



Fábio Alexandre Soromenho Pinto

Licenciado em Matemática Aplicada à Economia e à Gestão

Simulação em Sistemas de *Bonus Malus*

Dissertação para obtenção do Grau de
Mestre em Matemática e Aplicações
Ramo de Actuariado, Estatística e Investigação Operacional

Orientadoras: Maria Isabel Azevedo Rodrigues Gomes,
Professora Auxiliar, Universidade Nova de Lisboa
Gracinda Rita Diogo Guerreiro,
Professora Auxiliar, Universidade Nova de Lisboa

Júri:

Presidente: Manuel Leote Tavares Inglês Esquível

Arguente: Maria de Lourdes Belchior Afonso

Vogal: Gracinda Rita Diogo Guerreiro



FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

SETEMBRO, 2014

Simulação em Sistemas de *Bonus Malus*

Copyright © Fábio Alexandre Soromenho Pinto, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa

A Faculdade de Ciências e Tecnologia e a Universidade Nova de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objectivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

A quem sempre me apoiou

Agradecimentos

Um grande agradecimento às minhas orientadoras Isabel Gomes e Gracinda Guerreiro pela assistência dada ao longo da realização desta dissertação.

Deixo também um agradecimento aos meus familiares, amigos, colegas e professores que me deram ânimo e se mostraram sempre interessados pela evolução deste projecto.

Resumo

O objectivo desta dissertação é comparar os resultados obtidos para a evolução de uma carteira de seguro automóvel através do método tradicional de Sistemas de *Bonus Malus* com uma projecção do seu comportamento esperado, obtida recorrendo-se a métodos de Simulação Discreta.

Com este propósito, são descritas as duas metodologias: a primeira utiliza conceitos de Processos Estocásticos (Cadeias de Markov, distribuições estacionárias e processos de contagem), a segunda passa pela geração de Números Pseudo-Aleatórios e de amostras aleatórias.

Os resultados obtidos permitiram concluir que existe, efectivamente, uma concordância entre as duas formulações, mas o grau de concordância dependerá dos pressupostos utilizados.

Palavras-chave: Sistemas de *Bonus Malus*, Cadeias de Markov, Distribuição Estacionária, Processo de Contagem, Simulação, Números Pseudo-Aleatórios

Abstract

This dissertation's objective is to compare the results obtained for the evolution of an auto insurance portfolio through the *Bonus Malus* Systems' traditional method with a projection of its expected behaviour, which is obtained by using Discrete Simulation methods.

With this purpose in mind, both methodologies are described: the first one uses concepts from Stochastic Processes (Markov Chains, stationary distributions and counting processes), the second needs Pseudorandom Number generation and random sampling.

The results allowed to conclude that there is, effectively, an agreement between the two formulations, but the degree of agreement will depend on the assumptions used.

Keywords: *Bonus Malus* Systems, Markov Chains, Stationary Distribution, Counting Processes, Simulation, Pseudorandom Numbers

Conteúdo

1	Introdução	1
2	Modelação do Número de Sinistros	5
2.1	Processo de Poisson Homogéneo (Carteira Homogénea)	6
2.2	Processo de Poisson Heterogéneo (Carteira Heterogénea)	6
2.2.1	Distribuição Poisson - Gama	8
2.2.2	Distribuição Poisson - Inversa Gaussiana	9
2.3	Testes de Ajustamento	10
2.3.1	Teste do Qui-Quadrado de Pearson	10
2.3.2	Teste da Razão de Verossimilhanças	11
3	Sistemas de <i>Bonus Malus</i>	15
3.1	Características	16
3.2	Matrizes de Probabilidades de Transição	17
3.3	Distribuição Estacionária da Carteira	18
3.4	Distribuição Ponderada da Carteira	19
3.5	Escalas Óptimas de Prémios	20
3.5.1	Escala de Norberg - Modelo Assimptótico	21
3.5.2	Escala de Borgan, Hoem e Norberg - Modelo Não Assimptótico	22
3.5.3	Escala de Gilde e Sundt - Modelo Linear	23
4	Simulação Discreta	25
4.1	Geração de NPAs	26
4.1.1	Método Congruencial Misto	26
4.1.2	Método Estendido do Autómato Celular	27
4.2	Métodos de Geração de Amostras Aleatórias	28
4.2.1	Método da Inversão	29
4.2.2	Método da Convolução	29
4.2.3	Método da Aceitação-Rejeição	31

4.2.4	Método da Aceitação-Complemento	32
5	Resultados do SBM e da Simulação	35
5.1	Características do SBM Utilizado	35
5.2	Classes do Sistema, Vector de Prémios, Matriz de Regras de Transição e Número de Sinistros	36
5.3	Distribuição do Número de Sinistros	38
5.4	Geração das Amostras Aleatórias e Resultados da Simulação	39
5.5	Distribuição Estacionária e Prémio Médio	41
5.6	Escalas Óptimas de Prémios	44
5.7	Considerações Finais sobre os Modelos de Poisson Mistos	50
6	Conclusão	53
A	Evolução Anual da Distribuição das Apólices por Classe de <i>Bonus</i>	57
B	Evolução Anual da Distribuição das Apólices por grupos de Classes de <i>Bonus</i> próximas no longo-prazo	61
C	Comparação das Escalas de Prémios	65
D	Código criado no <i>Mathematica</i>	69

Lista de Figuras

5.1	Comparação entre as Distribuições Estacionárias obtidas através do SBM e de Simulação	42
5.2	Evolução da percentagem de apólices nas classes 1 e 20	43
5.3	Evolução Anual Esperada do Prémio Médio ao longo de 50 anos, em percentagem do Prémio Puro <i>a priori</i>	43
5.4	Evolução Anual Esperada do Prémio Médio, em percentagem do Prémio Puro <i>a priori</i>	44
5.5	Comparação entre as Distribuições Estacionária e Ponderada	45
5.6	Comparação das Distribuições Ponderadas obtidas através do SBM e de Simulação	47
5.7	Comparação das Escalas de Prémios	48
5.8	Comparação das Escalas de Norberg determinadas através do SBM e por Simulação	49
5.9	Comparação das Escalas de Borgan, Hoem e Norberg determinadas através do SBM e por Simulação	49
5.10	Comparação das Escalas de Gilde e Sundt determinadas através do SBM e por Simulação	50
A.1	Evolução da percentagem de apólices por classe	59
B.1	Evolução da percentagem de apólices nas classes 2 e 3	62
B.2	Evolução da percentagem de apólices nas classes 4 e 5	62
B.3	Evolução da percentagem de apólices nas classes 6 a 8	62
B.4	Evolução da percentagem de apólices nas classes 9 a 13	63
B.5	Evolução da percentagem de apólices nas classes 14 a 16	63
B.6	Evolução da percentagem de apólices nas classes 17 a 19	63
D.1	Matriz de Regras de Transição (Input)	70
D.2	Matriz de Regras de Transição (Forma Matricial)	70

D.3	Função para determinar a classe no ano seguinte de um segurado	71
D.4	Função para determinar a classe no ano seguinte de todos os segurados .	71
D.5	Função para determinar a distribuição das apólices na carteira num ano .	71
D.6	Aplicação das funções criadas	72

Lista de Tabelas

5.1	Classes e Prémios do SBM Utilizado	36
5.2	Distribuição do Número de Sinistros Observados, em 2006	36
5.3	Estimativas da Máxima Verosimilhança	38
5.4	Resultados dos testes de ajustamento	38
5.5	Distribuição Estacionária da Carteira (valores em percentagem do número total de apólices)	41
5.6	Distribuições Estacionária e Ponderada da Carteira (valores em percentagem do número total de apólices)	46
5.7	Eficiência das Escalas de Prémios	50
C.1	Escalas de Prémios segundo a Distribuição de Polya (valores em percentagem do prémio puro <i>a priori</i>)	66
C.2	Escalas de Prémios segundo a Distribuição de Sichel (valores em percentagem do prémio puro <i>a priori</i>)	67

Algoritmos

1	Método Congruencial Misto	27
2	Método da Inversão	29
3	Método da Aceitação-Rejeição	32
4	Método da Aceitação-Complemento	34

Siglas e Acrónimos

EMV	Estimador da Máxima Verosimilhança
EQM	Erro Quadrático Médio
fgm	Função Geradora de Momentos
IBNR	<i>Incurred But Not Reported</i>
iid	independentes e identicamente distribuídas
MCM	Método Congruencial Misto
NPA	Número Pseudo-Aleatório
SBM	Sistema de <i>Bonus Malus</i>
TLC	Teorema Limite Central
u.m.	unidade monetária
v.a.	variável aleatória



Símbolos

α	Parâmetro de forma de uma distribuição Gama (α, β)
A	Amostra
b	Escala de prêmios
β	Parâmetro de frequência de uma distribuição Gama (α, β)
C_i	Classe de <i>bonus</i> i
C_{i_0}	Classe de entrada do SBM
Δ	Diferença/Variação
g	Valor médio de uma distribuição Inversa Gaussiana (g, h)
$\Gamma(\alpha)$	Função Gama
h	Parâmetro de escala de uma distribuição Inversa Gaussiana (g, h)
K	Rácio de Verossimilhanças
λ	Intensidade de um processo de Poisson Homogêneo
Λ	Variável de estrutura de um processo de Poisson Misto
\mathcal{L}	Função de Verossimilhança
$M_{N(t)}(r)$	Função Geradora de Momentos do processo $N(t)$ calculada para o momento de ordem r
N	Variável aleatória que representa o número de sinistros total da carteira
n_k	Número de apólices que dão origem a k sinistros
$N(t)$	Processo de contagem ao fim de t unidades de tempo $(t \geq 0)$
$\pi_{\mathbf{T}, \lambda}$	Distribuição estacionária
$\pi_{\mathbf{T}}$	Distribuição estacionária incondicional
PP	Prémio Puro
$\mathbf{P}_{\mathbf{T}, \lambda}$	Matriz de probabilidades de transição num passo
$\rho_{\mathbf{T}, \lambda}$	Distribuição ponderada
$\rho_{\mathbf{T}}$	Distribuição ponderada incondicional
s	Número de classes de <i>bonus</i>
S	Variável aleatória que representa a severidade/custo dos sinistros agregados

t	Tempo decorrido num processo de Poisson
t_{ij}	Transição da classe i para a classe j num passo
\mathbf{T}	Matriz de regras de transição
U	Variável aleatória com distribuição Uniforme
u_{λ}^d	Vector das probabilidades da distribuição $U(\lambda)$ discretizada
$U(\lambda)$	Distribuição de estrutura de um processo de Poisson Misto
$U_{[0,1]}$	Distribuição Uniforme no intervalo $[0, 1]$
w_k	Ponderador do ano k para a distribuição ponderada
X	Variável aleatória que representa a severidade/custo de um sinistro
$Z_{S,m}$	Classe de <i>bonus</i> de uma determinada apólice no período m
\sim	Com distribuição



Introdução

Na actividade seguradora é de extrema importância determinar adequadamente qual o prémio a cobrar aos segurados uma vez que, se estes não reflectirem o risco real assumido pela Seguradora, o respectivo ramo da empresa poderá entrar em insolvência. Para prevenir uma situação destas, têm vindo a ser desenvolvidas e aperfeiçoadas várias técnicas de cálculo de prémios.

No caso do seguro automóvel, quando se dá a entrada de uma nova apólice numa carteira podem não existir dados sobre o segurado, pelo que o prémio a ser-lhe cobrado apenas pode ser determinado a partir de certos conjuntos de características observáveis do cliente. No caso particular do seguro automóvel, são geralmente observadas variáveis do condutor, como a idade ou o número de anos de carta; do veículo, como a marca e a potência; a zona de residência, entre muitas outras.

Sendo o seguro automóvel obrigatório, o volume de apólices em carteira deste ramo é geralmente elevado, pelo que um número significativo de segurados partilha das mesmas características. Desta forma, após identificadas as características relevantes em termos de mensuração do risco, agrupam-se indivíduos com características semelhantes, criando, assim, uma tarifa para classes “mais ou menos” homogéneas de risco. Estas técnicas de cálculo de prémios denominam-se por tarifas *a priori*, pois o prémio é calculado antes de se ter dados sobre a experiência do condutor, através do risco potencial.

Porém, nem sempre um prémio é correctamente afecto a uma apólice através destes métodos, pois nem sempre todos os segurados colocados numa determinada classe representam o mesmo risco para a Seguradora. O prémio deve, portanto, ser posteriormente corrigido, de forma que os prémios a pagar por cada segurado reflecta melhor a sua propensão ao risco de acordo com a sua sinistralidade observada. Com este propósito, foram criadas técnicas de tarifação *a posteriori*, ou seja, que corrigem o prémio aos

segurados de acordo com o seu histórico de sinistralidade, o que se entende como sendo o risco real. Os Sistemas de *Bonus Malus* (SBM) são uma destas técnicas, sendo a mais utilizada no mercado segurador.

Os Sistemas de *Bonus Malus* são uma técnica de ajustamento do prémio a cobrar aos segurados em cada classe de risco, ou classe de *bonus*, os quais a empresa seguradora utiliza para cobrir a sua responsabilidade com sinistros. Estes surgiram devido às técnicas de tarificação que determinam o prémio inicial, isto é, quando o tomador do seguro assina o contrato, não assentarem em pressupostos que permitam classificar correctamente todas as apólices da carteira. Naturalmente, visam bonificar o prémio dos segurados que não dão origem a sinistros, ao passo que aqueles que, pelo contrário, originam sinistros vêem o seu prémio agravado.

Nesta dissertação pretende-se estabelecer uma comparação entre as técnicas tradicionais de SBM com técnicas de Simulação: se o número de sinistros reportados é semelhante, se a distribuição das apólices pelas diferentes classes que classificam o seu risco coincide e qual deve ser o prémio a pagar por cada segurado, considerando sempre que a carteira é fechada, isto é, que não se admitem nem entradas nem saídas de apólices, apenas transições entre classes de *bonus* para aquelas que já existem.

O modelo tradicional de Sistemas de *Bonus Malus* assenta em princípios básicos de Processos Estocásticos, nomeadamente processos de contagem Poissonianos e Cadeias de Markov. Por outro lado, as técnicas de Simulação utilizadas para estimar o comportamento futuro da carteira passam pela geração de Números Pseudo-Aleatórios (NPAs), os quais serão transformados através de processos de geração de amostras aleatórias para dar origem à sinistralidade individual.

Toda a análise efectuada ao longo desta dissertação pressupõe um modelo fechado, isto é, assume-se que não existem entradas ou saídas de apólices da carteira. Apesar de ser um pressuposto limitador e de existirem modelos menos restritivos, este é o mais conhecido e utilizado quando se estuda Sistemas de *Bonus Malus* (ver, por exemplo, Lemaire (1995), entre a extensa bibliografia acerca de SBM).

Esta dissertação está estruturada da seguinte forma:

- No Capítulo 2 serão apresentados conceitos de Processos Estocásticos, que permitirão analisar o comportamento das apólices ao longo de vários anos;
- No Capítulo 3 introduz-se, mais aprofundadamente, o cerne dos Sistemas de *Bonus Malus*, como estes funcionam e os resultados que permitem obter;
- De seguida, no Capítulo 4 abordam-se os conceitos relacionados com a Simulação: os métodos de geração de NPAs e de amostras aleatórias, que são utilizados para replicar um sistema real;
- Por último, toda a exposição feita até aí culmina no Capítulo 5 com a aplicação dos conceitos apresentados aos dados da carteira de seguro automóvel de uma Seguradora portuguesa que, por motivos de confidencialidade, terá a designação de

Seguradora Y.

A exposição que se segue, e os resultados para esta obtidos, permitem não só estabelecer a tal comparação entre o método tradicional de SBM com as técnicas de Simulação, mas também dar mais um suporte à Seguradora de que pressupostos poderá assumir de forma que as análises que efectue sobre a sua carteira sejam mais robustas.



Modelação do Número de Sinistros

Conforme referido na Introdução, é essencial as seguradoras estimarem, o mais correctamente possível, qual o prémio que devem cobrar aos seus clientes. Esse prémio é estimado com base em duas variáveis de grande importância: o número de sinistros da carteira e a sua severidade (ou custo). Considerando que estas variáveis são mutuamente independentes, o prémio puro (livre de quaisquer encargos) a cobrar aos segurados é dado por:

$$PP = E[S] = E[X] \cdot E[N]$$

onde:

- $S = \sum_{i=1}^N X_i$ - montante das indemnizações agregadas (com $X_i \sim X$);
- X - montante de uma indemnização individual;
- N - número de sinistros da carteira.

Contudo, na maioria dos países não se toma em consideração o custo dos sinistros nas transições inerentes ao Sistema de *Bonus Malus* (SBM), isto é, as transições são efectuadas com base apenas no número de sinistros reportados (Lemaire (1995)). Assim, considere-se, sem perda de generalidade, que o custo médio por sinistros é de 1 u.m. A expressão anterior reduz-se, portanto, a

$$PP = E[N]. \tag{2.1}$$

Pela igualdade (2.1) compreende-se por que razão é importante conhecer a distribuição do número de sinistros, por forma a obter boas estimativas para o prémio a cobrar aos segurados. Desta forma, e sendo esta variável aleatória (v.a.) um processo de contagem, a suposição mais natural é a de que tem distribuição de Poisson. Comece-se, então, por analisar o caso mais simples.

2.1 Processo de Poisson Homogéneo (Carteira Homogénea)

Ao considerar um processo de Poisson Homogéneo $\{N(t), t \geq 0\}$ para representar a distribuição do número de sinistros da carteira, está a impor-se a condição da carteira ser homogénea, isto é, de todas as apólices apresentarem o mesmo risco,

$$N(t) \sim P(\lambda t) \Leftrightarrow f_{N(t)}(k) = \frac{e^{-\lambda t} (\lambda t)^k}{k!}, \quad k \in \mathbb{N}_0,$$

onde:

- $N(t)$ - número de sinistros reportados em t unidades de tempo;
- λ - frequência de sinistralidade.

O Estimador da Máxima Verosimilhança (EMV) do parâmetro deste processo é dado por:

$$\hat{\lambda} = \bar{\mathbf{A}} = \frac{1}{n} \sum_{i=1}^n \mathbf{A}_i$$

ou seja, pela média aritmética dos dados observados na amostra $\mathbf{A} = (\mathbf{A}_1, \dots, \mathbf{A}_n)$.

Conforme foi referido na Introdução, em Seguro Automóvel, o condutor, regra geral, tem influência na probabilidade de ocorrência de um sinistro, pelo que considerar que o número de sinistros segue um processo de Poisson Homogéneo é uma hipótese pouco realista, pois as diferenças no número de sinistros entre dois quaisquer segurados seria ditada meramente pelo acaso e nunca pelas características destes. Desta forma, o SBM aplicado pela seguradora serviria apenas para penalizar os condutores com má sorte.

Os processos de Poisson Homogéneos revelam-se, então, desadequados para ajustar o número de sinistros anual. Todavia, é possível utilizar modelos de Poisson cujo valor médio difira de segurado para segurado e que se ajustem melhor à realidade, designados como processos de Poisson Heterogéneos.

2.2 Processo de Poisson Heterogéneo (Carteira Heterogénea)

Nos processos de Poisson Heterogéneos, ou Mistos, considera-se que o risco é diferente de apólice para apólice, ou seja, que o número de sinistros segue um processo de

Poisson, mas o parâmetro envolvido é uma variável aleatória caracterizada por:

$$N(t) \sim P(\Lambda t), \Lambda \sim U(\lambda).$$

Assim, chama-se ao risco, representado por Λ , a variável de estrutura, e à sua distribuição, $U(\lambda)$, a distribuição de estrutura do processo Poissoniano.

A Função Geradora de Momentos (fgm) de um processo deste género é dada pela seguinte expressão:

$$M_{N(t)}(r) = M_{\Lambda}(t(e^r - 1)).$$

Seja $\{N(t), t \geq 0\}$ um processo de Poisson Misto que representa o número de sinistros reportados em t unidades de tempo. Este género de processos tem um conjunto de propriedades que se adequam à modelação do número de sinistros:

- $N(0) = 0$
(no momento em que se inicia a contagem dos sinistros, nenhum foi ainda reportado);
- $P[N(t+s) - N(t) = k] = \int_0^{+\infty} \frac{e^{-\lambda t} (\lambda t)^k}{k!} dU(\lambda)$
(os incrementos são estacionários, isto é, a probabilidade de ocorrência de um sinistro é igual quando se consideram intervalos de tempo de igual amplitude, não dependendo da sua localização específica no tempo);
- $E[N(t)] = t E[\Lambda]$
(o número médio de sinistros observados num período de tempo t é proporcional ao valor esperado da distribuição de estrutura, ou seja, à frequência de sinistralidade média esperada numa unidade de tempo);
- $\text{Var}(N(t)) = t E[\Lambda] + t^2 \text{Var}(\Lambda) > E[N(t)]$
(a variância do número de sinistros observados é maior do que o seu valor esperado).

As distribuições de estrutura que geralmente se adequam melhor (Grandell (1997)) são:

- Exponencial;
- Gama;
- Inversa Gaussiana;
- LogNormal;
- outras, com a condição de serem distribuições contínuas e que assumam valores não negativos, de preferência com assimetria positiva e, em muitos casos, cauda pesada.

As mais frequentemente utilizadas, porém, são a Gama e a Inversa Gaussiana, pelo que serão estas as estudadas nesta dissertação.

2.2.1 Distribuição Poisson - Gama

A função densidade de probabilidade de uma v.a. Λ com distribuição $G(\alpha, \beta)$ é:

$$u_{\Lambda}(\lambda) = \frac{1}{\Gamma(\alpha)} \beta^{\alpha} e^{-\beta\lambda} \lambda^{\alpha-1}, \quad \lambda, \alpha, \beta > 0$$

onde $\Gamma(\alpha)$ representa a conhecida função Gama, dada por:

$$\begin{aligned} \Gamma(\alpha) &= \int_0^{+\infty} t^{\alpha-1} e^{-t} dt \\ \Gamma(\alpha) &= (\alpha - 1) \Gamma(\alpha - 1) \underset{\alpha \in \mathbb{N}}{=} (\alpha - 1)! \end{aligned}$$

Desta forma, os dois primeiros momentos da v.a. $\Lambda \sim G(\alpha, \beta)$ são determinados, respectivamente, a partir de:

- $E[\Lambda] = \frac{\alpha}{\beta}$;
- $\text{Var}(\Lambda) = \frac{\alpha}{\beta^2}$.

