Faculty of Mathematics and Physics
Charles University

Marek Plesník

Katedra pravděpodobnosti a matematické statistiky

Use of Markov Chain Simulation in Long Term Care Insurance

Econometrics Seminar 1

16 March 2026

Table of Contents

1. Motivation
2. Markov chains
3. Monte Carlo simulation
4. Data & Model
5. Results
6. Premium Calculation
7. Conclusion

Motivation

- Population ageing
  - Increasing life expectancy
  - Rising share of individuals aged 65+
- Higher prevalence of age-related diseases
  - Dementia, Alzheimer's disease, Parkinson's disease, stroke, severe arthritis, ...
  - These conditions often lead to functional limitations and loss of independence
- Rising demand for long-term care (LTC)
  - Home care, nursing homes, assisted living, ...
- Rapid growth in LTC expenditures
  - Increasing financial pressure on insurers and public health systems
- Key questions
  - How do we model the evolution of health states over time?
  - How do we use this to price LTC insurance?

Literature

Main article:

- Mucha, V., Faybíková, I., Krčová, I. (2022). Use of Markov Chain Simulation in Long Term Care Insurance. *Statistika: Statistics and Economy Journal*, 102(4), 409–425.

Selected references:

- Fleischmann, A., Hirz, J., Sirianni, D. (2021). A long-term care multi-state Markov model revisited: a Markov chain Monte Carlo approach. *European Actuarial Journal.*
- Diz, E., Query, J. T. (2017). Applying a Markov model to a plan of social health provisions. *Insurance Markets and Companies.*
- Sugianto, V., Irsan, M. Y. T. (2024). Premium Calculations with Long Term Care Insurance using Markov Chain Method for HIV Inpatients. *Journal of Actuarial, Finance, and Risk Management.*

Markov Chains I

Probability space $(\Omega, \mathcal{A}, \mathrm{P})$, indexed by $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$, discrete states $S = \{s_1, s_2, \ldots\}$.

Definition.

A stochastic process $X = \{X_n, n \in \mathbb{N}_0\}$ with state space $S = \{s_1, s_2, \ldots\}$ is called a (discrete-time) **Markov chain**, if for every $n \in \mathbb{N}_0$ and $i_0, \ldots, i_{n-1}, i, j \in S$ such that

$$\mathrm{P}(X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) > 0,$$

the following holds:

$$\mathrm{P}(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = \mathrm{P}(X_{n+1} = j | X_n = i).$$

In our context:

- Annual time steps
- Finitely many health states
- State at age $n + 1$ depends only on state at age n, not on full health history

Definition.

Let $X = \{X_n, n \in \mathbb{N}_0\}$ be a Markov chain, $m \in \mathbb{N}$, $n \in \mathbb{N}_0$ and $i, j \in S$ such that $\mathrm{P}(X_n = i) > 0$. Then the conditional probability $p_{i,j}(n, n+m) := \mathrm{P}(X_{n+m} = j | X_n = i)$ is called the **transition probability** from state i at time n to state j at time $n+m$.

Definition.

Let $X = \{X_n, n \in \mathbb{N}_0\}$ be a Markov chain and $m \in \mathbb{N}$, $n \in \mathbb{N}_0$. Then the matrix

$$P(n, n+m) = \{p_{i,j}(n, n+m)\}_{i,j \in S}$$

is called the **transition probability matrix**.

Markov Chains III

Definition.

A Markov chain $X = \{X_n, n \in \mathbb{N}_0\}$ is said to be **homogeneous**, if for all $m \in \mathbb{N}$, $n_1, n_2 \in \mathbb{N}_0$ and $i, j \in S$ such that $\mathrm{P}(X_{n_1} = i), \mathrm{P}(X_{n_2} = i) > 0$ holds $p_{i,j}(n_1, n_1 + m) = p_{i,j}(n_2, n_2 + m)$. Otherwise, it is said to be **non-homogeneous**.

In our context:

- Transition probabilities depend on age
- The risk of illness and mortality increases with age
- Therefore, transition probabilities change over time

$\Rightarrow$ **Non-homogeneous Markov chain**

Definition.

Let $X = \{X_n, n \in \mathbb{N}_0\}$ be a Markov chain and $n \in \mathbb{N}_0$.

- For $j \in S$, the probability $p_j(n) := \mathrm{P}(X_n = j)$ is called the **absolute probability** at time n.
- The vector $\mathbf{p}(n) := \{p_i(n), i \in S\}$ is called the **absolute distribution** at time n.

