FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

Volodymyr Sahan

Use of Markov Chain Simulation in Long Term Care Insurance

2025

Research Objectives

- Model the distribution of insured individuals over time.
- Estimate the average time spent in healthy and ill states.
- Determine insurance premiums using Monte Carlo simulations.

What is a Markov Chain?

A Markov Chain satisfies:

$$X_{t_{h+1}} | X_{t_h}, X_{t_{h-1}}, \ldots, X_{t_0} \overset{(d)}{=} X_{t_{h+1}} | X_{t_h}$$

State Space and Time:

- The **state space** is a finite or countable set:

$$S = \{s_1, s_2, \ldots, s_m\}.$$

- Time evolves in discrete steps:

$$T = \mathbb{N}_0 = \{0, 1, 2, \ldots\}.$$

What is a Markov Chain?

A Markov Chain satisfies:

$$X_{t_{h+1}} | X_{t_h}, X_{t_{h-1}}, \ldots, X_{t_0} \overset{(d)}{=} X_{t_{h+1}} | X_{t_h}$$

State Space and Time:

- The **state space** is a finite or countable set:

$$S = \{s_1, s_2, \ldots, s_m\}.$$

- Time evolves in discrete steps:

$$T = \mathbb{N}_0 = \{0, 1, 2, \ldots\}.$$

Key Properties:

- The future state depends only on the present state.
- The history before t_h is irrelevant.

Transition Probabilities and Initial Distribution

Transition Probabilities:

$$p_{i,j}(t, t+1) = P(X_{t+1} = j \mid X_t = i)$$

Transition Probabilities and Initial Distribution

Transition Probabilities:

$$p_{i,j}(t, t+1) = P(X_{t+1} = j \mid X_t = i)$$

Time Homogeneity: A Markov chain $\{X_t\}_{t \in T}$ is called **non-homogeneous in time**, unless:

$$\forall i, j \in S, \forall k \in \mathbb{N}, \quad P(X_{t+k+1} = j \mid X_{t+k} = i) = P(X_{t+1} = j \mid X_t = i).$$

Transition Probabilities:

$$p_{i,j}(t,t+1) = P(X_{t+1} = j \mid X_t = i)$$

Time Homogeneity: A Markov chain $\{X_t\}_{t \in T}$ is called **non-homogeneous in time**, unless:

$$\forall i,j \in S, \forall k \in \mathbb{N}, \quad P(X_{t+k+1} = j \mid X_{t+k} = i) = P(X_{t+1} = j \mid X_t = i).$$

Initial Distribution: For a Markov chain $\{X_t\}_{t \geq 0}$, the probability distribution $\alpha = \{\alpha_k\}_{k \in S}$, where:

$$P(X_{t_0} = s_k) = \alpha_k = p_{s_k}(0), \quad s_k \in S,$$

is called the **initial distribution** of the chain.

Transition Probabilities:

$$p_{i,j}(t, t+1) = P(X_{t+1} = j \mid X_t = i)$$

Time Homogeneity: A Markov chain $\{X_t\}_{t\in T}$ is called **non-homogeneous in time**, unless:

$$\forall i, j \in S, \forall k \in \mathbb{N}, \quad P(X_{t+k+1} = j \mid X_{t+k} = i) = P(X_{t+1} = j \mid X_t = i).$$

Initial Distribution: For a Markov chain $\{X_t\}_{t\geq 0}$, the probability distribution $\alpha = \{\alpha_k\}_{k\in S}$, where:

$$P(X_{t_0} = s_k) = \alpha_k = p_{s_k}(0), \quad s_k \in S,$$

is called the **initial distribution** of the chain.
The absolute probability of the states of the Markov chain:

$$p_{k,j}(0, h) = p_j^{(k)}(h).$$

Chapman-Kolmogorov Equality

Chapman-Kolmogorov Equality (Fecenko, 2018) states that for a Markov chain:

$$p^{(k)}(h) = p^{(k)}(h-1) \cdot P(h-1; h) = p^{(k)}(0) \cdot \prod_{t=0}^{h-1} P(t; t+1).$$

Chapman-Kolmogorov Equality

Chapman-Kolmogorov Equality (Fecenko, 2018) states that for a Markov chain:

$$p^{(k)}(h) = p^{(k)}(h-1) \cdot P(h-1; h) = p^{(k)}(0) \cdot \prod_{t=0}^{h-1} P(t; t+1).$$

Interpretation:

- The probability of transitioning from state k at time 0 to a state at time h depends on the intermediate transition probabilities.
- The Chapman-Kolmogorov equation allows us to compute long-term probabilities by multiplying stepwise transitions.
- This is fundamental in analyzing Markov Chains over multiple time steps.