O caso particular da distribuição de Poisson cujo parâmetro de estrutura é uma v.a. com distribuição Gama é, na realidade, uma distribuição Binomial Negativa cujos parâmetros dependem dos da v.a. com distribuição Gama. Esta distribuição também é conhecida como distribuição de Polya (ver, por exemplo, Parzen (1965), Ross (1996) ou Centeno (2003)).

Assim, se

$$N(t) \sim P(\Lambda t), \quad \Lambda \sim G(\alpha, \beta)$$

mostra-se facilmente que

$$N(t) \sim \text{BN}\left(\alpha, \frac{\beta}{\beta + t}\right)$$

ou seja,

$$f_{N(t)}(k) = \binom{\alpha + k - 1}{k} \left(\frac{\beta}{\beta + t}\right)^{\alpha} \left(\frac{t}{\beta + t}\right)^k$$

obtendo-se, portanto, os dois primeiros momentos e a Função Geradora de Momentos (fgm) da seguinte forma (ver, por exemplo, Centeno (2003)):

- $E[N(t)] = \frac{\alpha t}{\beta}$;

- $\text{Var}(N(t)) = \frac{\alpha t}{\beta} + \frac{\alpha t^2}{\beta^2} = \frac{\alpha t}{\beta} \left(1 + \frac{t}{\beta}\right) = \frac{\alpha t}{\beta} \left(\frac{\beta + t}{\beta}\right);$
- $M_{N(t)}(r) = \left(1 - \frac{t(e^r - 1)}{\beta}\right)^{-\alpha}.$

Os EMV de α e β são, respectivamente, obtidos como a solução das seguintes equações (Lemaire (1995)):

- $\sum_{k=0}^m n_k \left(\frac{1}{\hat{\alpha}} + \dots + \frac{1}{\hat{\alpha} + k - 1}\right) = \sum_{k=0}^m n_k \log \left(1 + \frac{\bar{\mathbf{A}}}{\hat{\alpha}}\right);$
- $\hat{\beta} = \frac{\alpha}{\bar{\mathbf{A}}}.$

Pelas expressões acima referidas, compreende-se por que esta é uma das distribuições de estrutura mais utilizada para modelar o número de sinistros: além de geralmente apresentar bons resultados, tem expressões simples para as funções necessárias para determinação dos EMV dos seus parâmetros, da distribuição estacionária e das escalas de prémios¹, requerendo, portanto, pouca exigência computacional.

2.2.2 Distribuição Poisson - Inversa Gaussiana

Segundo Lemaire (1995), nos casos em que a distribuição do número de sinistros é um Processo de Poisson Misto sendo Λ uma v.a. com distribuição $IG(g, h)$, a função densidade de probabilidade de Λ será dada por

$$u_{\Lambda}(\lambda) = \frac{g}{\sqrt{2\pi h \lambda^3}} \exp \left\{ -\frac{1}{2h\lambda} (\lambda - g)^2 \right\}, \quad g, h > 0.$$

Assim, os momentos de $\Lambda \sim IG(g, h)$ tomam os valores

- $E[\Lambda] = g;$
- $\text{Var}(\Lambda) = gh.$

Esta distribuição de Poisson Mista, designada por distribuição de Sichel, tem função de probabilidade dada recursivamente por

$$\begin{aligned} f_{N(t)}(0) &= \exp \left\{ \frac{g}{h} \left(1 - (1 + 2h)^{1/2}\right) \right\} \\ f_{N(t)}(1) &= g \cdot f_{N(t)}(0) (1 + 2h)^{-1/2} \\ f_{N(t)}(k) &= \frac{h(k-1)(2k-3) f_{N(t)}(k-1) + g^2 f_{N(t)}(k-2)}{(1+2h)k(k-1)}, \quad k = 2, 3, \dots \end{aligned} \quad (2.2)$$

Os seus dois primeiros momentos são obtidos da seguinte forma:

- $E[N(t)] = g;$

¹Os conceitos de distribuição estacionária e escalas de prémios de um SBM serão abordados no próximo capítulo.

- $\text{Var}(N(t)) = g(1 + h)$.

O EMV de g é, então,

$$\hat{g} = \bar{\mathbf{A}}$$

e o de h é a solução positiva de

$$\frac{1}{n} \sum_{k=0}^n (n_k T_k(h)) = \bar{\mathbf{A}}$$

onde:

- $T_0(h) = \bar{\mathbf{A}}(1 + 2h)^{-1/2}$;
- $T_k(h) = \frac{1}{1 + 2h} \left((2k - 1)h + \frac{\bar{\mathbf{A}}^2}{T_{k-1}(h)} \right)$, $1 \leq k \leq m$.

A expressão anterior pode escrever-se, de forma mais simples, como

$$T_k(h) = (k + 1) \frac{f_{N(t)}(k + 1)}{f_{N(t)}(k)}.$$

Apesar da função densidade de probabilidade e os EMV terem expressões de cálculo aparentemente complexas, uma v.a. Poisson Mista com distribuição de estrutura de Sichel normalmente também se adequa à modelação do número de sinistros de uma carteira (Guerreiro (2001)), pelo que é bastante utilizada nos estudos de SBM (pode ver-se ser usada, por exemplo, em Tremblay (1992) ou Morillo e Bermúdez (2003)).

2.3 Testes de Ajustamento

Qualquer que seja a distribuição que se pretenda utilizar, é necessário primeiramente realizar testes de ajustamento de cada distribuição aos dados disponíveis, não só para entender se é estatisticamente aceitável, mas também para estabelecer uma base de comparação entre as distribuições em alternativa e decidir qual será a mais adequada. Para o efeito, são, de seguida, apresentados dois testes frequentemente usados para o tipo de dados em estudo: o teste do Qui-Quadrado de Pearson e o teste da Razão de Verossimilhanças.

2.3.1 Teste do Qui-Quadrado de Pearson

Este teste serve para averiguar se os dados observados são coerentes com uma determinada distribuição teórica. É levado a cabo separando os dados em n classes de valores. É importante ter em atenção a densidade de cada classe. Represente-se por p_k a probabilidade, de acordo com a distribuição teórica a ajustar, de numa amostra aleatória se observar um valor da k -ésima classe. Caso existam classes para as quais o produto np_k

é inferior a 5, deve proceder-se ao agrupamento das mesmas até que o referido produto seja maior ou igual a 5.

No caso particular da aplicação deste teste aos seguro automóvel é sensato fazer representar em cada classe cada possibilidade para o número de sinistros reportados numa unidade de tempo (0 sinistros, 1 sinistro, ...), isto é, guardar na classe n_k o número de apólices que reportaram k sinistros. Se for necessário agrupar classes, faz sentido que sejam aquelas associadas a uma maior frequência de sinistralidade (maiores valores de k), pois geralmente é aí que se verificam poucas observações.

Formalmente, o teste é representado por:

$$H_0 : N \sim F(\theta) \text{ vs. } H_1 : N \not\sim F(\theta)$$

em que:

- F - distribuição cujo ajustamento se pretende testar;
- θ - parâmetros da distribuição F .

Para um nível de significância de $(1 - \alpha) \times 100\%$, a estatística-teste é dada pela expressão

$$\chi_{obs}^2 = \sum_{k=1}^n \frac{(O_k - E_k)^2}{E_k}$$

onde:

- O_k - frequência observada na classe k ;
- E_k - frequência esperada na classe k , de acordo com a distribuição presente na hipótese nula.

Deve-se rejeitar H_0 , isto é, deve-se rejeitar que o número de sinistros seja modelado pela distribuição $F(\theta)$ quando o valor-p de χ_{obs}^2 seja menor do que α .

2.3.2 Teste da Razão de Verossimilhanças

Apesar de o teste anterior ser suficiente para concluir estatisticamente acerca da distribuição a ajustar, o teste da Razão de Verossimilhanças é mais robusto, pois indica o benefício obtido em considerar uma distribuição de Poisson Mista no lugar de uma distribuição de Poisson Homogénea, isto é, de utilizar uma distribuição com um número maior de parâmetros (Lemaire (1995), Guerreiro (2001)).

Represente-se por Ω o espaço dos parâmetros da distribuição Poisson Mista considerada e por ω um subconjunto deste espaço que contenha, como caso limite, a distribuição Poisson Homogénea. Seja \mathcal{L} a função de verossimilhança da função de distribuição conjunta dos dados:

$$\mathcal{L} = \prod_{k=1}^n (p_k)^{n_k}$$

onde:

- n - número de classes;
- p_k - probabilidade, de acordo com a distribuição teórica a ajustar, de numa amostra aleatória se observar um valor da k -ésima classe;
- n_k - número de observações na k -ésima classe.

Define-se como Razão de Verosimilhanças o seguinte rácio:

$$K = \frac{\max_{\omega} \{\mathcal{L}\}}{\max_{\Omega} \{\mathcal{L}\}}$$

onde:

- $\max_{\omega} \{\mathcal{L}\}$ é obtido utilizando as estimativas de máxima verosimilhança da distribuição Poisson Homogénea;
- $\max_{\Omega} \{\mathcal{L}\}$ é obtido utilizando as estimativas de máxima verosimilhança da distribuição Poisson Mista.

Se o valor de K for próximo de 1 então a distribuição de Poisson Heterogénea não acrescenta informação. Caso o valor de K seja baixo, então a distribuição de Poisson Mista é mais adequada do que a distribuição de Poisson Homogénea.

Caso seja mais fácil realizar os cálculos, também é possível utilizar a função de log-verosimilhança no lugar da função de verosimilhança:

$$\log K = -\log \max_{\omega} \{\mathcal{L}\} - \left(-\log \max_{\Omega} \{\mathcal{L}\} \right).$$

As equações (2.3) a (2.5) representam, respectivamente, as funções de log-verosimilhança das distribuições Poisson, Polya e Sichel (ver, por exemplo, Centeno (2003) e Guerreiro (2001)), que serão usadas aquando do cálculo deste rácio:

$$\bullet \quad -\log(\mathcal{L}(\lambda)) = \lambda \sum_{j=0}^m n_j - \log(\lambda) \sum_{j=0}^m j n_j + \sum_{j=0}^m n_j \log(j!) \quad (2.3)$$

$$\bullet \quad -\log(\mathcal{L}(\alpha, \beta)) = -\sum_{j=0}^m n_j \log\left(\binom{j+\alpha-1}{j}\right) - \alpha \log(p) \sum_{j=0}^m n_j - \log(1-p) \sum_{j=0}^m j n_j \quad (2.4)$$

$$\bullet \quad -\log(\mathcal{L}(g, h)) = -\frac{g}{h} \left(1 - \sqrt{1+2h}\right) \sum_{j=0}^m n_j - \log(g) \sum_{j=1}^m n_j + \log(1+2h) \sum_{j=1}^m \frac{n_j}{2} - \sum_{j=2}^m n_j \log(r_j) + \sum_{j=0}^m n_j \log(j!) \quad (2.5)$$

onde:

- $r_2 = h(1+2h)^{-1/2} + g$
- $r_3 = 3h(1+2h)^{-1} + 3gh(1+2h)^{-1/2} + g^2$
- $r_4 = 15h^3(1+2h)^{-3/2} + 15gh^2(1+2h)^{-1} + 6g^2h(1+2h)^{-1/2} + g^3$
- $r_5 = 105h^4(1+2h)^{-2} + 105gh^3(1+2h)^{-3/2} + 45g^2h^2(1+2h)^{-1} + 10g^3h(1+2h)^{-1/2} + g^4$
- $r_6 = 945h^5(1+2h)^{-5/2} + 945gh^4(1+2h)^{-2} + 420g^2h^3(1+2h)^{-3/2} + 105g^3h^2(1+2h)^{-1} + 15g^4h(1+2h)^{-1/2} + g^5$
- $r_7 = 10395h^6(1+2h)^{-3} + 10395gh^5(1+2h)^{-5/2} + 4725g^2h^4(1+2h)^{-2} + 1260g^3h^3(1+2h)^{-3/2} + 210g^4h^2(1+2h)^{-1} + 21g^5h(1+2h)^{-1/2} + g^6$
- ...

em que os r_k são obtidos como parte da expressão da função densidade de probabilidade da distribuição Poisson - Inversa Gaussiana. Com efeito, calculando $f_{N(t)}(k)$ através da equação (2.2), e escrevendo o resultado na forma (Guerreiro (2001))

$$f_{N(t)}(k) = \frac{1}{k!} g f_{N(t)}(0) (1+2h)^{-k/2} r_k, \quad k = 2, 3, \dots$$

determinam-se, então, os r_k por reciprocidade como

$$r_k = \frac{1}{g} \frac{k! f_{N(t)}(k)}{f_{N(t)}(0)} (1+2h)^{k/2}, \quad k = 2, 3, \dots$$



Sistemas de *Bonus Malus*

Os Sistemas de *Bonus Malus* (SBM) são uma técnica de tarifação *a posteriori* baseada em cadeias de Markov, que visa a movimentação das apólices entre classes de *bonus*. Com o conhecimento do número de sinistros reportados pelos segurados em cada unidade de tempo, as cadeias de Markov permitem fazer o ajustamento do prémio que deve ser pago, através das regras de transição entre classes, ou seja, o prémio que os segurados pagam em cada classe é uma percentagem fixa do prémio puro determinado *a priori*. Estas unidades de tempo devem ser de igual duração, tendo normalmente o prazo de um ano. Durante estes períodos, as apólices estão distribuídas por s classes de *bonus* e as transições entre classes coincidem com a transição de período, ou seja, só ocorrem no final de cada ano.

Na maioria dos países, incluindo Portugal, a classe para a qual uma apólice transita depende exclusivamente da classe onde se encontrava no período findo e do número de sinistros declarados nesse mesmo período, ignorando o histórico anterior (Lemaire (1995)). Noutros países, as transições podem depender também da severidade dos sinistros reportados (ver, por exemplo, Pinquet (1997)). Outros casos há em que se pode ter em conta um histórico mais alargado ou até em que as transições ocorrem só decorridos vários períodos, em lugar de apenas um, como é o caso da Bélgica (Lemaire (1995)). No entanto, estas situações exigem uma formulação matemática mais complexa e que está fora do âmbito desta dissertação, dado que nos iremos debruçar sobre um caso geral e não situações particulares ou especificidades.

No estudo teórico dos SBM deve-se ter o cuidado de filtrar os sinistros observados apenas a veículos de passageiros domésticos, ou seja, de excluir carros de aluguer, motorizadas, transportes públicos e outros veículos especiais, como carros de bombeiros ou de polícia ou ambulâncias, dado que a sinistralidade para estes tipos de veículos é, regra

geral, diferente (Lemaire (1995)).

3.1 Características

Um SBM com s classes caracteriza-se através de três elementos (ver, por exemplo, Centeno (2003)):

- uma escala de prémios, $\mathbf{b} = (b_1, b_2, \dots, b_s)$, ordenados de acordo com o risco associado à apólice. Representam a percentagem do prémio puro a pagar pelos segurados que se encontram na respectiva classe de *bonus*;
- uma classe de entrada, C_{i_0} , para a qual entram as novas apólices e onde os segurados pagam o valor correspondente ao Prémio Puro, sem descontos nem agravamentos. Apesar de não ser o que se pratica na realidade, assume-se, para efeitos teóricos, que esta classe é única, isto é, que a classe de entrada é a mesma para todas as novas apólices, independentemente do conhecimento que a Seguradora tem do tomador do seguro ou alguma eventual campanha que tenha em vigor;
- um conjunto de regras de transição que explicitam para que classe as apólices devem transitar na anuidade seguinte, dependendo da classe em que se encontram na anuidade actual e do número de sinistros reportados desde o início da mesma.

As regras de transição devem ser suficientemente claras e completas no que respeita a quem se aplicam e em que casos. Normalmente estão organizadas numa matriz

$$\mathbf{T} = [t_{ij}]_{s \times s}, \quad i, j \in \{1, \dots, s\} \quad (3.1)$$

de onde se depreende que um segurado que esteja actualmente na classe C_i deve transitar para a classe C_j caso reporte t_{ij} sinistros durante uma anuidade. Alternativamente, a matriz de regras de transição também pode ser definida por

$$\mathbf{T} = [t_{ij}(k)]_{s \times s}, \quad i, j \in \{1, \dots, s\}, k \in \mathbb{N}_0, \quad (3.2)$$

onde

$$t_{ij}(k) = \begin{cases} 1, & \text{se } k \text{ sinistros originam uma transição da classe } C_i \text{ para a classe } C_j \\ 0, & \text{caso contrário} \end{cases}.$$

De forma que não haja ambiguidade na transição a efectuar, deve-se certificar de que para cada par (i, k) ($i \in \{1, \dots, s\}, k \in \mathbb{N}_0$) se verifica que $t_{ij}(k) = 1$ para um e um só valor de j ($j \in \{1, \dots, s\}$). Caso contrário, ou todos os $t_{ij}(k)$ seriam nulos, o que não faz sentido, porque se não existe transição para uma classe diferente da actual, então existe uma transição para a mesma classe (uma transição degenerada), que implicaria

$t_{ii}(k) = 1$, ou haveria mais do que um valor de j para o qual $t_{ij}(k) = 1$, o que significa que a apólice transitaria simultaneamente para duas ou mais classes distintas.

Por último, as regras de transição devem ser construídas de modo a não ir contra o objectivo principal dos SBM: bonificar o prémio dos segurados que não originam sinistros e agravar o daqueles que originam, incentivando os segurados, em geral, a uma condução mais cuidadosa e, conseqüentemente, a uma menor ocorrência de sinistros.

3.2 Matrizes de Probabilidades de Transição

A partir da matriz de regras de transição, \mathbf{T} , estima-se outra matriz de extrema importância para os SBM: a matriz de probabilidades de transição (a um passo),

$$\mathbf{P}_{\mathbf{T},\lambda} = [p_{\mathbf{T},\lambda}(i,j)]_{s \times s},$$

que traduz a probabilidade de um dado segurado, com frequência de sinistralidade λ , que se encontra numa dada classe, transitar para qualquer uma das classes de *bonus* na anuidade seguinte. Estas probabilidades baseiam-se na distribuição do número de sinistros.

Matriz de Probabilidades de Transição a um passo

Represente-se por $Z_{S,m}$ a v.a. que representa a classe de *bonus* de uma determinada apólice no período m . A probabilidade dessa apólice transitar da classe C_i para a classe C_j na anuidade seguinte, tendo em conta que a sua frequência de sinistralidade é λ , é dada por (ver, por exemplo, Centeno (2003)):

$$p_{\mathbf{T},\lambda}(i,j) = \sum_{k=0}^{+\infty} P[N(t) = k] \cdot t_{ij}(k) = P[Z_{S,m+1} = j | Z_{S,m} = i, \Lambda = \lambda], \quad i, j = 1, \dots, s.$$

Matriz de Probabilidades de Transição a n passos

Para determinar a probabilidade de uma apólice, estando actualmente na classe C_i , transitar para a classe C_j ao fim de n anuidades, basta proceder a um cálculo análogo ao anterior:

$$p_{\mathbf{T},\lambda}^{(n)}(i,j) = P[Z_{S,m+n} = j | Z_{S,m} = i, \Lambda = \lambda], \quad i, j = 1, \dots, s, \quad (3.3)$$

onde a matriz de probabilidades de transição que descreve a probabilidade de uma apólice da classe i se situar na classe j daqui a n unidades de tempo, designada por matriz de probabilidades de transição a n passos, é dada por

$$\mathbf{P}_{\mathbf{T},\lambda}^{(n)} = \mathbf{P}_{\mathbf{T},\lambda}^n.$$

Se, por algum motivo, esta cadeia de Markov não for irredutível, então é necessário tomar alguma cautela, pois deverá escrever-se, nesse caso, a matriz de probabilidades de transição a um passo como uma matriz por blocos (Centeno (2003)) tal que as linhas que representam as m classes de estado transiente apareçam em primeiro lugar. Consegue representar-se da seguinte forma:

$$\mathbf{P}_{\mathbf{T},\lambda} = \begin{bmatrix} \mathbf{P}_{1,(\mathbf{T},\lambda)} & \mathbf{P}_{2,(\mathbf{T},\lambda)} \\ \mathbf{0} & \mathbf{P}_{3,(\mathbf{T},\lambda)} \end{bmatrix}$$

onde:

- $\mathbf{P}_{1,(\mathbf{T},\lambda)}$ - matriz quadrada de ordem m que representa as probabilidades de transição entre as classes de estado transiente;
- $\mathbf{P}_{2,(\mathbf{T},\lambda)}$ - matriz quadrada de ordem $(s - m)$ que representa as probabilidades de transição das classes de estado transiente para as classes de estado recorrente;
- $\mathbf{0}$ - matriz nula de dimensão $(s - m) \times m$ que representa as probabilidades de transição das classes de estado recorrente para as classes de estado transiente;
- $\mathbf{P}_{3,(\mathbf{T},\lambda)}$ - matriz de dimensão $m \times (s - m)$ que representa as probabilidades de transição entre as classes de estado recorrente.

Desta forma, prova-se que

$$\mathbf{P}_{\mathbf{T},\lambda}^{(n)} = \mathbf{P}_{\mathbf{T},\lambda}^n = \begin{bmatrix} \mathbf{P}_{1,(\mathbf{T},\lambda)}^{(n)} & (\mathbf{P}_{1,(\mathbf{T},\lambda)}\mathbf{P}_{2,(\mathbf{T},\lambda)} + \mathbf{P}_{2,(\mathbf{T},\lambda)}\mathbf{P}_{3,(\mathbf{T},\lambda)})\mathbf{P}_{2,(\mathbf{T},\lambda)}^{(n-2)} \\ \mathbf{0} & \mathbf{P}_{3,(\mathbf{T},\lambda)}^{(n)} \end{bmatrix}. \quad (3.4)$$

A correspondente probabilidade incondicional (probabilidade retirando uma apólice ao acaso da carteira) é dada por

$$p_{\mathbf{T}}^{(n)}(i,j) = \int_0^{+\infty} p_{\mathbf{T},\lambda}^{(n)}(i,j) dU(\lambda), \quad i, j = 1, \dots, s,$$

em que $U(\lambda)$, tal como referido no Capítulo 2, corresponde à distribuição da frequência de sinistralidade de um segurado retirado ao acaso da carteira.