For $n = 0$, we speak of the **initial probabilities** and the **initial distribution**.

Notation:

We denote $p_j^{(i)}(n) := p_{i,j}(0, n)$ and $\mathbf{p}^{(\mathbf{i})}(n) = \{p_{i,j}(0, n), j \in S\}$.

Proposition (Chapman-Kolmogorov).

Let $X = \{X_n, n \in \mathbb{N}_0\}$ be a Markov chain, $n \in \mathbb{N}$ and $i \in S$. Then

$$\mathbf{p}^{(\mathbf{i})}(n) = \mathbf{p}^{(\mathbf{i})}(n-1) \cdot P(n-1, n) = \mathbf{p}^{(\mathbf{i})}(0) \cdot P(0, 1) \cdots P(n-1, n).$$

Proposition (equivalent definition of a (discrete-time) Markov chain).

A stochastic process $X = \{X_t, t \in \mathbb{N}_0\}$ with state space $S = \{s_1, s_2, \dots\}$ is a Markov chain if and only if

$$\mathrm{P}(X_{t_{h+1}} = s_{t_{h+1}} | X_{t_h} = s_{t_h}, X_{t_{h-1}} = s_{t_{h-1}}, \dots, X_{t_0} = s_{t_0}) = \mathrm{P}(X_{t_{h+1}} = s_{t_{h+1}} | X_{t_h} = s_{t_h}),$$

for every $h \in \mathbb{N}_0$, all times $0 \leq t_0 < \dots < t_h < t_{h+1}$ and all $s_{t_0}, \dots, s_{t_{h+1}} \in S$ such that

$$\mathrm{P}(X_{t_h} = s_{t_h}, X_{t_{h-1}} = s_{t_{h-1}}, \dots, X_{t_0} = s_{t_0}) > 0.$$

Notation for the subsequent text:

- Indexing by $t \in \mathbb{N}_0$
- Finite state space, i.e. $S = \{s_1, \dots, s_m\}$ for some $m \in \mathbb{N}$

Monte Carlo Simulation I

Suppose we repeat a trial independently and under identical conditions.

Setup:

- n ... number of repeated independent simulation trials
- p ... theoretical probability of an event occuring in each trial
- f_n ... relative frequency of the observed event across n trials
- ε ... maximum error of estimation
- Y_n ... number of occurences of the observed event in n trials
- α ... significance level where $\alpha \in (0, 1)$

Observe that $Y_n = f_n \cdot n \sim Bi(n, p)$, hence $\mathrm{E}[Y_n] = n \cdot p$ and $\mathrm{var}[Y_n] = n \cdot p \cdot (1 - p)$.

By the **Law of Large Numbers**:

$$\lim_{n\to\infty} \mathrm{P}(|f_n - p| > \varepsilon^*) = 0, \quad \forall \varepsilon^* > 0.$$

Applying the **Central Limit Theorem**:

$$\sqrt{n} \cdot \frac{f_n - p}{\sqrt{p(1-p)}} \xrightarrow{d} N(0,1).$$

Therefore, we can derive a $(1-\alpha)$-accuracy bound $(p-\varepsilon, p+\varepsilon)$ for f_n from

$$f_n \in \left(p - u_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}, p + u_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} \right),$$

where the maximum (estimation) error ε satisfies

$$u_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} \leq u_{1-\frac{\alpha}{2}} \cdot \frac{1}{2} \cdot \sqrt{\frac{1}{n}} =: \varepsilon.$$

Monte Carlo Simulation III

Therefore:

- The error ε depends on the chosen significance level α
- The error ε decreases as n increases
- The worst case occurs when p is close or exactly 0.5

n	ε
100	0.0822
1,000	0.0260
10,000	0.0082
100,000	0.0026
1,000,000	0.0008

Table: Maximum error ε at significance level $\alpha = 0.1$

Generating Trajectories I

Using the **inverse transformation method**.

