Application of Markov Model in Long-Term Care Insurance

Shihao Xuanyuan[1, a], Shiang Xuanyuan[2, b]

[1]School of Economics Beijing Technology and Business University Jinan, China

[2]Graduate School of Arts and Sciences Columbia University in the City of New York Jinan, China

[a]Xysh1104@163.com, [b]sx2292@columbia.edu

†These authors contributed equally

Abstract. The Markov process is not only the actuarial basis of pricing of long term care (LTC) insurance but also the fundamental for predicting the future elder population and disabled population. This article aims to summarize how the Markov processes or Semi-Markov processes are used in the Long-term care risk and Long-term care insurance. We also discuss the models based on the time-homogeneous and time-inhomogeneous. Moreover, under the GLM framework, some studies show Tweedie-GLM would give more accurate predictions compared with other GLM models and additive models. However, these models, whether based on the Markov process or Semi-Markov process or GLM, have theoretical advantages due to the natural features, the researchers would quickly build the multi-state models, there are still exits many challenges, and they provoke the researchers into some tries of how to deal with the limitations of data, the development of medical technology, and the longer expectancy of life.

Keywords: Markov process, Semi-Markov process, Long-term care, time-homogeneous, time-inhomogeneous.

1. Introduction

With the increasing size of the elderly population, the problem of aging has already become a significant challenge that attaches great attention around the entire world. According to the United Nations' definition of population aging, a country or region is considered to be in the midst of population aging if the population aged 60 and over accounts for more than 10% of the total population, or if the population aged 65 and over accounts for more than 7% of the total population. With the continuous improvement in the total number and ratio of the senior people, the population aging process is intensifying, resulting in a large-scale increase in the disabled elderly. Moreover, society needs to pay attention not only to the elderly disabled group, but also to the middle-aged group. The risk of disability also needs to be paid attention to and concerns. Due to factors such as life pressure, environment, diet, work, and rest, this population with severe diseases begins to show the characteristics of rejuvenation, and major diseases such as malignant tumors, acute myocardial infarction, and brain diseases are gradually becoming common among middle-aged people.

The concept of Long-term Care was first proposed in 1986 at the Washington Institute of Medicine, where the meaning of LTC is that the person in need is unable to care for themselves due to psychological or physiological diseases and needs others to provide nursing services in order to maintain a basic life. The definition of dependency is therefore derived: a permanent and consolidated state of inability to execute daily activities without the assistance of others. At present, the definition of LTC is not uniform around the world, and many scholars try to give relevant definitions.

To determine the demand of long-term care is complicated. Its study not only takes into consideration of the incidence in LTC but also the probability of death for both health (can perform activities of daily living without other's help) and disabled people. However, it is hard to give a general pattern of this study, since, 1. The incidence in LTC would relate to several aspects, such as the level of education, living area, wealth level, former occupation, etc., and 2. First, the probability of death is linked to aging, and gender would also be taken into account; second, multiple causes may lead to death, like living style(smoke/intemperance), cancer, dementia, chronic diseases, etc. Apart from the reasons described above, data aggregation challenges may make it difficult to acquire a better picture of the risk[1].

The essence of these concerns, whether there is a requirement for long-term care insurance for the ole generations or when is the correct time to have LTC, is related to the beginning state of health and the forecast of the elderly's health status transfer. The standards of how to identify the initial state of health are different, and the common definition is based on Activities of Daily Living (ADL), includes estimating moving, eating, bathing, and feeding. Many studies use ADL to do the classification based on their dataset. Scholars created the multi-state model to anticipate the transfer of the elderly's health condition based on this premise, where the common models are the three-states transition model and four-states transition model, like Fong et al.[2] used the three-state model for LTC insurance applications, Rickayzen[3] applied a multiple state model with discrete time. Meanwhile, according to the feature that the prediction of a future state only depends on the current state, Markov processes could naturally be used in this study. As in the field of epidemiology, there already have many studies based on the application of the combination of the multi-state model and Markov process Semi-Markov processes. Also, Guillaume[1] indicated that, due to pathologies as the main cause of the long-term care states, Markov and Semi-Markov processes became natural candidates for these problems.

The transition probability based on Markov processes only depends on the present states. Since this model is simple and easy to apply, many insurance companies and academic papers still use this assumption recently, like Pitacco[4] or Fong et al.[2]. Hanewald, Li, and Shao[5] used a multi-state Markov model including age, temporal trends, and connections between them to model health transition, which performed in the framework of generalized linear models. The main shortcoming of this model is that it does not match the actual process of LTC since the Markov process does not take into account the duration of states, such as a higher mortality rate in the first year of LTC than in following years[1].