3.3 Distribuição Estacionária da Carteira

Caso as regras de transição estejam bem definidas, isto é, se o processo de Markov expresso pela matriz $\mathbf{P}_{\mathbf{T},\lambda}$ for irredutível, finito e aperiódico, então o limite

$$\pi_{\mathbf{T},\lambda}(j) = \lim_{n \rightarrow +\infty} p_{\mathbf{T},\lambda}^{(n)}(i,j), \quad i, j = 1, \dots, s$$

existe e não depende da classe inicial da apólice. Esta distribuição-limite coincide com a distribuição estacionária da carteira, isto é, com a distribuição das apólices pelas classes

de *bonus* passado tempo suficiente para estabilizar.

Caso o processo de Markov não apresente estas características, não é garantido que a distribuição-limite exista. No entanto, isso não implica que não exista distribuição estacionária. Nestas situações, a distribuição estacionária terá de ser determinada por definição, ou seja, como a única solução não negativa do seguinte sistema de equações:

$$\begin{cases} \pi_{\mathbf{T},\lambda} = \pi_{\mathbf{T},\lambda} \mathbf{P}_{\mathbf{T},\lambda} \\ \sum_{j=1}^s \pi_{\mathbf{T},\lambda}(j) = 1 \end{cases}.$$

No entanto, se a condição a ser violada for a da irredutibilidade, a distribuição estacionária da carteira deverá ser obtida recorrendo à matriz de regras de transição apresentada em (3.4), tendo em conta que

$$\lim_{n \rightarrow +\infty} \mathbf{P}_{\mathbf{T},\lambda}^{(n)} = \begin{bmatrix} \mathbf{0} & (\mathbf{P}_{1,(\mathbf{T},\lambda)} \mathbf{P}_{2,(\mathbf{T},\lambda)} + \mathbf{P}_{2,(\mathbf{T},\lambda)} \mathbf{P}_{3,(\mathbf{T},\lambda)}) \lim_{n \rightarrow +\infty} \mathbf{P}_{2,(\mathbf{T},\lambda)}^{(n-2)} \\ \mathbf{0} & \lim_{n \rightarrow +\infty} \mathbf{P}_{3,(\mathbf{T},\lambda)}^{(n)} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \pi_{3,(\mathbf{T},\lambda)} \end{bmatrix}.$$

Assim, arrançando a matriz novamente, de forma a que as classes tenham a ordenação inicial, obtém-se a distribuição estacionária da carteira.

A distribuição estacionária incondicional é obtida a partir de

$$\pi_{\mathbf{T}}(j) = \int_0^{+\infty} \pi_{\mathbf{T},\lambda}(j) dU(\lambda), \quad j = 1, \dots, s$$

e representa a probabilidade de um segurado retirado ao acaso da carteira estar na classe C_j ($j \in \{1, \dots, s\}$) quando a carteira apresentar um comportamento estacionário.

Por vezes não é fácil ou sequer possível calcular analiticamente o valor exacto destes integrais, pelo que uma estratégia para resolver este problema é a discretização da distribuição de estrutura, calculando a distribuição estacionária de acordo com a equação seguinte (Centeno (2003)):

$$\pi_{\mathbf{T}}(j) = \sum_{\lambda} \pi_{\mathbf{T},\lambda}(j) u_{\lambda}^d, \quad j = 1, \dots, s$$

onde u_{λ}^d representa o vector das probabilidades da distribuição $U(\lambda)$ discretizada.

3.4 Distribuição Ponderada da Carteira

A distribuição ponderada da carteira consiste em atribuir pesos não negativos, w_k , a cada matriz de probabilidades de transição a k passos, com $k \in \{1, 2, \dots, n\}$, definindo-se uma estrutura de importância do estado da carteira em cada ano k , durante n anos.

Normalmente o valor de n é suficientemente grande que permita à distribuição identificar-se com o real estado da carteira a médio prazo, mas não tanto que perca significado

prático (não se espera que os segurados se mantenham na carteira durante 50 anos, por exemplo).

Os ponderadores podem ser determinados de diferentes formas:

- atribuindo-se a cada ponderador w_k a frequência relativa de apólices com k anos de vigência (sugestão efectuada em Borgan et al. (1981) e Asmussen (2014));
- recorrendo-se ao sistema de equações seguinte (sugestão efectuada em Centeno e Andrade e Silva (2001)):

$$\begin{cases} w_k = \frac{w_{k-1}}{1+i}, & k \in \{2, \dots, n\} \\ \sum_{k=0}^n w_k = 1 \end{cases} \quad (3.5)$$

onde i representa como que uma taxa de juro entre anos consecutivos, cujo objectivo é reduzir a importância, na distribuição ponderada, de cada ano sucessivamente mais afastado do presente e sendo w_0 previamente definido;

- como uma mistura dos dois métodos anteriores;
- através de outros métodos que, actuarialmente, façam sentido.

É importante notar que a ponderação atribuída a w_0 deve ser interpretada como o peso da distribuição estacionária para a distribuição ponderada. Assim, quando a carteira tem uma convergência lenta para a estacionaridade, ou se o número de anos com ponderadores não nulos incluem o período em que se verifica que a distribuição do número de sinistros já estacionou, é usual definir-se $w_0 = 0$.

As distribuições ponderada $\rho_{\mathbf{T},\lambda}$ e ponderada incondicional $\rho_{\mathbf{T}}$ são, portanto, determinadas através das igualdades (3.6) e (3.7), respectivamente:

$$\rho_{\mathbf{T},\lambda}(j) = w_0 \cdot \pi_{\mathbf{T},\lambda}(j) + \sum_{k=1}^n w_k \cdot p_{\mathbf{T},\lambda}^{(k)}(j), \quad j = 1, \dots, s \quad (3.6)$$

onde $p_{\mathbf{T},\lambda}^{(k)}(j) = P[Z_{S,k} = j | \Lambda = \lambda]$, e

$$\rho_{\mathbf{T}}(j) = \int_0^{+\infty} \rho_{\mathbf{T},\lambda}(j) dU(\lambda), \quad j = 1, \dots, s. \quad (3.7)$$

3.5 Escalas Óptimas de Prémios

Como já foi referido anteriormente, a determinação dos prémios a cobrar aos segurados é um factor de extrema importância para qualquer empresa que opere na área dos seguros. Assim, foram propostos vários métodos para obter escalas de prémios que se adequem às necessidades de cada SBM, tendo em conta as suas restantes características (Centeno (2003)). Estas escalas são designadas por *escalas óptimas de prémios*. As mais

usualmente utilizadas serão apresentadas de seguida, juntamente com as respectivas medidas de eficiência, para que possam ser posteriormente comparadas.

Considere-se uma medida de eficiência $e(\cdot)$. Diz-se que uma escala de prémios T é mais eficiente do que uma escala \tilde{T} (e representa-se por $T \succ \tilde{T}$) se e só se $e(T) > e(\tilde{T})$.

Entenda-se que os valores obtidos pelas medidas de eficiência não têm significado por si só. Servem apenas para comparar diferentes escalas de prémios.

3.5.1 Escala de Norberg - Modelo Assimptótico

Segundo Norberg (1976), a escala de Norberg consiste em determinar o prémio que deverá ser cobrado às apólices em cada classe de *bonus* quando se atinge a estacionaridade da carteira, através da aplicação da estatística Bayesiana e de uma abordagem semelhante à utilizada em Teoria da Credibilidade. Esse prémio é o resultado da minimização do Erro Quadrático Médio,

$$\begin{aligned} Q_n(S) &= E [E [S_n|\lambda] - b_n(Z_{S,n})]^2 = \int_0^{+\infty} (E [S_n|\lambda] - b_n(Z_{S,n}))^2 dU(\lambda) = \\ &= \int_0^{+\infty} \sum_{j=1}^s (E [S_n|\lambda] - b_n(j))^2 \cdot p_{S,\lambda}^{(n)}(j) dU(\lambda) \end{aligned}$$

onde:

- $E [S_n|\lambda]$ - prémio que se espera cobrar a um segurado com frequência de sinistralidade λ ;
- $b_n(Z_{S,n})$ - estimador Bayesiano de $E [S_n|\lambda]$, designado por prémio de credibilidade, referente ao n -ésimo período da permanência da apólice na carteira;
- $E [b_n(Z_{S,n})] = E [E [S_n|\lambda]] = E [S_n]$ - em média, o prémio em credibilidade coincide com o valor das indemnizações agregadas;
- $p_{S,\lambda}^{(n)}(j) = P[Z_{S,n} = j|\lambda]$ - probabilidade de um segurado com frequência de sinistralidade λ se encontrar na classe C_j ao fim de n anos.

Uma vez que o prémio assim calculado depende do momento no tempo, n , Norberg propôs que se deveria assumir a estacionaridade, ou seja, que os prémios da carteira já não variam. Com este pressuposto, o objectivo é, agora, minimizar

$$Q_0(S) = E \left[(E [S^*|\lambda] - b(Z_T))^2 \right] = \int_0^{+\infty} \sum_{j=1}^s (E [S^*|\lambda] - b_T(j))^2 \cdot \pi_{T,\lambda}(j) dU(\lambda) \quad (3.8)$$

onde:

- $S^* \stackrel{iid}{\sim} S_n$;
- $b(Z_T) = E [E [S_n|Z_T]] = E [E [S^*|Z_T]] = E [S^*]$.

Desta forma, igualando a derivada da equação (3.8) a zero, obtém-se a escala óptima de prémios de Norberg:

$$\begin{aligned} b^N(j) &= \frac{1}{\pi_T(j)} \int_0^{+\infty} \sum_{j=1}^s E[S^*|\lambda] \cdot \pi_{T,\lambda}(j) dU(\lambda) = \\ &= \frac{1}{\pi_T(j)} \int_0^{+\infty} \sum_{j=1}^s \lambda \cdot \pi_{T,\lambda}(j) dU(\lambda), \quad j = 1, \dots, s. \end{aligned}$$

Note-se que se pode estabelecer a segunda igualdade dados os dois pressupostos assumidos no Capítulo 2:

- S_n depende de λ apenas através da distribuição do número de sinistros;
- o custo médio com sinistros é de 1 u.m.,

pelo que $E[S^*|\lambda] = E[X] \cdot E[N|\lambda] = 1 \cdot \lambda = \lambda$.

A medida de eficiência da escala de Norberg é dada pela expressão seguinte:

$$e(T) = E[b_T^2(Z_T)] = \sum_{j=1}^s b_T^2(j) \pi_T(j).$$

É ainda importante realçar que, por este critério ser assintótico, se a convergência do SBM para a estacionaridade for lenta, então os segurados com uma pequena frequência de sinistralidade poderão estar a pagar prémios demasiado elevados durante muito tempo e os segurados com frequência de sinistralidade alta a pagar prémios demasiado baixos. Esta situação vai contra o pressuposto de que o prémio a cobrar a cada apólice deve ser correspondente ao risco que esta representa para a seguradora.

3.5.2 Escala de Borgan, Hoem e Norberg - Modelo Não Assintótico

Segundo Borgan et al. (1981), esta escala surge como uma generalização da anterior. Considerando a distribuição da carteira ao longo de vários anos de observações, obtém-se, no final, uma distribuição ponderada conforme descrito na Secção 3.4. Para tal, é necessário definir um conjunto de ponderadores, w_n ($n \geq 1$), que representam o peso da distribuição da carteira ao fim de n anos, sendo w_0 o ponderador afecto à distribuição estacionária.

Ao contrário do que acontece com a escala de Norberg (1976), que é um caso particular desta, no qual o ponderador da distribuição estacionária assume o valor 1 e todos os restantes são nulos, esta escala não depende exclusivamente da distribuição estacionária da carteira, como os valores atribuídos aos ponderadores permitem conferir.

Assim, a função a minimizar é, agora, descrita por

$$Q(S) = \sum_{n=0}^{+\infty} w_n Q_n(S) = \sum_{n=0}^{+\infty} w_n \int_0^{+\infty} \sum_{j=1}^s (E[S^*|\lambda] - b(j))^2 \cdot p_{S,\lambda}^{(n)}(j) dU(\lambda) \quad (3.9)$$

onde $p_{S,\lambda}^{(0)}(j) = \pi_{T,\lambda}(j)$, com $j = 1, \dots, s$.

O seu minimizante, que se traduz como a escala óptima de prémios de Borgan, Hoem e Norberg, é dado por

$$b^B(j) = \frac{1}{p_S(j)} \sum_{n=0}^{+\infty} w_n \int_0^{+\infty} E[S_n|\lambda] \cdot p_{S,\lambda}^{(n)}(j) dU(\lambda), \quad j = 1, \dots, s$$

onde $p_S(j) = \sum_{n=0}^{+\infty} w_n \int_0^{+\infty} p_{S,\lambda}^{(n)}(j) dU(\lambda)$, com $j = 1, \dots, s$, é a distribuição ponderada.

A medida de eficiência desta escala expressa-se por

$$e(T) = \sum_{j=1}^s b_B^2(j) p_S(j).$$

Os autores da escala chamam a atenção para o facto de, por vezes, os valores obtidos através das escalas até agora apresentadas poderem não estar ordenados de acordo com a gravidade da classe de *bonus*, isto é, é possível que surjam apólices numa classe com risco maior com um prémio inferior ao de apólices numa classe com risco menor, o que vai contra o princípio básico dos SBM referido na Secção 3.1. A próxima escala apresenta uma solução para este problema.

3.5.3 Escala de Gilde e Sundt - Modelo Linear

A última escala que aqui se apresenta foi introduzida por Gilde e Sundt (1989) e é conhecida pelo nome dos autores, mas também por *A Escala Linear*, pois a proposta dos autores passa por uma linearização das escalas de Norberg (1976) e de Borgan et al. (1981). Desta forma, garante-se que à medida que vai aumentando o risco associado a uma classe, o correspondente prémio a cobrar será também superior.

A escala óptima é obtida, agora, minimizando a equação (3.9) sujeita ao seguinte conjunto de restrições:

$$b(j) = k + mj, \quad j = 1, \dots, s$$

onde m representa a diferença (constante) no prémio entre duas classes de *bonus* consecutivas.

Ao adicionar a informação da função a minimizar às restrições, obtém-se a seguinte função simplificada (Gilde e Sundt (1989)):

$$\begin{aligned} Q(S) &= \sum_{n=0}^{+\infty} w_n \int_0^{+\infty} \sum_{j=1}^s E^2[S_n|\lambda] dU(\lambda) + \sum_{j=1}^s (k + mj)^2 \cdot p_{S,\lambda}(j) \\ &- 2 \sum_{j=1}^s (k + mj) \cdot b^B(j) \cdot p_{S,\lambda}(j). \end{aligned}$$

Derivando e igualando a zero, conclui-se que o mínimo é obtido para o par (k, m) tal que (Gilde e Sundt (1989)):

$$\begin{cases} k = \sum_{j=1}^s b^B(j) \cdot p_S - m \sum_{j=1}^s j \cdot p_S(j) \\ m = \frac{\sum_{j=1}^s j \cdot b^B(j) \cdot p_S - \sum_{j=1}^s j \cdot p_S(j) \sum_{j=1}^s b^B(j) \cdot p_S(j)}{\sum_{j=1}^s j^2 \cdot p_S(j) - \left(\sum_{j=1}^s j \cdot p_S(j) \right)^2} \end{cases} .$$

A introdução de restrições em relação à escala anterior causa uma perda de eficiência, que é expressa através do rácio dos EQM das duas escalas,

$$\frac{e(T)}{e(\tilde{T})} .$$

onde:

- $e(T)$ é a eficiência da escala de Borgan, Hoem e Norberg;
- $e(\tilde{T})$ é a eficiência da escala de Gilde e Sundt.

Se o valor do rácio for próximo da unidade e a escala evoluir de acordo com a propensão ao risco, então as restrições podem ser introduzidas no modelo sem que haja perda de qualidade nas estimativas obtidas para os prémios.

Toda a exposição descrita até a este ponto refere-se ao modelo tradicional de SBM de evolução das apólices e respectivas escalas óptimas. Uma vez que o objectivo desta dissertação é estabelecer uma ligação entre este modelo e um modelo de Simulação, o capítulo seguinte tratará de apresentar algumas noções e algoritmos utilizados pela Simulação que são de grande utilidade para os resultados que se pretendem obter.

4

Simulação Discreta

A Simulação é um processo computacional que utiliza conceitos elementares de probabilidades e estatística para reproduzir um sistema real. No entanto, os resultados obtidos devem ser criteriosamente analisados, pois a Simulação não representa uma realidade, mas sim uma aproximação da realidade: existirá alguma distância a um sistema real devido à sua componente estocástica (van der Aalst e Voorhoeve (2000)).

Os problemas associados à aleatoriedade da Simulação podem ser amenizados fazendo crescer o tempo de simulação (ou número de simulações), dado que, de acordo com a Lei dos Grandes Números, quanto mais dados houver, mais rapidamente se observa a convergência para os resultados reais (van der Aalst e Voorhoeve (2000)).

Um dos ramos da Simulação, que tem especial interesse para o tópico a desenvolver nesta dissertação, designa-se por Simulação Discreta. Esta envolve processos úteis para a observação do comportamento de fenómenos como filas de espera ou processos de nascimento e morte, uma vez que as ocorrências dão-se em pontos discretos no tempo (Taha (2007)).

Recorde-se que o objectivo da presente dissertação é precisamente implementar um sistema o mais próximo possível da realidade, que pode ser “observado” através de Simulação. Visto que a variável de interesse, o número de sinistros reportados num ano numa carteira de seguros automóvel, se trata de um processo de contagem, este sistema pode ser modelado através de Simulação Discreta. Será, portanto, necessário gerar amostras de apólices, as quais terão associado um dado número de sinistros reportados.

A geração de amostras pode ser efectuada através de diversos métodos. Estes baseiam-se na determinação e utilização de Números Pseudo-Aleatórios (NPAs) gerados a partir de uma variável aleatória (v.a.) U , que segue uma distribuição Uniforme no intervalo $[0, 1]$ (doravante representada por $U_{[0,1]}$).

4.1 Geração de NPAs

Idealmente, a geração dos números a utilizar nos modelos de Simulação deveria ser verdadeiramente aleatória. Todavia, os computadores comuns não têm capacidade suficiente para obter números aleatórios em tempo útil, além de que não seriam adequados para detectar erros ou para replicar resultados, devido à sua aleatoriedade intrínseca, isto é, não possibilitariam a comparação entre diferentes corridas de simulação (Taha (2007)).

Assim, tornou-se necessária a criação de algoritmos que gerem números o mais aleatórios possível. Estes algoritmos são, por definição, determinísticos, o que significa que, na realidade, produzem ciclos, ou seja, os valores gerados produzem uma sequência finita de valores (van der Aalst e Voorhoeve (2000)). Assim, dá-se a estes valores a designação de Números Pseudo-Aleatórios pois não são verdadeiramente aleatórios.

4.1.1 Método Congruencial Misto

Um dos algoritmos mais simples e utilizados designa-se por Método Congruencial Misto (MCM). Neste, os NPAs são gerados através de operações aritméticas modulares.

O MCM necessita de um valor inicial, Y_0 , denominado *semente*, e de três parâmetros, b , c e m , os quais devem ser números inteiros não negativos. Os NPAs são, então, determinados através da sucessão (4.1) (Leemis e Park (2006)):

$$Y_i = (bY_{i-1} + c) \pmod{m}, \quad i \in \mathbb{N} \quad (4.1)$$

onde:

- b - parâmetro multiplicativo;
- c - parâmetro aditivo;
- $(bY_{i-1} + c) \pmod{m}$ - resto da divisão inteira de $(bY_{i-1} + c)$ por m .

Segundo Banks et al. (1998), as propriedades seguintes deverão ser verificadas de modo a ser produzida uma sequência com qualidade:

- $0 < b, c \leq m$;
- $m = 2^M - 1$ ou $m = 2^M$,

onde M é a precisão da máquina (8, 16, 32 ou 64 bits). Máquinas com diferentes precisões terão diferentes valores óptimos para a , c e m , de forma que a periodicidade do conjunto de valores gerados seja a maior possível, idealmente próxima do valor de m . Ao seleccionar valores nestas condições para os parâmetros, assegura-se que a sequência de NPAs gerada é o mais aleatória possível. Se $c = 0$, o método denomina-se Método Congruencial Multiplicativo pois apenas goza da operação de multiplicação (Banks et al. (1998)).

Uma vez que interessa que os NPAs provenham de uma distribuição $U_{[0,1]}$, os NPA terão de ser reescalados, de acordo com a equação (4.2).

$$U_i = \frac{Y_i}{m}, \quad i \in \mathbb{N}_0. \quad (4.2)$$

Desta forma, garante-se que os U_i ($i \in \mathbb{N}_0$) representam NPAs com distribuição $U_{[0,1]}$. O procedimento para determinação de NPAs através do MCM apresenta-se no Algoritmo 1.

