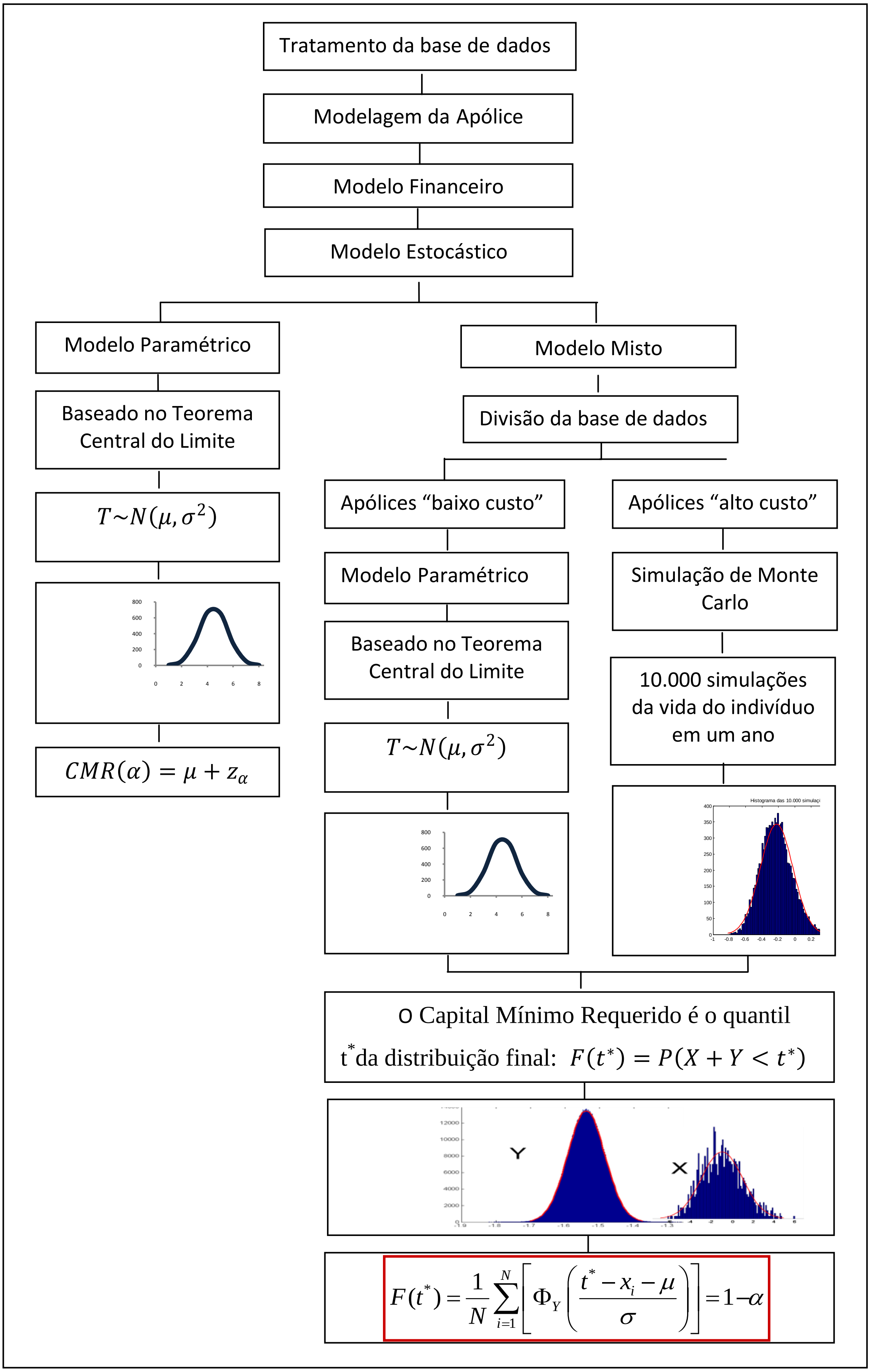
3 Metodologia

Um modelo interno é capaz de calcular o capital requerido que a seguradora necessita para se manter solvente por um período pré-fixado. Então, para este fim a metodologia foi baseada na determinação da distribuição de probabilidades do total líquido gasto com sinistros, pois o capital mínimo requerido é um quantil desta distribuição. Para isto cada apólice deve evoluir pelo período de um ano. Esta evolução será modelada como uma cadeia de Markov. Em seguida foi feita a derivação analítica da média e da variância do custo total com sinistros. Foram utilizados dois métodos o paramétrico e o não paramétrico. O método paramétrico foi baseado no teorema central do limite, desta forma o total líquido gasto com sinistros seguiria aproximadamente uma distribuição Normal. No entanto, esta aproximação pode ser inadequada para descrever a cauda da distribuição. Uma alternativa seria utilizar simulação de Monte Carlo. O problema de fazer uma simulação de Monte Carlo seria o custo computacional em simular uma carteira de mais de 250.000 apólices. Assim, outra proposta encontrada foi a de utilizar um método misto, dividindo a carteira em dois grupos. Neste ponto podemos classificar as apólices como apólices de "custo alto" e apólices de "baixo custo". As de "alto custo" são as apólices que provavelmente trariam grande custo com sinistro para empresa e as de "baixo custo" aquelas que provavelmente trariam menor custo. Assim, o grupo das apólices que são classificadas como "baixo custo" pode ser trato analiticamente pela aproximação pela Normal e o grupo das apólices de "alto custo" é tratado por Simulação de Monte Carlo. A distribuição final do custo total com sinistros é obtida pela convolução destas duas distribuições. O capital mínimo requerido é obtido pelo quantil 99,9%. A ilustração com os dados reais evidencia no método misto o peso da cauda da distribuição influenciando no resultado com um capital requerido maior. Com o objetivo de facilitar a compreensão do método segue o fluxograma para uma primeira visão geral do que será explicado em detalhes no decorrer deste capítulo.

Figura 1 - Fluxograma do método.



3.1
Base de dados

Três bases foram fornecidas para a ilustração do modelo. A base individual, a base grupo e a base de previdência. Estas serão chamadas, respectivamente, de base 1, 2 e 3 na dissertação. Estas bases precisaram de alguns pré-tratamentos. Dentre eles pode-se citar o problema da existência de algumas divergências nas relações de nome de cobertura e sinistro. Nas bases originais foram encontrados mais de um nome para um único tipo de sinistro e o mesmo ocorria para a nomenclatura das coberturas. Para solucionar a questão foi feita uma lista de sinistros e coberturas comuns a todas as apólices e a relação foi refeita de forma mais apurada. Novas nomenclaturas foram definidas para sinistros e coberturas. Outra observação importante é que dada a ocorrência de um sinistro algumas coberturas devem ser ativadas no sistema. Como cada cobertura possui um valor de capital segurado estes valores devem ser contabilizados formando o valor a ser pago pelo sinistro ocorrido. A tabela um apresenta a lista de coberturas identificadas nas bases e há a indicação com o número um para orientar em que base os tipos de coberturas aparecem. De forma semelhante encontramos na tabela dois a lista dos sinistros e as bases em que foram encontrados.

Tabela 1- Relação entre coberturas e bases.

Códigos coberturas	Coberturas	Base 1	Base 2	Base 3
1	Auxílio alimentação	1	0	0
2	Auxílio Funeral	1	0	0
3	Auxílio funeral dedutível	1	0	0
4	Despesas médicas hospitalares e odondológicas - AP	1	0	0
5	Despesas médicas hospitalares	1	0	1
6	Indenização especial acidente	1	1	1
7	Indenização especial por acidente do cônjuge	1	0	1
8	Invalidez funcional permanente total por doença - ANTEC	1	1	0
9	Invalidez permanente total por doença	1	1	1
10	Invalidez permantente total/parcial por acidente do cônjuge	1	0	1
11	Invalidez permanente total ou parcial por acidente	1	1	1
12	Invalidez permanente total por acidente	1	0	0
13	Morte	1	1	1
14	Morte cônjuge	1	1	1
15	Morte filhos	1	1	1
16	SAF familiar	1	1	0
17	SAF titular	1	1	0
18	Salário por incapacidade temporária	1	0	1
19	Invalidez permanente total por doença do cônjuge	0	0	1
20	Invalidez laboral permanente	0	0	1

Tabela 2- Relação entre sinistros e bases.

