INVERSE PROBLEMS FOR MARKOV MODELS IN HEALTH INSURANCE

Semyen I. Spivak

The Bashkir State University, Ufa, Russia

The paper deals with inverse problems for Markov models in health insurance. Two opposite problems arise when using Markov processes in modelling. Direct problem is estimating states probabilities and other process characteristics. It assumes that the model parameters are given. Inverse problem is based on estimating parameters using the experimental data. The current paper is about stating and solving inverse problems in Markov models used in healthcare insurance.

A number of well known applications use the approach that allows applying Markov processes theory to the modelling of multi-state systems. Examples can be found in such areas as queuing systems, advertising applications, reliability theory and various other applications in Biology, Chemistry, Physics, etc. This approach is also widely used in actuarial practice, where multi-state model is used for modelling various states of insured life.

Two opposite problems arise when using Markov processes in modelling. Direct problem is estimating states probabilities and other process characteristics. It assumes that the model parameters are given. Inverse problem is based on estimating the model parameters using the resulting characteristics for the process obtained from experimental data.

Direct problems of Markov processes are studied quite well, including their applications in actuarial modelling. The current paper analyses solution of an inverse problem for Markov processes. The problems of this type are not currently widely studied.

Input parameters in Markov model are intensities or forces of transition between states. In queuing systems or actuarial problems they can be estimated using statistical information on some input data.

Insurance premiums and reserves (especially in long-term lines of business) are based on present values of policy cash flows. An insured event, time and value of each cash flow are usually unknown in advance. They depend on random events. Usually the most important task is to estimate the expected value of this random variable which depends on probability of event occurrence.

As we noted earlier, a lot of traditional actuarial problems can be described using multiple-state models. We then assume that at any point in time an individual can be in one of the described states. An individual's current state can be related to some cash flows (payments). The problem is to give a quantitative estimate of how the individual's state affects the cash flows, i.e. to estimate probabilities of him occupying a particular state.

The best model for calculating probabilities of some event is Markov process scheme.

The simplest actuarial model can include just two states: “alive” and “dead”. See Figure 1 for the corresponding scheme.

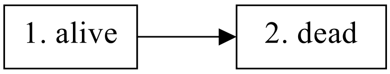

Figure 1 Two State Model

Transition is possible in only one direction. A simple annuity pays out while life remains in State 1 and ends upon transition into State 2. Whole life policy pays a premium while insured remains in State 1 and pays out a death claim upon transition into State 2. The methods of estimating actuarial probabilities in these cases are simple and well studied.

More complicated problem arises for models with of extra states. Figure 2 shows a scheme for three states model: “healthy”, “sick” and “dead”. It is usually used in disability or income protection insurance. In this case the premiums are paid while an individual is in State 1 and claims are paid while he is in State 2.

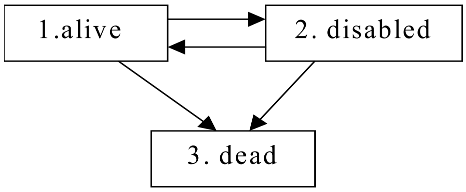

Figure 2 Three State Model.

Actuarial calculations here are more complicated as an individual can move between states 1 and 2 several times. For this reason, it is often assumed that transition into State 2 is impossible. If we assume that, the Markov problem is then easily solved in this case as well.

Let us introduce a general definition of Markov process and its properties.

Let us denote the state of individual aged t (t≥0) as X(t). Stochastic process is defined as {X(t), t≥0}. Let us assume that there is a finite number of states 1,2,...n, i.e. the process has a states universe {1,2,…,n}. Then {X(t), t≥0} is a Markov process if for every s, t ≥0 и i, j, x(u)∈{1,2,…n},

$$\Pr\{X(s+t)=j \mid X(s)=i, X(u)=x(u), 0\le u<s\} = \Pr\{X(s+t)=j \mid X(s)=i\}$$

This means that the future of the process (after time s) depends only on its state at time s and not on the history of the process prior to time s.

