Modeling the Number of Insureds' Cars Using Queuing Theory

Guillaume Couture-Piché

Université du Québec à Montréal

June 23, 2016

Motivation

A model of the number of cars per household allows the forecast :

- when the insured leaves the company (mid-term cancellation or non-renewal of contract)

Motivation

A model of the number of cars per household allows the forecast :

- when the insured leaves the company (mid-term cancellation or non-renewal of contract)
- additions and withdrawals cars ;

Motivation

A model of the number of cars per household allows the forecast :

- when the insured leaves the company (mid-term cancellation or non-renewal of contract)
- additions and withdrawals cars ;
- the amount of time insured (exposition) ;

Motivation

A model of the number of cars per household allows the forecast :

- when the insured leaves the company (mid-term cancellation or non-renewal of contract)
- additions and withdrawals cars ;
- the amount of time insured (exposition) ;
- the number of remaining cars ;

Motivation

A model of the number of cars per household allows the forecast :

- when the insured leaves the company (mid-term cancellation or non-renewal of contract)
- additions and withdrawals cars ;
- the amount of time insured (exposition) ;
- the number of remaining cars ;
- net present value (customer lifetime value).

Database used

The database has the following features :

- time period analized goes from 2003 to 2007 ;

Database used

The database has the following features :

- time period analized goes from 2003 to 2007 ;
- about 300,000 customers ;

Database used

The database has the following features :

- time period analized goes from 2003 to 2007 ;
- about 300,000 customers ;
- only province is Ontario.

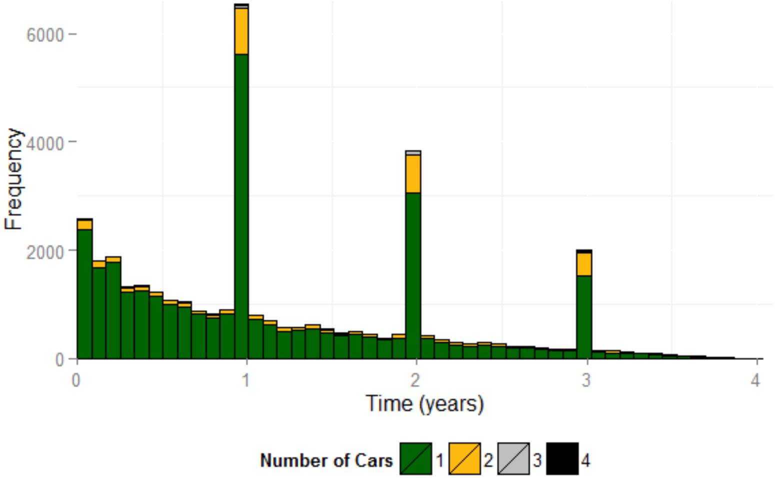


FIGURE – Life of an insurance policy

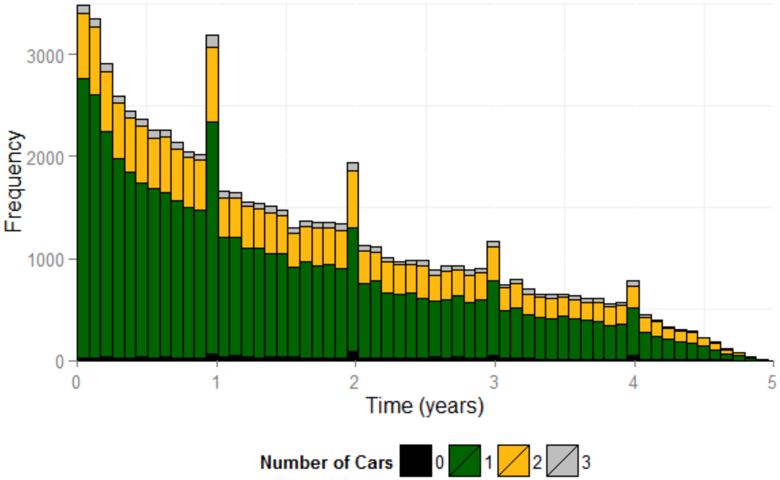


FIGURE – Time prior the addition of a car

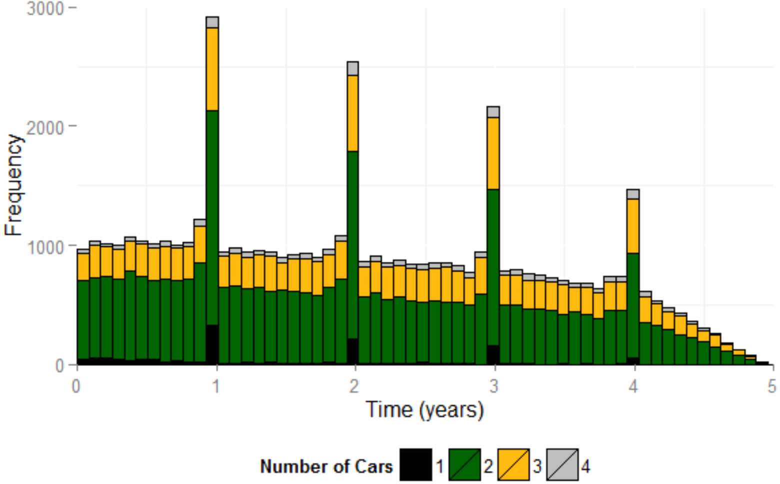


FIGURE – Time prior the removal of a car

Desired Model

0 1 2 3 4

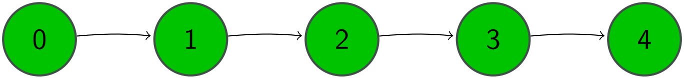
We therefore search for a model with these properties :

Desired Model

We therefore search for a model with these properties :

- the number of cars should mainly be between 0 and 4 ;

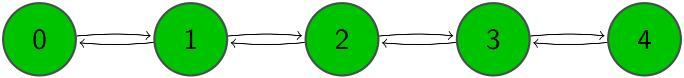
Desired Model



We therefore search for a model with these properties :