Algoritmo 1 : Método Congruencial Misto

Input: b - parâmetro multiplicativo

Input: c - parâmetro aditivo

Input: m - módulo

Input: dim - dimensão da amostra

Input: Y_0 - semente

1: $U_0 = \frac{Y_0}{m}$

2: $NPA \leftarrow U_0$

3: $i \leftarrow 1$

4: **while** $i < dim$ **do**

5: $Y_i \leftarrow (b \cdot Y_{i-1} + c) \bmod m$

6: $U_i = \frac{Y_i}{m}$

7: $NPA \leftarrow NPA \cup \{U_i\}$

8: $i \leftarrow i + 1$

9: **end while**

Output: NPA - conjunto de dim NPAs com distribuição $U_{[0,1]}$

4.1.2 Método Estendido do Autómato Celular

O método do Autómato Celular, apesar de não ser um processo estocástico, partilha de algumas das suas características. É definido através de um conjunto de células, estados, vizinhanças e regras de transição. Muito simplesmente, é utilizado para obter o estado de uma célula no período seguinte, utilizando, para o efeito, um conjunto de regras que ditam para que estado transitará, dependendo do estado das células na sua vizinhança (Feng et al. (2007)). Assumindo que existem p células no sistema, define-se este método como o grupo (García et al. (2008))

$$\{\mathbf{Z}^n, \mathbf{S}^{(k)}, \mathbf{V}^{(k)}, f\}, \quad k \in \mathbb{N}_0$$

onde:

- \mathbf{Z}^n - espaço Euclidiano de dimensão n , que representa o número de diferentes estados admissíveis no sistema;

- $\mathbf{S}^{(k)} = \{S_0^{(k)}, S_1^{(k)}, \dots, S_i^{(k)}, \dots, S_{p-1}^{(k)}\}$ - conjunto finito de células, em que $S_i^{(k)}$ representa o estado da i -ésima célula no período k . $\mathbf{S}^{(0)}$ indica o estado inicial de todas as células do sistema;
- $\mathbf{V}^{(k)} = \{\mathbf{V}_0^{(k)}, \mathbf{V}_1^{(k)}, \dots, \mathbf{V}_i^{(k)}, \dots, \mathbf{V}_{p-1}^{(k)}\}$ - conjunto de vizinhanças, onde $\mathbf{V}_i^{(k)}$, representa o estado das células na vizinhança da i -ésima célula no período k ;
- $f : \mathbf{V}_i^{(k)} \rightarrow S_i^{(k+1)}$ - função que implementa as regras de transição.

O método Estendido do Autômato Celular advém do anterior. Se o primeiro permite que apenas uma célula altere o seu estado por unidade de tempo, o método estendido possibilita uma transição de estado em múltiplas células durante o mesmo período (Shen et al. (1996)). Este é o método utilizado por defeito em alguns softwares, como é o caso do *Wolfram Mathematica*, software que foi usado no trabalho desenvolvido nesta dissertação.

À semelhança do MCM, este método permite obter um conjunto de NPAs com distribuição Uniforme no intervalo $[0, 1]$. Para tal, é necessário definir o número de células e de estados, as vizinhanças e as regras de transição. No software *Wolfram Mathematica* as predefinições para estes parâmetros são, respectivamente, $80 \times M$ células, em que M é a precisão da máquina, dois estados (0 ou 1), vizinhanças de cardinalidade 5 e a regra de transição corresponde a seleccionar os estados das células com índice de período 4, isto é, das células que se encontram nas posições 0, 4, 8, \dots , $\lfloor \frac{80M-1}{4} \rfloor$ (Wolfram Reference - RNG (2014)). Estes estados são, então, usados como uma sequência de bits a partir dos quais se geram os NPAs. Assim, utilizando a notação descrita anteriormente, tem-se

- $\mathbf{Z}^2 = \{0, 1\}$;
- $\mathbf{S}^{(k)} = \{S_0^{(k)}, S_1^{(k)}, \dots, S_{80M-1}^{(k)}\}$, $k \in \mathbb{N}_0$;
- $\mathbf{V}_i^{(k)} \subset \{S_i^{(k)}\} : \#\mathbf{V}_i^{(k)} = 5$, $i \in \{0, \dots, 80M-1\}$, $k \in \mathbb{N}_0$;
- $f(S_i^{(k)}) = \prod_{j=1}^5 V_{ij}^{(k)}$, $i \in \{0, \dots, 80M-1\}$, $k \in \mathbb{N}_0$,

em que $V_{ij}^{(k)}$ representa a j -ésima célula de $\mathbf{V}_i^{(k)}$.

É de notar que quanto maior for o período, melhor será a probabilidade dos NPAs gerados passarem em testes de aleatoriedade. Em contrapartida, a computação tornar-se-á mais lenta.

4.2 Métodos de Geração de Amostras Aleatórias

Existe uma grande variedade de métodos para gerar amostras aleatórias através de Simulação. Basta consultar, por exemplo, Banks et al. (1998) ou Leemis e Park (2006) para ter uma noção dos diversos tipos de métodos que existem. Nesta dissertação destacam-se quatro, não só por serem dos mais simples, mas também porque são os utilizados pelo software *Wolfram Mathematica*.

4.2.1 Método da Inversão

O método da Inversão é o mais adequado para gerar valores de variáveis aleatórias cuja função de distribuição seja facilmente invertível, como é o caso das distribuições Exponencial e Uniforme $[a, b]$, ou ainda distribuições em forma de escada.

Represente-se essa função de distribuição por $F(\cdot)$. A título de exemplo, considere-se uma v.a. $T \sim \text{Exp}(\lambda)$. A sua função de distribuição é facilmente determinada e é dada por

$$F_T(t) = 1 - e^{-\lambda t}, \quad t > 0$$

e a sua inversa por (ver, por exemplo, Murteira e Antunes (2012) ou Taha (2007))

$$F_T^{-1}(t) = -\left(\frac{1}{\lambda}\right) \log(1 - U), \quad U \sim U_{[0,1]}.$$

Note-se que no argumento da função logaritmo é indiferente ter $1 - U$ ou simplesmente U , pois se U é um NPA, o seu complementar também o é.

O procedimento a tomar para obter uma amostra aleatória utilizando este método está descrito no Algoritmo 2.

Algoritmo 2 : Método da Inversão

Input: $F^{-1}(\cdot)$ - função inversa de $F(\cdot)$

Input: dim - dimensão da amostra

```

1:  $\mathbf{A} \leftarrow \emptyset$ 
2:  $i \leftarrow 1$ 
3: while  $i \leq dim$  do
4:   Gerar um NPA  $U_i \sim U_{[0,1]}$ 
5:    $A_i \leftarrow F^{-1}(U_i)$ 
6:    $\mathbf{A} \leftarrow \mathbf{A} \cup \{A_i\}$ 
7:    $i \leftarrow i + 1$ 
8: end while

```

Output: \mathbf{A} - amostra de dimensão dim com distribuição $F(\cdot)$

Este método não é recomendado, porém, para variáveis aleatórias cuja função de distribuição não seja facilmente invertível ou não admita função inversa, como são exemplos as distribuições Binomial, Poisson ou Normal.

4.2.2 Método da Convolução

Para algumas distribuições, o método da Convolução resolve o problema da geração de amostras para as quais o método da Inversão não o consegue fazer, ou seja, para variáveis aleatórias cuja função de distribuição não seja facilmente invertível ou não admita função inversa. Este método consiste em utilizar a soma (convolução) de variáveis aleatórias cujas amostras podem ser obtidas através do método anterior, por exemplo, valores

que seguem uma distribuição Normal podem ser obtidos a partir de valores de uma v.a. com distribuição Uniforme.

De acordo com o Teorema Limite Central, à medida que a dimensão da amostra, n , aumenta, a v.a. resultante da soma das n v.a. independentes e identicamente distribuídas da amostra converge para a distribuição Normal, isto é, tendo $E[A] = \mu$ e $\text{Var}(A) = \sigma^2$, então

$$\mathbf{A}_n = \sum_{i=1}^n A_i \xrightarrow{n \rightarrow +\infty} N(n\mu, n\sigma^2), \quad A_i \stackrel{iid}{\sim} A$$

independentemente da distribuição da v.a. A , se bem que esta convergência é mais rápida para distribuições que, à semelhança da Normal, sejam simétricas (Murteira e Antunes (2012)). Desta forma, tem-se que

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \rightarrow N(0, 1).$$

Assim, considere-se que Y é uma v.a. com distribuição Normal de média μ e variância σ^2 e seja Z a v.a. resultante de proceder à soma (convolução) de n NPAs, ou seja,

$$Z = \sum_{i=1}^n U_i, \quad U_i \stackrel{iid}{\sim} U_{[0,1]}.$$

Sabendo que a média e a variância de uma v.a. $U \sim U_{[0,1]}$ são, respectivamente, $E[U] = \frac{1}{2}$ e $\text{Var}(U) = \frac{1}{12}$, então podem-se obter valores de Y , a partir de Z , através da equação seguinte:

$$Y = \mu + \sigma \left(\frac{Z - \frac{n}{2}}{\sqrt{\frac{n}{12}}} \right).$$

Utilizando uma amostra de dimensão $n = 12$, a expressão simplifica-se para

$$Y = \mu + \sigma (Z - 6).$$

Desta forma, conclui-se que são necessários $12 \times m$ NPAs para gerar uma amostra de dimensão m de uma v.a. $Y \sim N(\mu, \sigma^2)$, o que é bastante ineficiente. Isto não significa que este método seja inadequado para geração de variáveis aleatórias. Significa, sim, que não é o mais indicado para gerar variáveis aleatórias com distribuição Normal. Existem outras distribuições para as quais o método da Convolução é mais eficiente, como o caso da distribuição de Poisson (Taha (2007)).

4.2.3 Método da Aceitação-Rejeição

Este método tem a vantagem de poder ser aplicado a qualquer v.a. Y independentemente da sua distribuição, sendo apenas necessário o conhecimento da sua função densidade de probabilidade, $f_Y(\cdot) = F'_Y(\cdot)$, e que o seu domínio seja limitado e finito, isto é, que

$$y \in D = [a, b] \subset \mathbb{R}.$$

Com a exigência desta segunda condição, este método exclui, à partida, várias distribuições cujo domínio é ilimitado, como a distribuição Normal. Contudo, se se considerar uma distribuição Normal truncada em pontos cuja densidade é muito baixa, por exemplo, 3 desvios-padrão além da média, então não deverá haver diferenças significativas nos resultados obtidos (Hillier e Lieberman (2005)).

O método da Aceitação-Rejeição consiste em tomar uma decisão sobre a aceitação ou a rejeição de um valor simulado de Y , y , como pertencente à distribuição $f_Y(\cdot)$. Para a sua concretização são necessários dois NPAs com distribuição $U_{[0,1]}$, u_1 e u_2 . Adicionalmente, defina-se

$$y = a + (b - a) \times u_1$$

e observe-se a relação

$$u_2 \leq \frac{f_Y(y)}{f_Y(c)},$$

onde

- $c = \max(f_Y(y))$;
- $\frac{f_Y(y)}{f_Y(c)}$ representa a probabilidade de aceitar y .

Caso a relação se confirme, então aceita-se y , caso contrário repetem-se a geração dos NPAs e a definição de y até que a inequação dê uma resposta verdadeira (consultar Hillier e Lieberman (2005)). O método termina quando se tiverem aceite tantos valores de y como os que se pretende ter na amostra.

Este método tem também problemas de eficiência, uma vez que será necessário gerar, pelo menos, 2 NPAs para que se obtenha um valor da distribuição pretendida. Assim, estabelece-se à partida que o método da Aceitação-Rejeição terá, na melhor das hipóteses, metade da eficiência do método da Inversão (onde apenas é necessário gerar 1 NPA), pelo que este último método será preferível sempre que se reunirem as condições para o usar.

O Algoritmo 3 apresenta o pseudo-código que ilustra o método da Aceitação-Rejeição.

O vector **A** assim obtido constitui a amostra aleatória com a dimensão desejada e o valor de k indica a proporção de NPAs gerados que permitiram dar origem a um valor

Algoritmo 3 : Método da Aceitação-Rejeição**Input:** $f(\cdot)$ - função densidade**Input:** a - limite inferior do domínio da v.a.**Input:** b - limite superior do domínio da v.a.**Input:** dim - dimensão da amostra

```

1:  $c \leftarrow \max(f(\cdot))$ 
2:  $\mathbf{A} \leftarrow \emptyset$ 
3:  $i \leftarrow 1$ 
4:  $j \leftarrow 0$ 
5: while  $i \leq dim$  do
6:    $aceite \leftarrow \text{False}$ 
7:   repeat
8:     Gerar dois NPAs  $U_1, U_2 \sim U_{[0,1]}$ 
9:      $Y_i \leftarrow a + (b - a) \times U_1$ 
10:    if  $U_2 \leq \frac{f(Y_i)}{f(c)}$  then
11:       $aceite \leftarrow \text{True}$   $\triangleright Y_i$  é aceite como um valor de  $f(\cdot)$ 
12:    end if
13:     $j \leftarrow j + 2$ 
14:  until  $aceite == \text{True}$ 
15:   $\mathbf{A} \leftarrow \mathbf{A} \cup \{Y_i\}$ 
16:   $i \leftarrow i + 1$ 
17: end while
18:  $k \leftarrow \frac{dim}{j}$ 

```

Output: \mathbf{A} - amostra de dimensão dim com distribuição $F(\cdot)$ **Output:** k - medida de eficiência do método

da amostra. Quanto maior for o valor de k (tendo em conta que nunca será superior a 0,5), maior terá sido a eficiência do método. Por outro lado, quanto mais se aproximar de 0, maior terá sido a quantidade de valores de y rejeitados quando comparado com o número de valores aceites.

Há outros autores, por exemplo Taha (2007), que em vez de utilizarem $f_Y(c)$, tomam a escolha de definir uma função $g_Y(\cdot)$ tal que

$$g_Y(y) \geq f_Y(y), \quad y \in D.$$

Esta alteração ao método permite que, se a definição da função g_Y tiver sido adequada (caso g_Y seja próxima de f_Y), a proporção de valores de y rejeitados seja menor relativamente aos aceites. Apenas se exige que g_Y domine f_Y no domínio de Y e que tenha uma expressão simples, por forma a não tornar lenta a computação do método.

4.2.4 Método da Aceitação-Complemento

O último dos métodos de geração de amostras aleatórias apresentado é o da Aceitação-Complemento. Este deriva do método anterior e introduz algumas alterações de modo a

melhorar a sua eficácia. Com efeito, a ideia base é a mesma, sendo que apenas o processo de aceitação de valores de y tem algumas diferenças. A descrição do método que se segue baseia-se em Kronmal e Peterson Jr (1981) e Kronmal e Peterson Jr. (1984).

Em primeiro lugar, será necessário decompor a função densidade de probabilidade em duas subdensidades, isto é, definir $f_Y^{(1)}(\cdot)$ e $f_Y^{(2)}(\cdot)$ tais que

$$f_Y(\cdot) = f_Y^{(1)}(\cdot) + f_Y^{(2)}(\cdot),$$

definir uma nova função densidade de probabilidade, $g_Y(\cdot)$ que verifique a condição

$$g_Y(\cdot) \geq f_Y^{(1)}(\cdot),$$

isto é, que domina a primeira subdensidade, e, ainda, uma outra função $h_Y(\cdot)$ dada por

$$h_Y(y) = \frac{f_Y^{(2)}(y)}{\int_a^b f_Y^{(2)}(y) dy}, \quad y \in D.$$

Adicionalmente, defina-se a variável auxiliar

$$r = \max \left(\frac{f_Y^{(1)}(y)}{h_Y(y)} \right).$$

Cada nova iteração do método inicia-se com a geração de uma v.a. $U \sim U_{[0,1]}$ e respectiva comparação com r . Caso a condição

$$U \geq r$$

se verifique, então gera-se uma v.a. $Y \sim h_Y(\cdot)$ e aceita-se Y . Caso contrário, gera-se $Y_1 \sim g_Y(\cdot)$ e observa-se a relação

$$U \leq \frac{f_Y^{(1)}(Y_1)}{g_Y(Y_1)}. \quad (4.3)$$

Se esta condição for verdadeira, aceita-se Y_1 , senão gera-se $Y_2 \sim h_Y(\cdot)$ e será este o valor de Y aceite.

A comparação de U com r permite evitar ter de se calcular e observar a relação (4.3) em alguns casos. No entanto, este passo não é essencial para o método e pode ser excluído. Com efeito, se o valor de r for próximo de 1, a eliminação deste passo pode, na verdade, melhorar a rapidez de computação do método.

No que diz respeito à eficiência, o método da Aceitação-Complemento permite obter um valor de Y para cada 2, no mínimo, ou 3, no máximo, NPAs gerados, conforme o resultado da relação (4.3). Assim, determina-se que a eficiência do método situa-se entre os 33,33% e os 50%.

O Algoritmo 4 apresenta uma síntese deste método, em forma de pseudo-código.

Algoritmo 4 : Método da Aceitação-Complemento

Input: $f(\cdot)$ - função densidade

Input: a - limite inferior do domínio da v.a.

Input: b - limite superior do domínio da v.a.

Input: dim - dimensão da amostra

1: $f_1(\cdot), f_2(\cdot) : f(\cdot) = f_1(\cdot) + f_2(\cdot)$

2: $g(\cdot) : g(\cdot) \geq f_1(\cdot) \triangleright g(\cdot)$ é uma densidade

3: $h(\cdot) \leftarrow \frac{f_2(\cdot)}{\int_a^b f_2(\cdot) dy}$

4: $r \leftarrow \max \left(\frac{f_1(\cdot)}{h(\cdot)} \right)$

5: $\mathbf{A} \leftarrow \emptyset$

6: $i \leftarrow 1$

7: $j \leftarrow 0$

8: **while** $i \leq dim$ **do**

9: Gerar um NPA $U \sim U_{[0,1]}$

10: $j \leftarrow j + 1$

11: **if** $U \geq r$ **then**

12: $\mathbf{A} \leftarrow \mathbf{A} \cup Y_i \sim h(\cdot) \triangleright Y_i$ é aceite como um valor de $f(\cdot)$

13: $j \leftarrow j + 1$

14: **else**

15: Gerar $Y_{i,1} \sim g(\cdot)$

16: $j \leftarrow j + 1$

17: **if** $U \leq \frac{f_1(Y_{i,1})}{g(Y_{i,1})}$ **then**

18: $\mathbf{A} \leftarrow \mathbf{A} \cup Y_{i,1} \triangleright Y_{i,1}$ é aceite como um valor de $f(\cdot)$

19: **else**

20: $\mathbf{A} \leftarrow \mathbf{A} \cup Y_{i,2} \sim h(\cdot) \triangleright Y_{i,2}$ é aceite como um valor de $f(\cdot)$

21: $j \leftarrow j + 1$

22: **end if**

23: **end if**

24: $i \leftarrow i + 1$

25: **end while**

26: $k \leftarrow \frac{dim}{j}$

Output: \mathbf{A} - amostra de dimensão dim com distribuição $F(\cdot)$

Output: k - medida de eficiência do método

Com a apresentação destes métodos, termina a exposição teórica dos conteúdos necessários para a aplicação do modelo tradicional da evolução de carteiras de Sistemas de Bonus Malus e do modelo de Simulação à carteira de Seguro Automóvel de uma Seguradora portuguesa. A caracterização da referida carteira, a descrição do respectivo SBM e os resultados obtidos podem ser lidos no Capítulo 5.



Resultados do SBM e da Simulação

Terminada a exposição teórica dos Sistemas de *Bonus Malus* e das componentes de Simulação necessárias para o âmbito desta dissertação, ir-se-á aplicar o que foi referido nos capítulos anteriores a uma carteira de Seguro Automóvel.

Esta aplicação foi retirada de Guerreiro (2001), actualizada pela informação disponível em Guerreiro et al. (2014). Os dados utilizados referem-se à carteira de responsabilidade civil automóvel de 2006 de uma Seguradora portuguesa que, por motivos de confidencialidade, será referida como Seguradora Y.

5.1 Características do SBM Utilizado

O SBM que a Seguradora Y, à data, tinha em prática apresenta as seguintes características:

- 20 classes de *bonus*, ordenadas de acordo com a propensão à sinistralidade;
- as novas apólices entram directamente para a classe C_{10} ;
- as regras de transição são tais que:
 - por cada ano em que não seja reportado qualquer sinistro, a apólice desce uma classe;
 - pelo primeiro sinistro reportado numa anuidade, a apólice sobe duas classes;
 - por cada sinistro reportado na anuidade além do primeiro, a apólice sobe cinco classes.

5.2 Classes do Sistema, Vector de Prémios, Matriz de Regras de Transição e Número de Sinistros

A Tabela 5.1 exhibe os descontos e agravamentos atribuídos em cada classe de *bonus*.

Tabela 5.1: Classes e Prémios do SBM Utilizado

Classe	Desconto	Classe	Agravamento
C ₁	50%	C ₁₁	10%
C ₂	45%	C ₁₂	20%
C ₃	40%	C ₁₃	30%
C ₄	35%	C ₁₄	40%
C ₅	30%	C ₁₅	55%
C ₆	25%	C ₁₆	70%
C ₇	20%	C ₁₇	85%
C ₈	15%	C ₁₈	100%
C ₉	10%	C ₁₉	125%
C ₁₀	0%	C ₂₀	150%

O vector de prémios do SBM, \mathbf{b} , apresentado adiante na equação (5.1) é um vector que contém a informação descrita na tabela anterior. Estes descontos e agravamentos correspondem à escala comercial da Seguradora Y.

De acordo com as regras de transição enunciadas, apresenta-se, adiante na equação (5.2), a respectiva matriz de regras de transição, que indica, dependendo da classe de *bonus* em que a apólice se encontra, para qual transita ao reportar t_{ij} sinistros.

A Tabela 5.2 estabelece uma relação entre o número de sinistros reportados e o número de apólices na carteira em 2006.