Semi-Markov processes introduce this additional variable, the length of time between states (time spent in the current state[1]). In addition, Pitacco[3] developed an illness-death model, and semi-Markov processes have been applied to disability insurance. Nevertheless, the duration of disability insurance is terminated at the age of retirement. On the other hand, there should be no expiration time of long-term care. Therefore, there are two issues came along with applying the illness-death model to the long-term care, the higher age and the longer time spent in the disabled state. Using data from the French long-term care system, Guillaume[6] established a construction technique for a Semi-Markov model with four degrees of dependency, and then applied it to the pricing of LTC insurance using the Monte Carlo method, and then in 2016, Guillaume described a continuous-time Semi-Markov framework with pathologies. Fuino and Wagner[7] used a Semi-Markov framework with two dependent transition probabilities, state and time. Hence, non-parametric methods based on the Nelson-Aslen estimator[8] could be applied for long-term care, and parametric methods would also be associated, for instance, the example in Guibert and Planchet[9].

There still have other methods to construct models for prediction in Long-term care insurance, such as Lally and Hartman[10] used data from a large LTCI provider, they constructed random forest models, GLM models, an additive model with zero-inflated Poisson, negative binomial, and Tweedie errors, and compared the predictive capacity.

2. Theoretical Model

2.1. Multi-state model

According to the definition given by Hougaard[11], the multi-state model was used to describe the process. This model can describe the possible events that occur on a single person and could show the dependency between several people. The common models used in long-term care insurance are the three-state and four-state models, which regard 'death' as an absorbing state. Typically, the models can be built as follows:

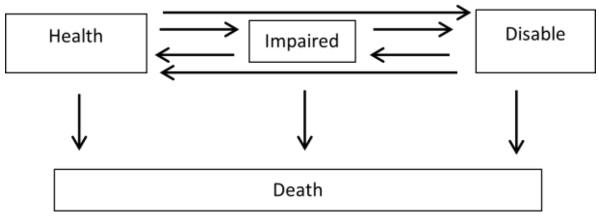

Fig. 1 4-state model: with three transitory states and one absorbing state (death)

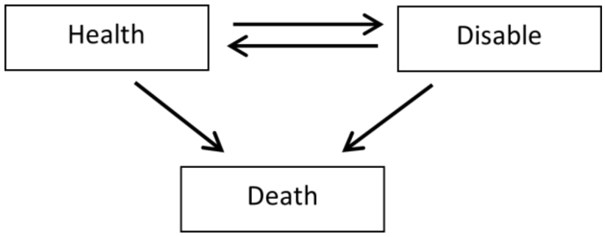

Fig. 2 3-state model: with two transitory states and one absorbing state (death)

Generally, the transition between two states, excluding the death state, is bidirectional. In these situations, the recovery rate is brought into consideration. Fuino and Wagner [7]demonstrated a model as similar as figure 1 but with more specific states (mild, moderate, severe). However, some scholars just omitted the effect of recovery rate and took the unidirectional model to make the model simpler, like Guillaume[6] constructed a unidirectional model with four different states of dependency, and the three states Markov model built by Hanewald et al.[5].

2.2. Markov Process / Semi-Markov Process

The main differentiation between the Markov and semi-Markov processes is that the transition probabilities are determined by the current state and its duration. To make the computation easily, if assuming that the transition probabilities are the same at each step, or independent of time n, this model is said to have a time-homogeneous Markov chain, on the other words, the assumption of time-homogeneous regards the transition intensity as a constant, the transition probability only relates to the time interval, but not to the current time point. The transition matrix is denoted as follows,

$$P(X_{n+1} = j|X_n = i) = P(X_1 = j|X_0 = i) \tag{1}$$

$$p_{ij} = P(X_1 = j|X_0 = i) \tag{2}$$

where i, j are the states, p_{ij} is the probability of one transition from state i to state j. With these variables, construct a transition matrix $P = (p_{ij})$. Thus, a time-homogeneous Markov process model or semi-Markov process model could be constructed during the research.