- the number of cars should mainly be between 0 and 4 ;
- a car arrivals process ;

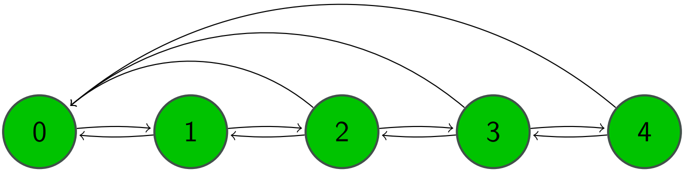
Desired Model



We therefore search for a model with these properties :

- the number of cars should mainly be between 0 and 4 ;
- a car arrivals process ;
- a car removal process ;

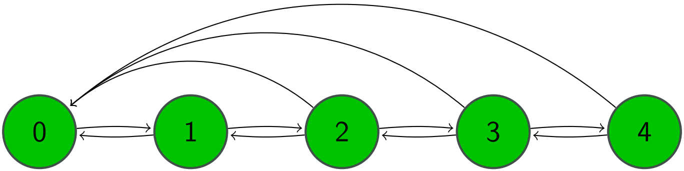
Desired Model



We therefore search for a model with these properties :

- the number of cars should mainly be between 0 and 4 ;
- a car arrivals process ;
- a car removal process ;
- an insured departure process ;

Desired Model



We therefore search for a model with these properties :

- the number of cars should mainly be between 0 and 4 ;
- a car arrivals process ;
- a car removal process ;
- an insured departure process ;
- a shock at contract renewal.

Definitions

Some definitions :

Definitions

Some definitions :

- $\mathcal{N}(t)$: number of cars at time t ;

Definitions

Some definitions :

- $\mathcal{N}(t)$: number of cars at time t ;
- $\mathcal{Z}(t)$: number of cars addition up to time t ;

Definitions

Some definitions :

- $\mathcal{N}(t)$: number of cars at time t ;
- $\mathcal{Z}(t)$: number of cars addition up to time t ;
- λ : car arrivals rate ;

Definitions

Some definitions :

- $\mathcal{N}(t)$: number of cars at time t ;
- $\mathcal{Z}(t)$: number of cars addition up to time t ;
- λ : car arrivals rate ;
- μ : car removal rate ;

Definitions

Some definitions :

- $\mathcal{N}(t)$: number of cars at time t ;
- $\mathcal{Z}(t)$: number of cars addition up to time t ;
- λ : car arrivals rate ;
- μ : car removal rate ;
- γ : mid-term cancellation rate ;

Definitions

Some definitions :

- $\mathcal{N}(t)$: number of cars at time t ;
- $\mathcal{Z}(t)$: number of cars addition up to time t ;
- λ : car arrivals rate ;
- μ : car removal rate ;
- γ : mid-term cancellation rate ;
- p : probability of insurance renewal.

Poisson Process

$$0 \xrightarrow{\lambda} 1 \xrightarrow{\lambda} 2 \xrightarrow{\lambda} 3 \xrightarrow{\lambda} \cdots$$

The first assumption is that the addition of cars in a insurance contract follows a Poisson process :

Poisson Process

$$0 \xrightarrow{\lambda} 1 \xrightarrow{\lambda} 2 \xrightarrow{\lambda} 3 \xrightarrow{\lambda} \cdots$$

The first assumption is that the addition of cars in a insurance contract follows a Poisson process :

$$\mathcal{Z}(t) \sim \text{Poisson}(\lambda t),$$

with the distribution :

$$\Pr(\mathcal{Z}(t) = j) = e^{-\lambda t} \frac{(\lambda t)^j}{j!}.$$

Kendall's notation

A queue process can be noted using the Kendall notation $(A/S/C/K/N/D)$ depending on system capacity and distributions :

- A : time distribution between **A**rrivals ;

Kendall's notation

A queue process can be noted using the Kendall notation $(A/S/C/K/N/D)$ depending on system capacity and distributions :

- A : time distribution between **A**rrivals ;
- S : time distribution of **S**ervice ;

Kendall's notation

A queue process can be noted using the Kendall notation $(A/S/C/K/N/D)$ depending on system capacity and distributions :

- A : time distribution between **A**rrivals ;
- S : time distribution of **S**ervice ;
- c : number of servers (*service* **C***hannels*) ;

Kendall's notation

A queue process can be noted using the Kendall notation $(A/S/C/K/N/D)$ depending on system capacity and distributions :

- A : time distribution between **A**rrivals ;
- S : time distribution of **S**ervice ;
- c : number of servers (*service* **C***hannels*) ;
- K : queue capacity ;

Kendall's notation

A queue process can be noted using the Kendall notation $(A/S/C/K/N/D)$ depending on system capacity and distributions :

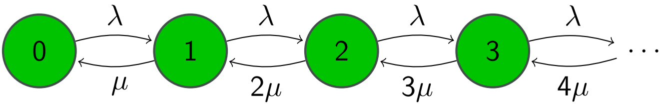
- A : time distribution between **A**rrivals ;
- S : time distribution of **S**ervice ;
- c : number of servers (*service* **C***hannels*) ;
- K : queue capacity ;
- N : size of the population ;

Kendall's notation

A queue process can be noted using the Kendall notation $(A/S/C/K/N/D)$ depending on system capacity and distributions :

- A : time distribution between **A**rrivals ;
- S : time distribution of **S**ervice ;
- c : number of servers (*service* **C***hannels*) ;
- K : queue capacity ;
- N : size of the population ;
- D : queuing **D**iscipline (ex : first come, first served).