Generating a Discrete Random Variable

Inverse Transformation Method: To generate a discrete random variable Z with distribution

$$P(Z = z_k) = p_k, \quad k = 1, \ldots, m,$$

we use the following steps:

Generating a Discrete Random Variable

Inverse Transformation Method: To generate a discrete random variable Z with distribution

$$P(Z = z_k) = p_k, \quad k = 1, \ldots, m,$$

we use the following steps:

- Generate $u \sim \text{Unif}(0, 1)$.
- If $u < p_1$, then return z_1.
- Otherwise, find the smallest k such that:

$$\sum_{i=1}^{k-1} p_i < u \leq \sum_{i=1}^{k} p_i.$$

  Then return z_k.

Generating Trajectories of a Non-Homogeneous Markov Chain

For a previously defined non-homogeneous Markov Chain, the process of generating one trajectory of length h follows these steps:

Generating Trajectories of a Non-Homogeneous Markov Chain

For a previously defined non-homogeneous Markov Chain, the process of generating one trajectory of length h follows these steps:

- Generate x_0 from the initial distribution, determining the starting state X_{t_0}.
- At time k, generate x_k from the transition distribution:

$$\left(p_{x_{k-1},j}(t_{k-1},t_k)\right)_{j\in S}.$$

- Repeat until $k = h$.

Trajectory matrix

The results of multiple simulations are stored in a trajectory matrix M:

$$^{(z)}M = [m_{ij}]_{n \times h}, \quad z \in S.$$

where:

- n = number of simulated trajectories.
- h = number of time steps.
- Each entry m_{ij} represents the state of the trajectory i at the time step j.
- z is the initial state.

Accuracy of Estimation

For $Y_n \sim \text{Binom}(n, p)$, the approximate confidence interval for p at level α is given by:

$$\left(\frac{Y_n}{n} - u_{1-\alpha/2} \cdot \sigma, \frac{Y_n}{n} + u_{1-\alpha/2} \cdot \sigma \right),$$

where:

$$\sigma = \sqrt{\frac{p(1-p)}{n}} \leq \frac{1}{2\sqrt{n}},$$

and $u_{1-\alpha/2}$ is the quantile of the standard normal distribution.
Thus, we obtain the uniform confidence intervals:

$$\left(\frac{Y_n}{n} - u_{1-\alpha/2} \cdot \frac{1}{2\sqrt{n}}, \frac{Y_n}{n} + u_{1-\alpha/2} \cdot \frac{1}{2\sqrt{n}} \right).$$

Accuracy of Estimation

Table 1 Accuracy of the probability estimate p for a given number of simulations n with confidence $1 - \alpha = 0.9$

n	$\varepsilon_{0.9}$
1 000	0.0260
10 000	0.0082
100 000	0.0026

Multi-State Model for Long-Term Care Insurance

We consider a **unidirectional** multi-state model with three states:

- **Healthy/Active (A)**: Can transition to **Ill (I)** or **Dead (D)**.
- **Ill (I)**: Can only transition to **Dead (D)**.
- **Dead (D)**: Absorbing state.

Multi-State Model for Long-Term Care Insurance

We consider a **unidirectional** multi-state model with three states:

- **Healthy/Active (A)**: Can transition to **Ill (I)** or **Dead (D)**.
- **Ill (I)**: Can only transition to **Dead (D)**.
- **Dead (D)**: Absorbing state.