Let us define the transition probability function as

$$p_{ij}(s,s+t) \equiv \Pr\{X(s+t)=j \mid X(s)=i\}, i,j\in\{1,2,...n\},$$

and assume that

$$\sum_{j=1}^{n} p_{ij}(s,s+t)=1$$ for every t≥0.

Let us also assume the existence of the following limits:

$$\mu_{ij}(t)=\lim_{h\to 0}\frac{p_{ij}(t,t+h)}{h},$$

i,j∈{1,2…n}, i≠j.

When i≠j μij is transition intensity from state i into state j. It is easily seen that when s, t, u ≥0

$$p_{ij}(s,s+t+u)=\sum_{l=1}^{n}p_{il}(s,s+t)p_{lj}(s+t,s+t+u), \qquad (1)$$

i,j∈{1,2,…n}.

Equations (1) are known as Chapman-Kolmogorov equations.

We also need to derive transition probability functions to calculate the actuarial values. Transition intensities and transition probability functions are linked through direct and inverse Kolmogorov equations:

$$\frac{\partial}{\partial t}p_{ij}(s,s+t)=\sum_{l=1}^{n}p_{il}(s,s+t)\mu_{lj}(s+t) \qquad (2)$$

$$\frac{\partial}{\partial t}p_{ij}(s,s+t)=-\sum_{l=1}^{n}\mu_{li}(s)p_{lj}(s,s+t) \qquad (3)$$

In general case, these systems of differential equations should be sold numerically to obtain the transition probability functions.

The explicit expressions for transition probability functions can be obtained when $\mu_{ij}(t)=\mu_{ij}$ for all t. This kind of Markov process is known as homogeneous in time, or stationary. The assumption that transition intensities are constant, implies that time spent in each state is exponentially distributed. It also implies that all functions $p_{ij}(s,s+t)$ are the same for all values of s and can be denoted simply as p_{ij}.

When transition intensities are given, the problem becomes the direct case of solving Chapman-Kolmogorov equations. For models with constant intensities it is straightforward for every number of possible system states. Nevertheless, if the transition intensities are not given, the problem is the inverse one, i.e. the problem of estimating transition intensities using statistical data.

Solution of the inverse problem is linked to the analysis of the infomativeness of measurements.

When building models using Markov process schemes it is natural to use the concept of a quality. In a common (general) sense a quality is a system’s ability to possess a certain property. A quality can be seen as a numeric value of a certain variable falling inside a pre-determined interval.

For our problem it means that the transition intensity from one state into the next is in a certain interval, the borders of which are determined by the characteristics of the process.

Solving the inverse problem means to build a system of quality control. By a quality control we understand some criteria and its verification algorithm which gives an answer to a question whether a system satisfies a certain quality level or not. Using our understanding of quality, this means whether the intensity value falls within a given interval or not.

This immediately implies a very important problem of quality regulation: to build a procedure of organizing a process in such a way that quality falls within a certain interval.

The problem of transition intensities is quite complicated one, especially in the multivariate cases. That is, if there n states, we can form C_n^2 different pairs (i,j), i,j=1..n, i≠j. Each pair (i,j) has two corresponding transition intensities μij and μji. Therefore, in total we need to determine $2C_n^2 = n(n-1)$ statistical estimates of transition intensities that are not parameters, but functions of age *x*.

Therefore, this problem should be attempted using certain assumptions, namely:

1. We are going to approximate transition intensities μ_{ij} with partially-constant functions. In order to do this, we divide the observation interval on to time intervals such that within each of them we could consider transition intensities to be constant. Let <…> be one of these intervals and $\mu_{ij} \cong \mu_{ij}$ for all of them.
2. We observe individuals whose age x falls within the interval.
3. Every individual has a corresponding to random realization of the stochastic process.
4. Individual’s behavior does not depend on how his observation commenced and why it finished.