For a given state $s_r \in S$ at time t, define a random variable Z_r whose distribution corresponds to the r-th row of the transition matrix $P(t, t+1)$:

$$\mathrm{P}(Z_r = j) = p_{s_r,j}(t, t+1), \quad \forall j \in S.$$

Algorithm for generating a value of Z_r:

1. Generate a value u from $U(0,1)$
2. Transform u to a state by applying

$$Z_r = \begin{cases} s_1, & u \leq \mathrm{P}(Z_r = s_1), \\ j, & \sum_{l=s_1}^{j-1} \mathrm{P}(Z_r = l) < u \leq \sum_{l=s_1}^{j} \mathrm{P}(Z_r = l). \end{cases}$$

Generating trajectories II

Algorithm for constructing a trajectory of the chain up to time t_h (for some $h \in \mathbb{N}_0$), starting from initial distribution $\boldsymbol{\alpha} = \{\alpha_k\}_{k \in S}$:

1. **Initialise:** Generate the initial state s_{t_0} from the initial distribution $\boldsymbol{\alpha}$
2. **First step:** Generate the value s_{t_1} from the distribution $\{p_{s_{t_0},j}(t_0, t_1)\}_{j \in S}$
3. **Iterate while** $t_c < t_h$**:** If X_{t_c} is generated, then $\{p_{s_{t_c},j}(t_c, t_{c+1})\}_{j \in S}$ generates $s_{t_{c+1}}$
4. **End if** $t_c = t_h$**.**

The result will be a realisation of a set of $h + 1$ states after h steps. We repeat this algorithm n times, receiving n trajectories of this Markov chain.

Data & Model I

1. Permanent disability model
   - Models illnesses with no possibility of recovery
2. Three states
   - **A** ... healthy, **I** ... ill, **D** ... dead
3. Non-homogeneous Markov chain
   - Transition probabilities $p_{i,j}(t, t+1)$ depend on age t
   - Time advances in discrete steps of one year
4. Key assumptions
   - Transitions out of **A** are irreversible
   - From **I**, the only admissible transition is to **D**
   - **D** is an absorbing state

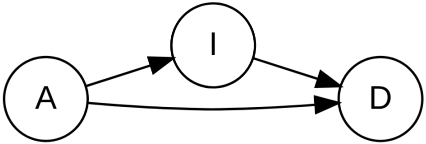

Diagram: Health-state transitions.

$$\begin{pmatrix} p_{AA}(t, t+1) & p_{AI}(t, t+1) & p_{AD}(t, t+1) \\ 0 & p_{II}(t, t+1) & p_{ID}(t, t+1) \\ 0 & 0 & 1 \end{pmatrix}$$

Equation: Transition matrix $P(t, t+1)$ for $t \in \mathbb{N}_0$

Data & Model II

Data Source & Structure:

- de Angelis, P., di Falco, L. (2016). *Assicurazioni Sulla Salute: Caratteristiche, modelli attuariali e basi tecniche*
- Male population in Italy, aged 20 - 100 years (dataset extends up to age 120 years)
- Each age $t \in \{20, \dots, 100\}$ yields one transition matrix $P(t, t+1)$
- Obtained via `markovchain` package in R

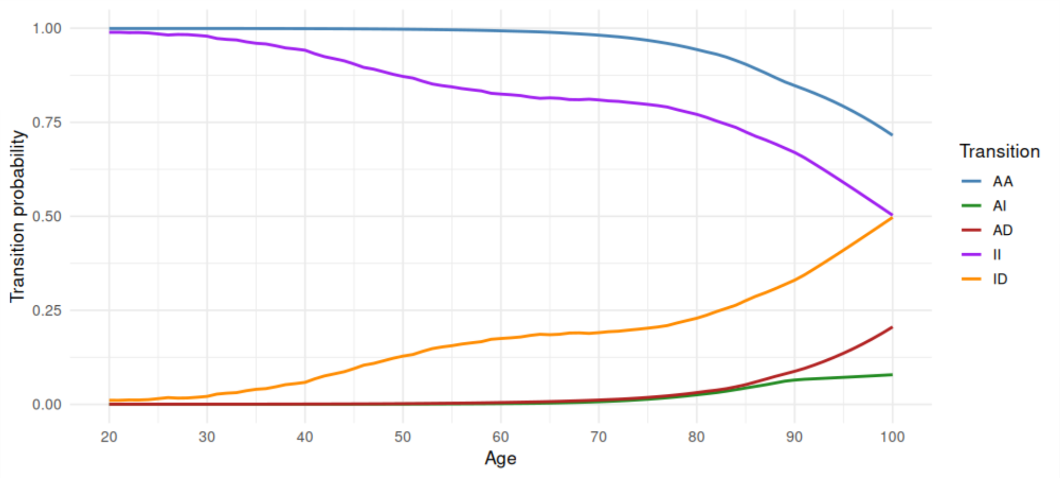

Figure: Transition probabilities p_{AA}, p_{AI}, p_{AD}, p_{II}, p_{ID} of the model as a function of age