Transition Probability Matrix:

$$P(t, t+1) = \begin{bmatrix} p_{A;A}(t;t+1) & p_{A;I}(t;t+1) & p_{A;D}(t;t+1) \\ 0 & p_{I;I}(t;t+1) & p_{I;D}(t;t+1) \\ 0 & 0 & 1 \end{bmatrix}$$

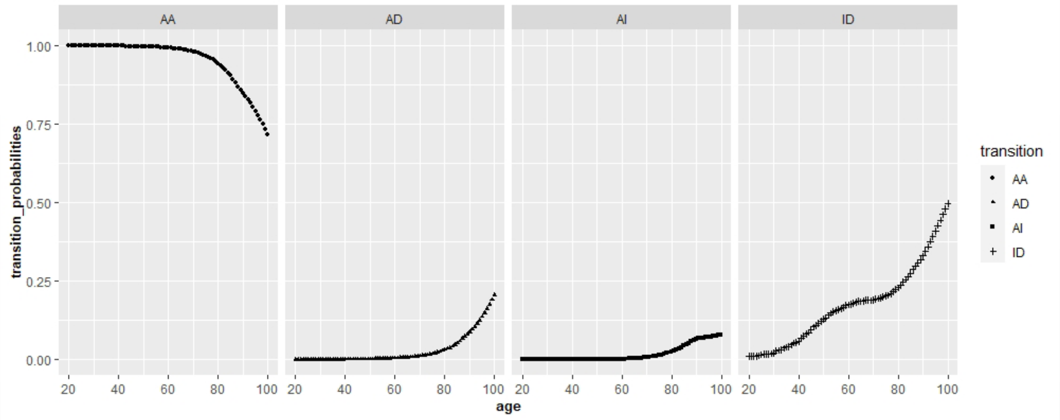

Figure 1 Transition probabilities $p_{i,j}(t, t+1), i, j \in \{A, I, D\}$ by age t of males in Italy.

Estimated percentage distribution

$$perc_g^{(z)}(h) = p_g^{(z)}(h) \cdot 100 \approx \frac{\sum_{i=1}^{n} I[m_{ij} = g]}{n} \cdot 100, \quad g \in \{A, I, D\}, z \in \{A, I\}.$$

Portfolio Composition

The portfolio is composed of K insured lives:

$$K = K_A + K_I$$

where K_s denotes the number of insured lives in the state s.
The total expected number of insured lives in all states is given by the formula:

$$\mathbf{k} = K_A \cdot \mathbf{p}^{(A)}(h) + K_I \cdot \mathbf{p}^{(I)}(h).$$

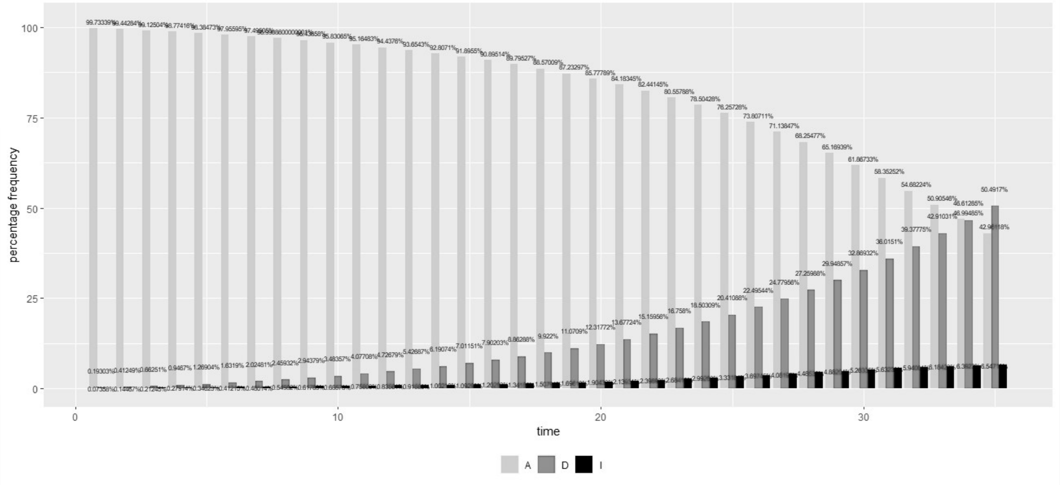

Figure 2 Percentage distribution of initially healthy lives aged 50 in the different states A, I, D over time