Conversely, if the transition probabilities are different at each transition, i.e., the transition probability relates to the current time point, it is regarded as time-inhomogeneous.

$$P_t(x, y) = P(X_t = j|X_{t-1} = I) \tag{3}$$

This is one step transition matrix. To compute the probability that takes several steps transition, we could use:

$$P_{t,t+r}(x, y) = P_{t+1}P_{t+2}P_{t+3} \dots P_{t+r}(x, y) \tag{4}$$

$$\forall\, t \geq 0, r \geq 1$$

since the future state only depends on the current point independent of the previous status, the above formula, the production rule of independent events, is derived. If taking the time spent on the current states into consideration, then, it would be a time-inhomogeneous semi-Markov model, otherwise, the Markov process with time inhomogeneity should come into mind. Assume that the transitions between health statues are abode by the time-inhomogeneous Markov process, the transition probabilities depend on the moment when the transition occurs,

$$P_{ij}(x,t,h) = P(S(x+h,t+h) = j|S(x,t) = i),$$
$$i,j \in \{\text{health status exclude death}\} \quad (5)$$
$$q_{ij}(x,t) = \lim_{h \to 0^+} \frac{P_{ij}(x,t,h)}{h} \quad (6)$$

where x is the age, t represents the time, h is the time interval, $h \geq 0$, S(x,t) reflects an individual's health at age x and time t. $P_{ij}(x,t,h)$ is the transition probability which the health state i at age x and time t moves to the health state j at age (x+h) and time (t+h), and $q_{ij}(x,t)$ is the intensity of the transition.

3. Recent Research

Guillaume Biessy (2015) used data from the "Allocation Personnalisée d'Autonomie" (APA) for groups over 60 years old, a French public aid program for long-term care. This study created a four-state semi-Markov model, utilized likelihood maximization to account for the parameter, and used a Monte Carlo approach to price a hypothetical LTC insurance product. Scholars used the "Autonomie Gérontologue Groupes Iso-Resources" (AGGIR) grid to identify four states. The AGGIR grid is used to divide persons with comparable requirements into groups depending on their level of reliance.Only people in Gir 1 to Gir 4 states have the right to receive the benefit from public assistance. As a result, only these 4 dependency levels have been considered in this model, and this article considered a model with these 4 different dependency states Gender has a considerable impact on the probability of duration laws, according to the findings. From the risk of the transition from Gir 4 to the dead state, males were 2.5 times higher than females. Therefore, females have a certain survival advantage over males and survive longer. In addition, age also has a certain effect on the transition probability. The probability of disability increases with age. People who are over 95 were four times more likely to transfer from Gir 4 to Gir 2 than those over 65. From the life expectancy, for the 60 years old group, the average is 84 for men and 87 for women. In addition, for the disabled group, women survived an average of four years and men less than three years. Life expectancy for both men and women declines with age.

Joelle H. Fong et al. (2015) used data from the Health and Retirement Study (HRS) from 1998 to 2010, a representative survey of the United States over the age of 50 and their spouses. Firstly, this research used a GLM in order to evaluate the disability transition rate. Secondly, a three-state model was used for long-term care insurance applications. The total number of valid samples is 19,547, and the sample individuals lived in 1998. Moreover, each sampled individual gave valid answers to all ADL questions, and the questioning methods for ADL questions were consistent. Scholars derived four groups of sex- and ages- of transition probabilities: onset of disability, recovery, mortality (non-disabled state), and mortality (disabled state) for older people from 50 to 100. The results of this study show that, first of all, from the perspective of the probability of disability transfer that requires LTC, the elderly only have a 10% possibility of reaching the disability state of LTC when they are over 90 years old, rather than at their 80s. Furthermore, women have a higher probability than males. Second, the study found that there are significant differences in the probability of disability and recovery between the sexes and age. Women, in particular, are substantially more likely than males to be handicapped. The rate of disability outpaced the risk of mortality for females of all ages. However, this conclusion has the opposite situation for men over 60. Males' recovery rate decreased with age. On the contrary, for females, who are under 65, there was a greater chance of recovery. For the rest of the females, who are over 65, this rate would be declined with age. Finally, from the perspective of mortality, the mortality rate of the disabled group was higher than that of the non-disabled group at all ages. In terms of gender, male groups have a higher risk of death than female groups.

Guillaume Biessy (2016) studied the mortality in long-term care related with four main pathology groups: cancer, dementia, neurological illnesses, and others, with data from a French LTC portfolio.