Tabela 5.2: Distribuição do Número de Sinistros Observados, em 2006

Número de Sinistros	Número de Apólices
0	191 449
1	12 170
2	913
3	80
4	8
5	2
6	1
Total	204 623

Os dados da Tabela 5.2 levam a suspeitar que o número de sinistros poderá não seguir um processo de Poisson Homogéneo, visto que a variância do número de sinistros,

$$\mathbf{b}' = \begin{bmatrix} 50 & 55 & 60 & 65 & 70 & 75 & 80 & 85 & 90 & 100 & 110 & 120 & 130 & 140 & 155 & 170 & 185 & 200 & 225 & 250 \end{bmatrix} \quad (5.1)$$

$$\mathbf{T} = \begin{bmatrix} C_1 & C_2 & C_3 & C_4 & C_5 & C_6 & C_7 & C_8 & C_9 & C_{10} & C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} & C_{17} & C_{18} & C_{19} & C_{20} \\ \{0\} & - & \{1\} & - & - & - & - & \{2\} & - & - & - & - & \{3\} & - & - & - & - & \{4\} & - & \{5, 6, \dots\} \\ \{0\} & - & - & \{1\} & - & - & - & - & \{2\} & - & - & - & - & \{3\} & - & - & - & \{4\} & - & \{5, 6, \dots\} \\ - & \{0\} & - & - & \{1\} & - & - & - & - & \{2\} & - & - & - & - & \{3\} & - & - & - & - & \{4, 5, \dots\} \\ - & - & \{0\} & - & - & \{1\} & - & - & - & - & \{2\} & - & - & - & - & \{3\} & - & - & - & \{4, 5, \dots\} \\ - & - & - & \{0\} & - & - & \{1\} & - & - & - & - & \{2\} & - & - & - & - & \{3\} & - & - & \{4, 5, \dots\} \\ - & - & - & - & \{0\} & - & - & \{1\} & - & - & - & - & \{2\} & - & - & - & \{3\} & - & - & \{4, 5, \dots\} \\ - & - & - & - & - & \{0\} & - & - & - & \{1\} & - & - & - & \{2\} & - & - & - & \{3\} & - & \{4, 5, \dots\} \\ - & - & - & - & - & - & \{0\} & - & - & - & \{1\} & - & - & - & \{2\} & - & - & - & \{3, 4, \dots\} \\ - & - & - & - & - & - & - & \{0\} & - & - & - & \{1\} & - & - & - & \{2\} & - & - & \{3, 4, \dots\} \\ - & - & - & - & - & - & - & - & \{0\} & - & - & - & \{1\} & - & - & - & \{2\} & - & \{3, 4, \dots\} \\ - & - & - & - & - & - & - & - & - & \{0\} & - & - & - & \{1\} & - & - & - & - & \{2, 3, \dots\} \\ - & - & - & - & - & - & - & - & - & - & \{0\} & - & - & - & \{1\} & - & - & - & \{2, 3, \dots\} \\ - & - & - & - & - & - & - & - & - & - & - & \{0\} & - & - & - & \{1\} & - & - & \{2, 3, \dots\} \\ - & - & - & - & - & - & - & - & - & - & - & - & \{0\} & - & - & - & \{1\} & - & \{2, 3, \dots\} \\ - & - & - & - & - & - & - & - & - & - & - & - & - & \{0\} & - & - & \{1\} & - & \{2, 3, \dots\} \\ - & - & - & - & - & - & - & - & - & - & - & - & - & - & \{0\} & - & - & \{1\} & - & \{1, 2, \dots\} \\ - & - & - & - & - & - & - & - & - & - & - & - & - & - & - & \{0\} & - & - & \{1, 2, \dots\} \\ - & - & - & - & - & - & - & - & - & - & - & - & - & - & - & - & \{0\} & - & \{1, 2, \dots\} \end{bmatrix} \quad (5.2)$$

$$\text{Var}(N) = 0,0770146,$$

é superior ao seu valor esperado,

$$E[N] = 0,0698064,$$

fenómeno característico das distribuições de Poisson Mistas.

5.3 Distribuição do Número de Sinistros

Utilizando os Estimadores da Máxima Verosimilhança determinados no Capítulo 2, obtiveram-se as estimativas apresentadas na Tabela 5.3, para cada uma das distribuições consideradas.

Tabela 5.3: Estimativas da Máxima Verosimilhança

Poisson	Polya	Sichel
$\hat{\lambda} = 0,06981$	$\hat{\alpha} = 0,69583$	$\hat{g} = 0,06981$
	$\hat{\beta} = 9,96793$	$\hat{h} = 0,04747$

Os resultados da aplicação dos testes do Qui-Quadrado de Pearson e da Razão de Verosimilhanças são apresentados na Tabela 5.4.

Tabela 5.4: Resultados dos testes de ajustamento

Distribuição	gl	$\chi_{gl; 0,95}^2$	Qui-Quadrado de Pearson	Razão de Verosimilhanças	Estatística K
Poisson	1	3,841	687,313	53 126,2	
Polya	2	5,991	1,346	52 776,0	8×10^{-153}
Sichel	2	5,991	0,325	52 773,9	1×10^{-153}

Pode-se concluir que as distribuições de Polya e de Sichel ajustam-se aos dados analisados, o que não acontece com o processo de Poisson Homogéneo, como seria de esperar. Note-se que a partir dos valores do quadro anterior, não se pode concluir qual das distribuições será a mais adequada aos dados, pois apresentam resultados muito idênticos. Por esta razão, daqui em diante, todas as análises efectuadas terão em consideração ambas as distribuições de Poisson Mistas: Poisson com distribuição de estrutura Gama $(\hat{\alpha}, \hat{\beta})$ e distribuição de estrutura Inversa Gaussiana (\hat{g}, \hat{h}) . A distribuição

Gama $(\hat{\alpha} = 0,6958, \hat{\beta} = 9,9679)$ tem a função densidade de probabilidade expressa por

$$u_{\Lambda}(\lambda) = \frac{1}{\Gamma(0,6958)} \times 9,9679^{0,6958} \times e^{-9,9679\lambda} \times \lambda^{0,3042}, \quad \lambda > 0,$$

sendo a função de probabilidade para o número de sinistros reportados dada por

$$f_N(k) = \binom{k-0,3042}{k} \times \left(\frac{9,9679}{10,9679}\right)^{0,6958} \times \left(\frac{1}{10,9679}\right)^k, \quad k \in \mathbb{N}_0, \quad (5.3)$$

Por seu lado, a função densidade de probabilidade da distribuição Inversa Gaussiana $(\hat{g} = 0,0698, \hat{h} = 0,0475)$ é definida como

$$u_{\Lambda}(\lambda) = \frac{0,0698}{\sqrt{0,0949\pi\lambda^3}} \exp\left\{-\frac{1}{0,0949\lambda}(\lambda - 0,0698)^2\right\}, \quad \lambda > 0,$$

sendo a função de probabilidade para o número de sinistros reportados definida por recorrência como

$$\begin{aligned} f_N(0) &= \exp\left\{\frac{0,0698}{0,0475} \times \left(1 - (1,0949)^{1/2}\right)\right\} \\ f_N(1) &= 0,0698 \times f_N(0) \times 1,0949^{-1/2} \\ f_N(k) &= \frac{0,0475 \times (k-1)(2k-3)f_N(k-1) + 0,0698^2 \times f_N(k-2)}{1,0949 \times k(k-1)}, \quad k = 2, 3, \dots \end{aligned} \quad (5.4)$$

Sempre que possível, serão estabelecidas comparações entre os resultados obtidos com o modelo de SBM e com o modelo de Simulação considerando as distribuições (5.3) e (5.4) para a modelação do número de sinistros.

5.4 Geração das Amostras Aleatórias e Resultados da Simulação

Nos modelos de Simulação é necessário partir de alguns pressupostos para se conseguir efectuar o estudo pretendido. Os pressupostos assumidos foram de que existem $204\,623 \times 100$ apólices na carteira (cuja justificação se encontra no parágrafo seguinte) e que o número de sinistros reportados é descrito através de uma distribuição de Poisson Mista com uma de duas distribuições de estrutura. A primeira a ser considerada foi uma Gama $(\hat{\alpha} = 0,6958, \hat{\beta} = 9,9679)$ e a outra uma IG $(\hat{g} = 0,0698, \hat{h} = 0,0475)$.

O pressuposto de que existem $204\,623 \times 100$ apólices na carteira parte da necessidade de proceder à discretização das distribuições consideradas. Utilizando o método de discretização que mantém o Valor Esperado (ver, por exemplo, Centeno (2003)), foi considerado um tecto de valor 1 para a frequência de sinistralidade, λ , uma vez que o integral

$$\int_0^1 u_{\Lambda}(\lambda) d\lambda,$$

para ambas as distribuições de estrutura, se encontra suficientemente próximo da unidade, e um passo de 0,01. Assim, foram geradas 204 623 amostras de segurados com $\lambda = 0,01i$ ($i \in \{1, \dots, 100\}$), isto é, com 100 diferentes valores para a sua frequência de sinistralidade.

A partir destes pressupostos, com recurso ao software *Wolfram Mathematica* e tendo validado a independência e aleatoriedade dos NPAs gerados¹ através de um simples teste de autocorrelação, foram geradas amostras aleatórias² de números de sinistros reportados anualmente à Seguradora Y por segurado. Tendo em conta as regras de transição de apólices entre classes da Seguradora Y e assumindo que no momento inicial todas as apólices se encontravam na classe de entrada, obteve-se a distribuição anual da carteira³, durante um horizonte temporal de 50 anos. Este período foi escolhido de acordo com as conclusões previamente obtidas pelo modelo tradicional de evolução das carteiras de SBM. As distribuições assim determinadas permitem não só concluir acerca da estacionaridade da carteira, mas também da evolução do prémio médio anual.

Com efeito, uma vez que se observou, como será descrito na Secção seguinte, que a estacionaridade da carteira será alcançada num horizonte de aproximadamente 45 anos, foram considerados 50 anos de Simulações. Depois de todas estas corridas foi feita uma análise à estacionaridade da carteira. Esta análise revelou-se positiva, isto é, permitiu que se verificasse também para o modelo de Simulação a estacionaridade dentro do período considerado, portanto não foi necessário proceder à Simulação de mais anos de observações.

O facto de ter o mesmo número de apólices para cada valor de λ não é alarmante, pois em todos os passos se tem em conta a probabilidade de haver segurados com a respectiva frequência de sinistralidade.

Por último, foram obtidas escalas óptimas de prémios. O único cuidado necessário nesta parte dos resultados é que, ao passo que para o modelo tradicional de Sistemas de *Bonus Malus* foram consideradas distribuições que funcionam em tempo contínuo, agora foi necessário ajustar as funções de modo a darem resultados para distribuições em tempo discreto. Esta prática é recorrente nos SBM quando, por alguma razão, não é possível utilizar as distribuições originais, isto é, sem estarem discretizadas (ver, por exemplo, Centeno (2003)).

Todos os resultados obtidos através deste modelo de Simulação estão apresentados nas Secções que se seguem, onde é realizada, paralelamente, uma comparação com os resultados obtidos através do modelo tradicional de Sistemas de *Bonus Malus*.

¹O *Mathematica* utiliza o método Estendido do Autómato Celular apresentado na Subsecção 4.1.2 para geração de NPAs.

²Para a geração de amostras aleatórias, o *Mathematica* utiliza o método da Aceitação-Rejeição para modelar comportamentos que seguem uma distribuição Gama ou de Poisson e o método da Aceitação-Complemento caso sigam uma distribuição Inversa Gaussiana.

³Não tendo conhecimento de qualquer função nativa do *Mathematica* que faça este procedimento, foram criadas, no âmbito desta dissertação, novas funções para o realizarem. O código criado está disponibilizado no Anexo D.

5.5 Distribuição Estacionária e Prémio Médio

Utilizando a matriz de regras de transição e a expressão da função densidade de probabilidade das distribuições usadas para modelação do número de sinistros, obtêm-se as distribuições estacionárias da carteira apresentadas na Tabela 5.5 e na Figura 5.1, onde também estão presentes as distribuições estacionárias obtidas através de 5000 corridas de simulação. Por uma questão de escala, a classe de maior bonificação está excluída dos gráficos, para facilitar a comparação.

Tabela 5.5: Distribuição Estacionária da Carteira (valores em percentagem do número total de apólices)

Classe	Percentagem de Apólices através da			
	Distribuição de Polya		Distribuição de Sichel	
	SBM	Simulação	SBM	Simulação
C ₁	82,80	79,99	83,36	83,47
C ₂	4,21	4,85	4,36	4,32
C ₃	4,70	5,41	4,80	4,76
C ₄	1,28	1,52	1,15	1,16
C ₅	1,05	1,20	0,92	0,90
C ₆	0,72	0,85	0,61	0,62
C ₇	0,66	0,80	0,57	0,58
C ₈	0,63	0,72	0,54	0,52
C ₉	0,43	0,52	0,36	0,37
C ₁₀	0,38	0,46	0,31	0,32
C ₁₁	0,32	0,36	0,26	0,26
C ₁₂	0,30	0,37	0,25	0,26
C ₁₃	0,28	0,34	0,24	0,24
C ₁₄	0,27	0,31	0,24	0,23
C ₁₅	0,27	0,33	0,24	0,25
C ₁₆	0,28	0,32	0,26	0,25
C ₁₇	0,29	0,34	0,28	0,28
C ₁₈	0,32	0,38	0,33	0,33
C ₁₉	0,37	0,42	0,40	0,39
C ₂₀	0,43	0,50	0,52	0,50

De uma forma geral, pode-se afirmar, por análise à Tabela 5.5 e à Figura 5.1 que, a longo prazo, se estima que a carteira esteja mais concentrada nas classes extremas (quer de *bonus* quer de *malus*) se for considerada a distribuição de Sichel. Observando os resultados do modelo de Simulação, a afirmação anterior apenas é verdadeira para o extremo das classes de *bonus* que conferem uma maior bonificação ao prémio dos segurados.

Estima-se que 82,80% das apólices da carteira estarão na classe mais bonificada do Sistema (Classe 1) se se assumir que o número de sinistros segue a distribuição de Polya, ou 83,36% olhando para os valores obtidos através da distribuição de Sichel. Os valores

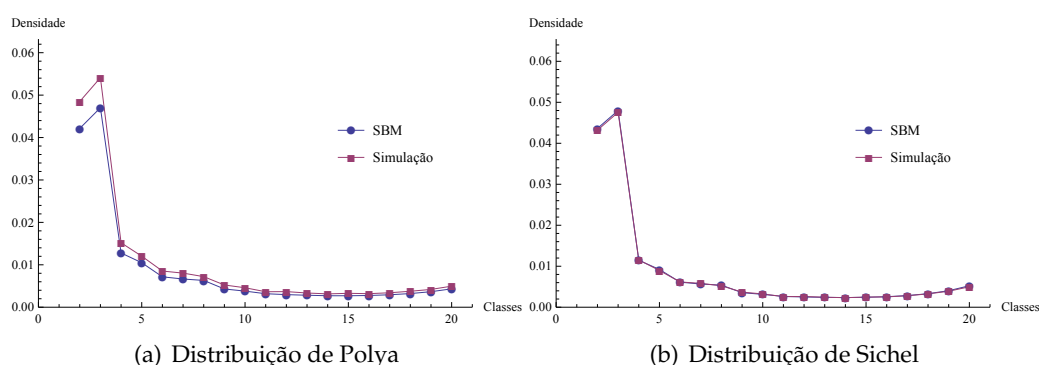


Figura 5.1: Comparação entre as Distribuições Estacionárias obtidas através do SBM e de Simulação

obtidos através de Simulação apontam para os 79,99% e os 83,47%, respectivamente. A percentagem estimada de apólices que se situa nas três classes que garantem um desconto maior é, no total, de, respectivamente, 91,71% e 92,52%. Os correspondentes valores simulados são de 90,25% e 92,55%.

No outro extremo das classes de *bonus* estima-se que apenas 0,43% do número total de apólices estejam na classe com maior agravamento (Classe 20) se a distribuição de estrutura utilizada for a Gama, ou 0,52% se, pelo contrário, se optar pela distribuição de estrutura Inversa Gaussiana. Independentemente da distribuição de estrutura escolhida para a modelação do número de sinistros, os valores simulados situam-se nos 0,50%. Segundo a primeira distribuição, estima-se que haja 1,12% do total das apólices no conjunto das três classes com maior agravamento, e 1,25% se for considerada a segunda. As proporções obtidas por Simulação são de 1,30% e 1,22%, respectivamente.

Como seria de esperar, uma vez que nenhum segurado origina sinistros propositadamente, dadas as regras de transição da Seguradora Y e devido à frequência de sinistralidade associada a cada apólice, estas, no longo-prazo, tenderão a concentrar-se nas classes mais bonificadas. Na Figura 5.2 pode-se analisar a evolução anual da percentagem de apólices da carteira nas classes 1 e 20, segundo os modelos de Simulação considerados. No Anexo A podem-se ver gráficos análogos para as restantes classes de *bonus*. Adicionalmente, no Anexo B pode-se identificar quais são as classes que, entre si, têm comportamentos a longo-prazo semelhantes.

Uma simples análise à Figura 5.3 permite concluir que, embora os valores em ambas as classes sejam próximos, a distribuição de Polya apresenta valores mais elevados para o prémio médio do que se for utilizada a distribuição de Sichel.

Tendo em conta as estimativas da distribuição das apólices pelas classes de *bonus* em cada ano, obtidas através da expressão (3.3), na Figura 5.3 é possível analisar a evolução anual esperada do prémio médio a cobrar aos segurados durante um período de 50 anos.

Apesar de se conseguir observar o comportamento geral da evolução do prémio médio na figura anterior, não é possível distinguir as séries temporais, dada a proximidade dos valores com escala utilizada. Para se perceber melhor as diferenças apresentam-se,

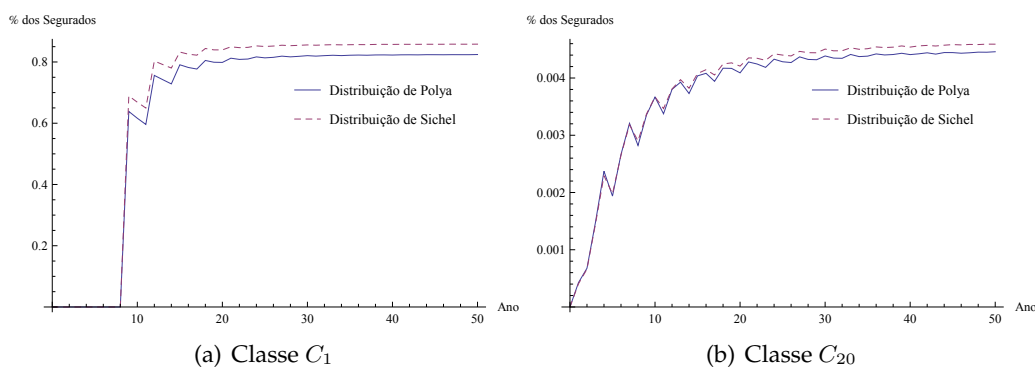


Figura 5.2: Evolução da percentagem de apólices nas classes 1 e 20

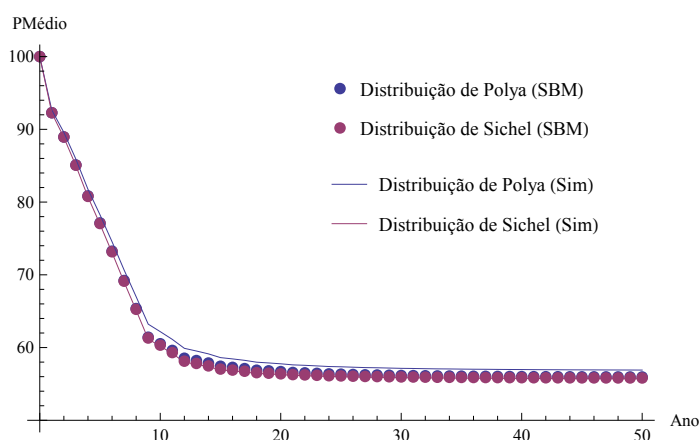


Figura 5.3: Evolução Anual Esperada do Prémio Médio ao longo de 50 anos, em percentagem do Prémio Puro *a priori*

na Figura 5.4 três ampliações da anterior, respeitantes aos períodos compreendidos entre o 1.º e o 10.º, o 11.º e o 25.º e o 26.º e o 50.º anos de permanência das apólices na carteira.

A divisão entre as duas primeiras subfiguras foi feita devido à escala: na primeira analisam-se os anos em que a variação do prémio médio é maior, depois observam-se os valores quando as variações começam a baixar. A última subfigura contém mais anos do que as restantes, para se conseguir ver mais facilmente quando é que a estacionaridade é alcançada, dado que as variações no prémio médio, nestes anos, são muito pequenas.

Ao mesmo tempo que as classes mais bonificadas tendem a ficar cada vez mais populadas com o passar dos anos, o prémio médio apresenta um comportamento decrescente até atingir o seu valor estacionário. As Figuras 5.3 e 5.4 permitem afirmar que se espera que este tenha um decréscimo rápido durante os primeiros 10 a 15 anos, tempo que os segurados demoram a atingir as classes mais bonificadas, e, a partir daí, começa a estabilização até atingir a estacionaridade, próximo dos 45 anos.

A Figura 5.4 permite ver que, em média e no longo-prazo, a estimativa do prémio pago por cada segurado é de 55,92% ou 55,80% do valor inicial, dependendo de qual a distribuição escolhida: se a Polyá ou a de Sichel. As razões observadas através de

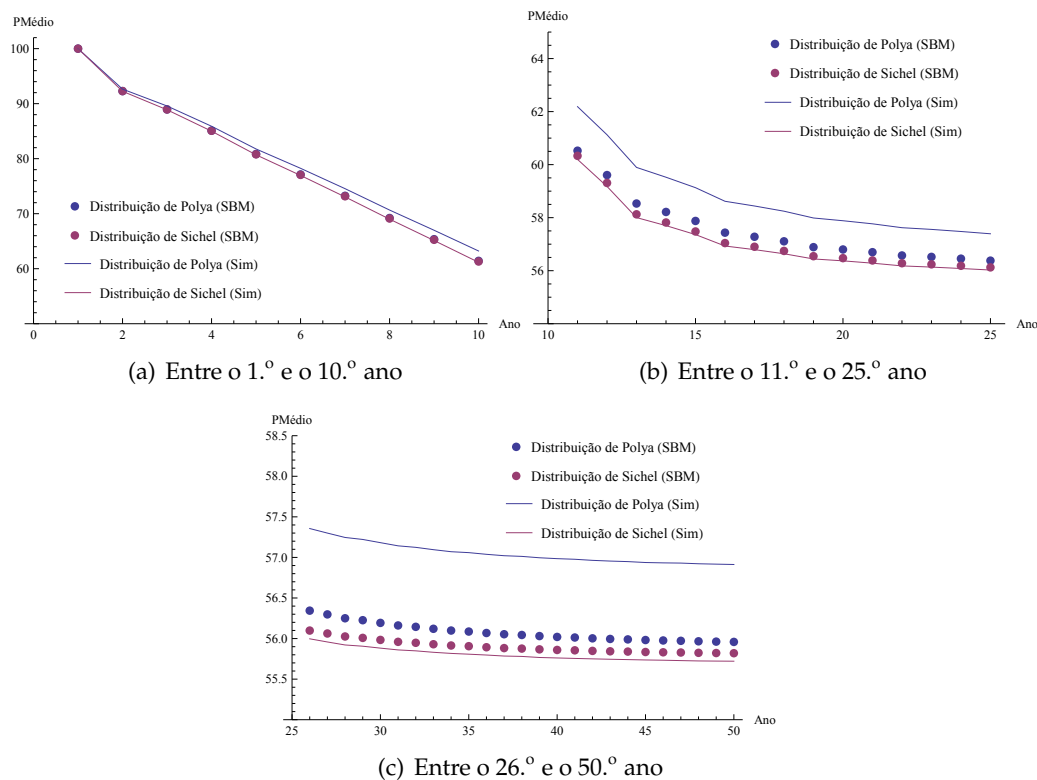


Figura 5.4: Evolução Anual Esperada do Prémio Médio, em percentagem do Prémio Puro *a priori*

Simulação são de, respectivamente, 56,90% e 55,72%. À semelhança do que aconteceu com a distribuição estacionária, verifica-se que os valores do prémio médio determinados pelo modelo de SBM e pelo modelo de Simulação estão mais próximos entre si quando se considera a distribuição de Sichel do que se se assumir que o número de sinistros reportados segue a distribuição de Polya.