Estimated Distribution of Insured Lives After 5 Years

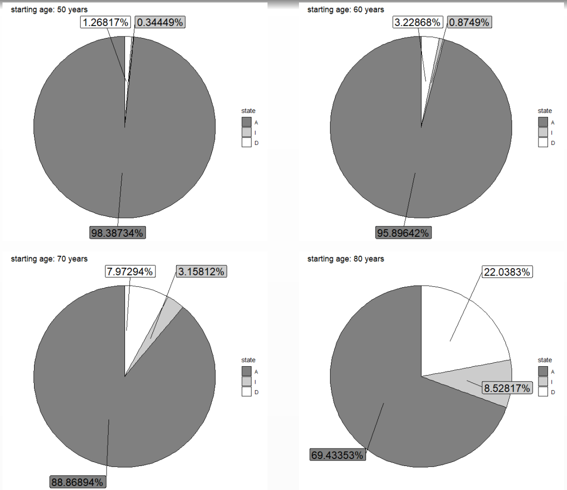

Figure 3 Distribution of insured lives after 5 years based on 100,000 simulations.

Table

Time	Healthy	Ill	Dead	Total
1	299,211 (93.503%)	17,649 (5.515%)	3,140 (0.982%)	320,000 (100%)
2	298,339 (93.231%)	15,551 (4.860%)	6,110 (1.909%)	320,000 (100%)
3	297,386 (92.933%)	13,628 (4.259%)	8,986 (2.808%)	320,000 (100%)
4	296,329 (92.603%)	11,900 (3.719%)	11,771 (3.678%)	320,000 (100%)
5	295,166 (92.239%)	10,407 (3.252%)	14,427 (4.509%)	320,000 (100%)
6	293,895 (91.842%)	9,140 (2.856%)	16,965 (5.302%)	320,000 (100%)
7	292,519 (91.412%)	8,072 (2.523%)	19,409 (6.065%)	320,000 (100%)
8	291,014 (90.942%)	7,185 (2.245%)	21,801 (6.813%)	320,000 (100%)
9	289,364 (90.426%)	6,460 (2.019%)	24,176 (7.555%)	320,000 (100%)
10	287,541 (89.857%)	5,871 (1.834%)	26,588 (8.309%)	320,000 (100%)

Distribution of the number of insured lives during 10 years for males aged 50, of which at the beginning $K_A = 300000$ were in the healthy state and $K_I = 20000$ in the ill state

Estimation of Time Remaining Healthy

Objective: Estimate the duration an insured individual remains in the healthy state using simulated trajectories of a non-homogeneous Markov chain.

Estimation of Time Remaining Healthy

Objective: Estimate the duration an insured individual remains in the healthy state using simulated trajectories of a non-homogeneous Markov chain.

Methodology:

- Simulated 1,000 trajectories using transition probability matrices.
- Initial ages considered: 50, 60, 70, 80.
- Model states up to age 120.

Visualization of Estimated Healthy Years

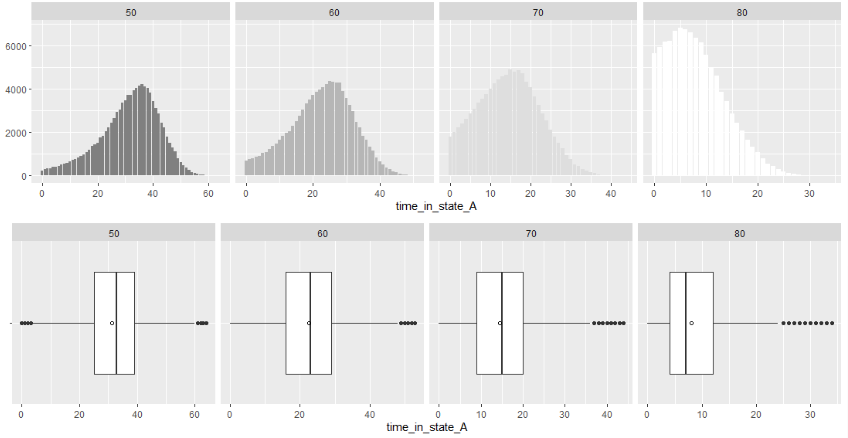

Figure 4 Analysis of the number of years the insured life remained healthy for the initial ages 50, 60, 70, and 80 using a bar plot and a box plot.

Estimated Healthy Years at Different Initial Ages

Age	$x_{0.25}$	Median	$x_{0.75}$	Mean
50	25	33	39	31.39801
60	16	23	29	22.53211
70	9	15	20	14.54286
80	4	7	12	8.10224

Table 4 Estimated values for the number of years the insured life remains in the healthy state.