Códigos Sinistro	Sinistros	Base 1	Base 2	Base 3
1	Invalidez funcional permanente total por doença	1	1	0
2	Invalidez permanente parcial do cônjuge por acidente	1	0	1
3	Invalidez permanente parcial por acidente	1	1	1
4	Invalidez permanente total do cônjuge por acidente	1	0	1
5	Invalidez permanente total por acidente	1	1	1
6	Invalidez permanente total por doença	1	1	1
7	Invalidez temporária por acidente	1	0	1
8	Morte por acidente	1	1	1
9	Morte por acidente do cônjuge	1	0	1
10	Morte qualquer causa	1	1	1
11	Morte qualquer causa do cônjuge	1	1	1
12	Morte qualquer causa do filho	1	1	1
13	Qualquer tipo de acidente	1	0	1
14	Invalidez permanente total por doença do cônjuge	0	0	1
15	Invalidez laboral permanente	0	0	1

Estas tabelas foram úteis para a separação e adequação das nomenclaturas entre as bases, uma vez que elas formarão uma única base a ser trabalhada. Com o fim de unificar as bases de forma satisfatória foi realizado um estudo detalhado e uma adequação da base de dados para o modelo. É importante verificar se as variáveis que serão necessárias para o cálculo estão bem definidas e se os valores carregados estão corretos, ou se precisam de algum tipo de tratamento para que reflita o valor real que deve ser utilizado para o cálculo do capital mínimo requerido. Por exemplo, em uma das bases utilizadas, a variável prêmio estava duplicada, levando a uma superestimação da realidade. Este problema nos levaria a constatar que o valor recebido de prêmio seria muito maior do que o real valor.

As coberturas estão relacionadas com os tipos de sinistro. Isto acontece porque com a ocorrência de um sinistro os valores, em dinheiro, referentes às coberturas acionadas são pagos aos segurados. Assim, podemos dizer que um percentual do capital segurado é pago ao segurado pela ocorrência do sinistro. Um exemplo disto é que nos casos de sinistros por invalidez parcial estima se em média um pagamento de 32% do valor segurado para a invalidez total. Este tipo de informação deve ser levado em consideração ao calcular os valores segurados para a base de dados. Uma vez que cada cobertura contratada possui um valor de capital segurado o total em dinheiro a ser pago pela ocorrência de determinado sinistro corresponde à soma dos capitais segurados referentes a cada cobertura que for acionada. Estes valores devem ser calculados pelo capital unitário, isto serve para os casos em que a apólice for referente ao seguro em grupo, para que o valor a ser pago possa ser contabilizado corretamente por indivíduo. A base unificada de dados deve ser preparada para ter o layout apresentado na tabela três. Levando em consideração todos os tratamentos citados até então e alguma peculiaridade que a base de dados usada possa trazer para o caso de uma aplicação do modelo a outro tipo de base de dados.

Tabela 3- Layout da base de dados preparada.

Variáveis para o layout
Indivíduo
Idade
Sexo
Valor pago mensalmente de prêmio pelo indivíduo
Valor que deverá ser pago se o indivíduo se encontrar no estado fim (extinção do contrato)
Valor que deverá ser pago se o indivíduo se encontrar no estado válido
Valor que deverá ser pago se o indivíduo se encontrar no estado inválido
Valor que deverá ser pago se o sinistro de Invalidez funcional permanente total por doença ocorrer
Valor que deverá ser pago se o sinistro de Invalidez permanente parcial do cônjuge por acidente ocorrer
Valor que deverá ser pago se o sinistro de Invalidez permanente parcial por acidente ocorrer
Valor que deverá ser pago se o sinistro de Invalidez permanente total do cônjuge por acidente ocorrer
Valor que deverá ser pago se o sinistro de Invalidez permanente total por acidente ocorrer
Valor que deverá ser pago se o sinistro de Invalidez permanente total por doença ocorrer
Valor que deverá ser pago se o sinistro de Invalidez temporária por acidente ocorrer
Valor que deverá ser pago se o sinistro de Morte por acidente ocorrer
Valor que deverá ser pago se o sinistro de Morte por acidente do cônjuge ocorrer
Valor que deverá ser pago se o sinistro de Morte qualquer causa ocorrer
Valor que deverá ser pago se o sinistro de Morte qualquer causa do cônjuge ocorrer
Valor que deverá ser pago se o sinistro de Morte qualquer causa do filho ocorrer
Valor que deverá ser pago se o sinistro de Qualquer tipo de acidente ocorrer
Valor que deverá ser pago se o sinistro de Invalidez permanente total por doença do cônjuge ocorrer
Valor que deverá ser pago se o sinistro de Invalidez laboral permanente ocorrer

Os campos criados com valores para estado fim, estado válido e estado inválido devem ser preenchidos com valor zero, pois estas variáveis ajudarão na construção da Cadeia de Markov para os indivíduos, mas não há de ser pago nenhum valor aos segurado pelo acontecimento destes eventos. Já os outros campos devem ser preenchidos com as devidas informações retiradas da base de dados previamente tratada. As variáveis sexo e idade são muito importantes para o processo, pois os cálculos dependerão destas informações. Será observado mais adiante que para o cálculo da média e variância do total líquido gasto com sinistro a base deve ser agrupada por sexo e idade. Assim, retira-se a variável indivíduo somando os valores por sexo e idade. É importante ressaltar que no cálculo da média da distribuição este layout será utilizado somando o capital segurado, sem ser utilizado o capital unitário. O programa foi preparado para fazer o cálculo corretamente desta forma, mas no caso do cálculo da variância é importante verificar que se não utilizar o capital unitário a conta será feita superestimando o valor final, por causa da quantidade de segurados dentro de uma mesma apólice, nos casos de apólice em grupo. Por este motivo foram dados tratamentos ligeiramente diferenciados na programação para a base que servirá para o cálculo da média e a base que servirá para o cálculo da variância. A diferença esta no uso do capital segurado para uma e no uso do capital unitário para a outra.

3.2 Modelagem da Apólice

A apólice deve ser modelada como uma relação entre tipos de sinistro e tipos de cobertura. Este é o primeiro passo, após o tratamento da base, do modelo. É importante que os sinistros sejam separados como eventos disjuntos. Este procedimento foi visto na seção anterior. Após a distinção de todos os tipos de sinistros que podem ser cobertos torna-se possível relacioná-los com os tipos de cobertura oferecidos. Lembrando que uma cobertura pode ser acionada por mais de um sinistro. O exemplo dado abaixo é para o caso específico da base utilizada para aplicação do método. Então, a tabela quatro mostra a lista de coberturas possíveis acionadas por cada sinistro possível da apólice utilizada pela seguradora.

Tabela 4- Relação entre sinistros e coberturas.

	Coberturas																			
	Auxílio alimentação	Auxílio Funeral	Auxílio funeral dedutível	Despesas médicas hospitalares e odondológicas - AP	Despesas médicas hospitalares	Indenização especial acidente	Indenização especial por acidente do cônjuge	Invalidez funcional permanente total por doença - ANTEC	Invalidez permanente total por doença	Invalidez permantente total/parcial por acidente do cônjug	Invalidez permanente total ou parcial por acidente	Invalidez permanente total por acidente	Morte	Morte cônjuge	Morte filhos	SAF familiar	SAF titular	Salário por incapacidade temporária	Invalidez permanente total por doença do cônjuge	Invalidez laboral permanente
Sinistros	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Invalidez funcional permanente total por doença								x												
Invalidez permanente parcial do cônjuge por acidente										x										
Invalidez permanente parcial por acidente											x									
Invalidez permanente total do cônjuge por acidente										x										
Invalidez permanente total por acidente											x	x								
Invalidez permanente total por doença									x											
Invalidez temporária por acidente																		x		
Morte por acidente	x	x	x			x							x				x			
Morte por acidente do cônjuge							x							x		x				
Morte qualquer causa	x	x	x										x				x			
Morte qualquer causa do cônjuge														x		x				
Morte qualquer causa do filho															x	x				
Qualquer tipo de acidente				x	x															
Invalidez permanente total por doença do cônjuge																			x	
Invalidez laboral permanente																				x

A tabela acima é uma tabela geral, onde todos os sinistros possíveis estão relacionados com todas as apólices possíveis. Porém, neste método, modelar a apólice significa construir uma tabela destas para cada apólice. A relação do tipo de sinistro com a cobertura deve ser definida para cada segurado. Para isto, observar-se no documento contratual, ou na própria base de dados as coberturas escolhidas pelo segurado, marcando em sua tabela os sinistros aos quais ela dá cobertura. Esta modelagem auxiliará na construção dos vetores de estado e cobertura do titular, do cônjuge e dos filhos conforme será descrito nas seções 3.5 e 3.6.

3.3
Pressupostos do Modelo

A construção de um modelo requer o uso de simplificações e suposições da realidade. Descrever fielmente todos os aspectos envolvidos e todas as variáveis internas e externas que afetam os sistemas modelados é uma tarefa intangível. Assim, algumas simplificações são feitas de forma a se obter uma melhor compreensão do fenômeno. Geralmente, um modelo utiliza informações extraídas do banco de dados da companhia para os cálculos. Algumas informações sobre as apólices individuais e em grupo não estão disponíveis nas bases de dados estudadas. Por este motivo algumas premissas foram adotadas.