The question whether some characteristic falls within some interval that characterizes a given quality level is solved by analysis corresponding experimental data. This particular fact brings us to necessity of building corresponding mathematical models and their use as a way of processing large amounts of information.

The existence of such models allows us to use a possibility of describing measurements within a model as a quality control criterion for this model.

Regulating quality in this case is about planning and organising the corresponding experiment.

As an example let us use mortality estimates for smokers and non-smokers. Statistical data used comes from source [1].

See [1] for mortality tables for four groups of people: male smokers, male non-smokers, female smokers and female non-smokers. A part of one of these tables is shown below.

Mortality Table for Male Non-Smokers

Age x	l_x	d_x	q_x
25	96570	65	0.00067
26	96505	63	0.00066
27	96442	62	0.00065
28	96380	62	0.00065
29	96318	64	0.00066
30	96254	65	0.00067
31	96189	67	0.00070
32	96122	70	0.00072
33	96052	73	0.00076
34	95979	77	0.00080
35	95902	82	0.00086

We are looking at smoking and non-smoking males aged 30-45 for the purpose of this paper.

The tables show mortality probabilities for a given age using the formulas described above.

An alternative method would be described by a three-state process with transition probabilities shown schematically in Figure 3.

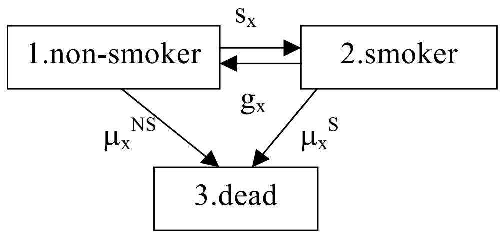

Figure 3 Model for Smokers/Non-Smokers

The following assumptions are made for this scheme:

1) There are three possible states. Each individual can occupy any one of these states. There are two ‘live’ states: smoker (State 1) and non-smoker (State 2) and ‘dead’ (State 3) which is an absorbing state.
2) We assume that non-smokers can take up smoking and then give up smoking and also those who were smokers from start can give up smoking as well.
3) For every ‘live’ state an individual can die either from smoking-related cause (whether a current smoker or used to smoke in the past) or from other reasons.
4) In the age interval between 30 and 45 we consider, all transition intensities are considered fixed.

There are four unknown parameters in the described model – transition intensities s_x , g_x, μ_x^{NS} , μ_x^{S}

The mortality tables give us mortality probabilities for age x. For our model shown in Figure 3 these are transition probabilities from the state 1 to 3, and from the state 2 to 3.

Chapman-Kolmogorov equations for this scheme are as follows:

$$\begin{cases} \frac{d}{dt}p_{11} = p_{12}\mu_{21} + p_{11}\mu_{11} \\ \frac{d}{dt}p_{12} = p_{12}\mu_{22} + p_{11}\mu_{12} \\ \frac{d}{dt}p_{13} = p_{11}\mu_{13} + p_{12}\mu_{23} \\ \frac{d}{dt}p_{21} = p_{21}\mu_{11} + p_{22}\mu_{21} \\ \frac{d}{dt}p_{22} = p_{21}\mu_{12} + p_{22}\mu_{22} \\ \frac{d}{dt}p_{23} = p_{21}\mu_{13} + p_{22}\mu_{23} \end{cases} \qquad (4)$$

Norm condition:

$$\sum_{i,j} p_{ij} = 1 .$$

In this situation the initial conditions of two types can be given:

1. An individual does not smoke initially, i.e. occupies State 1:
2. An individual smokes initially, i.e. occupies State 2:

$$p_{11}(0) = 1,\ p_{12}(0) = 0, p_{13}(0) = 0,$$
$$p_{21}(0) = 1,\ p_{22}(0) = 0, p_{23}(0) = 0.$$

Let us assume that there is sufficient statistical data available to estimate $p_{ij}(t)$ and therefore we consider them known. The model describes measurements with the sufficiently acceptable level of accuracy if the following system of inequalities holds:

$$| \quad p_{ij}^{э}(t) \quad - \quad p_{ij}^{P}(t) \quad | \quad < \quad \varepsilon \qquad (5)$$

Where $p_{ij}^{э}(t)$ are experimental probabilities of given values from statistical estimates, $p_{ij}^{P}(t)$ are probabilities estimated using the model and e is the maximum acceptable measurement error.