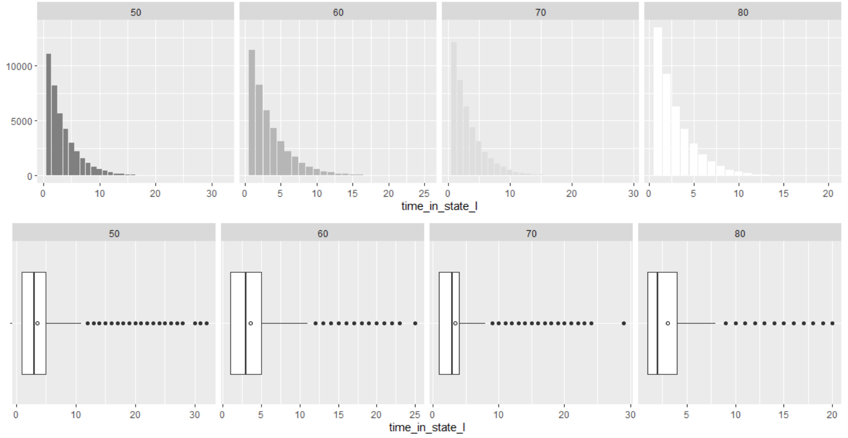

Figure 5 Analysis of the number of years during which insured life remained ill for ages 50, 60, 70 and 80 using a bar graph and a box plot

Estimated Ill Years at Different Initial Ages

Age	$x_{0.25}$	Median	$x_{0.75}$	Mean
50	1	3	5	3.599501
60	1	3	5	3.532215
70	1	3	4	3.365925
80	1	2	4	3.002761

Table 5 Estimated values for the number of years during which the insured life was in the ill state.

Calculation of Long-Term Care Insurance Premiums

We determine the single premium P that an insured life aged x must pay to receive an annual benefit C when in a state of non-self-sufficiency.

- The insured is initially in the healthy state at the start of the policy.
- We use generated non-homogeneous Markov chain trajectories stored in matrix $^{(A)}M = [m_{ij}]_{n \times h}$.
- To calculate the premium, we transform ill-state (I) elements to C and other elements to zero, resulting in matrix M_C.

Premium Calculation Formula

The single premium P is determined as:

$$P = M_C \cdot U$$

where:

$$U = [u_{ij}]_{h \times 1}, \quad u_{ij} = (1 + u)^{-i}$$

with u as the annual interest rate.

Example:

- A healthy life aged 50 would pay €12,583.42 as a single premium to receive C=12000 annually with interest rate u = 0.01.
- The premium was estimated using the arithmetic average over 1,000 repeated scenarios.

Standard Life Insurance Formula

$$P = \sum_{t=1}^{\omega - x} p^{AA}_{x,t-1} \cdot q^{AI}_{x+t-1} \cdot v^t \cdot \pi(\ddot{a}^{(I)}_{x+t})$$

where:

- ω – highest age in the relevant mortality table.
- $p^{AA}_{x,t-1}$ – probability that a life aged x remains healthy for $t-1$ years.
- q^{AI}_{x+t-1} – probability that a life aged $x+t-1$ in the healthy state becomes ill within one year.
- $v = (1+u)^{-1}$ – discount factor, where u is the annual interest rate.
- $\pi(\ddot{a}^{(I)}_{x+t})$ – whole life annuity-due for a life aged $x+t$ if they are in the ill state, defined as:

$$\pi(\ddot{a}^{(I)}_{x+t}) = C \sum_{k=1}^{\omega-(x+t)} p^{II}_{x+t,k-1} \cdot v^{k-1}$$