- Toda apólice cobre: um titular, um cônjuge e dois filhos
- A idade do cônjuge será a do titular mais quatro anos, se este for do sexo feminino e menos quatro anos, se este for do sexo masculino;
- Os filhos são do sexo masculino e de catorze anos de idade;
- Todos os titulares, cônjuges e filhos iniciam o período de análise com a apólice ativa e no estado válido;
- A ocorrência de sinistros por segurado é mutuamente excludente. Ou seja, em um dado mês, apenas um sinistro poderá ocorrer por segurado;

3.4
Fatores de risco

Os fatores de risco no modelo proposto serão:

- A ocorrência de sinistro com o titular, cônjuge ou filhos;
- Idade e sexo

- A decisão de interromper o pagamento dos prêmios, extinguindo a apólice;
- Taxa de desconto utilizada para trazer os fluxos financeiros, sinistro menos prêmio, a valor presente.

3.5
Estados do indivíduo

Em cada apólice pode existir apenas um titular, caso a apólice seja individual ou mais de um titular, caso a apólice seja de seguro em grupo. Cada titular possui um cônjuge e dois filhos, como foi definido nas premissas. Portanto, a partir deste momento, quando fizer referência aos segurados deve ficar claro que são: titular, cônjuge e filhos. Um segurado (titular, cônjuge ou filhos) deverá estar a cada final de mês em um determinado estado, o qual indicará como se encontram os segurados desta apólice. Cada um dos segurados deve estar em uma das seguintes situações: válido, inválido ou em algum tipo de invalidez específica ou, no caso do titular, em fim de vigência. Desta forma, os possíveis estados do titular são definidos pela tabela cinco.

Tabela 5- Estados do titular.

Código do estado	Código do sinistro	Tipos de estados
1	-	Fim
2	-	Válido
3	-	Inválido
4	1	Invalidez funcional permanente total por doença
5	3	Invalidez permanente parcial por acidente
6	5	Invalidez permanente total por acidente
7	6	Invalidez permanente total por doença
8	7	Invalidez temporária por acidente
9	8	Morte por acidente
10	10	Morte qualquer causa
11	13	Qualquer tipo de acidente
12	15	Invalidez laboral permanente

Esta informação pode ser representada pelo vetor aleatório abaixo, que representará o estado do titular da apólice ***i*** no mês ***t***.

$$e'_{i,t} = \left(e_{i,t,1}, e_{i,t,2}, \ldots, e_{i,t,12}\right) \quad (3.1)$$

Cada componente deste vetor aleatório é uma variável indicadora que vale zero ou um, indicando ou não a ocorrência do evento (estado) para um titular. Como os estados são mutuamente excludentes em cada mês t apenas uma das componentes valerá um. Analogamente, os estados possíveis do cônjuge são definidos na tabela seis.

Tabela 6- Estados do cônjuge.

Código do estado	Código do sinistro	Tipos de estados
1	-	Fim
2	-	Válido
3	-	Inválido
4	2	Invalidez permanente parcial do cônjuge por acidente
5	4	Invalidez permanente total do cônjuge por acidente
6	9	Morte por acidente do cônjuge
7	11	Morte do cônjuge por qualquer causa
8	14	Invalidez permanente total por doença do cônjuge

O vetor que representa o estado do cônjuge do titular da apólice i no mês t é dado por:

$$\boldsymbol{f}'_{i,t} = \left(\boldsymbol{f}_{i,t,1}, \boldsymbol{f}_{i,t,2}, \dots, \boldsymbol{f}_{i,t,8}\right) \qquad (3.2)$$

A mesma definição foi aplicada aos estados possíveis dos filhos como pode ser observado na tabela sete.

Tabela 7- Estados dos filhos.

Código do estado	Código do sinistro	Tipos de estados
1	-	Fim
2	-	Válido
3	-	Inválido
4	12	Morte do filho por qualquer causa

A partir dos quais se definem os vetores de estado a seguir:

$$\boldsymbol{g}'_{i,t} = \left(\boldsymbol{g}_{i,t,1}, \boldsymbol{g}_{i,t,2}, \dots, \boldsymbol{g}_{i,t,4}\right) \ \boldsymbol{e} \ \boldsymbol{h}'_{i,t} = \left(\boldsymbol{h}_{i,t,1}, \boldsymbol{h}_{i,t,2}, \dots, \boldsymbol{h}_{i,t,4}\right) \qquad (3.3)$$

Estes vetores representam, respectivamente, o estado do filho um e o estado do filho dois do titular da apólice i no mês t.

3.6 Coberturas e Prêmios

Existe um possível valor a ser a ser pago ao segurado relacionado com cada estado que ele pode assumir. Este valor é formado pela soma dos capitais segurados, referente às coberturas que cada tipo de sinistro aciona ao ocorrer. Esta relação pode ser observada na tabela quatro, como foi visto anteriormente. Este valor será nulo, caso o estado não esteja representando um tipo de sinistro como é o caso dos estados válido, inválido ou fim de vigência. Foram criados vetores que serão compostos por estes valores em dinheiro correspondentes a cada estado. Estes vetores de valores a serem pagos, que serão chamados de vetores de coberturas são definidos a seguir. O vetor correspondendo aos valores relacionados ao titular *i* é dado por:

$$\boldsymbol{c}_i' = \left(\boldsymbol{c}_{i,1}, \boldsymbol{c}_{i,2}, \dots, \boldsymbol{c}_{i,12}\right) \tag{3.4}$$

Correspondendo ao cônjuge:

$$\boldsymbol{b}_i' = \left(\boldsymbol{b}_{i,1}, \boldsymbol{b}_{i,2}, \dots, \boldsymbol{b}_{i,8}\right) \tag{3.5}$$

Correspondendo aos filhos, tem-se:

$$\boldsymbol{a}_i' = \left(\boldsymbol{a}_{i,1}, \boldsymbol{a}_{i,2}, \dots, \boldsymbol{a}_{i,4}\right) \tag{3.6}$$

Conseqüentemente, cada apólice i, possui um valor corresponde ao recebimento de um prêmio a ser pago a cada mês, representado pela variável r_i enquanto a apólice estiver em vigência.

3.7
Modelo financeiro

Cada elemento das fórmulas que serão apresentadas a seguir está definido na lista de símbolos, nos elementos pré-textuais desta dissertação. Assim como os cálculos mais detalhados estão no anexo 8.1.

O modelo financeiro será uma função matemática que calculará o capital necessário para fazer frente a uma lista de sinistros ocorridos em dada carteira de contratos. O montante necessário hoje para pagar todos os sinistros é igual ao valor presente dos sinistros pagos ao longo do ano, subtraído do valor presente dos prêmios recebidos. Assim, o montante T será o valor total líquido a ser gasto com sinistros, o qual é a soma para cada um dos doze meses descontados pelo fator beta:

$$T = \sum_{t=1}^{12} \beta_t (S_t - P_t) \tag{3.7}$$

$$\beta_t = \frac{1}{\prod_{i=1}^{t} (1 + x_i)} \tag{3.8}$$

Onde x_i indica os juros anual. O total de sinistros menos os prêmios em um dado mês pode ser composto como abaixo, onde i indica cada apólice e N é o total de apólices:

$$S_t - P_t = \sum_{i=1}^{N} (S_{it} - P_{it}) \tag{3.9}$$

O montante em cada mês é a soma dos sinistros ocorridos para todas as apólices, subtraindo os respectivos prêmios pagos à seguradora:

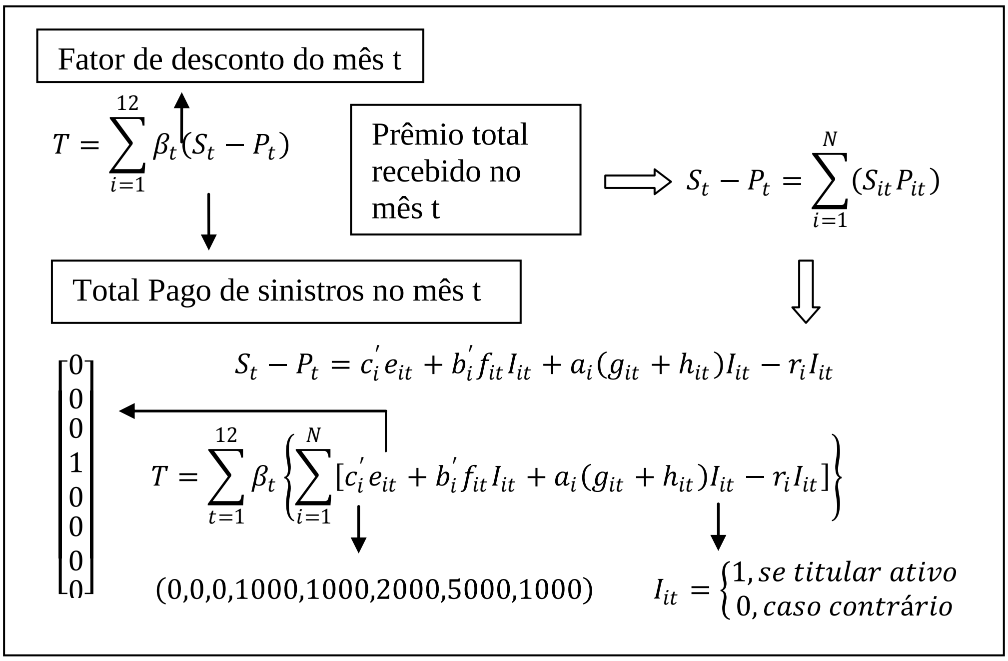
$$S_{it} - P_{it} = c'_i e_{it} + b'_i f_{it} I_{it} + a'_i (g_{it} + h_{it}) I_{it} - r'_i \mathrm{I}_{it} \tag{3.10}$$

O valor mensal gasto com sinistro será dado pela multiplicação dos vetores indicadores de estado com seus respectivos vetores de valores a serem pagos. Nos casos dos dependentes do titular existirá a multiplicação pelo vetor indicador de atividade da apólice, que só permitira a contabilização do valor a ser pago de sinistro caso o titular ainda esteja com a apólice ativa naquele mês. O mesmo ocorre para o vetor de valores recebidos mensalmente de prêmio. Assim, a expressão para avaliação do capital mínimo é dada por:

$$T = \sum_{t=1}^{12} \beta_t \left\{ \sum_{i=1}^{N} [c_i^{'} e_{it} + b_i^{'} f_{it} I_{it} + a_i^{'} (g_{it} + h_{it}) I_{it} - r_i^{'} I_{it}] \right\} \tag{3.11}$$

Esta expressão permite calcular o total a ser reservado hoje para assegurar o pagamento dos sinistros ao longo de doze meses a partir de um cenário futuro dos sinistros, representado nos estados dos segurados, além da evolução futura da taxa de juros. Observe que esta mesma expressão será também utilizada na etapa de obtenção da *valuation* estocástica. Nesta etapa utiliza-se o modelo estocástico para as variáveis de risco, o qual será desenvolvido na seção a seguir. Este modelo fornece expressões analíticas para a média e a variância de T, a partir das quais se obtém uma aproximação para a distribuição de probabilidade de T, valor total líquido a ser gasto com sinistros. A figura abaixo pretende facilitar a compreensão do modelo financeiro exposto acima.

Figura 2 – Modelo financeiro para o total líquido gasto com sinistro.



3.8 Modelo estocástico

Os modelos estocásticos são mais informativos e capturam não somente a aleatoriedade de acontecimentos futuros, mas também as incertezas nas suposições e processos. Porém, o modelo deve ser parcimonioso porque quanto maior o número de variáveis estocásticas envolvido, mais complexo será o modelo e maior será o tempo de processamento. Este modelo nos permitirá tratar as probabilidades e a aleatoriedade relacionadas à carteira trabalhada.

3.8.1 Cadeia de Markov

Como foi dito anteriormente, os estados que os segurados assumem a cada instante de tempo serão modelados utilizando cadeias de Markov. Uma cadeia de Markov é definida como um processo estocástico cujos estados são discretos e possui a propriedade Markoviana. A memória markoviana indica que os estados anteriores são irrelevantes para a predição dos estados seguintes, para isto o estado atual deve ser conhecido. A cadeia de Markov será uma sequência de variáveis aleatórias indexadas no tempo. O conjunto de valores que estas variáveis podem assumir é chamado de espaço de estados. O espaço de estados será representado pela matriz de transição. Para o caso de um espaço de estados discretos, as probabilidades de transição de *t* passos a frente são obtidas pelo cálculo da *t*-ésima potência da matriz de transição. Isto é, se $\mathrm{P}_{\mathrm{s(i),d(i)}}$ é a matriz de transição para um passo, então $P^{t}_{s(i),d(i)}$ é a matriz de transição para a transição de *t* passos. Desta forma, sejam $\mathrm{E}_{\mathrm{i,t}}, \mathrm{F}_{\mathrm{i,t}}, \mathrm{G}_{\mathrm{i,t}}, \mathrm{H}_{\mathrm{i,t}}$, variáveis aleatórias discretas que representam os estados dos segurados da apólice *i* no instante *t* com domínios respectivamente definidos por, $R_E = \{1, \dots, 12\}, R_F = \{1, \dots, 8\}\ e\ R_G = R_H = \{1, \dots, 4\}$. Estas variáveis indicam o estado do indivíduo ao receber o número "um" em determinada componente dos vetores $e_{i,t}, f_{i,t}, g_{i,t}, h_{i,t}$, como abaixo, onde k é o número que indica o estado:

$$\mathrm{E}_{\mathrm{i,t}} = \mathrm{k} \in \mathrm{R}_{\mathrm{E}} \leftrightarrow \mathrm{e}_{\mathrm{i,t,k}} = 1 \text{ e } \mathrm{E}_{\mathrm{i,t}} \neq \mathrm{k} \in \mathrm{R}_{\mathrm{E}} \leftrightarrow \mathrm{e}_{\mathrm{i,t,k}} = 0$$

$$\mathrm{F}_{\mathrm{i,t}} = \mathrm{k} \in \mathrm{R}_{\mathrm{F}} \leftrightarrow \mathrm{f}_{\mathrm{i,t,k}} = 1 \text{ e } \mathrm{F}_{\mathrm{i,t}} \neq \mathrm{k} \in \mathrm{R}_{\mathrm{F}} \leftrightarrow \mathrm{f}_{\mathrm{i,t,k}} = 0$$

$$G_{i,t} = k \in R_G \leftrightarrow g_{i,t,k} = 1 \text{ e } G_{i,t} \neq k \in R_G \leftrightarrow g_{i,t,k} = 0$$

$$H_{i,t} = k \in R_H \leftrightarrow h_{i,t,k} = 1 \text{ e } H_{i,t} \neq k \in R_H \leftrightarrow h_{i,t,k} = 0$$

Assim, a evolução temporal dos estados $E_{i,t}, F_{i,t}, G_{i,t}, H_{i,t}$ ao longo do horizonte de doze meses será modelado por quatro cadeias de Markov independentes entre si com estados iniciais iguais a válido para todos os segurados, definidos por:

$$E_{i,0} = F_{i,0} = G_{i,0} = H_{i,0} = 2 \tag{3.12}$$

O número dois corresponde à linha da matriz de probabilidades de transição que conterá as probabilidades de um indivíduo válido mudar de estado no próximo instante de tempo.

Figura 3 – Probabilidades de transição do estado válido.

Código do estado	Código do sinistro	Tipos de estados		1	2	3	4	5	6	7	8	9	10	11	12
1	-	Fim	1	1											
2	-	Válido	2												
3	-	Inválido	3												
4	1	Invalidez funcional permanente total por doença	4			1									
5	3	Invalidez permanente parcial por acidente	5		1										
6	5	Invalidez permanente total por acidente	6	1											
7	6	Invalidez permanente total por doença	7	1											
8	7	Invalidez temporária por acidente	8												
9	8	Morte por acidente	9	1											
10	10	Morte qualquer causa	10	1											
11	13	Qualquer tipo de acidente	11		1										
12	15	Invalidez laboral permanente	12			1									

Assume-se que as matrizes de transição de estado dependem apenas do sexo e da idade do segurado. Seja:

$$s(i) = \begin{cases} 0, \text{se o sexo for desconhecido} \\ -1, \text{se o sexo for masculino} \\ 1, \text{se o sexo for feminino} \end{cases}$$

e $d(i)$ é a idade do titular.

Desta forma, a matriz de probabilidades de transição do titular é dada por $P_{s(i),d(i)}$ enquanto a matriz do cônjuge é dada por $Q_{\{-s(i),d(i)+4s(i)\}}$, para respeitar a premissa sobre a diferença de idade dos cônjuges. A idade deve ser limitada a um e para o caso em que o titular é do sexo masculino e limitada a cento e vinte se o titular for do sexo feminino. No caso do sexo ser desconhecido a idade e o sexo serão genéricos para os dois, titular e cônjuge. O código genérico associado a eles servirá de chave para o tratamento dos dados faltante. Nestes casos o cônjuge e o titular terão uma matriz de probabilidades ponderadas como veremos ainda neste capítulo na seção 3.8.3. A matriz de transição do filho é a mesma para qualquer segurado e é denotada por U. Devido à premissa de que os filhos são do sexo masculino e têm sempre catorze anos para qualquer que seja o titular.