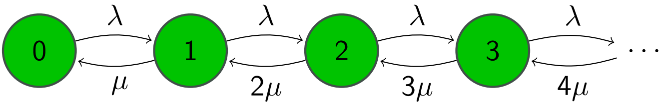
$M/M/\infty$ Model



A $M/M/\infty$ model will be used, which according to Kendall's queueing notation, means :

- first M : Poisson process arrival (**M**arkovian) ;

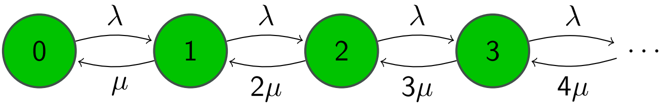
$M/M/\infty$ Model



A $M/M/\infty$ model will be used, which according to Kendall's queueing notation, means :

- first M : Poisson process arrival (**M**arkovian) ;
- second M : exponential service time (**M**arkovian) ;

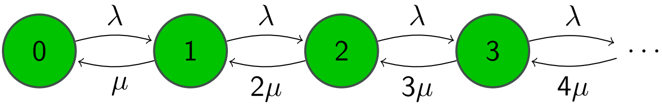
$M/M/\infty$ Model



A $M/M/\infty$ model will be used, which according to Kendall's queueing notation, means :

- first M : Poisson process arrival (**M**arkovian) ;
- second M : exponential service time (**M**arkovian) ;
- ∞ : Infinite number of servers.

$M/M/\infty$ Model

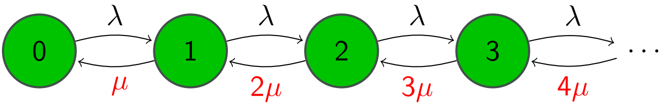


A $M/M/\infty$ model will be used, which according to Kendall's queueing notation, means :

- first M : Poisson process arrival (**M**arkovian) ;
- second M : exponential service time (**M**arkovian) ;
- ∞ : Infinite number of servers.

If they are not specified, the other parameters are assumed as : ∞,∞, FIFO *(first in, first out)*.

$M/M/\infty$ Model



A $M/M/\infty$ model will be used, which according to Kendall's queueing notation, means :

- first M : Poisson process arrival (**M**arkovian) ;
- second M : exponential service time (**M**arkovian) ;
- ∞ : Infinite number of servers.

If they are not specified, the other parameters are assumed as : ∞,∞, FIFO *(first in, first out)*.

$M/M/\infty$ Model

In our case, new automobiles insured enter an imaginary queue

$M/M/\infty$ Model

In our case, new automobiles insured enter an imaginary queue and are insured (or "served") some random time before leaving.

$M/M/\infty$ Model

In our case, new automobiles insured enter an imaginary queue and are insured (or "served") some random time before leaving.
From ∞, there is no limit to the number of cars insured.

$M/M/\infty$ Model distribution

Using the law of total probability :

$$\Pr(\mathcal{N}(t) = i) = \sum_{j=i} \Pr(\mathcal{N}(t) = i | \mathcal{Z}(t) = j) \Pr(\mathcal{Z}(t) = j)$$

For the next step, define :

- Q_j : time before the removal of vehicle j ;

$M/M/\infty$ Model distribution

Using the law of total probability :

$$\Pr(\mathcal{N}(t) = i) = \sum_{j=i} \Pr(\mathcal{N}(t) = i | \mathcal{Z}(t) = j) \Pr(\mathcal{Z}(t) = j)$$

For the next step, define :

- Q_j : time before the removal of vehicle j ;
- I_j : time when the vehicle j was added ;

$M/M/\infty$ Model distribution

Using the law of total probability :

$$\Pr(\mathcal{N}(t) = i) = \sum_{j=i} \Pr(\mathcal{N}(t) = i | \mathcal{Z}(t) = j) \Pr(\mathcal{Z}(t) = j)$$

For the next step, define :

- Q_j : time before the removal of vehicle j ;
- I_j : time when the vehicle j was added ;
- $S(\cdot)$: fonction de survie du temps de service.

$M/M/\infty$ Model distribution

Pour $i = j = 1$ et $\mathcal{N}(0) = 0$, nous avons :

$$\Pr(\mathcal{N}_0(t) = 1 | \mathcal{Z}(t) = 1) = \int_0^t \Pr(Q_1 > t - x | I_1 = x) \times \Pr(I_1 = x) dx$$

$M/M/\infty$ Model distribution

Pour $i = j = 1$ et $\mathcal{N}(0) = 0$, nous avons :

$$\begin{aligned} \Pr(\mathcal{N}_0(t) = 1 | \mathcal{Z}(t) = 1) &= \int_0^t \Pr(Q_1 > t - x | I_1 = x) \times \Pr(I_1 = x) dx \\ &= \int_0^t \Pr(Q_1 > t - x | I_1 = x) \frac{1}{t} dx \end{aligned}$$

$M/M/\infty$ Model distribution

Pour $i = j = 1$ et $\mathcal{N}(0) = 0$, nous avons :

$$\begin{aligned}
\Pr(\mathcal{N}_0(t) = 1 | \mathcal{Z}(t) = 1) &= \int_0^t \Pr(Q_1 > t - x | I_1 = x) \times \Pr(I_1 = x) dx \\
&= \int_0^t \Pr(Q_1 > t - x | I_1 = x) \frac{1}{t} dx \\
&= \int_0^t \frac{S(t - x)}{t} dx
\end{aligned}$$

$M/M/\infty$ Model distribution

Pour $i = j = 1$ et $\mathcal{N}(0) = 0$, nous avons :

$$\begin{aligned}
\Pr(\mathcal{N}_0(t) = 1 | \mathcal{Z}(t) = 1) &= \int_0^t \Pr(Q_1 > t - x | I_1 = x) \times \Pr(I_1 = x) dx \\
&= \int_0^t \Pr(Q_1 > t - x | I_1 = x) \frac{1}{t} dx \\
&= \int_0^t \frac{S(t-x)}{t} dx \\
&= \int_0^t \frac{S(x)}{t} dx
\end{aligned}$$

$M/M/\infty$ Model distribution

Pour $i = j = 1$ et $\mathcal{N}(0) = 0$, nous avons :

$$
\begin{aligned}
\Pr(\mathcal{N}_0(t) = 1 | \mathcal{Z}(t) = 1) &= \int_0^t \Pr(Q_1 > t - x | I_1 = x) \times \Pr(I_1 = x) dx \\
&= \int_0^t \Pr(Q_1 > t - x | I_1 = x) \frac{1}{t} dx \\
&= \int_0^t \frac{S(t-x)}{t} dx \\
&= \int_0^t \frac{S(x)}{t} dx \\
&\equiv q_t,
\end{aligned}
$$

$M/M/\infty$ Model distribution

Using the binomial distribution and independence between the arrivals, we can generalize this result to i arrivals so that : $\mathcal{N}_0(t)|\mathcal{Z}(t) = j \sim \text{Binomial}(j, q_t)$.