Two factors affect mortality in a continuous-time semi-Markov framework developed by researchers, which are inception age and temporal length. Researchers considered two models. The first one is a three states model, including autonomy, disability, and death. Another is a multi-state model with autonomy, four levels of incapacity that correspond to the dataset's four classes of pathologies, and death. Transition intensities were used to define both models: non-disabled mortality, disability rate, and mortality (disabled). The results indicate that the cancer group is primarily responsible for the observed peak in mortality following entrance into LTC. The impact of pathology groups on mortality in long-term care was confirmed substantially. Due to a considerable variation in the basal level of death for cancer compared to the general population, directly smoothing all-cause mortality in long-term care leads in over-smoothing the mortality surface in low-duration regions in long-term care. Furthermore, the cancer group is primarily responsible for the unique shape of the disability mortality curve, with a significant death rate shortly after commencement, according to this paper. This conclusion strongly suggests that the cancer group's mortality should be estimated independently from the other groups' mortality.

Data from the Chinese Longitudinal Healthy Longevity Survey from 1998 to 2012 was used by Katja Hanewald et al. (2017). To depict the health transition state, a multi-state Markov model including age effects, time effects, and age-time interactions was used. Seniors aged 65 and over made up the survey's sample. The model is implemented in a generalized linear model. The evaluation criteria for the disability state in this article were based on the ADL regulations. The analysis indicated that, in addition to age effects, temporal trends and age-time interactions played a role in explaining the changes in the health status. Second, in China, there are large variations in disability and mortality rates among the elderly in urban and rural areas. Additionally, the existing urban-rural differences in life expectancy for healthy males and healthy and impaired females grew over time. However, for disabled males, this gap was slowly closing. For the life expectancy in health groups, the results showed that trends were broadly similar for male groups between urban and rural areas, while there were urban-rural differences between female groups. Third, the decrease in the mortality rate of the elderly over 65 years old in China is mainly due to the decline in the mortality rate of the disabled elderly group and has nothing to do with the mortality rate of healthy individuals. Finally, in terms of life expectancy, urban inhabitants have a longer life expectancy than rural ones, while female groups have a longer life expectancy than male groups, and the life expectancy of healthy groups is higher than that of disabled groups.

M Fuino, Jol Wagner (2018) used a longitudinal dataset of all long-term care needs in Switzerland from 1995 to 2015 and utilized healthy state and time-dependent transition probabilities in a semi-Markov framework. This model takes into account the type of care in two forms of care, home care and institutional care. Meanwhile, this model also divides health status into four states: mild disability, moderate disability, severe disability, and death. The findings of this study revealed that the probability of transition was dependent on sex, age, and the duration of the disabled state. Second, in models of home care and institutional care, women persisted longer than men in the incapacitated state. In addition, from the analysis of the types of care patterns, it could be found that the disabled group, which received long-term care from their kinship, was more likely to transfer to professional institutions for the specialized care after one year. Therefore, the authors concluded that the propensity to receive professional care compared with home care was related with a lower death rate due to the availability of competent nursing services. This conclusion, combined with the underlying conditions of the disabled population, may explain why, at specific times, individuals who are somewhat impaired have a greater death rate than those who are moderately or severely disabled.

4. Challenges

According to all articles, it is not difficult to find the following problems that scholars use Markov chains and transition matrices to study long-term care needs or the prediction of disabled populations.

First of all, a lot of the databases used by scholars are non-professional disability data. Due to the lack of reliable disability sample data in countries, the results obtained by scholars may vary to some extent. Although some data were provided by long-term care insurance companies, they lack a certain degree of representativeness and comprehensiveness. In addition, owing to the small sample group of elderly, especially those over 85 years old, there are some differences in the disability rate and duration of disability in this part of the population. If more professional and reliable disability sample data can be formed, more accurate empirical analysis results may be obtained.

Secondly, generally speaking, in the Markov model, in addition to the death state, there are mutual transitions between other health states. However, when some researchers use the Markov model to establish a healthy state transition model, they only consider the unidirectional model, which means that they did not consider the transition from the disabled state to the healthy state. This is inconsistent with the actual situation, therefore, the obtained results may also be controversial.

Thirdly, some studies utilized a time-homogeneous Markov process. The transition intensity is a scalar quantity due to the assumption of temporal homogeneity, and the transition probability is simply related to the time interval rather than the time point when it happened. That is, the health state changes at the same pace regardless of how old the inception is. Clearly, this hypothesis is better suited to serious conditions in which the likelihood of recovery is unrelated to age or for short-term imprecise estimations. Furthermore, some researchers also consider the time-inhomogeneous Markov process. Under the time-inhomogeneous Markov framework, the transition likelihood is connected to the point in time, but not to the period of the point in time. Because population changes occur rapidly, it is vital to account for time effects in health state transitions. Besides, articles on the Semi-Markov process have begun to emerge in recent years. The Semi-Markov process implies that the transition probability is associated not only with the points in time, but also with the duration of the points in time.