Comparison of Premium Calculation

Using this formula, we calculated the single premium for a male life aged 50 as:

$$P = 12,584.37$$

In comparison, the premium obtained from the simulation was:

$$P = 12,583.42$$

Premium Estimates with 1,000 Simulations per Scenario

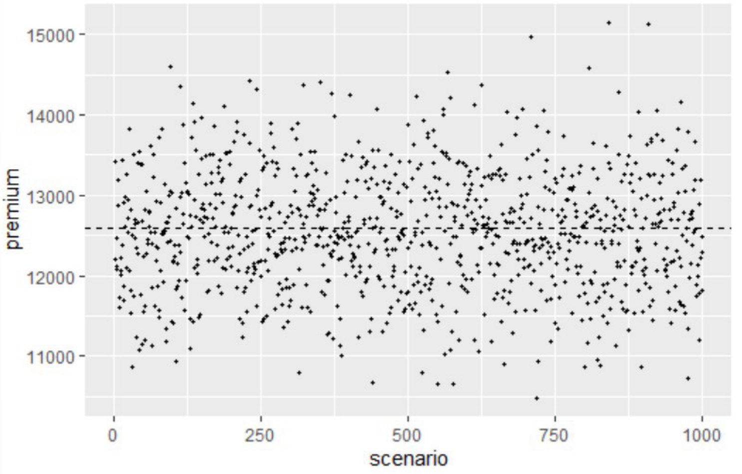

Figure 6 Modelling of the premiums based on 1 000 scenarios for 1 000 non-homogeneous Markov chain simulations

Premium Estimates with 100,000 Simulations per Scenario

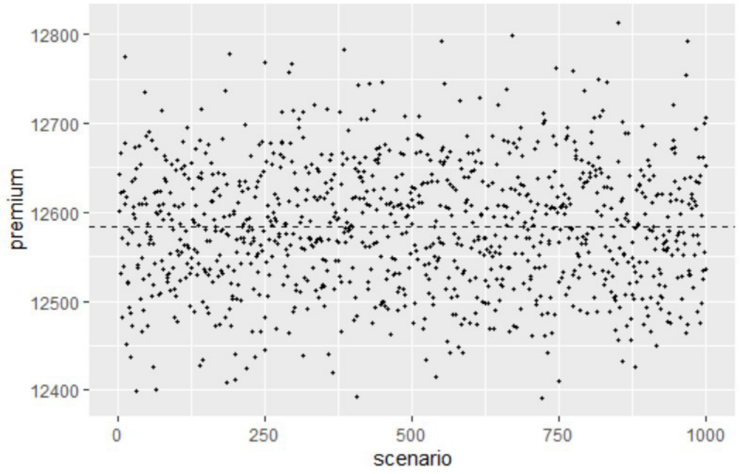

Figure 7 Modelling of the premiums based on 1 000 scenarios for 100 000 non-homogeneous Markov chain simulations

Variation in Premium Estimates Across Scenarios

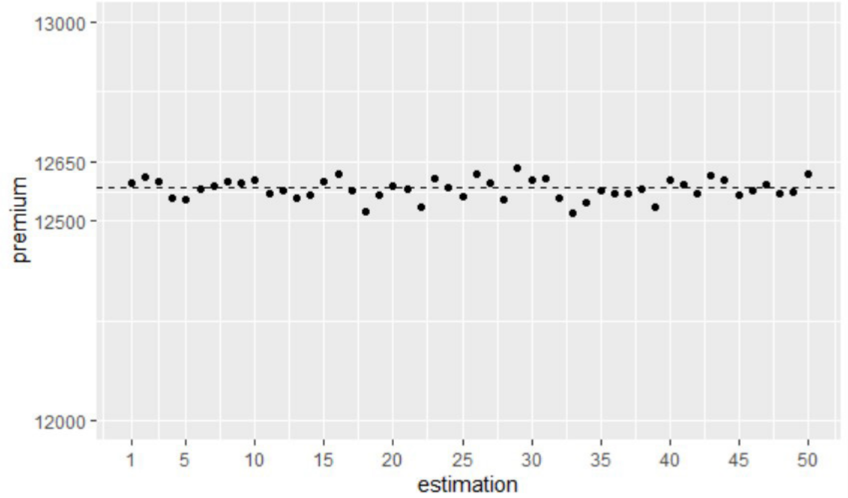

Figure 8 50 premium estimates for 1 000 scenarios for 1 000 non-homogeneous Markov chain simulations.

References

1. Mucha, V., Faybíková, I., Krčová, I. (2022). *Use of Markov Chain Simulation in Long Term Care Insurance*. Statistika, 102(4), 409-425. https://doi.org/10.54694/stat.2022.20

Thank you for your attention!