$M/M/\infty$ Model distribution

Using the binomial distribution and independence between the arrivals, we can generalize this result to i arrivals so that : $\mathcal{N}_0(t)|\mathcal{Z}(t) = j \sim \text{Binomial}(j, q_t)$.
We can there arrive to :

$$\Pr(\mathcal{N}_0(t) = i) \quad = \quad \sum_{j=i}^{\infty} \binom{j}{i} q_t^i (1 - q_t)^{j-i} \frac{e^{-\lambda t}(\lambda t)^j}{i!}$$

$M/M/\infty$ Model distribution

Using the binomial distribution and independence between the arrivals, we can generalize this result to i arrivals so that :
$\mathcal{N}_0(t)|\mathcal{Z}(t) = j \sim \text{Binomial}(j, q_t)$.
We can there arrive to :

$$\begin{aligned}
\Pr(\mathcal{N}_0(t) = i) &= \sum_{j=i}^{\infty} \binom{j}{i} q_t^i (1-q_t)^{j-i} \frac{e^{-\lambda t}(\lambda t)^j}{i!} \\
&= \frac{(\lambda t q_t)^i e^{-\lambda t}}{i!} \sum_{j=i}^{\infty} \frac{[\lambda t(1-q_t)]^{j-i}}{(j-i)!}
\end{aligned}$$

$M/M/\infty$ Model distribution

Using the binomial distribution and independence between the arrivals, we can generalize this result to i arrivals so that : $\mathcal{N}_0(t)|\mathcal{Z}(t) = j \sim \text{Binomial}(j, q_t)$.
We can there arrive to :

$$\begin{aligned}
\Pr(\mathcal{N}_0(t) = i) &= \sum_{j=i}^{\infty} \binom{j}{i} q_t^i (1-q_t)^{j-i} \frac{e^{-\lambda t}(\lambda t)^j}{i!} \\
&= \frac{(\lambda t q_t)^i e^{-\lambda t}}{i!} \sum_{j=i}^{\infty} \frac{[\lambda t(1-q_t)]^{j-i}}{(j-i)!} \\
&= \frac{(\lambda t q_t)^i e^{-q_t \lambda t}}{i!}.
\end{aligned}$$

$M/M/\infty$ Model distribution

In the $M/M/\infty$ model, the service time distribution is $S(t) = e^{-\mu t}$ and therefore :

$M/M/\infty$ Model distribution

In the $M/M/\infty$ model, the service time distribution is $S(t) = e^{-\mu t}$ and therefore :

$$q_t = \frac{1 - e^{\mu t}}{\mu t}.$$

$M/M/\infty$ Model distribution

In the $M/M/\infty$ model, the service time distribution is $S(t) = e^{-\mu t}$ and therefore :

$$q_t = \frac{1 - e^{\mu t}}{\mu t}.$$

Therefore, the Probability Generating Function (PGF) is :

$M/M/\infty$ Model distribution

In the $M/M/\infty$ model, the service time distribution is $S(t) = e^{-\mu t}$ and therefore :

$$q_t = \frac{1 - e^{\mu t}}{\mu t}.$$

Therefore, the Probability Generating Function (PGF) is :

$$P_{\mathcal{N}_0(t)}(z, t) = e^{(z-1)(1-e^{-\mu t})\frac{\lambda}{\mu}}.$$

$M/M/\infty$ Model distribution

In the $M/M/\infty$ model, the service time distribution is $S(t) = e^{-\mu t}$ and therefore :

$$q_t = \frac{1 - e^{\mu t}}{\mu t}.$$

Therefore, the Probability Generating Function (PGF) is :

$$P_{\mathcal{N}_0(t)}(z, t) = e^{(z-1)(1-e^{-\mu t})\frac{\lambda}{\mu}}.$$

A PGF is defined as :

$$P_X(z) = \sum \Pr(X = x) \times z^x,$$

and allows to easily find moments (expectation, variance) as well as probabilities by mathematical transformation.

Initial cars

A problem with the previous PGF is that it suppose that the number of initial cars insured is 0.

Initial cars

A problem with the previous PGF is that it suppose that the number of initial cars insured is 0.
If we suppose we have a initial cars, we can define the number of these a cars left at time t as $\mathcal{D}_a(t)$.

Initial cars

A problem with the previous PGF is that it suppose that the number of initial cars insured is 0.
If we suppose we have a initial cars, we can define the number of these a cars left at time t as $\mathcal{D}_a(t)$. Since the service time is independant and exponential for the cars, we have that $\mathcal{D}_a(t) \sim \text{Binomial}(a, e^{\mu t})$ and therefore :

$$P_{\mathcal{D}_a(t)} = \left[(z-1)e^{-\mu t} + 1\right]^a$$

Initial cars

A problem with the previous PGF is that it suppose that the number of initial cars insured is 0.
If we suppose we have a initial cars, we can define the number of these a cars left at time t as $\mathcal{D}_a(t)$. Since the service time is independant and exponential for the cars, we have that $\mathcal{D}_a(t) \sim \text{Binomial}(a, e^{\mu t})$ and therefore :

$$P_{\mathcal{D}_a(t)} = \left[(z-1)e^{-\mu t} + 1\right]^a$$

Since $\mathcal{N}_a(t) = \mathcal{D}_a(t) + \mathcal{N}_0(t)$, with independance we know that :