É interessante notar que, tanto para a distribuição estacionária como para o prémio médio, os valores estimados e simulados são mais próximos quando se considera a distribuição de estrutura Inversa Gaussiana: o prémio médio em estacionaridade difere em apenas 0,08 pontos percentuais, ao passo que se se utilizar a distribuição de estrutura Gama esta diferença aumenta para os 0,98 pontos percentuais.

5.6 Escalas Óptimas de Prémios

Comece-se por comparar as distribuições estacionária e ponderada da carteira na Tabela 5.6 e na Figura 5.5, tendo em atenção que para estudar as escalas óptimas de prémios recorreu-se à escala de Borgan et al. (1981), tendo-se obtido a distribuição ponderada de acordo com a definição 3.5, assumindo-se um horizonte de 20 anos de permanência na carteira, que a distribuição estacionária não é significativa e que o ponderador atribuído

a cada ano é 5% menor do que o ponderador associado ao ano anterior. Assim, tem-se

$$\begin{cases} w_k = \frac{w_{k-1}}{1,05}, & k \in \{2, \dots, 20\} \\ w_k = 0, & k = 0 \vee k > 20 \\ \sum_{k=0}^{+\infty} w_k = 1 \end{cases}$$

Na Tabela 5.6 e na Figura 5.5 mostram-se os resultados obtidos conjuntamente com os resultados da Tabela 5.5, para facilitar a comparação. A mesma tabela mostra, também, a diferença, em pontos percentuais, que separa a distribuição ponderada da distribuição estacionária.

Por motivos de escala e análise visual das distribuições, as percentagens de apólices referentes à classe 1 foram excluídas dos gráficos da Figura 5.5.

Por análise destes resultados, constata-se que a distribuição ponderada, ao contrário da distribuição estacionária, tem uma densidade crescente para as classes 3 a 9, o que indica, juntamente com a óbvia disparidade na densidade da classe com maior bonificação, que estas, ao fim de 20 anos, ainda estão longe de alcançar a estacionaridade.

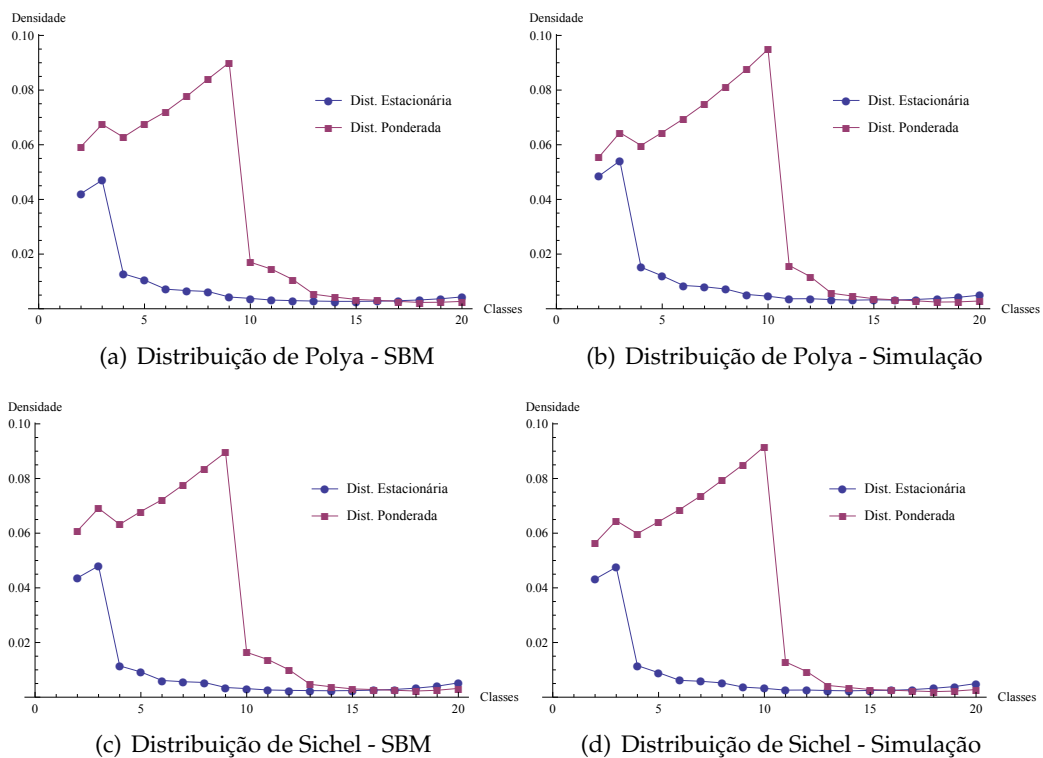


Figura 5.5: Comparação entre as Distribuições Estacionária e Ponderada

O comportamento, em geral, de ambas as distribuições aparenta ter apenas ligeiras diferenças quando determinada pelo Sistema de *Bonus Malus* com Cadeias de Markov ou

Tabela 5.6: Distribuições Estacionária e Ponderada da Carteira (valores em percentagem do número total de apólices)

Classe	Distribuição de Polya						Distribuição de Sichel					
	SBM			Simulação			SBM			Simulação		
	Est.	Pond.	Δ	Est.	Pond.	Δ	Est.	Pond.	Δ	Est.	Pond.	Δ
C ₁	82,80	35,03	-47,77	79,99	29,20	-50,79	83,36	35,05	-48,31	83,47	31,16	-52,31
C ₂	4,21	5,93	1,72	4,85	5,54	0,69	4,36	6,08	1,72	4,32	5,64	1,32
C ₃	4,70	6,75	2,04	5,41	6,44	1,03	4,80	6,91	2,11	4,76	6,46	1,70
C ₄	1,28	6,28	5,00	1,52	5,97	4,45	1,15	6,32	5,17	1,16	5,98	4,82
C ₅	1,05	6,77	5,72	1,20	6,45	5,25	0,92	6,79	5,87	0,90	6,42	5,52
C ₆	0,72	7,21	6,49	0,85	6,95	6,10	0,61	7,22	6,61	0,62	6,87	6,25
C ₇	0,66	7,78	7,11	0,80	7,50	6,70	0,57	7,77	7,20	0,58	7,37	6,79
C ₈	0,63	8,40	7,77	0,72	8,13	7,41	0,54	8,37	7,83	0,52	7,94	7,42
C ₉	0,43	9,00	8,57	0,52	8,77	8,25	0,36	8,96	8,60	0,37	8,52	8,15
C ₁₀	0,38	1,70	1,32	0,46	9,49	9,03	0,31	1,64	1,33	0,32	9,17	8,85
C ₁₁	0,32	1,45	1,13	0,36	1,58	1,22	0,26	1,38	1,12	0,26	1,29	1,04
C ₁₂	0,30	1,07	0,77	0,37	1,17	0,80	0,25	1,00	0,75	0,26	0,94	0,68
C ₁₃	0,28	0,53	0,25	0,34	0,57	0,23	0,24	0,47	0,23	0,24	0,43	0,19
C ₁₄	0,27	0,43	0,16	0,31	0,46	0,15	0,24	0,38	0,14	0,23	0,35	0,12
C ₁₅	0,27	0,34	0,07	0,33	0,37	0,04	0,24	0,30	0,06	0,25	0,28	0,03
C ₁₆	0,28	0,31	0,03	0,32	0,33	0,01	0,26	0,28	0,02	0,25	0,25	0,00
C ₁₇	0,29	0,27	-0,02	0,34	0,29	-0,05	0,28	0,25	-0,03	0,28	0,23	-0,05
C ₁₈	0,32	0,23	-0,09	0,38	0,24	-0,14	0,33	0,23	-0,10	0,33	0,20	-0,13
C ₁₉	0,37	0,24	-0,13	0,42	0,25	-0,17	0,40	0,25	-0,15	0,39	0,22	-0,17
C ₂₀	0,43	0,27	-0,16	0,50	0,28	-0,22	0,52	0,32	-0,20	0,50	0,28	-0,22

com Simulação. No entanto, nos modelos de Simulação é claramente visível que as proporções de apólices nas classes 2 e 3 estão mais próximas entre distribuição estacionária e distribuição ponderada.

Na Figura 5.6 pode-se analisar a diferença entre as distribuições ponderadas determinadas a partir do SBM e da Simulação para cada uma das distribuições ajustadas para o número de sinistros reportados. Apresenta-se, também, uma ampliação destas comparações, desta vez sem os valores da classe 1, para, por motivos de escala, melhor se observar estas diferenças.

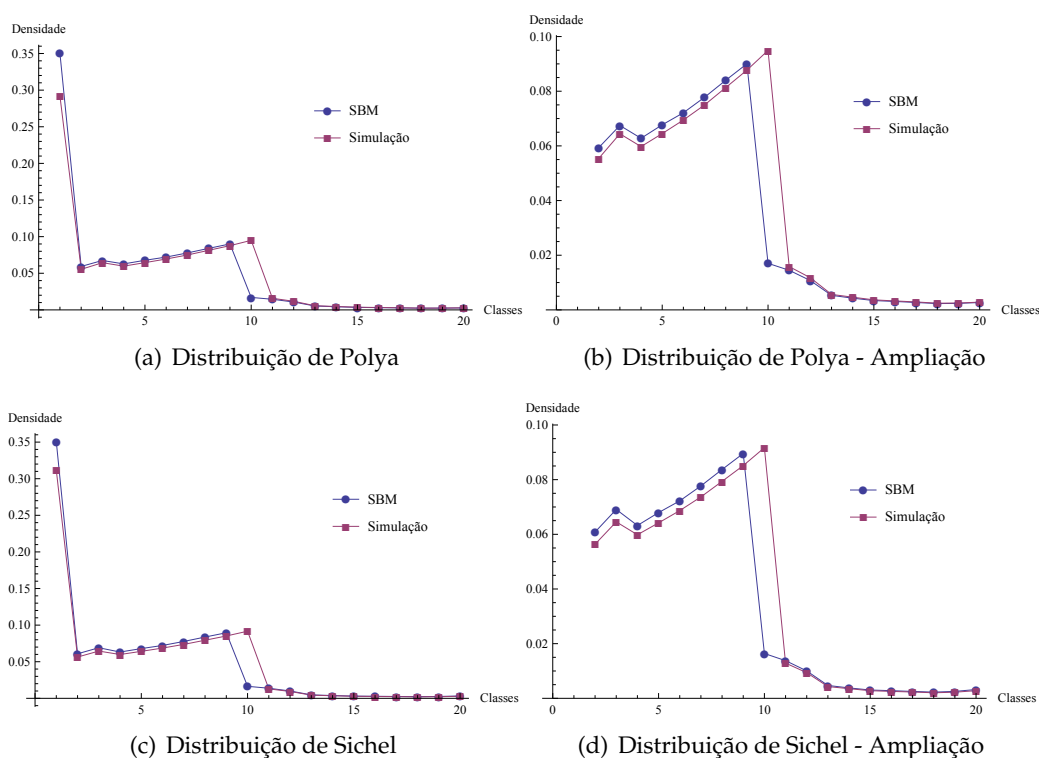


Figura 5.6: Comparação das Distribuições Ponderadas obtidas através do SBM e de Simulação

A partir desta figura apenas há a referir que as diferenças na distribuição ponderada entre SBM e Simulação são menores, à semelhança do que se observa com a distribuição estacionária, quando se considera o ajustamento pela distribuição de Sichel.

A partir destas distribuições é, então, possível determinar as escalas óptimas de prémios para esta carteira de seguro automóvel. As tabelas no Anexo C e a Figura 5.7 permitem comparar a escala comercial da Seguradora Y com as três escalas óptimas de prémios apresentadas no Capítulo 3. É importante referir que as escalas de Gilde e Sundt (1989) apresentadas linearizam a escala de Borgan et al. (1981) do respectivo modelo.

A Figura 5.7 permite observar que a relação entre as escalas é imutável em qualquer um dos quatro casos considerados. A escala de Borgan et al. (1981) é a que apresenta as maiores bonificações, mas também as maiores penalizações - repare-se que o leque de prémios estimados varia entre os 26,49% e os 256,17% do prémio puro *a priori* para

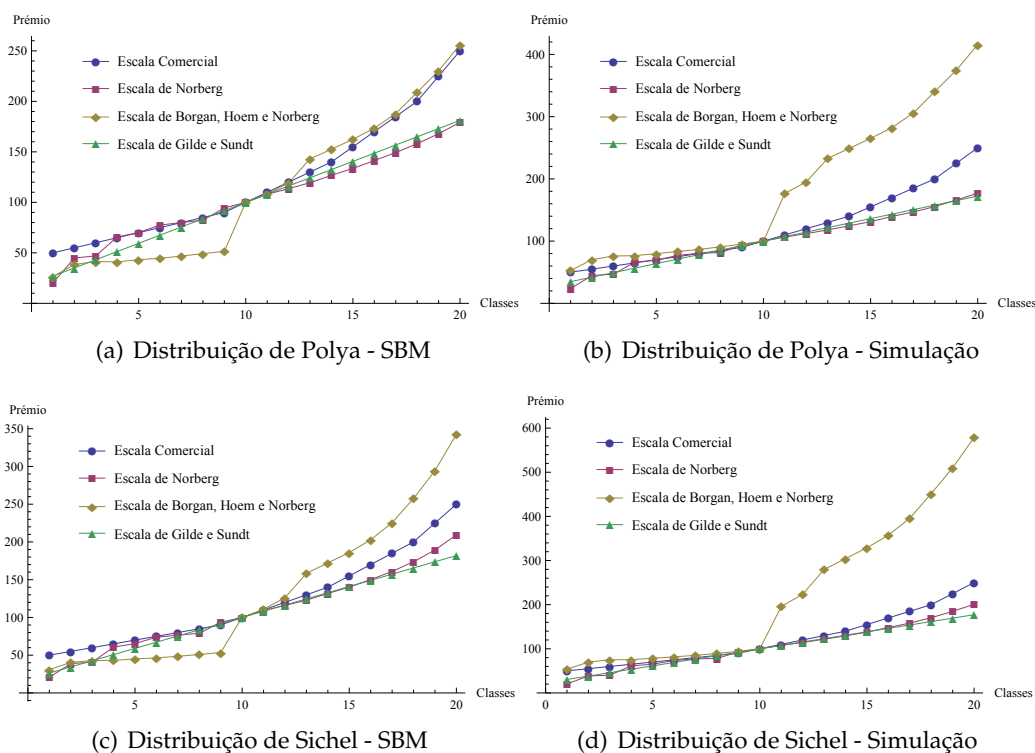


Figura 5.7: Comparação das Escalas de Prémios

a distribuição de Polya e os 30,86% e 342,66% para a distribuição de Sichel. Os valores simulados correspondentes variam, respectivamente, entre os 47,93% e 473,31% e os 54,06% e os 578,37%. A escala de Norberg aproxima-se da escala comercial, no entanto para as classes mais penalizadoras tende a atribuir prémios mais baixos, a rondar os 20% contra os 50% da escala comercial. Por fim, a escala de Gilde e Sundt (1989) é a que se apresenta com valores mais baixos, independentemente da classe de *bonus* em que os segurados se encontrem. Verifica-se, também, que, no caso das escalas determinadas para o modelo teórico (SBM) considerando a distribuição de Polya, a escala de Borgan et al. (1981) é a que mais se aproxima da escala comercial nas classes mais altas. No entanto, nos restantes casos, a que mais se aproxima é a escala de Norberg.

Analise-se, agora, as Figuras 5.8, 5.9 e 5.10, onde é possível estabelecer uma comparação entre as escalas determinadas através do modelo de Sistema de *Bonus Malus* e o modelo de Simulação. Para melhor se estudar estas escalas, foi incluída nestas figuras a escala comercial da Seguradora Y.

Com efeito, observa-se que para a escala de Norberg (1976), tanto a escala estimada como a escala gerada através da distribuição de Sichel são as que mais se aproximam da escala comercial. Pode-se afirmar, ainda, para ambas as distribuições, os valores estimados e gerados apresentam uma muito grande proximidade entre si.

Partindo para a escala de Borgan et al. (1981), é, como já foi referido, a escala obtida pela distribuição de Polya para o modelo de SBM que mais se acerca da escala comercial. Porém, ao contrário do que se observa com a escala de Norberg (1976), neste caso

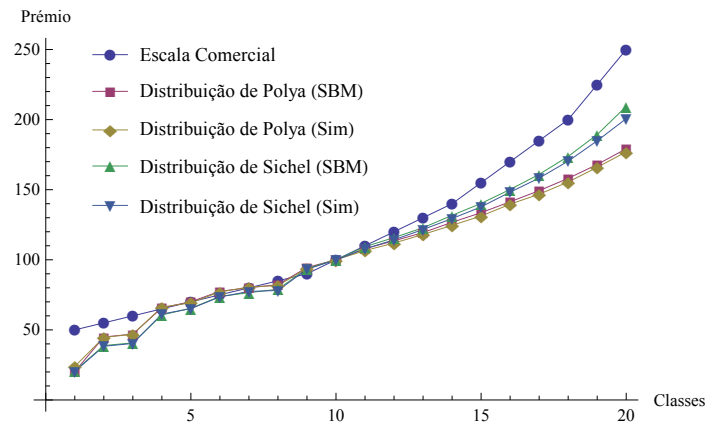


Figura 5.8: Comparação das Escalas de Norberg determinadas através do SBM e por Simulação

não existe qualquer semelhança entre as escalas estimadas e geradas para as classes com agravamento.

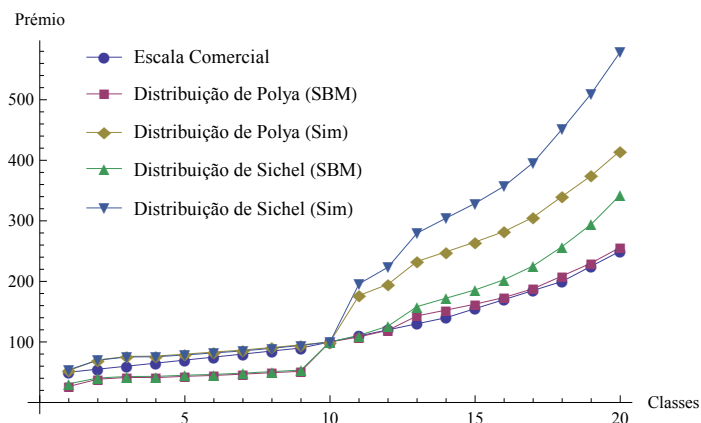


Figura 5.9: Comparação das Escalas de Borgan, Hoem e Norberg determinadas através do SBM e por Simulação

Por último, na escala de Gilde e Sundt (1989) é difícil comparar os vários resultados obtidos, pois são todos muito próximos uns dos outros. Ainda assim, consegue-se ver que os valores estimados e gerados considerando a distribuição de Sichel encontram-se substancialmente mais próximos do que os obtidos considerando a distribuição de Polya.

Estando próximas ou afastadas da escala comercial, qual das escalas ótimas determinadas é a melhor escolha para substituir a da Seguradora Y, por forma a tornar o seu SBM mais eficiente? Atente-se a Tabela 5.7, a qual apresenta a eficiência de cada uma das escalas ótimas de prémios aplicadas ao SBM em estudo.

Para qualquer um dos modelos investigados, é a escala de Norberg que se apresenta sempre como a mais eficiente. Assim conclui-se que, ao alterar a sua escala de prémios,

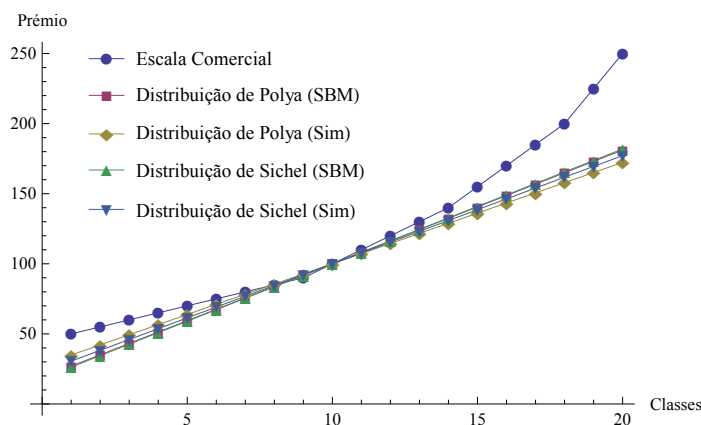


Figura 5.10: Comparação das Escalas de Gilde e Sundt determinadas através do SBM e por Simulação

Tabela 5.7: Eficiência das Escalas de Prémios

Escala	Eficiência das Escalas de Prémios determinadas através da			
	Distribuição de Polya SBM	Simulação	Distribuição de Sichel SBM	Simulação
Norberg	0,0085	95,6337	0,0202	88,1277
Borgan, Hoem e Norberg	0,0069	70,9252	0,0072	68,5349
Gilde e Sundt	0,0063	61,9939	0,0063	59,3974

a Seguradora Y deverá optar pela escala de Norberg, apesar de apenas ter em conta a distribuição das apólices em estacionaridade.