Simulation:

Simulated trajectories are written into a matrix ${}^{(z)}M = \{m_{i,j}\}_{n\times h}$, where

- $n \in \mathbb{N}$ ... number of simulated trajectories
- $h \in \mathbb{N}_0$ ... number of time steps
- $z \in \{A, I\}$ ... initial state of the insured life

Percentage distribution after h steps, for initial state z, and final state $g \in \{A, I, D\}$:

$$perc_g^{(z)}(h) = p_g^{(z)}(h) \cdot 100 \approx \frac{\sum_{i=1}^{n} \mathbb{1}_{[m_{i,h}=g]}}{n} \cdot 100$$

Portfolio of $K = K_A + K_I$ insured lives (absolute distribution after h steps):

$$\boldsymbol{k} = K_A \cdot \boldsymbol{p}^A(h) + K_I \cdot \boldsymbol{p}^I(h) = \left(K_A \cdot \boldsymbol{p}_A^A(h), K_A \cdot \boldsymbol{p}_I^A(h) + K_I \cdot \boldsymbol{p}_I^I(h), K_A \cdot \boldsymbol{p}_D^A(h) + K_I \cdot \boldsymbol{p}_D^I(h)\right)^T$$

Key remarks:

- $p_g^{(z)}(h)$ estimated as relative frequency of state g in the h-th column of ${}^{(z)}M$
- Valid by the **Law of Large Numbers** when K_A, K_I are sufficiently large
- By modelling general population, this condition is automatically satisfied

Percentage distribution:

Weighted average using K_A, K_I as weights and $\boldsymbol{perc} = (perc_1, perc_2, perc_3)^T$

$$perc_1 = \frac{K_A}{K} \cdot perc_A^{(A)}(h), \quad perc_2 = \frac{K_A}{K} \cdot perc_I^{(A)}(h) + \frac{K_I}{K} \cdot perc_I^{(I)}(h)$$

$$perc_3 = \frac{K_A}{K} \cdot perc_D^{(A)}(h) + \frac{K_I}{K} \cdot perc_D^{(I)}(h)$$

Note: $\mathrm{perc}_A^{(I)}(h) = 0$ since transition from I to A is not admissible.

Results I

Simulation setup:

- 100,000 simulations per scenario, repeated 100 times and averaged for greater accuracy
- Time steps are annual, modelled up to age 100
- Results stored in matrices ${}^{(z)}M = \{m_{i,j}\}_{n \times h}$, $z \in \{A, I\}$

We analyse:

- Distribution of insured lives across states A, I, D over time
- Time spent in the healthy state A
- Time spent in the ill state I
- Insurance premium estimation

Results II

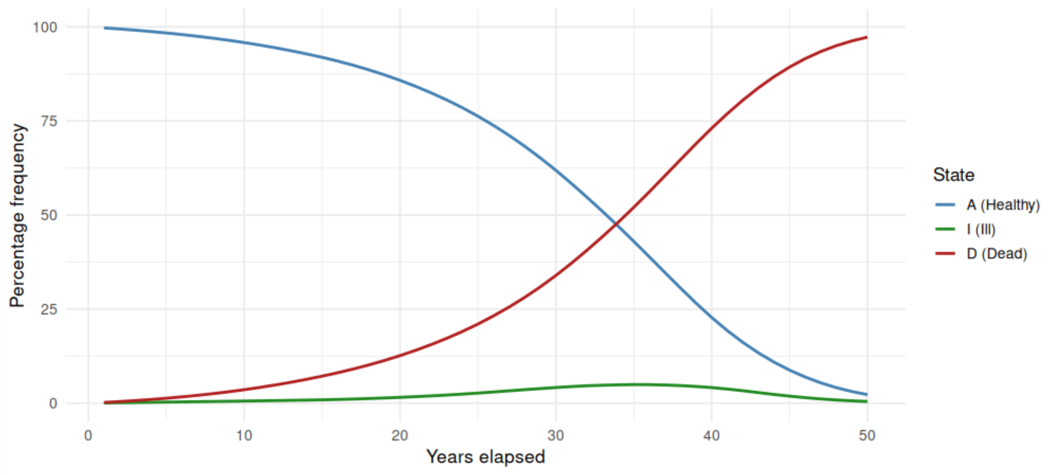

Figure: Percentage distribution of initially healthy lives aged 50 across states **A**, **I**, **D** over time