Initial cars

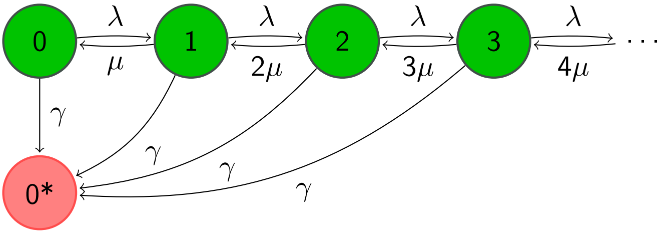
A problem with the previous PGF is that it suppose that the number of initial cars insured is 0.
If we suppose we have a initial cars, we can define the number of these a cars left at time t as $\mathcal{D}_a(t)$. Since the service time is independant and exponential for the cars, we have that $\mathcal{D}_a(t) \sim \text{Binomial}(a, e^{\mu t})$ and therefore :

$$P_{\mathcal{D}_a(t)} = \left[(z-1)e^{-\mu t} + 1\right]^a$$

Since $\mathcal{N}_a(t) = \mathcal{D}_a(t) + \mathcal{N}_0(t)$, with independance we know that :

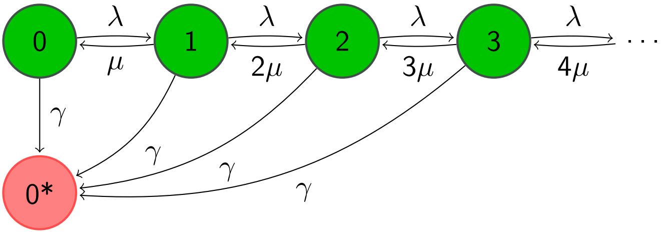
$$P_{\mathcal{N}_a(t)}(z,t) = \left[(z-1)e^{-\mu t} + 1\right]^a e^{(z-1)(1-e^{-\mu t})\frac{\lambda}{\mu}}$$

Two dimensions model



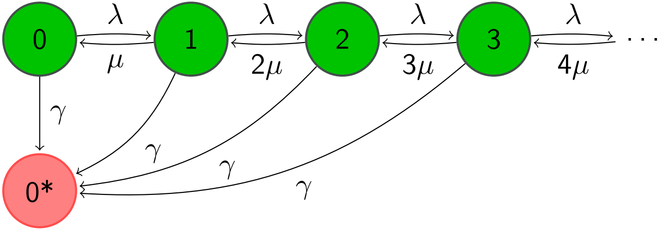
Empirical data suggests some sort of dependance where all cars insurance is removed at once.

Two dimensions model



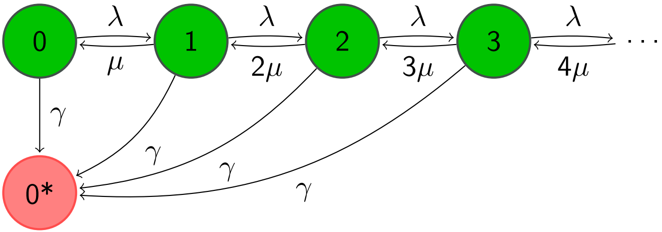
Empirical data suggests some sort of dependance where all cars insurance is removed at once. Hence, we adjust the notation to include $M(t) = m$, a binary variable where $m = 0$ means that the insured has definitively left the company (0* in the graph).

Two dimensions model



Empirical data suggests some sort of dependance where all cars insurance is removed at once. Hence, we adjust the notation to include $M(t) = m$, a binary variable where $m = 0$ means that the insured has definitively left the company (0* in the graph).
Inversely, $m = 1$ means that the insured is still loyal to the company.

Two dimensions model



Empirical data suggests some sort of dependance where all cars insurance is removed at once. Hence, we adjust the notation to include $M(t) = m$, a binary variable where $m = 0$ means that the insured has definitively left the company (0* in the graph).

Inversely, $m = 1$ means that the insured is still loyal to the company.

The special case, $m = 1, n = 0$, could be, for example, an insured person without any car but is home insured or an insured who stored his car for a while.

Two dimensions model

In the first graph, policyholders can leave the company at any time,

Two dimensions model

In the first graph, policyholders can leave the company at any time, but obviously do it particularly on contract renewal.

Two dimensions model

In the first graph, policyholders can leave the company at any time, but obviously do it particularly on contract renewal. With a constant rate of mid term cancellation γ and p the probability of renewal, we find that :

Two dimensions model

In the first graph, policyholders can leave the company at any time, but obviously do it particularly on contract renewal. With a constant rate of mid term cancellation γ and p the probability of renewal, we find that : $M(t) \sim \text{Bernoulli}(e^{-\gamma t} p^{\lfloor t+c \rfloor})$

Two dimensions model

In the first graph, policyholders can leave the company at any time, but obviously do it particularly on contract renewal. With a constant rate of mid term cancellation γ and p the probability of renewal, we find that : $M(t) \sim \text{Bernoulli}(e^{-\gamma t} p^{\lfloor t+c \rfloor})$ where c is a constant for placing the renewal at the right time

Two dimensions model

In the first graph, policyholders can leave the company at any time, but obviously do it particularly on contract renewal. With a constant rate of mid term cancellation γ and p the probability of renewal, we find that : $M(t) \sim \text{Bernoulli}(e^{-\gamma t} p^{\lfloor t+c \rfloor})$ where c is a constant for placing the renewal at the right time and $\lfloor \cdot \rfloor$ the floor function.