Finally, most literature definitions of disability are based on the ADL criteria, which only assess activities of daily living. However, it is well known that people with dementia, Alzheimer's disease, and psychiatric disorders also need LTC, and even belong to the disabled group. Therefore, the combination of ADL, instrumental activities of daily living (IADL), cognitive function, and other indicators, introduced multi-dimensional issues such as social participation and activity limitation, to make up for defects such as single-dimensional evaluation bias.

5. Conclusion

Markov process is a random process composed of multiple states and can describe the transition probability between states. The realization of the Markov process and the construction of the transition matrix between health states are the basis of LTC insurance pricing. Moreover, the calculation of the transition probability matrix is not only the actuarial basis for LTC insurance pricing but also the basis for predicting the future nursing state population size and healthy survival time. Besides, the estimation of health life involves multiple states and the possibility of transition between states. Therefore, the construction of the Markov process-based transition probability matrix has theoretical advantages. In addition, the model also has certain advantages in other aspects. For example, it can be used to anticipate the future growth of the disabled population, the duration of disability and long-term care costs, or to examine the heterogeneity of elderly health with different education levels, income levels, social status, or elderly health, etc., especially for situations where individual information is not abundant and the time span of the longitudinal survey is not significant.

Of course, there are still some issues worth exploring with the model. First, the model is not appropriate for the influence of future medical technology advancements on morbidity and mortality, which might result in an underestimating. The second is sample selection bias. To be specific, individuals with poor health status are more likely to be untracked or challenging to be investigated, resulting in high estimates, especially for the remaining life expectancy of the elderly.

Reference

[1] Biessy Guillaume, “A Semi-Markov Model with Pathologies for Long-Term Care Insurance,” Working Paper, Universit'e d'Evry Val d’Essonne, hal-01356063, 2016.

[2] Fong, J. H., A. W. Shao, and M. Sherris, “Multi-state Actuarial Models of Functional Disability,” North American Actuarial Journal, vol. 19, issue 1, pp. 41-59, February 2015.

[3] Pitacco. E, “Health Insurance - Basic Actuarial Models,” Springer International Publishing, ISBN: 3319122347, 2014.

[4] Pitacco. E, “Actuarial values for long-term care insurance products. a sensitivity analysis,” Working Paper, ARC Centre of Excellence in Population Ageing Research, 2015.

[5] Hanewald Katja, Han Li, and Adam Wenqiang Shao, “Modeling Multi-State Health Transitions in China: A Generalized Linear Model with Time Trends,” Social Science Electronic Publishing, vol. 13, issue 1, pp. 145-165, July 2018.

[6] Biessy Guillaume, “Long-Term Care Insurance: a Multi-State Semi-Markov Model to Describe the Dependency Process for Elderly People,” Bulletin Francais d’Actuariat, vol.15, No.29, 2015.

[7] Fuino M, and Wagner J, “Long-Term Care Models and Dependence Probability Tables by Acuity Level: New Empirical Evidence from Switzerland,” Insurance: Mathematics and Economics, vol.81, pp.51-70, June 2018.

[8] Klein, J. P, “Small sample moments of some estimators of the variance of the kaplan-meier and nelson-aalen estimators,” Scandinavian Journal of Statistics, vol.13, pp.333–340, 1991.

[9] Guibert. Q. and F. Planchet, “Non-parametric inference of transition probabilities based on aalenjohansen integral estimators for semi-competing risks data: application to ltc insurance,” In Conference of the LIFE Section of the International Actuarial Association, **hal-01183542,** 2015.

[10] Lally Nathan R., and Brian M.Hartman, “Predictive modeling in long-term care insurance,” North American Actuarial Journal, vol. 20, No. 2, pp. 160-183, May 2016.

[11] Hougaard P, “Multi-state models: a review,” Lifetime data analysis, vol. 5(3), pp. 239–264, 1999.

[12] Eling M, and Ghavibazoo O, “Research on Long-term Care Insurance: Status Quo and Directions for Future Research,” The Geneva Papers on Risk and Insurance-Issues and Practice 44, pp. 303-356, December 2018.

[13] B. D. Rickayzen, “A Multi-state Model of Disability for the United Kingdom: Implications for Future Need for Long-term Care for the Elderly,” British Actuarial Journal, vol.8, No.2, pp. 341-393, 2002.