For every constant value μ_{ij} the uncertainty interval

$$\mathrm{Dij} = [\ \min\ \mu_{ij},\ \max \mu_{ij}\ ] \qquad (6)$$

is defined as an interval within each the variance Dij keeps system (5) true.

The problems of determining intervals (6) given the system of linear restriction (5) were pioneered by the inventor of linear programming Leonid V. Kantorovich and were described in his report to Siberian Mathematical Society in 1962 [2].

It is important that the formulation above was a new method in the theory of mathematical methods in processing experimental results. Classical results in the area of mathematical methods in experiments processing goes back to works of Legandre and Gauss and consists of finding mathematical parameter values such that they minimize some criteria of calculation correspondence to measurements, for example the sum of squares of deviations in observed experimental values and modeled values. Such method is widely known as the least squares method. This method is based on probabilities as well. In case of experiment error being distributed normally (normal distribution curve and Gauss' portrait even made their way to German banknotes), the values of parameters calculated using the least squares method are in some sense the most probable.

The question is how appropriate it is to consider the normal distribution of errors hypothesis as true. As early as in the beginning of this century the famous mathematician Puankare made a witty comment on this subject: "Everyone believes in the exponential law in his own way: a mathematician believes that the law is confirmed by experiments, and the one who conducts the experiments considers it a consequence of mathematical proof". In fact, establishing the distribution law for experimental errors is very complicated problem on its own. At the same time, the formulation of the problem of processing experimental results as suggested by Kantorovich does not require any knowledge about statistical properties of experimental errors distribution. Values ε in the system of inequalities (5) are the characteristics of maximum acceptable experimental errors. The information about the value of maximum acceptable experimental errors is usually given in some form. Then if conditions (5) hold, it means that the model describes the measurements within the limits determined by the maximum acceptable experimental error, which is perfectly natural.

In our previous papers we progressed in applying Kantorovich's method to solving the inverse problems in such areas as chemical kinetics and thermodynamics [3].

In the case considered parameters of the model will depend on age x. In the more general case they will depend on the duration of stay in the particular state (i.e. they will differ for individuals of the same age but with the different past information). This case is obviously more complicated and requires more data about observed individuals.

So, let us assume that the intensities of transition between the states differ for different ages, i.e. $\mu ij \cong \mu ij(x)$.

If we consider an individual's age as a continuous parameter and not discrete, we can try to build "smooth" transition intensities. It is possible to calculate the values of the transition functions in the discrete points from a given data and then interpolate these points by a smooth enough curve.

We can consider the [0, 1] interval as a subclass on which we are looking for an approximate solution (this is justified by the prior considerations). The following estimates for the intervals in question were obtained:

$$\mu_{12} \in [0{,}000171;0{,}00036],$$
$$\mu_{21} \in [0{,}000121;0{,}00051],$$
$$\mu_{13} \in [0{,}000671;0{,}00086],$$
$$\mu_{23} \in [0{,}001180;0{,}00193].$$

Age *x*	$P_{13\ stat}$	$P_{13\ res}$	$P_{23\ stat}$	$P_{23\ res}$
25	0.00067	0.000670	0.00118	0.001180
26	0.00066	0.000661	0.00117	0.001169
27	0.00065	0.000650	0.00118	0.001181
28	0.00065	0.000651	0.00121	0.001211
29	0.00066	0.000661	0.00128	0.001281
30	0.00067	0.000670	0.00135	0.001351
31	0.00070	0.000700	0.00145	0.001450
32	0.00072	0.000720	0.00154	0.001541
33	0.00076	0.000760	0.00165	0.001651
34	0.00080	0.000800	0.00178	0.001781
35	0.00086	0.000859	0.00193	0.001931

As the above solution is not unique, we have to estimate the interval of values of $\mu_{ij,}$ in which we obtain a good approximation of the statistical data. The analysis of the corresponding Markov scheme and the set of equations shows that the substantial changes in $\mu_{13,}$ μ_{23} will cause the large changes in the values of $p_{13,}$ p_{23}. The changes in $\mu_{12,}$ μ_{21} will influence these variables in the much smaller extend.