Two dimensions model

The number of cars $\mathcal{H}_a(t)$ insured at time t is simply :

Two dimensions model

The number of cars $\mathcal{H}_a(t)$ insured at time t is simply :

$$\mathcal{H}_a(t) = \sum^{\mathcal{M}(t)} \mathcal{N}_a(t)$$

Two dimensions model

The number of cars $\mathcal{H}_a(t)$ insured at time t is simply :

$$\mathcal{H}_a(t) = \sum^{\mathcal{M}(t)} \mathcal{N}_a(t)$$

Using the properties of nested variables with generating functions, we have :

Two dimensions model

The number of cars $\mathcal{H}_a(t)$ insured at time t is simply :

$$\mathcal{H}_a(t) = \sum^{\mathcal{M}(t)} \mathcal{N}_a(t)$$

Using the properties of nested variables with generating functions, we have :

$$P_{\mathcal{H}_a(t)}(z,t) \quad = \quad P_{\mathcal{M}(t)}(P_{\mathcal{N}_a(t)}(z,t),t)$$

Two dimensions model

The number of cars $\mathcal{H}_a(t)$ insured at time t is simply :

$$\mathcal{H}_a(t) = \sum^{\mathcal{M}(t)} \mathcal{N}_a(t)$$

Using the properties of nested variables with generating functions, we have :

$$\begin{aligned} P_{\mathcal{H}_a(t)}(z,t) &= P_{\mathcal{M}(t)}(P_{\mathcal{N}_a(t)}(z,t),t) \\ &= 1 - e^{-\gamma t} p^{\lfloor t+c \rfloor} \left[1 - \left((z-1)e^{-\mu t} + 1 \right)^a e^{(z-1)(1-e^{-\mu t})\frac{\lambda}{\mu}} \right] \end{aligned}$$

Likelihood function

By independance of the various distributions, the likelihood function is simply the multiplication of exponential and Bernoulli distributions.

$$\mathcal{L}(E, A, S, Q, T, V|\lambda, \gamma, \mu, p) \propto e^{-\lambda T}\lambda^{E}e^{-\mu V}\mu^{S}e^{-\gamma T}\gamma^{A}p^{Q}(1-p)^{\lfloor T\rfloor - Q}$$

Likelihood function

By independance of the various distributions, the likelihood function is simply the multiplication of exponential and Bernoulli distributions.

$$\mathcal{L}(E, A, S, Q, T, V|\lambda, \gamma, \mu, p) \propto e^{-\lambda T}\lambda^{E}e^{-\mu V}\mu^{S}e^{-\gamma T}\gamma^{A}p^{Q}(1-p)^{\lfloor T\rfloor - Q}$$

By maximizing this function with respect to specific parameters, we can obtain estimates for them.

Likelihood function

By independance of the various distributions, the likelihood function is simply the multiplication of exponential and Bernoulli distributions.

$$\mathcal{L}(E, A, S, Q, T, V|\lambda, \gamma, \mu, p) \propto e^{-\lambda T}\lambda^{E}e^{-\mu V}\mu^{S}e^{-\gamma T}\gamma^{A}p^{Q}(1-p)^{\lfloor T \rfloor - Q}$$

By maximizing this function with respect to specific parameters, we can obtain estimates for them. We can also modify the function to allow for censored data (done in the thesis, not shown here).

Parameter estimates

Without censored data, we have closed formulas for every parameters :

Parameter estimates

Without censored data, we have closed formulas for every parameters :

$$\hat{\lambda} \quad = \quad \frac{\sum^{\xi} E_i}{\sum^{\xi} T_i} \text{mean of inter arrival time,}$$

Parameter estimates

Without censored data, we have closed formulas for every parameters :

$$\hat{\lambda} = \frac{\sum^{\xi} E_i}{\sum^{\xi} T_i} \text{mean of inter arrival time,}$$

$$\hat{\gamma} = \frac{\sum^{\xi} A_i}{\sum^{\xi} T_i} \text{mean of mid term cancellation rate,}$$

Parameter estimates

Without censored data, we have closed formulas for every parameters :

$$
\begin{aligned}
\hat{\lambda} &= \frac{\sum^{\xi} E_i}{\sum^{\xi} T_i} \text{mean of inter arrival time,} \\
\hat{\gamma} &= \frac{\sum^{\xi} A_i}{\sum^{\xi} T_i} \text{mean of mid term cancellation rate,} \\
\hat{\mu} &= \frac{\sum^{\xi} S_i}{\sum^{\xi} V_i} \text{mean of service time,}
\end{aligned}
$$

Parameter estimates

Without censored data, we have closed formulas for every parameters :

$$\begin{aligned}
\hat{\lambda} &= \frac{\sum^{\xi} E_i}{\sum^{\xi} T_i} \text{mean of inter arrival time,} \\
\hat{\gamma} &= \frac{\sum^{\xi} A_i}{\sum^{\xi} T_i} \text{mean of mid term cancellation rate,} \\
\hat{\mu} &= \frac{\sum^{\xi} S_i}{\sum^{\xi} V_i} \text{mean of service time,} \\
\hat{p} &= \frac{\sum^{\xi} \lfloor T_i \rfloor - Q_i}{\sum^{\xi} \lfloor T_i \rfloor} \text{proportion of renewals.}
\end{aligned}$$

Values of the general parameters

Using the data, we find the following estimates :

Car addition rate λ	Car removal rate μ	Mid-term canc. rate γ	Renewal prob. p
0,0730 (0,0003)	0,0828 (0,0003)	0,0400 (0,0003)	0,9179 (0,0003)

TABLE – Parameter estimates (standard deviation)