Caso opte por substituir a escala comercial pela escala de Norberg, o prémio médio da carteira, quando esta está estacionária, está estimado em 29,28% do prémio puro *a priori* no caso de ser utilizada a distribuição de Polya para a modelação do número de sinistros, ou em 28,99% caso se opte pelo modelo de Poisson Misto com distribuição de estrutura Inversa Gaussiana. Os valores gerados são de, respectivamente, 28,73% e 28,50%.

5.7 Considerações Finais sobre os Modelos de Poisson Mistos

Depois de analisados os resultados obtidos através tanto do modelo de SBM tradicional e do modelo de Simulação, pode-se afirmar que o ajustamento pela distribuição de Sichel parece ser mais indicado do que o obtido pela distribuição de Polya, devido à proximidade entre resultados teóricos e resultados práticos.

Na comparação das distribuições estacionárias e da evolução do prémio médio, os valores obtidos pelo modelo tradicional de SBM coincidem, em efeitos práticos, com aqueles gerados pelo modelo de Simulação.

Ao comparar as distribuições ponderadas, a conclusão a retirar não é muito diferente. Apesar de em algumas classes a distância entre distribuição estimada e distribuição gerada ser menor quando analisada a distribuição de Polya, essa diferença não é significativa. Porém, para a classe 1, que é aquela que apresenta uma maior distância, é através da distribuição de Sichel que esta se atenua.

Passando para uma análise às escalas óptimas de prémios, verificou-se que a escala mais eficiente é a de Norberg. Aqui não há dúvidas de que é a distribuição de Sichel que apresenta os valores mais adequados para substituir a escala comercial da Seguradora Y.

Tendo todos estes aspectos em conta, e caso fosse esse o objectivo, teria sido prudente aconselhar a Seguradora Y, na altura em que os dados fornecidos ficaram disponíveis, a utilizar um ajustamento à distribuição de Sichel, descrita em (5.4), para o número de sinistros reportados ao efectuar este género de estudos sobre a sua carteira de seguro automóvel.

6

Conclusão

O objectivo desta dissertação era comparar as estimativas obtidas para a evolução de uma carteira de seguro automóvel através de Sistemas de *Bonus Malus* com uma replicação da realidade sustentada por métodos de Simulação. Os resultados a que se chegou permitiram concluir que existe, de facto, uma grande concordância entre os valores estimados e simulados. No entanto, deve-se ressaltar que esta concordância depende em grande medida dos pressupostos assumidos, que nesta dissertação foram a utilização de uma distribuição de Polya e de uma distribuição de Sichel para a modelação do número de sinistros reportados. Ambos apresentam bons resultados, mas o segundo, pelas razões apresentadas na Secção 5.7, revela-se mais adequado.

A implementação da Simulação neste estudo foi, sem dúvida, trabalhosa e demorada, dado que as corridas são lentas e houve ocasionais problemas computacionais, geralmente devido a falta de memória. Todavia, revelou-se ser um investimento proveitoso, pois permitiu estabelecer uma comparação dos pressupostos mais próxima da realidade, tendo, por fim, permitido proceder a uma selecção mais criteriosa de quais seriam os melhores a adoptar.

O modelo de Simulação traz ainda mais algumas vantagens em relação ao modelo tradicional de evolução de carteiras de SBM. Com efeito, através dele é possível estimar o número de sinistros futuro não só da carteira como um todo, mas também de cada apólice individualmente, o que permite uma maior flexibilidade. Com estas estimativas é, ainda, possível criar intervalos de confiança para a distribuição anual da carteira, prémio anual e escalas de prémios.

Como trabalho futuro, este estudo deveria ser estendido considerando-se um Sistema de *Bonus Malus* em carteira aberta (por exemplo, tal como proposto em Guerreiro et al. (2014) ou Centeno e Andrade e Silva (2001)), de modo a analisar que implicações

traz a entrada e saída de apólices na estacionaridade, nas escalas de prémios e nas diferenças observadas entre modelo aberto de SBM *vs.* Simulação. Uma vez que o número de apólices simuladas é facilmente manipulável, esta análise seria particularmente interessante quando aplicada a carteiras com poucas apólices, quer sejam carteiras jovens ou com perspectivas de crescimento, podendo-se criar, de modo igualmente fácil, análises de sensibilidade.

Durante a execução deste projecto surgiram outros tópicos que seriam interessantes explorar, como uma análise “sinistros ocorridos *vs.* sinistros reportados”, também conhecida como a sede de *bonus* (Lemaire (1995)), que inclui um estudo sobre os sinistros ocorridos mas não reportados (IBNR). Outra extensão possível seria estudar, também, o que aconteceria se se considerasse que o número de sinistros fosse modelado por um processo de Poisson Misto com distribuição de estrutura Log-Normal.

Bibliografia

- Asmussen, S. (2014). "Modeling and Performance of Bonus-Malus Systems: Stationarity versus Age-Correction". *Risks* 2(1), pp. 49–73.
- Banks, J. et al. (1998). *Handbook of Simulation*. Wiley Online Library.
- Borgan, Ø., J. M. Hoem e R. Norberg (1981). "A Non Asymptotic Criterion for the Evaluation of Automobile Bonus System". *Scandinavian Actuarial Journal* 1981(3), pp. 165–178.
- Centeno, M. L. (2003). "Teoria do Risco na Actividade Seguradora". *Celta Editora, Oeiras*.
- Centeno, M. L. e J. Andrade e Silva (2001). "Bonus Systems in Open Portfolio". *Insurance: Mathematics and Economics* 28, pp. 341–350.
- Feng, Y., X. Tong e M. Liu (2007). "Extended Cellular Automata Based Model for Simulating Multi-Scale Urban Growth using GIS". *International Conference on Intelligent Systems and Knowledge Engineering 2007*. Atlantis Press.
- García, F. J., V. Requena e V. Tomás (2008). "Pseudo-Random Sequence Generators Based on Cellular Automata and Bent Functions". *WSEAS Transactions on Computer Research* 3(2), pp. 92–97.
- Gentle, J. E. (1998). *Random Number Generation and Monte Carlo Methods*. Statistics and Computing. Springer: New York.
- Gilde, V. e B. Sundt (1989). "On Bonus Systems with Credibility Scales". *Scandinavian Actuarial Journal* 1989(1), pp. 13–22.
- Grandell, J. (1997). *Mixed Poisson Processes*. Chapman & Hall.
- Guerreiro, G. R., J. T. Mexia e M. de Fátima Miguens (2014). "Statistical Approach for Open Bonus Malus". *Astin Bulletin* 44(01), pp. 63–83.
- Guerreiro, G. R. D. (2001). *Uma Abordagem Alternativa para Bonus Malus*. Tese de Mestrado, Instituto Superior de Economia e Gestão, Universidade Técnica de Lisboa.
- Hillier, F. S. e G. J. Lieberman (2005). *Introduction to Operations Research*. 8ª ed. McGraw-Hill Education: New York.
- Kronmal, R. A. e A. V. Peterson Jr (1981). "A Variant of the Acceptance-Rejection Method for Computer Generation of Random Variables". *Journal of the American Statistical Association* 76(374), pp. 446–451.

- Kronmal, R. A. e A. V. Peterson Jr. (1984). "An Acceptance-Complement Analogue of the Mixture-Plus-Acceptance-Rejection Method for Generating Random Variables". *ACM Trans. Math. Softw.* 10(3), pp. 271–281. ISSN: 0098-3500. DOI: 10.1145/1271.1272. URL: <http://doi.acm.org/10.1145/1271.1272>.
- Leemis, L. M. e S. K. Park (2006). *Discrete-Event Simulation: A first course*. Pearson Prentice Hall Upper Saddle River, NJ.
- Lemaire, J. (1995). *Bonus-malus Systems in Automobile Insurance*. Vol. 19. Springer.
- Morillo, I. e L. Bermúdez (2003). "Bonus-Malus System using an Exponential Loss Function with an Inverse Gaussian Distribution". *Insurance: Mathematics and Economics* 33(1), pp. 49–57.
- Murteira, B. e M. Antunes (2012). *Probabilidades e Estatística*. Vol. I. Escolar Editora.
- Norberg, R. (1976). "A Credibility Theory for Automobile Bonus Systems". *Scandinavian Actuarial Journal* 1976(2), pp. 92–107.
- Parzen, E. (1965). *Stochastic Processes*. Holden-Day.
- Pinquet, J. (1997). "Allowance for Cost of Claims in Bonus-Malus Systems". *Astin Bulletin* 27, pp. 33–58.
- Ross, S. M. (1996). *Stochastic Processes*. Vol. 2. John Wiley & Sons New York.
- Shen, H. C., H. L. Chau e K. K. Wong (1996). "An Extended Cellular Automaton Model for Flexible Manufacturing Systems". *The International Journal of Advanced Manufacturing Technology* 11(4), pp. 258–266.
- Taha, H. A. (2007). *Operations Research: An introduction*. Pearson/Prentice Hall.
- Tremblay, L. (1992). "Using the Poisson Inverse Gaussian in Bonus-Malus Systems". *Astin Bulletin* 22(1), pp. 97–106.
- van der Aalst, W. M. P. e M. Voorhoeve (2000). "Simulation Handbook". *BPM Center Report BPM-00-04*, BPMcenter.org.
- Wolfram Reference - RNG (2014). *Random Number Generation*. URL: <http://reference.wolfram.com/language/tutorial/RandomNumberGeneration.html>.

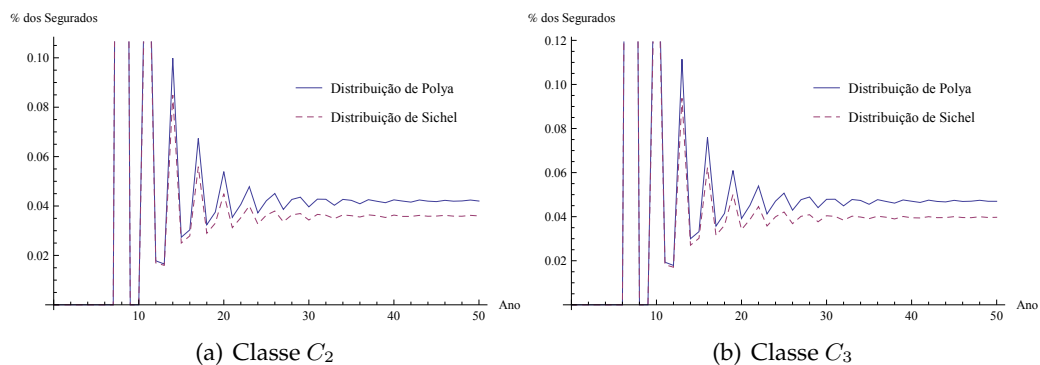


Evolução Anual da Distribuição das Apólices por Classe de *Bonus*

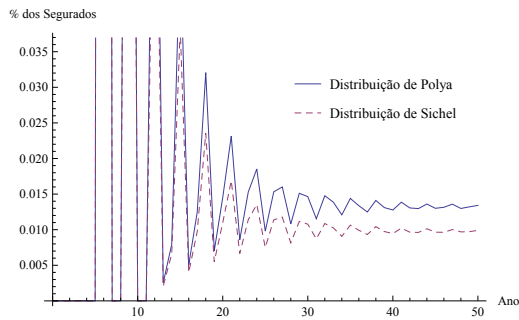
Os gráficos que se seguem mostram a evolução anual da distribuição das apólices da carteira de seguro automóvel da Seguradora Y por classe de *bonus* de acordo com os modelos de Simulação implementados. Os gráficos referentes às classes C_1 e C_{20} não surgem aqui, pois já foram apresentados no corpo principal da dissertação.

Foram considerados, para a distribuição de Polya, um processo de Poisson Misto com distribuição de estrutura Gama ($\hat{\alpha} = 0.6958, \hat{\beta} = 9.9679$) e, para a distribuição de Sichel, um processo de Poisson Misto com distribuição de estrutura IG ($\hat{g} = 0.0698, \hat{h} = 0.0475$).

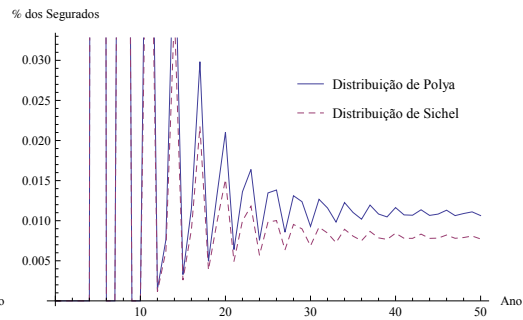
É de salientar que os gráficos não têm todos a mesma escala. Isto acontece devido ao simples facto de cada classe convergir para um valor diferente.



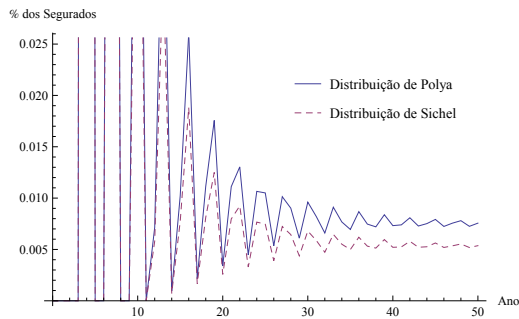
A. EVOLUÇÃO ANUAL DA DISTRIBUIÇÃO DAS APÓLICES POR CLASSE DE *Bonus*



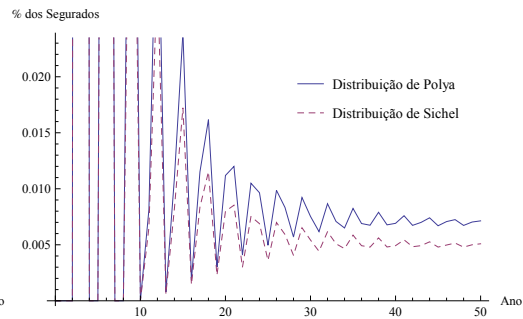
(c) Classe C_4



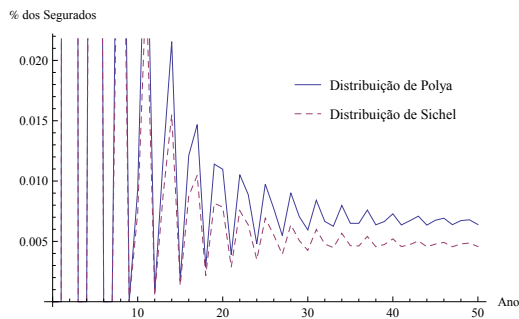
(d) Classe C_5



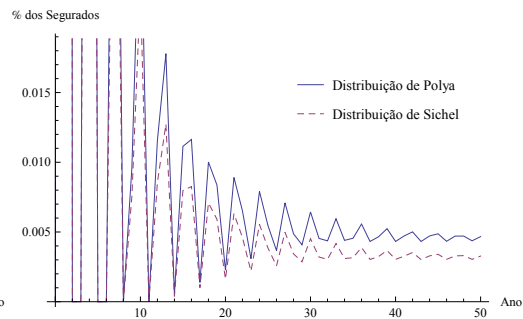
(e) Classe C_6



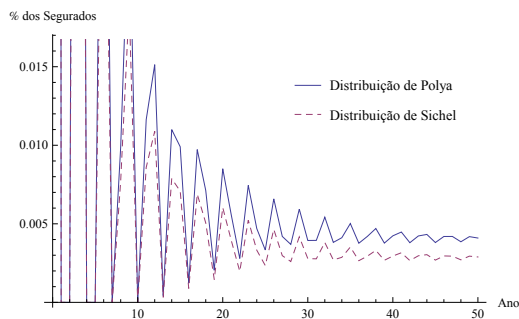
(f) Classe C_7



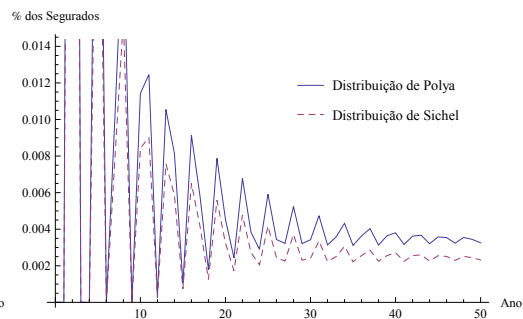
(g) Classe C_8



(h) Classe C_9



(i) Classe C_{10}



(j) Classe C_{11}

A. EVOLUÇÃO ANUAL DA DISTRIBUIÇÃO DAS APÓLICES POR CLASSE DE *Bonus*

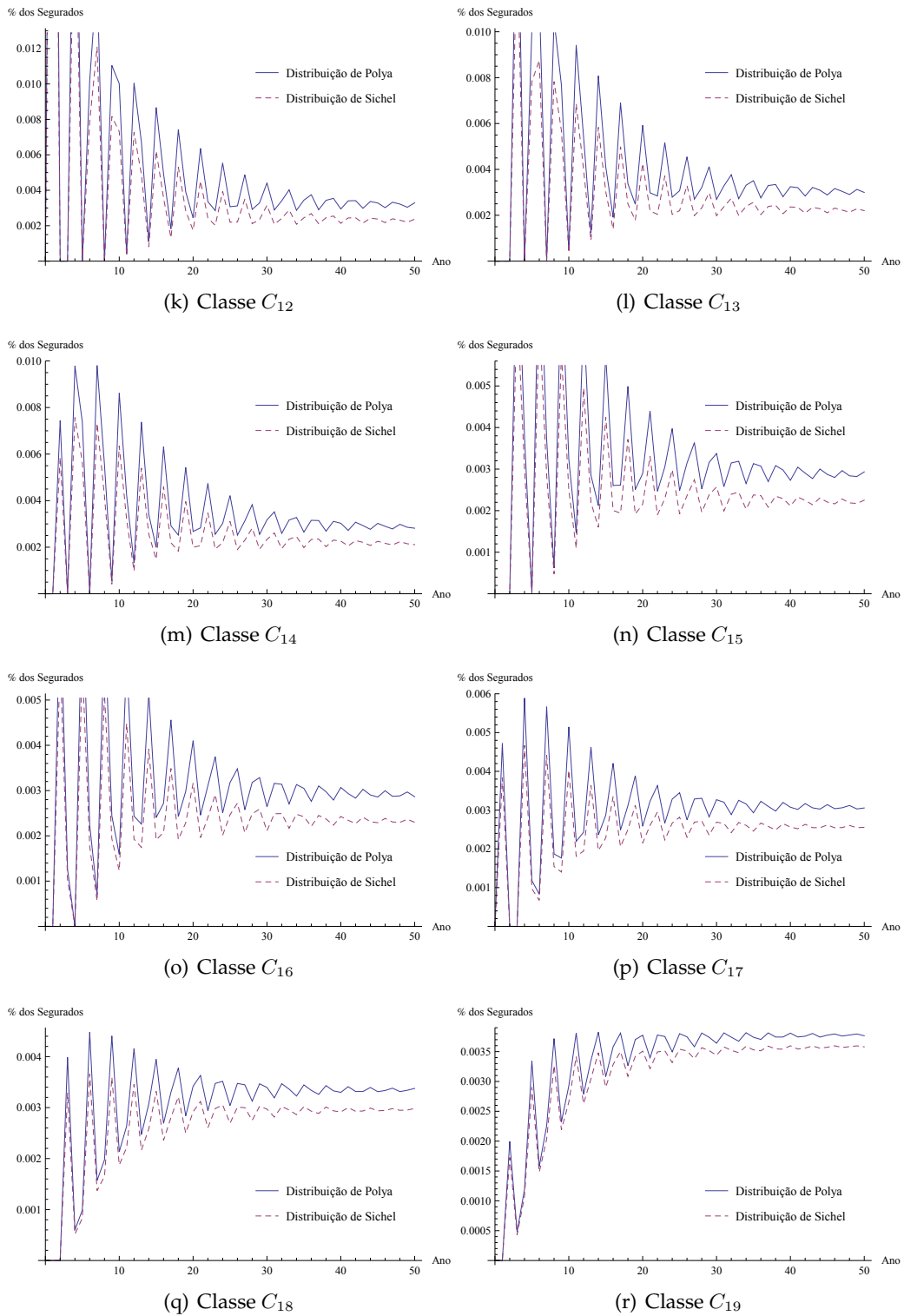


Figura A.1: Evolução da percentagem de apólices por classe



Evolução Anual da Distribuição das Apólices por grupos de Classes de *Bonus* próximas no longo-prazo

Os gráficos apresentados neste anexo ilustram a evolução anual da proporção de apólices em grupos de classes de *bonus* que, no longo-prazo, apresentam valores próximos. Por não serem semelhantes a mais nenhuma e já terem sido analisadas no corpo principal da dissertação, os gráficos referentes às classes C_1 e C_{20} apresentam-se omissos deste anexo.

Para conveniência de interpretação, apresentam-se em pares: do lado esquerdo vê-se a representação gráfica para a Simulação que tomou a distribuição de Polya para modelar o número de sinistros reportados, e do lado direito a Simulação com recurso à distribuição de Sichel.

É de salientar que os gráficos não têm todos a mesma escala. Isto acontece devido ao simples facto de cada grupo de classes convergir para valores pouco similares.