3.8.2
Matriz de probabilidade de transição de estados

O modelo via cadeia de Markov para titular e dependentes considera cadeias de Markov independentes uma para o titular, uma para o cônjuge e duas para os filhos. Os estados dos indivíduos são associados aos sinistros e dois estados adicionais foram criados, o estado sem sinistro ou válido e o estado de extinção da apólice denominado “fim”.

Forma geral das matrizes de probabilidade de transição de estado

A matriz de transição possui o *(i,j)* - ésimo elemento, dado por:

$$P_{i,j} = Pr(X_{t+1} = j|X_t = i) \tag{3.13}$$

Esta é a probabilidade de ir do estado i para o estado j em um passo. Assim, a matriz de probabilidades de transição de estados do titular, $P_{s(i),d(i)}$, deve obedecer a seguinte estrutura, onde as células em branco valem zero e as azuis devem ser preenchidas a partir de tabelas atuariais ou através de outros tipos de estimativas de uso interno das seguradoras.

Tabela 8- Matriz de transição de estados do titular.

Código do estado	Código do sinistro	Tipos de estados		1	2	3	4	5	6	7	8	9	10	11	12
1	-	Fim	1	1											
2	-	Válido	2												
3	-	Inválido	3												
4	1	Invalidez funcional permanente total por doença	4			1									
5	3	Invalidez permanente parcial por acidente	5		1										
6	5	Invalidez permanente total por acidente	6	1											
7	6	Invalidez permanente total por doença	7	1											
8	7	Invalidez temporária por acidente	8												
9	8	Morte por acidente	9	1											
10	10	Morte qualquer causa	10	1											
11	13	Qualquer tipo de acidente	11		1										
12	15	Invalidez laboral permanente	12			1									

O preenchimento da matriz de probabilidades do titular segue o seguinte critério:

1. **Fim -** O estado um é um estado absorvente, logo ao chegar neste estado não há possibilidade de transitar para outro estado. Ele sinaliza o fim do processo para a apólice. Por este motivo com probabilidade um, ele permanece no estado um no instante seguinte.
2. **Válido -** Dado que o indivíduo está no estado dois existe possibilidade de ir para qualquer outro estado, exceto para o estado inválido, fato que será explicado a seguir. Para preenchimento destas probabilidades serão usadas as probabilidades de uma tábua de vida para pessoa válida, no caso da aplicação desta dissertação serão usadas probabilidades baseadas na experiência da companhia cuja carteira esta sendo estudada na aplicação do modelo.
3. **Inválido -** Dado que o indivíduo está no estado três tem a possibilidade de ir para qualquer outro estado, exceto o estado válido. Este estado foi criado como um artifício para que ao entrar em qualquer estado de invalidez o indivíduo possa passar a ter as probabilidades de transição de uma pessoa invalida e não mais dada pela tábua de pessoa válida. Uma vez entrando neste estado de invalidez o indivíduo carregará o agravo da ocorrência da invalidez no seu processo de evolução no tempo não retornando ao estado de validez no mês imediatamente posterior.
4. **Invalidez funcional permanente total por doença –** dado que o indivíduo está no estado quatro passará para o estado inválido com probabilidade um no período seguinte. E as probabilidades para o próximo estado serão probabilidades de inválido.

5. **Invalidez permanente parcial por acidente –** Este estado cinco indica que o indivíduo sofreu a perda permanente, mas que não comprometeu sua funcionalidade. Desta forma, o indivíduo irá para o estado válido com probabilidade um para ter probabilidades de válido para seleção de seu próximo estado.
6. **Invalidez permanente total por acidente –** Ao alcançar o estado seis o indivíduo tem o seu contrato finalizado por se tratar de um dano irreversível e por ser pago por isto toda a quantia segurada.
7. **Invalidez permanente total por doença –** Da mesma forma que o item seis desta lista, no caso o item anterior.
8. **Invalidez temporária por acidente -** A invalidez parcial significa que o indivíduo não está incapaz de realizar seu trabalho, ou de continuar uma vida normal. Logo, ele tem as probabilidades de uma pessoa válida para prosseguir na evolução da cadeia de Markov no tempo.
9. **Morte por acidente –** Da mesma forma que o item seis desta lista. O fato é que a morte por acidente é um acréscimo em dinheiro para o tipo de sinistro de morte qualquer causa. Mas, como pode ser contratada junto ou separado da morte qualquer causa, ela teve que ser considerada como um tipo de sinistro, estado em separado.
10. **Morte qualquer causa –** Da mesma forma que o item seis desta lista.
11. **Qualquer tipo de acidente –** Após o recebimento do valor devido pelo acontecimento do sinistro o indivíduo retorna ao seu estado válido.
12. **Invalidez laboral permanente –** Necessariamente a pessoa passa para o estado de inválido no estado seguinte, para que possa no mês seguinte ter as probabilidades de invalido como base para seleção do seu próximo estado na cadeia.

A matriz de probabilidades de transição de estados $Q_{\{-s(i),d(i)+\,4s(i)\}}$ do cônjuge segue a estrutura abaixo:

Tabela 9- Matriz de transição de estados do cônjuge.

Código do estado	Código do sinistro	Tipos de estados		1	2	3	4	5	6	7	8
1	-	Fim	1	1							
2	-	Válido	2								
3	-	Inválido	3								
4	2	Invalidez permanente parcial do cônjuge por acidente	4			1					
5	4	Invalidez permanente total do cônjuge por acidente	5	1							
6	9	Morte por acidente do cônjuge	6	1							
7	11	Morte do cônjuge por qualquer causa	7	1							
8	14	Invalidez permanente total por doença do cônjuge	8	1							

O preenchimento da matriz de probabilidades do cônjuge foi feito seguindo o seguinte critério:

1, 2 e 3 - Os estados: fim, válido e invalido têm o mesmo significado para todas as matrizes.

4. **Invalidez permanente parcial do cônjuge por acidente -** Como é um sinistro permanente o indivíduo fica impedido de receber qualquer quantia no mês seguinte e também passa a ter suas probabilidades de transição agravadas indo com probabilidade um para o estado de inválido.
5. **Invalidez permanente total do cônjuge por acidente -** Como se trata de evento irreversível toda a quantia segurada deve ser recebida finalizando o contrato para esta cobertura referente ao dependente.
6. **Morte por acidente do cônjuge -** Ao alcançar este estado o indivíduo tem o seu contrato finalizado por se tratar de um dano irreversível e por ser pago por isto toda a quantia segurada. Funciona como um adicional de dinheiro para a morte qualquer causa. Como pode ser contratado junto ou separado, este item foi colocado como um possível estado.
7. **Morte qualquer causa do cônjuge** - Ao alcançar este estado o indivíduo tem o seu contrato finalizado por se tratar de um dano irreversível e por ser pago por isto toda a quantia segurada.
8. **Invalidez permanente total por doença do cônjuge -** Como se trata de um evento irreversível toda quantia é paga e o contrato é finalizado para esta cobertura referente ao dependente.

Tabela 10- Matriz de transição de estados do filho.

Código do estado	Código do sinistro	Tipos de estados
1	-	Fim
2	-	Válido
3	-	Inválido
4	12	Morte do filho por qualquer causa

	1	2	3	4
1	1			
2				
3				
4	1			

Assim como a outras matrizes de probabilidades de transição de estado a matriz U dos filhos tem a seguinte estrutura:

1, 2 e 3 - Os estados: fim, válido e inválido têm a mesma função e o mesmo significado das outras matrizes. O estado inválido permanece para fins de padronização da estrutura e para que haja possibilidade de uso, em algum estudo futuro.

4. **Morte qualquer causa do filho -** É um evento irreversível. Toda quantia segurada deve ser recebida finalizando o contrato para esta cobertura.