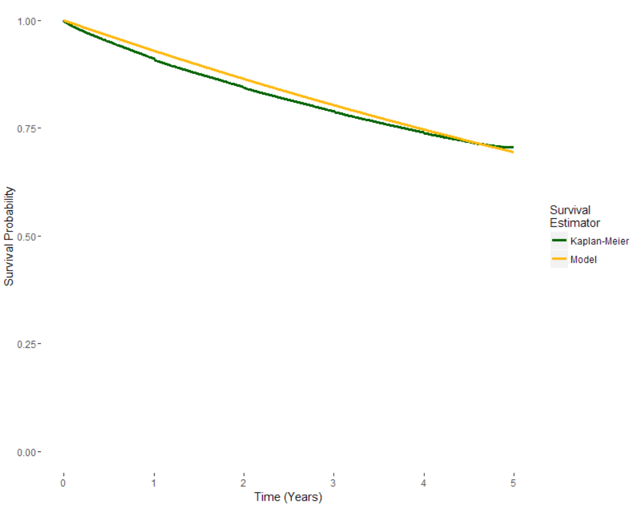


FIGURE – Arrival rate of a new car

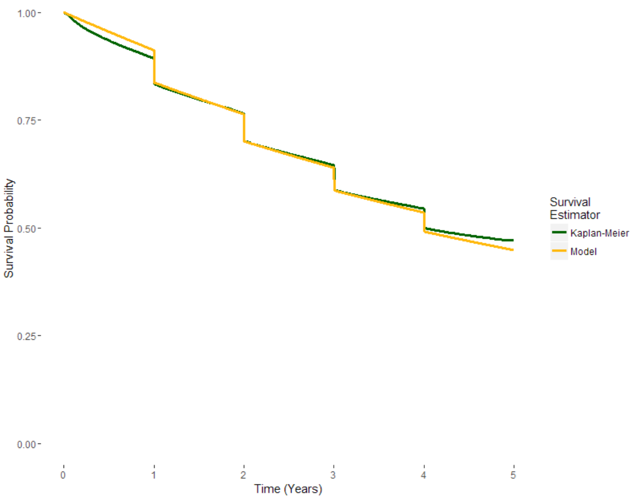


FIGURE – Departure rate of a client (model without censored data)

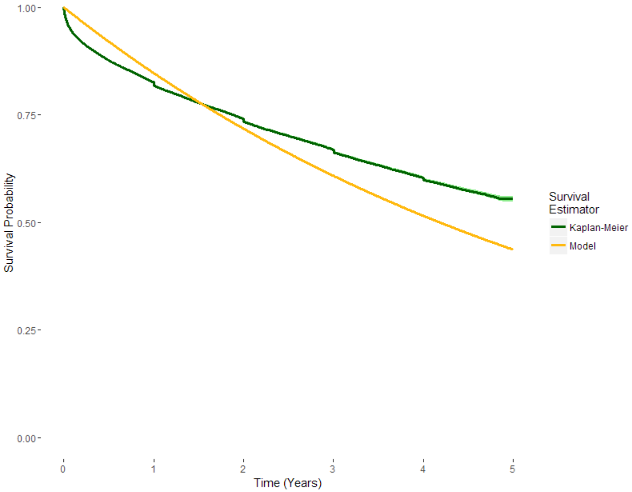


FIGURE – Removal rate with 2 cars (model without censored data)

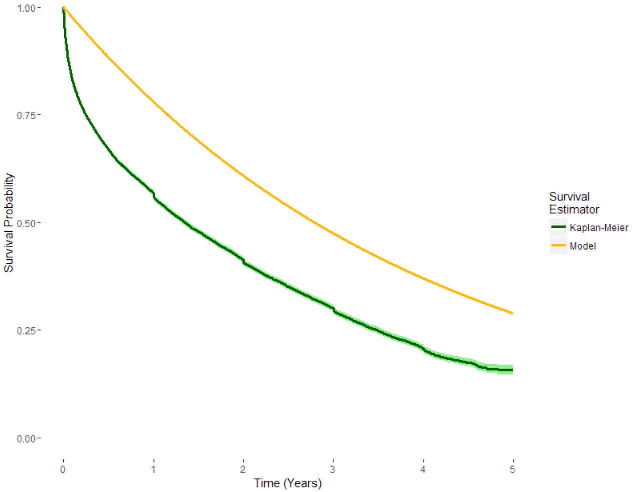


FIGURE – Removal rate with 3 cars (model without censored data)

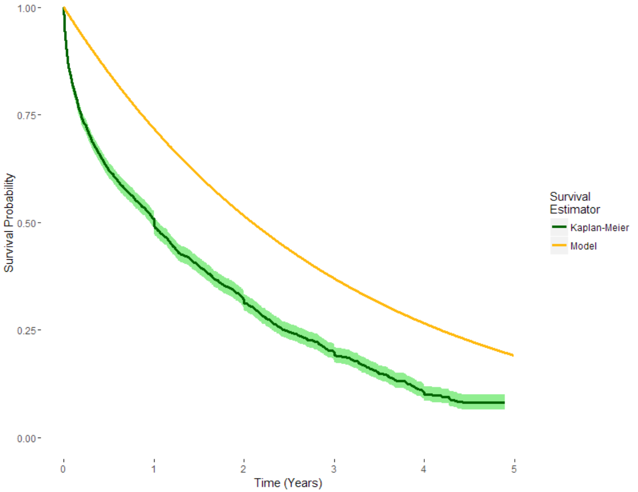


FIGURE – Removal rate with 4 cars (model without censored data)

Regression with explanatory variables

The next step is to generalize this model with explanatory variables.

Regression with explanatory variables

The next step is to generalize this model with explanatory variables. Since all distributions are from the GLM framework, the regression is straightforward and will easily converge (the likelihood function is concave).

Selected explanatory variables

The characteristics chosen to define the insured are, in order, if the insured :

- is from direct or affinity market,

Selected explanatory variables

The characteristics chosen to define the insured are, in order, if the insured :

- is from direct or affinity market,
- has at least one leased vehicle,

Selected explanatory variables

The characteristics chosen to define the insured are, in order, if the insured :

- is from direct or affinity market,
- has at least one leased vehicle,
- is not married,

Selected explanatory variables

The characteristics chosen to define the insured are, in order, if the insured :

- is from direct or affinity market,
- has at least one leased vehicle,
- is not married,
- is insured for less than 9 years (at the beginning),

Selected explanatory variables

The characteristics chosen to define the insured are, in order, if the insured :

- is from direct or affinity market,
- has at least one leased vehicle,
- is not married,
- is insured for less than 9 years (at the beginning),
- has its effective contract date between the month of January and June,

Selected explanatory variables

The characteristics chosen to define the insured are, in order, if the insured :

- is from direct or affinity market,
- has at least one leased vehicle,
- is not married,
- is insured for less than 9 years (at the beginning),
- has its effective contract date between the month of January and June,
- has its effective contract date in the month of July,

Selected explanatory variables

The characteristics chosen to define the insured are, in order, if the insured :

- is from direct or affinity market,
- has at least one leased vehicle,
- is not married,
- is insured for less than 9 years (at the beginning),
- has its effective contract date between the month of January and June,
- has its effective contract date in the month of July,
- has its effective contract date the first day of the month.