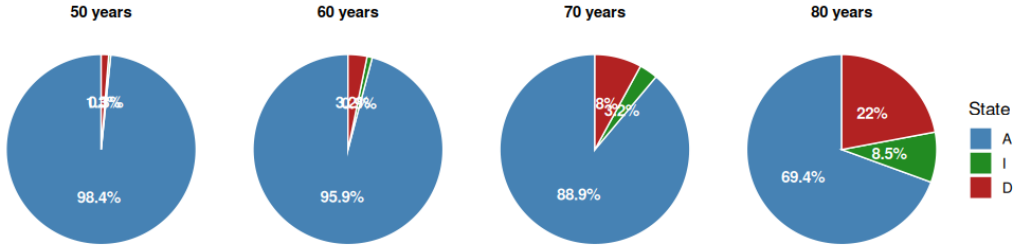

Figure: Percentage distribution after 5 years for initially healthy males aged 50, 60, 70 and 80

Results IV

Portfolio:

Composed of $K_A = 300{,}000$, $K_I = 20{,}000$ insured lives intially aged 50 years.

Year	A	I	D
1	299,211	17,649	3,140
2	298,339	15,551	6,110
5	295,166	10,407	14,427
10	287,541	5,871	26,588

Table: Absolute distribution

Year	A	I	D
1	93.50%	5.52%	0.98%
2	93.23%	4.86%	1.91%
5	92.24%	3.25%	4.51%
10	89.86%	1.83%	8.31%

Table: Percentage distribution

Remarks:

- For percentage distribution, only the ratio K_A/K_I matters, not the absolute values
- Simulation results consistent with values from absolute probability vectors

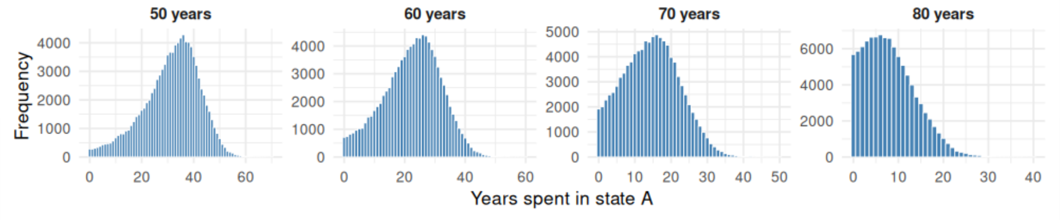

Figure: Years spent in healthy state A by starting age

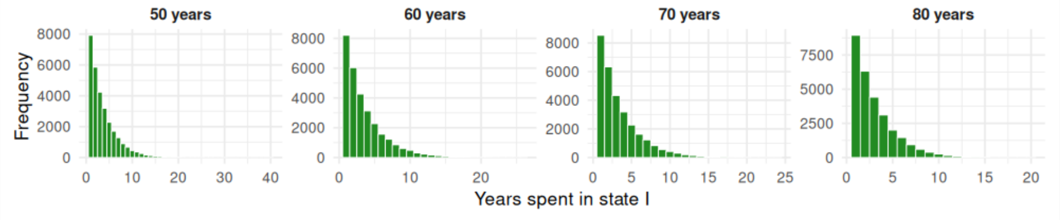

Figure: Years spent in ill state I by starting age (given illness occurs)

Results VI

Age	$\chi_{0.25}$	Median	$\chi_{0.75}$	Mean
50	25	33	39	31.40
60	16	23	29	22.53
70	9	15	20	14.54
80	4	7	12	8.10

Table: Years remaining healthy

Age	$\chi_{0.25}$	Median	$\chi_{0.75}$	Mean
50	1	3	5	3.60
60	1	3	5	3.53
70	1	3	4	3.37
80	1	2	4	3.00

Table: Years spent ill (given illness occurs)

Remarks:

- Many lives transition directly from healthy to dead
- At age 50, nearly 60 % lives never entered the ill state
- Once ill, the average duration in state **I** is only 3 to 4 years

Premium Calculation I

Goal:

Determine the single premium P a healthy life aged x must pay to receive an annual payment C for each year spent in the ill state **I**

Simulation-based approach:

1. Run $n = 100{,}000$ simulations, storing trajectories in ${}^{(A)}M = \{m_{i,j}\}_{n\times h}$, where $h = 120 - x$
2. Transform ${}^{(A)}M$ into the matrix $M_C = \{m^C_{i,j}\}_{n\times h}$ by substituting
$$m^C_{i,j} = \begin{cases} C, & \text{if } m_{i,j} = I, \\ 0, & \text{otherwise.} \end{cases}$$
3. Compute the discounted value for each trajectory using the vector $U = \{u_{i,1}\}_{h\times 1}$, where $u_{i,1} = (1+u)^{-i}$ and u is the annual interest rate:
$$P = M_C \cdot U$$

Standard actuarial formula:

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

- ω ... highest age in the mortality table
- ${}_{t-1}p_x^{AA}$ ... probability of remaining in state A from age x for $t-1$ years
- q_{x+t-1}^{AI} ... probability of transitioning from A to I within one year at age $x+t-1$
- $v = \frac{1}{1+u}$ ... annual discount factor, where u is the annual interest rate
- $\pi(\ddot{a}_{x+t}^{(I)}) = C \cdot \sum_{k=1}^{\omega-(x+t)} {}_{k-1}p_{x+t}^{II} \cdot v^{k-1}$ ... whole life annuity-due for a life aged $x+t$ in state **I**, with annual payment C payable in advance
  - ${}_{k-1}p_{x+t}^{II}$ ... probability of remaining in state **I** from age $x+t$ for $k-1$ years

Premium Calculation III

Example:

Healthy male aged 50 years, annual payment $C = €12{,}000$, interest rate $u = 0.01$

	Premium P
Simulation	€12,583.42
Standard actuarial formula	€12,584.37

Table: Simulation based on $n = 100{,}000$ trajectories averaged over $1{,}000$ scenarios

Key remarks:

- Results comparable — simulation **validated** against closed-form formula
- Simulation offers greater **flexibility** — no need to re-derive expressions when model changes
- Premium increases with x and p_{AI}, decreases with u

Premium Calculation IV

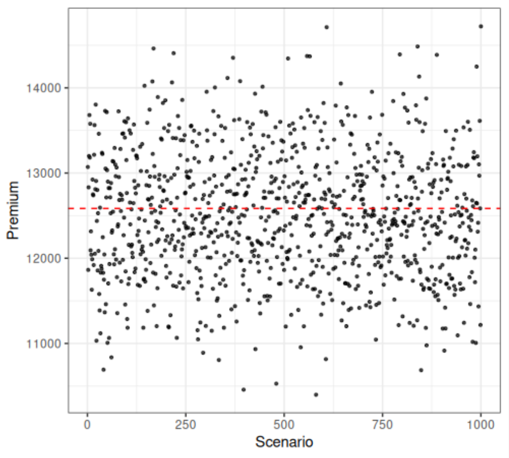

Figure: 1,000 scenarios, $n = 1{,}000$ simulations

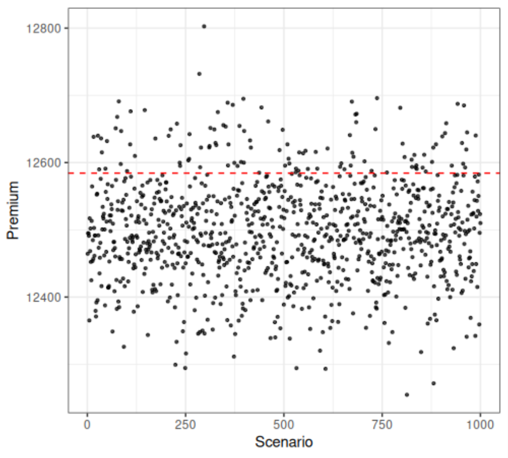

Figure: 1,000 scenarios, $n = 100{,}000$ simulations

Conclusion

Methodology & Results:

- Non-homogeneous Markov chains — natural framework for LTC insurance
- Simulation validated: €12,583.42 vs. €12,584.37 (closed-form)
- Flexible — no need to re-derive expressions when model changes

Limitations:

- Deterministic transition probabilities, fixed interest rate
- Irreversibility reasonable for Alzheimer's — not universally valid
- Italian male population only

Extensions:

- Continuous-time or semi-Markov chains
- Bayesian estimation of transition probabilities
- Stochastic interest rates, richer state spaces

Thank You for Your Attention!