Probabilidade de sinistros

Seguindo a definição da probabilidade de transição dada pela fórmula 3.13 serão apresentadas nesta seção as expressões para as probabilidades de sinistros para os diversos segurados. Estas probabilidades vão compor as matrizes de probabilidade de transição $P_{s(i),d(i)}$, $Q_{\{-s(i),d(i)+4s(i)\}}$ e U. Vetores da matriz $P_{s(i),d(i)}$ para o titular:

$$p_{i,t} \equiv E\left[e_{i,t}|E_0\right] = \begin{bmatrix} P\{e_{i,t,1} = 1|E_0\} \\ \vdots \\ P\{e_{i,t,12} = 1|E_0\} \end{bmatrix} = \begin{bmatrix} P\{E_{i,t} = 1|E_0\} \\ \vdots \\ P\{E_{i,t} = 12|E_0\} \end{bmatrix} = \left(P'_{s(i),d(i)}\right)^t 1_{e0} \quad (3.14)$$

Sendo o vetor de probabilidade inicial dado por:

$$1'_{e0} \equiv [0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0] \quad (3.15)$$

Vetores da matriz $Q_{\{-s(i),d(i)+4s(i)\}}$ para o cônjuge:

$$q_{i,t} \equiv E[f_{i,t}|F_0] = \begin{bmatrix} P\{f_{i,t,1} = 1|F_0\} \\ \vdots \\ P\{f_{i,t,8} = 1|F_0\} \end{bmatrix} = \begin{bmatrix} P\{F_{i,t} = 1|F_0\} \\ \vdots \\ P\{F_{i,t} = 8|F_0\} \end{bmatrix} = \tag{3.16}$$

$$= \left(Q'_{-s(i),d(i)+4s(i)}\right)^t 1_{f0}$$

Sendo o vetor de probabilidade inicial dado por:

$$1'_{f0} \equiv [0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ ] \tag{3.17}$$

Vetores da matriz U para o filho:

$$u_{i,t} \equiv E[g_{i,t}|G_0] = E[h_{i,t}|H_0] = \begin{bmatrix} P\{h_{i,t,1} = 1|H_0\} \\ \vdots \\ P\{h_{i,t,4} = 1|H_0\} \end{bmatrix} = \begin{bmatrix} P\{H_{i,t} = 1|H_0\} \\ \vdots \\ P\{H_{i,t} = 4|H_0\} \end{bmatrix} \tag{3.18}$$

$$= (U')^t 1_{h0}$$

Sendo o vetor de probabilidade inicial dado por:

$$1'_{h0} \equiv [0\ 1\ 0\ 0\ ] \tag{3.19}$$

O vetor que indica a probabilidade da apólice estar ativa é dado por:

$$p^A_{i,t} \equiv E[I_{i,t}|E_0] = P\{e_{i,t,1} \neq 1|E_0\} = 1'_A\left(P^t_{s(i),d(i)}\right)' 1_{e0} \tag{3.20}$$

Com,

$$1'_A = [0\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1] \tag{3.21}$$

As expressões anteriores dependem, segundo o modelo probabilístico adotado, apenas do sexo e da idade do segurado. Sendo assim, os vetores p, q e u de probabilidades de eventos para os vários sinistros calculados acima, podem ser expressos por:

$$p_{i,t} \equiv p_{s(i),d(i)}, q_{i,t} \equiv q_{sc(i),dc(i)} \ e \ u_{i,t} \equiv u_{sf(i),df(i)} \qquad (3.22)$$

Sendo *s(i)* e *d(i)* o sexo e a idade do titular da apólice i, *sc(i)* e *dc(i)* representam o sexo e a idade do cônjuge do titular da apólice i e *sf(i) e df(i)* o sexo e a idade dos filhos.

Expressão Final do Modelo Estocástico

O modelo estocástico apresentado para o total líquido a ser gasto com sinistros considera que os segurados titulares dentro de cada apólice e entre as apólices são independentes entre si. Neste caso, vale o Teorema Central do Limite, o qual estabelece que, sob determinadas condições, a distribuição da soma de um grande número de variáveis aleatórias converge para uma distribuição Normal, ou seja, por aproximação segue que:

$$T \sim N\big(E(T), V(T)\big) \qquad (3.23)$$

O Teorema Central do Limite diz que a soma de um número de efeitos aleatórios, sem dominância de nenhum deles sobre o resultado total, produz uma variável aleatória com distribuição Normal. Por este motivo a divisão da base de dados será necessária para que o modelo misto seja testado como alternativa a aproximação pela Normal. Pois, na base inteira existe uma grande diversidade de valores.

3.8.3
Tratamento para dados faltantes

Para algumas apólices as informações de sexo e idade não estavam disponíveis na base de dados. Como a matriz de probabilidades de transição de estados depende do sexo e idade do indivíduo para estes casos optou-se por fazer uma estimativa da matriz. A estimação da matriz de probabilidades de transição de estados é calculada pela média ponderada das matrizes de transição de estados de todos os sexos e idades que se tem informação. Os pesos são obtidos pelas proporções observadas na base de dados. Assim, cada matriz ponderada será estimada pela seguinte fórmula:

$$\hat{P} = \sum_{sx,id} \hat{p}rop_{sx,id} * P_{sx,id} \qquad (3.24)$$

Ao estabelecer o modelo financeiro e o modelo estocástico é possível calcular analiticamente a média do total líquido gasto com sinistro, E(T) e a sua variância, V(T).

3.8.4 Média do total líquido gasto com sinistro

Utilizando a expressão de T dada pela fórmula 3.11, pode-se obter o seu valor esperado da seguinte forma:

$$E(T) = \sum_{t=1}^{12} \beta_t \left\{ \sum_{i=1}^{N} [c_i' E(e_{it}) + b_i' E(f_{it}) E(I_{it}) + a_i' E(g_{it} + h_{it}) E(I_{it}) - r_i' E(I_{it})] \right\} \qquad (3.25)$$

Substituindo pelos respectivos valores esperados obtém-se

$$E(T) = \sum_{t=1}^{12} \beta_t \left\{ \sum_{i=1}^{N} [c_i' p_{it} + b_i' q_{it} p_{it}^A + 2a_i' u_{it} p_{it}^A - r_i' p_{it}^A] \right\} \qquad (3.26)$$

Separando por sexo e idade tem-se:

$$E(T) = \sum_{t=1}^{12} \beta_t \sum_{sx=-1,0,1} \sum_{id=0}^{120} \sum_{i \epsilon A(sx,id)} \begin{bmatrix} c_i' \left(P_{sx,id}'\right)^t 1_{e_0} + \\ b_i' \left(Q_{-sx,id+4sx}'\right)^t 1_{f_0} 1_A' \left(P_{sx,id}'\right)^t 1_{e_0} + \\ 2a_i' (U')^t 1_{g_0} 1_A' \left(P_{sx,id}'\right)^t 1_{e_0} - \\ r_i' 1_A' \left(P_{sx,id}'\right)^t 1_{e_0} \end{bmatrix} \qquad (3.27)$$

Onde,

$A(sx, id) = \{i, s(i) = sx \ e \ d(i) = id\}$

$\Sigma_{i\epsilon A(sx,id)}$ - representa os valores que repetem, parte mais agregada do cálculo

Como veremos o cálculo será feito na base agregada, pois os perfis de idade e sexo se repetem na base de dados.

$$E(T) = \sum_{t=1}^{12} \beta_t \sum_{sx=-1,0,1} \sum_{id=0}^{120} \begin{bmatrix} \sum_{i\epsilon A(sx,id)} c_i' \left(P_{sx,id}'\right)^t 1_{e_0} + \\ \sum_{i\epsilon A(-sx,id+4sx)} b_i' \left(Q_{-sx,id+4sx}'\right)^t 1_{f_0} 1_A' \left(P_{sx,id}'\right)^t 1_{e_0} + \\ \sum_{i\epsilon A(sx,id)} 2a_i' (U')^t 1_{g_0} 1_A' \left(P_{sx,id}'\right)^t 1_{e_0} - \\ \sum_{i\epsilon A(sx,id)} r_i' 1_A' \left(P_{sx,id}'\right)^t 1_{e_0} \end{bmatrix} \tag{3.28}$$

Onde,

$$\bar{c}_{sx,id}' = \sum_{i\epsilon A(sx,id)} c_i'$$

$$\bar{b}_{-sx,id+4sx}' = \sum_{i\epsilon A(-sx,id+4sx)} b_i'$$

$$\bar{a}_{sx,id}' = \sum_{i\epsilon A(sx,id)} a_i'$$

$$\bar{r}_{sx,id}' = \sum_{i\epsilon A(sx,id)} r_i'$$

Assim, a fórmula final para a média de T segue abaixo:

$$\mu \equiv E(T) = \sum_{t=1}^{12} \beta_t \sum_{sx=-1,0,1} \sum_{id=0}^{100} \begin{bmatrix} \overline{c'}_{sx,id} \left(P'_{sx,id}\right)^t 1_{e_0} + \\ \overline{b'}_{-sx,id+4sx} \left(Q'_{-sx,id+4sx}\right)^t 1_{f_0} 1'_A \left(P'_{sx,id}\right)^t 1_{e_0} + \\ \overline{2a'}_{sx,id} (U')^t 1_{g_0} 1'_A \left(P'_{sx,id}\right)^t 1_{e_0} - \\ \overline{r'}_{sx,id} 1'_A \left(P'_{sx,id}\right)^t 1_{e_0} \end{bmatrix} \tag{3.29}$$