Insured examples

For purposes of simplification, I will show only the results of 5 examples of insureds from the many possible combinations.

Parameter	Direct Market	Leased	Not married	0-9 yrs	January-June	July	1st of month
Household A	1	0	1	1	0	0	0
Household B	1	1	1	0	0	0	0
Household C	0	0	1	0	0	0	1
Household D	1	1	0	0	0	0	0
Household E	0	1	0	0	0	1	1

TABLE – Features of the insured examples selected

Insureds parameters

Car addition rate	Car removal rate λ	Mid-term canc. rate γ	Renewal p μ
Household A	0,0492	0,0557	0,1270
Household B	0,0349	0,0283	0,1063
Household C	0,0460	0,0159	0,0868
Household D	0,0550	0,0228	0,0705
Household E	0,0570	0,0107	0,0530

TABLE – Values of the parameters for each profile

Number of cars expected

With the PGF, we can calculate the expected number of cars at the time t using derivatives :

Number of cars expected

With the PGF, we can calculate the expected number of cars at the time t using derivatives :

$$E[\mathcal{H}_a(t)] = \frac{\partial P_{H_a(t)}(z=1,t)}{\partial z}$$

Number of cars expected

With the PGF, we can calculate the expected number of cars at the time t using derivatives :

$$
\begin{aligned}
E[\mathcal{H}_a(t)] &= \frac{\partial P_{H_a(t)}(z=1,t)}{\partial z} \\
&= e^{-\gamma t} p^{\lfloor t+c \rfloor} \left(a e^{-\mu t} + \left(1 - e^{-\mu t}\right) \frac{\lambda}{\mu} \right).
\end{aligned}
$$

Number of cars expected, 5 years

For $t = 5$ years, after the renewal shock, we have the following results :

Number of initial vehicles	1	2	3	4
Household A	0,281	0,490	0,700	0,909
Household B	0,417	0,756	1,096	1,435
Household C	0,541	0,961	1,381	1,801
Household D	0,604	1,058	1,513	1,967
Household E	0,812	1,424	2,036	2,648

TABLE – Expected number of cars after 5 years

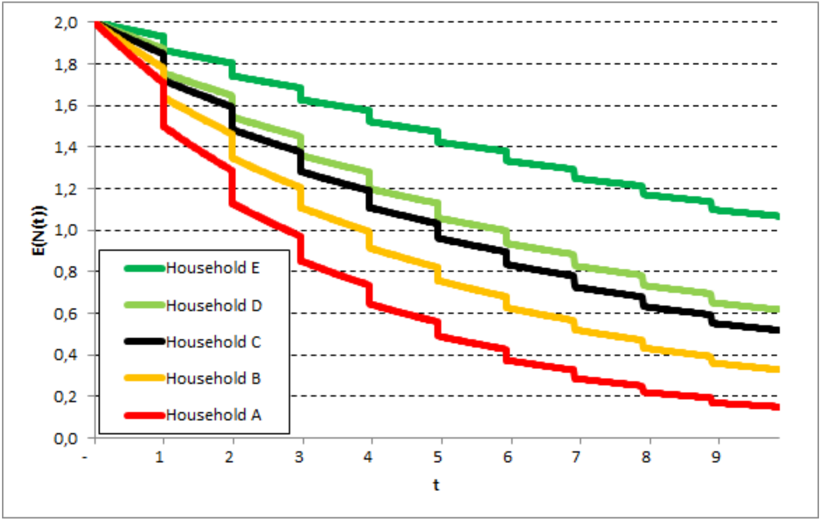


FIGURE – Expected number of cars at time t

Customer lifetime value

With the assumption that insured cars give 1 $ profit per year, we can calculate the sum of these future profits, discounting with the effective rate of δ :

Customer lifetime value

With the assumption that insured cars give 1 \$ profit per year, we can calculate the sum of these future profits, discounting with the effective rate of δ :

$$\text{Profits} = \int_0^\infty \mathbb{E}(\mathcal{H}_a(t))e^{-\delta t}dt$$

Customer lifetime value

With the assumption that insured cars give 1 \$ profit per year, we can calculate the sum of these future profits, discounting with the effective rate of δ :

$$\begin{aligned}
\text{Profits} &= \int_0^\infty \mathbb{E}(\mathcal{H}_a(t))e^{-\delta t}dt \\
&= \frac{\left(1-e^{-(\gamma+\delta+\mu)}\right)\left(a-\frac{\lambda}{\mu}\right)}{(\gamma+\delta+\mu)\left(1-e^{-(\gamma+\delta+\mu)}p\right)} \\
&\quad + \frac{(1-e^{-(\gamma+\delta)})\lambda}{(\gamma+\delta)\mu(1-e^{-(\gamma+\delta)}p)},
\end{aligned}$$

Customer lifetime value with $\delta = 0.02$

Number of initial cars	1	2	3	4
Household A	3,964	7,161	10,358	13,555
Household B	5,588	9,992	14,395	18,799
Household C	7,661	13,011	18,361	23,711
Household D	8,782	14,586	20,390	26,194
Household E	16,130	24,735	33,340	41,945

Conclusion

To advance this model, the following points should be explored :

- test better explanatory variables ;

Conclusion

To advance this model, the following points should be explored :

- test better explanatory variables ;
- model with seasonality and growth of the company ;

Conclusion

To advance this model, the following points should be explored :

- test better explanatory variables ;
- model with seasonality and growth of the company ;
- effect of claims in customer retention ;

Conclusion

To advance this model, the following points should be explored :

- test better explanatory variables ;
- model with seasonality and growth of the company ;
- effect of claims in customer retention ;
- generalization of constant parameters.

Thank you everyone