The numerical experiment gave the following results:

Large (of the 2 orders) increase and decrease of the parameter values for μ_{12} and μ_{21} causes insignificant changes in the calculated probabilities (of the order 1-3%), probabilities p_{12} and p_{21} change a lot, proportionally to changes of μ_{12} and μ_{21}.

Large changes in μ_{13} and μ_{23} cause proportional changes of p_{13}, and p_{23}. Such dependence is explained by the characteristics of the model.

All the above allows to conclude that solving the inverse problem for the given model requires additional information apart from statistical data about mortality probabilities. Such information could be the data on duration of stay in the states 1 and 2. More precise solution could be found by estimating the probabilities p_{12} and p_{21} for various ages. The source of such information could be a study of smoker and non-smoker population.

There will be more than one solution of the inverse problem corresponding to this Markov scheme. This is typical for most of inverse tasks.

There are different methods of finding the intervals for estimates that is selection of a subspace in the space of admissible solutions. Each point of such subspace satisfies a set of equations, initial conditions and system of limitations.

Thus, it is possible to assign different values from solutions subspace to the parameters and obtain equally good description of the statistical data. Such "variation" of parameters can be done arbitrarily or using some functional connections between parameters.

For example, for the three states process in Figure 4,

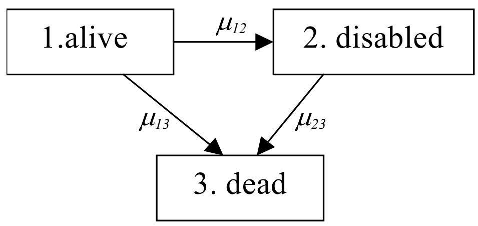

Figure 4.

the force of mortality for people of the given age group will depend on transition forces as follows:

$$\mu(t) = \frac{(\mu_{23} - \mu_{13})(\mu_{12} + \mu_{13})e^{-(\mu_{12}+\mu_{13})t} - \mu_{12}\mu_{23}e^{-\mu_{23}t}}{(\mu_{23} - \mu_{13})e^{-(\mu_{12}+\mu_{13})t} - \mu_{12}e^{-\mu_{23}t}}.$$

It is seen from here, that for any chosen set of values for three parameters, there is another choice, which gives exactly the same value for μ(t). If $\mu_{12} = \tilde{\mu}_{12}, \mu_{13} = \tilde{\mu}_{13}, \mu_{23} = \tilde{\mu}_{23}$, then the same μ(t) can be obtained, assuming $\hat{\mu}_{12} = \tilde{\mu}_{23} - \tilde{\mu}_{13}, \hat{\mu}_{13} = \tilde{\mu}_{13}, u\ \hat{\mu}_{23} = \tilde{\mu}_{12} + \tilde{\mu}_{13}$. In lack of the prior information about parameter values, we can choose any values. Our purpose is simply to find the best value of μ(t), based on the three-state model.

The presence of ambiguity in a solution of the inverse task leads to the following question: Is it possible to realize such a choice from the subspace of solutions, which whenever possible optimizes the properties of the given process?

This question will be the topic of our further research.

References

1. Benjamin B., Michaelson R. (1988) Mortality differences between smokers and non-smokers, *Journal of the Institute of Actuaries*. 115(III), 461-476;
2. Kantorovoch L.V. (1962) *Siberian Mathematical Journal (in Russian)*. 3(5), 701-709;
3. Akhunov I.R., Spivak S.I. (19840. *React. Kinet. Catal. Lett.* 24(2), 305-311.