B. EVOLUÇÃO ANUAL DA DISTRIBUIÇÃO DAS APÓLICES POR GRUPOS DE CLASSES DE *Bonus* PRÓXIMAS NO LONGO-PRAZO

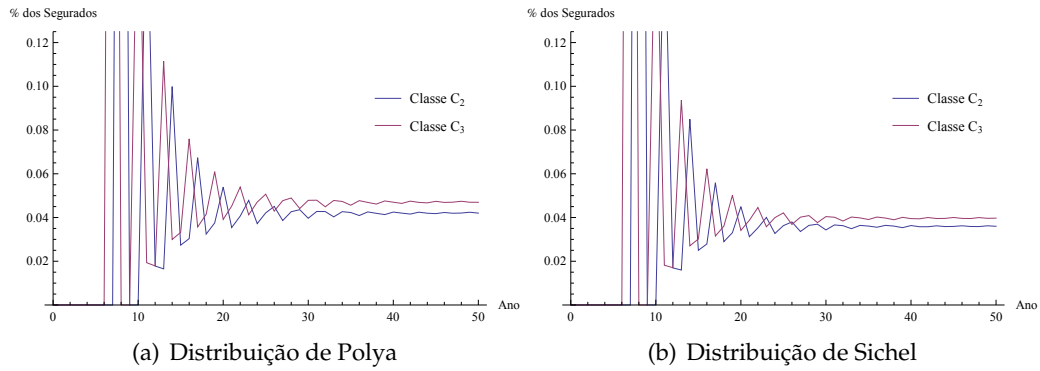


Figura B.1: Evolução da percentagem de apólices nas classes 2 e 3

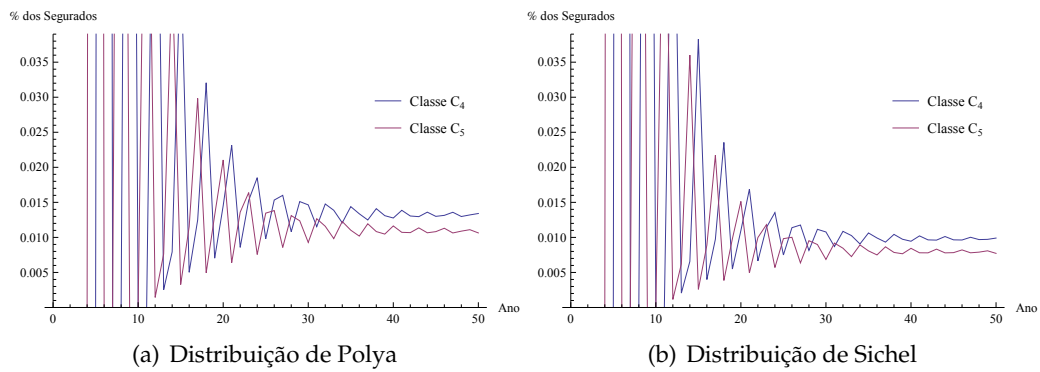


Figura B.2: Evolução da percentagem de apólices nas classes 4 e 5

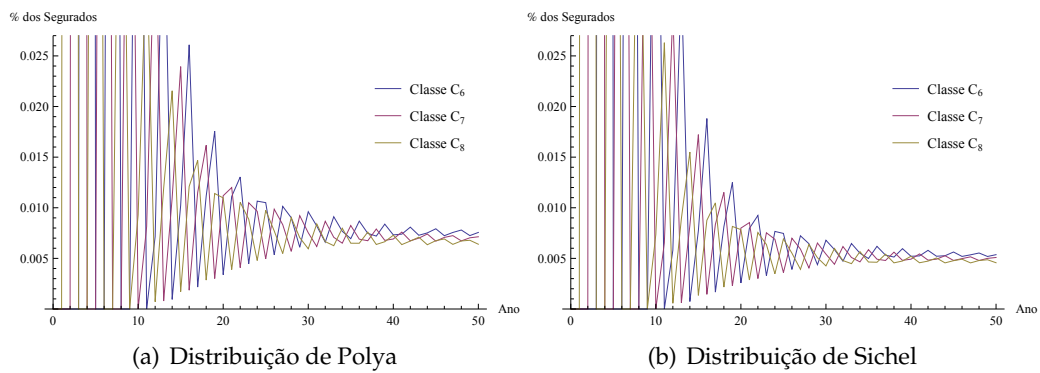


Figura B.3: Evolução da percentagem de apólices nas classes 6 a 8

B. EVOLUÇÃO ANUAL DA DISTRIBUIÇÃO DAS APÓLICES POR GRUPOS DE CLASSES DE *Bonus* PRÓXIMAS NO LONGO-PRAZO

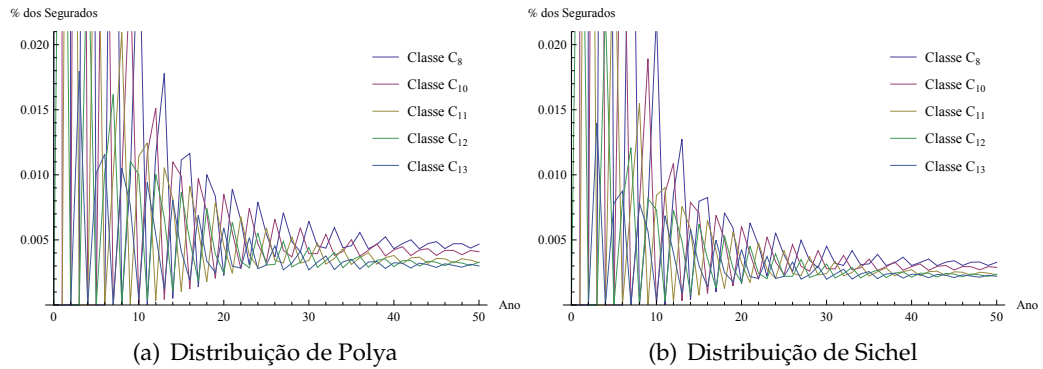


Figura B.4: Evolução da percentagem de apólices nas classes 9 a 13

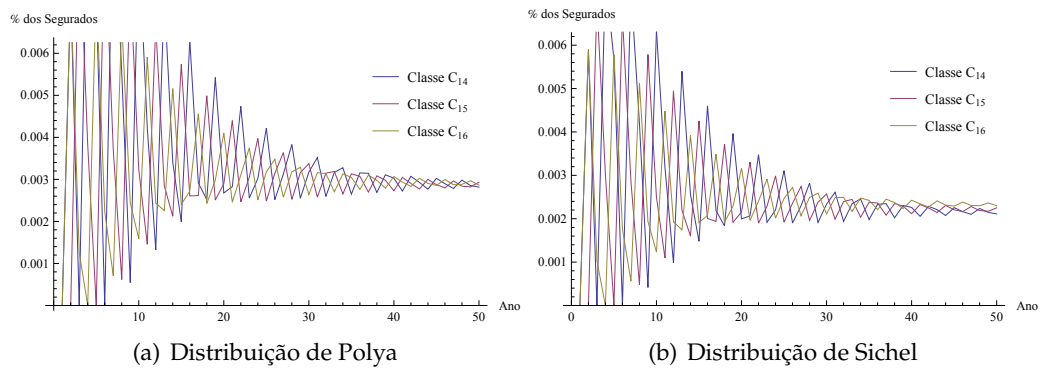


Figura B.5: Evolução da percentagem de apólices nas classes 14 a 16

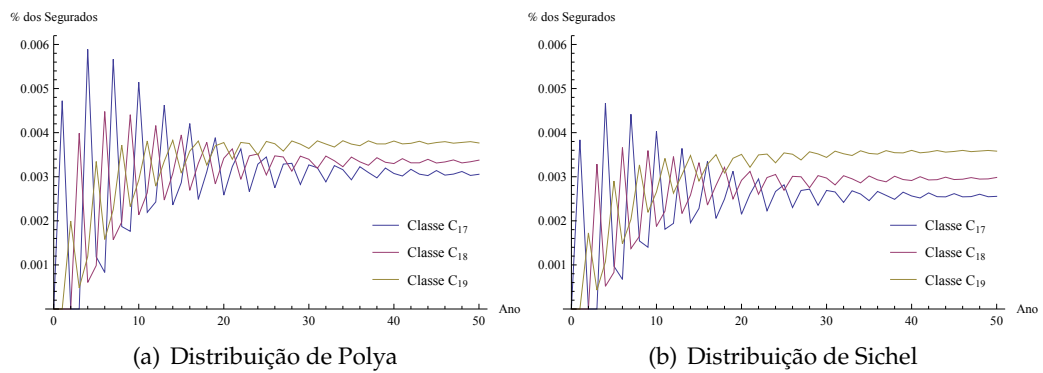


Figura B.6: Evolução da percentagem de apólices nas classes 17 a 19



Comparação das Escalas de Prémios

As tabelas apresentadas de seguida revelam as escalas de prémios obtidas, tanto pelo modelo tradicional de SBM como pelo modelo de Simulação, para ambas as distribuições utilizadas para modelar o número de sinistros na presente dissertação.

Tabela C.1: Escalas de Prêmios segundo a Distribuição de Polya (valores em percentagem do prêmio puro *a priori*)

Classe	Escala Comercial	Escala de Norberg		Escala de Borgan, Hoem e Norberg		Escala de Gilde e Sundt	
		SBM	Sim	SBM	Sim	SBM	Sim
C ₁	50,00	20,24	19,72	26,49	47,93	26,96	30,14
C ₂	55,00	44,92	44,63	38,44	68,92	35,07	37,90
C ₃	60,00	47,00	46,88	41,38	75,60	43,19	45,66
C ₄	65,00	65,74	66,24	41,09	74,96	51,31	53,43
C ₅	70,00	70,09	69,34	43,21	78,35	59,42	61,19
C ₆	75,00	77,53	77,10	44,65	82,09	67,54	68,95
C ₇	80,00	80,04	80,57	46,80	85,69	75,65	76,71
C ₈	85,00	82,31	81,59	49,25	90,20	83,77	84,48
C ₉	90,00	94,51	93,98	51,49	94,83	91,88	92,24
C ₁₀	100,00	100,00	100,00	100,00	100,00	100,00	100,00
C ₁₁	110,00	108,01	106,58	108,29	201,14	108,12	107,76
C ₁₂	120,00	113,54	111,94	119,37	222,27	116,23	115,53
C ₁₃	130,00	119,44	118,08	143,03	265,30	124,35	123,29
C ₁₄	140,00	126,76	124,59	152,62	283,34	132,46	131,05
C ₁₅	155,00	133,60	131,23	162,41	301,81	140,58	138,81
C ₁₆	170,00	141,39	139,67	173,48	321,48	148,70	146,57
C ₁₇	185,00	149,31	146,79	187,93	348,37	156,81	154,34
C ₁₈	200,00	158,11	155,49	208,96	387,64	164,93	162,10
C ₁₉	225,00	167,96	165,99	230,26	426,47	173,04	169,86
C ₂₀	250,00	179,23	176,97	256,17	473,31	181,16	177,62

Tabela C.2: Escalas de Prémios segundo a Distribuição de Sichel (valores em percentagem do prémio puro *a priori*)

Classe	Escala Comercial	Escala de Norberg		Escala de Borgan, Hoem e Norberg		Escala de Gilde e Sundt	
		SBM	Sim	SBM	Sim	SBM	Sim
C ₁	50,00	20,74	20,42	30,86	54,06	26,27	30,57
C ₂	55,00	38,82	38,33	40,45	69,85	34,46	38,29
C ₃	60,00	40,73	40,30	42,92	75,18	42,65	46,00
C ₄	65,00	60,65	61,29	43,12	75,53	50,85	53,71
C ₅	70,00	65,41	65,19	45,06	78,49	59,04	61,43
C ₆	75,00	73,84	73,52	46,45	81,84	67,23	69,14
C ₇	80,00	76,53	77,43	48,54	85,19	75,42	76,86
C ₈	85,00	78,99	78,09	51,01	89,51	83,62	84,47
C ₉	90,00	93,63	93,27	53,50	94,27	91,81	92,29
C ₁₀	100,00	100,00	100,00	100,00	100,00	100,00	100,00
C ₁₁	110,00	109,35	107,74	110,87	195,78	108,19	107,71
C ₁₂	120,00	115,84	114,27	126,17	223,34	116,39	115,43
C ₁₃	130,00	122,86	121,62	158,38	279,47	124,58	123,14
C ₁₄	140,00	131,70	129,42	172,27	304,22	132,77	130,86
C ₁₅	155,00	140,22	137,74	185,90	328,63	140,96	138,57
C ₁₆	170,00	150,18	148,28	202,86	357,01	149,15	146,29
C ₁₇	185,00	160,95	158,28	225,52	395,73	157,35	154,00
C ₁₈	200,00	173,73	170,37	257,52	451,27	165,54	161,71
C ₁₉	225,00	189,34	184,87	294,38	509,62	173,73	169,43
C ₂₀	250,00	209,41	200,71	342,66	578,37	181,92	177,14



Código criado no *Mathematica*

O código de *Mathematica* referido na nota de rodapé 3 da Secção 5.4 encontra-se, aqui, disponibilizado.

Primeiramente foi necessário redefinir a matriz de regras de transição, de forma que o *Mathematica* seja capaz de realizar as transições. Assim, foi criada a matriz **TSim**, apresentada nas Figuras D.1 (input) e D.2 (forma matricial).

Tendo a matriz definida desta forma, as transições ocorrerão onde se encontram valores não negativos, os quais representam o número de sinistros que devem ocorrer para dar lugar a essa transição. No caso de haver apólices a reportar um número de sinistros maior do que aquele que se encontra na última coluna, a transição é efectuada para a classe 20. Nas células onde existe -1 não ocorrerão transições.

Por exemplo, alguma apólice que se encontre na classe 16 transitará para a classe 15 caso não origine sinistros, para a classe 18 caso origine 1 sinistro ou para a classe 20 caso dê origem a 2 ou mais sinistros. A mesma apólice não poderá, no ano seguinte, transitar para as classes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 17, ou 19 nem se manterá na classe 16.

Tendo, agora, a matriz de regras de transição definida, é necessário dar início às transições das apólices entre classes de *bonus*.

Selecione-se, ao acaso, uma qualquer apólice da carteira e sejam

- *classe* - classe onde a apólice se encontra actualmente;
- *nSinistros* - número de sinistros reportados pela apólice até ao final da anuidade corrente;
- **matrizT** - matriz de regras de transição. Por defeito corresponde a **TSim**.

```

TSim = {{0, -1, 1, -1, -1, -1, -1, 2, -1, -1, -1, -1, 3, -1, -1, -1, -1, 4, -1, 5},
        {0, -1, -1, 1, -1, -1, -1, -1, 2, -1, -1, -1, -1, 3, -1, -1, -1, -1, 4, 5},
        {-1, 0, -1, -1, 1, -1, -1, -1, -1, 2, -1, -1, -1, -1, 3, -1, -1, -1, -1, 4},
        {-1, -1, 0, -1, -1, 1, -1, -1, -1, -1, 2, -1, -1, -1, -1, 3, -1, -1, -1, 4},
        {-1, -1, -1, 0, -1, -1, 1, -1, -1, -1, -1, 2, -1, -1, -1, -1, 3, -1, -1, 4},
        {-1, -1, -1, -1, 0, -1, -1, 1, -1, -1, -1, -1, 2, -1, -1, -1, -1, 3, -1, 4},
        {-1, -1, -1, -1, -1, 0, -1, -1, 1, -1, -1, -1, -1, 2, -1, -1, -1, -1, 3, 4},
        {-1, -1, -1, -1, -1, -1, 0, -1, -1, 1, -1, -1, -1, -1, 2, -1, -1, -1, -1, 3},
        {-1, -1, -1, -1, -1, -1, -1, 0, -1, -1, 1, -1, -1, -1, -1, 2, -1, -1, -1, 3},
        {-1, -1, -1, -1, -1, -1, -1, -1, 0, -1, -1, 1, -1, -1, -1, -1, 2, -1, -1, 3},
        {-1, -1, -1, -1, -1, -1, -1, -1, -1, 0, -1, -1, 1, -1, -1, -1, -1, -1, 2, 3},
        {-1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, 0, -1, -1, 1, -1, -1, -1, -1, 2},
        {-1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, 0, -1, -1, 1, -1, -1, -1, 2},
        {-1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, 0, -1, -1, 1, -1, -1, 2},
        {-1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, 0, -1, -1, -1, -1, 2},
        {-1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, 0, -1, -1, -1, 1},
        {-1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, 0, -1, 1},
        {-1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, 0, -1, 1},
        {-1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, 0, 1}} // MatrixForm

```

Figura D.1: Matriz de Regras de Transição (Input)

$$\begin{pmatrix}
 0 & -1 & 1 & -1 & -1 & -1 & -1 & 2 & -1 & -1 & -1 & -1 & 3 & -1 & -1 & -1 & -1 & 4 & -1 & 5 \\
 0 & -1 & -1 & 1 & -1 & -1 & -1 & -1 & 2 & -1 & -1 & -1 & -1 & 3 & -1 & -1 & -1 & -1 & 4 & 5 \\
 -1 & 0 & -1 & -1 & 1 & -1 & -1 & -1 & -1 & 2 & -1 & -1 & -1 & -1 & 3 & -1 & -1 & -1 & -1 & 4 \\
 -1 & -1 & 0 & -1 & -1 & 1 & -1 & -1 & -1 & -1 & 2 & -1 & -1 & -1 & -1 & 3 & -1 & -1 & -1 & 4 \\
 -1 & -1 & -1 & 0 & -1 & -1 & 1 & -1 & -1 & -1 & -1 & 2 & -1 & -1 & -1 & -1 & 3 & -1 & -1 & 4 \\
 -1 & -1 & -1 & -1 & 0 & -1 & -1 & 1 & -1 & -1 & -1 & -1 & 2 & -1 & -1 & -1 & -1 & 3 & -1 & 4 \\
 -1 & -1 & -1 & -1 & -1 & 0 & -1 & -1 & 1 & -1 & -1 & -1 & -1 & 2 & -1 & -1 & -1 & -1 & 3 & 4 \\
 -1 & -1 & -1 & -1 & -1 & -1 & 0 & -1 & -1 & 1 & -1 & -1 & -1 & -1 & 2 & -1 & -1 & -1 & -1 & 3 \\
 -1 & -1 & -1 & -1 & -1 & -1 & -1 & 0 & -1 & -1 & 1 & -1 & -1 & -1 & -1 & 2 & -1 & -1 & -1 & 3 \\
 -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 0 & -1 & -1 & 1 & -1 & -1 & -1 & -1 & 2 & -1 & -1 & 3 \\
 -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 0 & -1 & -1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & 2 \\
 -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 0 & -1 & -1 & 1 & -1 & -1 & -1 & -1 & -1 & 2 \\
 -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 0 & -1 & -1 & 1 & -1 & -1 & -1 & -1 & 2 \\
 -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 0 & -1 & -1 & 1 & -1 & -1 & -1 & 2 \\
 -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 0 & -1 & -1 & 1 & -1 & -1 & 1 \\
 -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 0 & -1 & -1 & 1 \\
 -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 0 & 1
 \end{pmatrix}$$

Figura D.2: Matriz de Regras de Transição (Forma Matricial)

A função *ClasseAnoSeguinteInd*, apresentada na Figura D.3, devolve, com base em *classe* e *nSinistros* e utilizando a matriz **matrizT**, a classe de *bonus* em que a apólice a ser avaliada se encontrará na anuidade seguinte.

```

ClasseAnoSeguinteInd[classe_, nSinistros_, matrizT_ : TSim] :=
Module[{transicoes},
  transicoes = matrizT[[1, classe, All]];
  Position[transicoes, Min[nSinistros, Max[transicoes]]][[1, 1]]
]

```

Figura D.3: Função para determinar a classe no ano seguinte de um segurado

A função *ClasseAnoSeguinte*, exibida na Figura D.4, devolve a classe de *bonus* em que todas as apólices da carteira se encontrarão na anuidade seguinte.

```

ClasseAnoSeguinte[classe_, nSinistros_, matrizT_ : TSim] := Module[{i, classe1},
  For[i = 1, i ≤ Length[classe], i++,
    classe1[i] = ClasseAnoSeguinteInd[classe[[i]], nSinistros[[i]]];
  Table[classe1[i], {i, 1, Length[classe]}]
]

```

Figura D.4: Função para determinar a classe no ano seguinte de todos os segurados

Sejam

- *classes* - vector que contém informação sobre a classe onde cada apólice da carteira se encontra actualmente;
- *nClasses* - número de classes no SBM. Por defeito assume o valor na variável global *s* que, para o caso do SBM da Seguradora Y, tem o valor 20;
- *nSegurados* - número total de segurados na carteira. Por defeito assume o valor na variável global *nSegur* que, para o caso da carteira da Seguradora Y, tem o valor 204 623.

A função *DistAno*, que se pode ver na Figura D.5, efectua uma contagem das apólices em cada classe, devolvendo, assim, a distribuição da carteira num dado ano.

```

DistAno[classes_, nClasses_ : s, nSegurados_ : nSegur] := Module[{j, k, classe},
  Do[classe[k] = Count[classes, k], {k, 1, nClasses}];
  Table[classe[k], {k, 1, nClasses}] / nSegurados // N
]

```

Figura D.5: Função para determinar a distribuição das apólices na carteira num ano

Agora que todas as funções necessárias estão definidas, é necessário aplicá-las aos dados.

Sejam

- *CEntrada* - classe de entrada no SBM. Para o caso do SBM da Seguradora Y, corresponde à classe 10;

- **nSinTotalLambdaIG** - matriz que contém informação sobre o número total de sinistros reportados por cada apólice com uma dada frequência de sinistralidade até ao fim de cada anuidade, considerando que o número de sinistros segue uma distribuição de Poisson Mista com distribuição de estrutura IG ($\hat{g} = 0.0698, \hat{h} = 0.0475$);
- *nAnosTotal* - número total de anos de sinistros. Para o caso das análises efectuadas nesta dissertação, são considerados 50 anos.

A Figura D.6 ilustra de que forma as funções criadas podem ser usadas para produzir as transições das apólices entre classes de *bonus* e obter a distribuição anual da carteira.

```
classesLambdaIG[1] = Table[CEntrada, {i, nSegur}];  
distLambdaIG[1] = DistAno[classesLambdaIG[1]];  
Do[  
  classesLambdaIG[j] =  
    ClasseAnoSeguinte[classesLambdaIG[j - 1], nSinTotalLambdaIG[[All, j - 1]]];  
  distLambdaIG[j] = DistAno[classesLambdaIG[j]];  
  , {j, 2, nAnosTotal}];
```

Figura D.6: Aplicação das funções criadas

Neste exemplo são criadas as transições de todas as apólices com uma determinada frequência de sinistralidade entre classes de *bonus* num horizonte de 50 anos, assumindo que a distribuição de estrutura do processo de Poisson Misto é a Inversa Gaussiana. A classe em que as apólices se encontram em cada ano é guardada na matriz **classesLambdaIG**. Simultaneamente é também guardada a distribuição anual das apólices pela carteira na matriz **distLambdaIG**.

Evidentemente, escusado será dizer que este código deverá ser corrido num ciclo que inclua todas as apólices da carteira, isto é, que percorra todos os valores considerados para a frequência de sinistralidade dos segurados existentes na carteira de seguro automóvel da Seguradora.