3.8.5
Variância do total líquido gasto com sinistro

Utilizando a expressão de T dada na fórmula 3.11, obtém-se:

$$V(T) = E(T^2) - \mu_T^2$$
$$V(T) = E(T^2) - E(T)^2$$

Definindo:

$$y_{it}^{'} = [e_{it}^{'}, I_{it} f_{it}^{'}, I_{it} g_{it}^{'}, I_{it} h_{it}^{'}, I_{it}]$$
$$\beta^{'} = [\beta_1, \beta_2, \dots, \beta_{12}]$$
$$w_i^{'} = [c_i^{'}, b_i^{'}, a_i^{'}, a_i^{'}, r_i^{'}]$$

Segue que se pode reescrever T e a sua variância através das seguintes expressões:

$$T = \sum_{t=1}^{12} \beta_t \left\{ \sum_{i=1}^{N} [c_i^{'} e_{it} + b_i^{'} f_{it} I_{it} + a_i^{'} (g_{it} + h_{it}) I_{it} - r_i^{'} I_{it}] \right\} \tag{3.30}$$
$$T = \sum_{t=1}^{12} \beta_t \sum_{i=1}^{N} w_i^{'} y_{it}$$
$$V(T) = \sum_{i=1}^{N} V\left(\sum_{t=1}^{12} \beta_t w_i^{'} y_{it} \right)$$

Assim, a fórmula final para a variância de T é dada por:

$$\sigma^2 \equiv V(T) = \sum_{sx=-1,0,1} \sum_{id=1}^{120} \sum_{i \in A(sx,id)} (w'_i \otimes w_i) \sum_{t=1}^{12} \sum_{s=1}^{12} \beta_t \beta_s \, vec\left(\Sigma_{t,s}^{(sx,id)} \right) \tag{3.31}$$

Utiliza-se a seguinte propriedade do operador VEC:

$$vec(ABC) = (C^{'} \otimes A) vec(B) \tag{3.32}$$

$$\Sigma_{t,s}^{(sx,id)} = E\left[\left(y_{it} - E(y_{it})\right)\left(y_{is} - E(y_{is})\right)' | sx(i) = sx, id(i) = id, e_{i,0} = 1_{e_0}\right]$$

$$E\left[y_{it} | sx(i) = sx, id(i) = id, e_{i,0} = 1_{e_0}\right] = \begin{bmatrix} p_{it} \\ q_{it} \\ u_{it} \\ u_{it} \end{bmatrix} \quad (3.33)$$

$$E\left[y_{it} y_{is} | sx(i) = sx, id(i) = id, e_{i,0} = 1_{e_0}\right] \quad (3.34)$$

Onde, $\Sigma_{t,s}^{(sx,id)}$ é a matriz de matrizes de variância e covariância entre os estados dos segurados.

$$\Sigma_{t,s}^{(sx,id)} = \begin{bmatrix} \Sigma_{e,e,t,s}^{(sx,id)} & \Sigma_{e,f,t,s}^{(sx,id)} & \Sigma_{e,g,t,s}^{(sx,id)} & \Sigma_{e,h,t,s}^{(sx,id)} & \Sigma_{e,I,t,s}^{(sx,id)} \\ \Sigma_{f,e,t,s}^{(sx,id)} & \Sigma_{f,f,t,s}^{(sx,id)} & \Sigma_{f,g,t,s}^{(sx,id)} & \Sigma_{f,h,t,s}^{(sx,id)} & \Sigma_{f,I,t,s}^{(sx,id)} \\ \Sigma_{g,e,t,s}^{(sx,id)} & \Sigma_{g,f,t,s}^{(sx,id)} & \Sigma_{g,g,t,s}^{(sx,id)} & \Sigma_{g,h,t,s}^{(sx,id)} & \Sigma_{g,I,t,s}^{(sx,id)} \\ \Sigma_{h,e,t,s}^{(sx,id)} & \Sigma_{h,f,t,s}^{(sx,id)} & \Sigma_{h,g,t,s}^{(sx,id)} & \Sigma_{h,h,t,s}^{(sx,id)} & \Sigma_{h,I,t,s}^{(sx,id)} \\ \Sigma_{I,e,t,s}^{(sx,id)} & \Sigma_{I,f,t,s}^{(sx,id)} & \Sigma_{I,g,t,s}^{(sx,id)} & \Sigma_{I,h,t,s}^{(sx,id)} & \Sigma_{I,I,t,s}^{(sx,id)} \end{bmatrix} \quad (3.35)$$

Os elementos desta matriz são as matrizes que serão calculadas na seção 3.8.6 a seguir. Como foi exposto acima para o cálculo da variância foi necessário utilizar o produto de kronecker e uma propriedade do operador VEC, por este motivo as seções abaixo buscam esclarecem a funcionalidade destas ferramentas no cálculo da variância. Ferramentas importantes utilizadas:

- Produto de Kronecker
- Operador VEC

O produto de kronecker, ou tensorial é uma ferramenta útil que permite juntar espaços vetoriais para formar espaços vetoriais maiores. Suponha que A é uma matriz m x n e que B é uma matriz p x q. Então a representação do produto de kronecker entre elas é a seguinte

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix} \tag{3.36}$$

O VEC é um operador que retorna o vetor obtido pelo empilhamento das colunas da matriz X uma acima da outra. Este operador permite um pré-cálculo dos produtos cruzados por sexo e idade para o cálculo da variância.

3.8.6 Matrizes de covariância e probabilidades conjuntas

Nesta seção serão apresentadas as matrizes de covariância que serão utilizadas no cálculo da variância como vimos na seção anterior, suas derivações estão no anexo 8.1. Estas matrizes vão por sua vez depender de um conjunto de probabilidades conjuntas envolvendo estados dos segurados. As variáveis de interesse são:

$$y'_{it} = [e'_{it}, I_{it}f'_{it}, I_{it}g'_{it}, I_{it}h'_{it}, I_{it}] \tag{3.37}$$

Deve-se observar que, fora o primeiro elemento, correspondente ao vetor de estado do titular, os demais elementos estão sendo multiplicados pela variável indicadora de vigência I_{it}. Lembrando que isto ocorre porque estes elementos dependem da apólice estar ativa para serem contabilizados. As matrizes de covariância a seguir estão apresentadas para uma dada combinação de sexo e idade do segurado. Para simplificar a notação, os desenvolvimentos serão feitos omitindo os sobrescritos sx e id. Veja a linha de raciocínio com a figura abaixo:

Figura 4 – Matriz de matrizes de variância e covariância.

$$
\begin{array}{c}
 \\ e \\ f \\ g \\ h \\ I
\end{array}
\begin{array}{c}
\begin{array}{ccccc} \quad e \quad & \quad f \quad & \quad g \quad & \quad h \quad & \quad I \quad \end{array} \\
\begin{bmatrix}
\boxed{\Sigma^{(sx,id)}_{e,e,t,s}} & \boxed{\Sigma^{(sx,id)}_{e,f,t,s}} & \boxed{\Sigma^{(sx,id)}_{e,g,t,s}} & \boxed{\Sigma^{(sx,id)}_{e,h,t,s}} & \boxed{\Sigma^{(sx,id)}_{e,I,t,s}} \\
\Sigma^{(sx,id)}_{f,e,t,s} & \boxed{\Sigma^{(sx,id)}_{f,f,t,s}} & \boxed{\Sigma^{(sx,id)}_{f,g,t,s}} & \boxed{\Sigma^{(sx,id)}_{f,h,t,s}} & \boxed{\Sigma^{(sx,id)}_{f,I,t,s}} \\
\Sigma^{(sx,id)}_{g,e,t,s} & \Sigma^{(sx,id)}_{g,f,t,s} & \boxed{\Sigma^{(sx,id)}_{g,g,t,s}} & \boxed{\Sigma^{(sx,id)}_{g,h,t,s}} & \boxed{\Sigma^{(sx,id)}_{g,I,t,s}} \\
\Sigma^{(sx,id)}_{h,e,t,s} & \Sigma^{(sx,id)}_{h,f,t,s} & \Sigma^{(sx,id)}_{h,g,t,s} & \boxed{\Sigma^{(sx,id)}_{h,h,t,s}} & \boxed{\Sigma^{(sx,id)}_{h,I,t,s}} \\
\Sigma^{(sx,id)}_{I,e,t,s} & \Sigma^{(sx,id)}_{I,f,t,s} & \Sigma^{(sx,id)}_{I,g,t,s} & \Sigma^{(sx,id)}_{I,h,t,s} & \boxed{\Sigma^{(sx,id)}_{I,I,t,s}}
\end{bmatrix}
\end{array}
$$

Deve-se calcular a matriz acima. Como se trata de uma matriz simétrica, basta calcular as expressões para as matrizes em destaque, pois as matrizes não destacadas são obtidas pela transposta de sua correspondente matriz destacada. O cálculo das matrizes com mais detalhes pode ser visto no Anexo 8.1. Após os devidos cálculos as matrizes correspondem aos resultados que são mostrados a seguir.

