下, Markov 模型并不适合, 可拟合非Markov 模型, 同时解决短期依赖和长期依赖问题。全模型估计完成后, 我们不仅可以针对实际资料进行预测, 还可以通过改变一些参数计算假想模型的转移概率。

## 附录 C: 相关肿瘤数据

表 C.1 1983-2001 年上海杨浦区居民恶性肿瘤年龄段发病率(1/10 万)

年龄段 (岁)	合计	男性	女性
0~	12.98	15.25	10.64
5~	9.02	9.35	8.67
10~	11.06	11.61	10.45
15~	11.3	12.54	9.89
20~	11.97	11.86	12.11
25~	25.34	22.19	28.91
30~	49.44	39.68	61.11
35~	90.61	76.25	106.95
40~	152.87	128.52	181.09
45~	196.37	173.57	222.05
50~	278.6	275.33	281.97
55~	428.77	510.56	351.12
60~	598.15	754.15	455.55
65~	861.43	1124.36	618.62
70~	1263.58	1669.53	895.42
75~	1424.75	1937.8	1018.44
80~	1470.75	2235.44	996.1
85~	1119.53	1884.39	792.03

表 C.2 分年龄段累计生存率（不分性别）

生存时间 (年)	<35 岁	35~45 岁	45~55 岁	55~65 岁	65~75 岁	75~85 岁	85 岁以 上
0.5 年	0.793	0.8053	0.7859	0.6982	0.6067	0.4887	0.3893
1.0 年	0.6907	0.7103	0.6830	0.5657	0.4650	0.3349	0.2327
1.5 年	0.6406	0.6459	0.6181	0.4845	0.3854	0.2526	0.1597
2.0 年	0.6129	0.6229	0.5815	0.4433	0.3408	0.2140	0.1303
2.5 年	0.5795	0.6003	0.5514	0.4039	0.3091	0.1850	0.1018
3.0 年	0.5604	0.5854	0.5330	0.3842	0.2880	0.1694	0.0856
3.5 年	0.5463	0.5692	0.5179	0.3665	0.2717	0.1516	0.0723
4.0 年	0.5396	0.5583	0.5024	0.3564	0.2579	0.1435	0.0661
4.5 年	0.5359	0.5450	0.4936	0.3417	0.2452	0.1327	0.0548
5 年	0.5300	0.5369	0.4860	0.3335	0.2389	0.1294	0.0548
5 年平均 余命	3.01 年	3.09 年	2.88 年	2.19 年	1.70 年	1.10 年	0.67 年

表 C.3 某医院男性年龄别住院情况统计（肿瘤）

年龄段	住院费用（一年期）	
	均值	方差
30~34	47271.79	0.00
35~39	58284.09	0.00
40~44	20662.02	18441.04
45~49	19037.70	20494.57
50~54	18212.35	18210.01
55~59	18859.74	16517.65
60~64	21859.19	13885.32
65~69	26359.10	20788.02
70~74	37445.58	43024.21
75~79	28414.42	26172.84
80~84	65687.79	15728.10

表 C.4 某医院女性年龄别住院情况统计（肿瘤）

年龄段	住院费用（一年期）	
	均值	方差
30~34	20631.63	4208.33
35~39	1202.78	0.00
40~44	25141.96	7312.87
45~49	23435.76	8141.42
50~54	34227.57	8692.78
55~59	23435.62	5931.08
60~64	29355.73	4971.78
65~69	22112.23	9121.13
70~74	25113.85	5082.58
75~79	54760.67	8403.08
80~84	340206.95	0.00



## 致 谢

本研究及学位论文是在我的导师方积乾教授的亲切关怀和悉心指导下完成的。他严肃的科学态度，严谨的治学精神，精益求精的工作作风，深深地感染和激励着我。从论文的选题到论文的最终完成，方积乾教授都始终给予我细心的指导和不懈的支持。两年来，方积乾教授不仅在学业上给我以精心指导，同时还在思想、生活上给我以无微不至的关怀，在此谨向方老师致以诚挚的谢意和崇高的敬意。方教授拥有高尚的人格魅力，高超的学术水平，是我们青年学子学习的楷模。

同时感谢中山大学数学与计算科学学院的各位老师，特别是统计科学系的任佳刚教授，区景祺教授，余锦华教授等老师，几位老师所授予的知识和指导是我在中山大学求学阶段最大的收获之一。感谢中山大学公共卫生学院流行病与医学统计系的各位老师专家，各位老师敬业向上的精神给我留下了深刻的印象，并在这里让我认识到一个新的领域。中山大学数学与计算科学学院和公共卫生学院给我的熏陶和教导，将使我终生受益。

感谢暨南大学伍超标老师给予我的亲切指导。

同时感谢我的同学，他们在我学习和生活中都给予了我无私的帮助。

感谢我的父母家人还有我大学阶段的各位老师，特别是郑州轻工业学院的赵玲玲教授，是你们催我奋进，给我支持和力量。

## 论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

学位论文作者签名：李瑞林

日期：2006年5月27日