1. Covariância entre os estados do titular nos períodos t e s

Este resultado é obtido utilizando a definição de covariância e a informação do estado inicial do titular, portanto tem-se que:

$$\Sigma^{(sx,id)}_{e,e,t,s} \equiv \Sigma_{e,e,t,s}$$

$$\Sigma_{e,e,t,s} = E\{[e_{it} - E(e_{it})][e_{is} - E(e_{is})]' | e_{i0} = 1_{e_0}\}$$

$$= E\{(e_{it} - p_{it})(e_{is} - p_{is})' | e_{i0} = 1_{e_0}\}$$

Ao desenvolver esta expressão chega-se a fórmula final para qualquer período t, s = 1,2,...,12, que é dada por:

$$\Sigma_{e,e,t,s} = \left(P'_{s(i),d(i)}\right)^{|t-s|} diag\left(p_{i,\min(t,s)}\right) - p_{it}p'_{is} \qquad (3.38)$$

2. Covariância entre os estados do titular e do cônjuge nos períodos t e s

Este resultado é obtido utilizando a definição de covariância e a informação do estado inicial do titular e do cônjuge, portanto tem-se que:

$$\Sigma_{e,f,t,s}^{(sx,id)} \equiv \Sigma_{e,f,t,s}$$

$$\Sigma_{e,f,t,s} = E\{[e_{it} - E(e_{it})][I_{is}f_{is} - E(I_{is}f_{is})]'|e_{i0} = 1_{e_0}, f_{i0} = 1_{f_0}\}$$

$$= E\{[e_{it} - p_{it}][I_{is}f_{is} - (p_{is}^{A}q_{is})]'|e_{i0} = 1_{e_0}, f_{i0} = 1_{f_0}\}$$

Ao desenvolver esta expressão chega-se a fórmula final para qualquer período t, s = 1,2,...,12, que é dada por:

$$\Sigma_{e,f,t,s} = \begin{cases} \left(P'_{s(i),d(i)}\right)^{t-s} \begin{bmatrix} 0 & 0 \\ 0 & I_{(11x11)} \end{bmatrix} p_{is}q'_{is} - p_{it}p_{is}^{A}q'_{is}, se\ t \geq s \\ diag(p_{it})\left(P'_{s(i),d(i)}\right)^{s-t} 1_A q'_{is} - p_{it}p_{is}^{A}q'_{is}, se\ t < s \end{cases} \quad (3.39)$$

3. Covariância entre os estados do titular e dos filhos entre os instantes t e s

Este resultado é obtido utilizando a definição de covariância e a informação do estado inicial do titular e dos filhos, portanto tem-se que:

$$\Sigma_{e,g,t,s}^{(sx,id)} = E\{[e_{it} - E(e_{it})][I_{is}g_{is} - E(I_{is}g_{is})]'|e_{i0} = 1_{e_0}, g_{i0} = 1_{g_0}\}$$

$$\Sigma_{e,h,t,s}^{(sx,id)} = E\{[e_{it} - E(e_{it})][I_{is}h_{is} - E(I_{is}h_{is})]'|e_{i0} = 1_{e_0}, h_{i0} = 1_{h_0}\}$$

$$\Sigma_{e,g,t,s}^{(sx,id)} = \Sigma_{e,h,t,s}^{(sx,id)} = \Sigma_{e,g,t,s} = \Sigma_{e,h,t,s}$$

$$\Sigma_{e,g,t,s} = E\{[e_{it} - p_{it}][I_{is}g_{is} - E(I_{is}g_{is})]'|e_{i0} = 1_{e_0}, g_{i0} = 1_{g_0}\}$$

Ao desenvolver esta expressão chega-se a fórmula final para qualquer período t, s = 1,2,...,12, que é dada por:

$$\Sigma_{e,g,t,s} = \begin{cases} \left(P'_{s(i),d(i)}\right)^{t-s} \begin{bmatrix} 0 & 0 \\ 0 & I_{(11x11)} \end{bmatrix} p_{is}u'_{is} - p_{it}p_{is}^{A}u'_{is}\text{ , se } t \geq s \\ diag(p_{it})\left(P'_{s(i),d(i)}\right)^{s-t} 1_{A}u'_{is} - p_{it}p_{is}^{A}u'_{is}\text{ , se } t < s \end{cases} \quad (3.40)$$

4. Covariância entre os estados do titular e a variável de vigência entre os instantes t e s

Este resultado é obtido utilizando a definição de covariância e a informação do estado inicial do titular, portanto tem-se que:

$$\Sigma_{e,I,t,s}^{(sx,id)} = E\{[e_{it} - E(e_{it})][I_{is} - E(I_{is})]' | e_{i0} = 1_{e_0}\}$$

$$\Sigma_{e,I,t,s} = E\{[e_{it} - p_{it}][I_{is} - E(I_{is})]' | e_{i0} = 1_{e_0}\}$$

Ao desenvolver esta expressão chega-se a fórmula final para qualquer período t, s = 1,2,...,12, que é dada por:

$$\Sigma_{e,I,t,s} = \begin{cases} \left(P'_{s(i),d(i)}\right)^{t-s} \begin{bmatrix} 0 & 0 \\ 0 & I_{(11x11)} \end{bmatrix} p_{is} - p_{it}p_{is}^{A}\text{ , se } t \geq s \\ diag(p_{it})\left(P'_{s(i),d(i)}\right)^{s-t} 1_{A} - p_{it}p_{is}^{A}\text{ , se } t < s \end{cases} \quad (3.41)$$

5. Covariância entre os estados do cônjuge entre os instantes t e s

Este resultado é obtido utilizando a definição de covariância e a informação do estado inicial do titular e cônjuge, portanto tem-se que:

$$\Sigma_{f,f,t,s}^{(sx,id)} = E\{[I_{it}f_{it} - E(I_{it}f_{it})][I_{is}f_{is} - E(I_{is}f_{is})]' | e_{i0} = 1_{e_0}, f_{i0} = 1_{f_0}\}$$

$$\Sigma_{f,f,t,s} = E\{[I_{it}f_{it} - p_{it}^{A}q_{it}][I_{is}f_{is} - p_{is}^{A}q_{is}]' | e_{i0} = 1_{e_0}, f_{i0} = 1_{f_0}\}$$

Ao desenvolver esta expressão chega-se a fórmula final para qualquer período t, s = 1,2,...,12, que é dada por:

$$\Sigma_{f,f,t,s} = \begin{cases} 1'_A\left(P'_{sx,id}\right)^{|t-s|}\begin{bmatrix} 0 & 0 \\ 0 & I_{(11x11)} \end{bmatrix} p_{i,min(t,s)}\left(Q'_{-s(i),d(i)+4s(i)}\right)^{|s-t|} \\ diag\left(q_{i,min(t,s)}\right) - p_{it}^A q_{it} p_{is}^A q'_{is} \end{cases} \tag{3.42}$$

6. Covariância entre os estados do cônjuge e dos filhos entre os instantes t e s

Este resultado é obtido utilizando a definição de covariância e a informação do estado inicial do titular, cônjuge e dos filhos, portanto tem-se que:

$$\Sigma_{f,g,t,s}^{(sx,id)} = E\{[I_{it}f_{it} - E(I_{it}f_{it})][I_{is}g_{is} - E(I_{is}g_{is})]'|e_{i0} = 1_{e_0}, f_{i0} = 1_{f_0}, g_{i0} = 1_{g_0}$$

$$\Sigma_{f,h,t,s}^{(sx,id)} = E\{[I_{it}f_{it} - E(I_{it}f_{it})][I_{is}h_{is} - E(I_{is}h_{is})]'|e_{i0} = 1_{e_0}, f_{i0} = 1_{f_0}, h_{i0} = 1_{h_0}\}$$

$$\Sigma_{f,g,t,s}^{(sx,id)} = \Sigma_{f,h,t,s}^{(sx,id)} = \Sigma_{f,g,t,s} = \Sigma_{f,h,t,s}$$

$$\Sigma_{f,g,t,s}^{(sx,id)} = E\{[I_{it}f_{it} - E(I_{it}f_{it})][I_{is}g_{is} - E(I_{is}g_{is})]'|e_{i0} = 1_{e_0}, f_{i0} = 1_{f_0}, g_{i0} = 1_{g_0}\}$$

Ao desenvolver esta expressão chega-se a fórmula final para qualquer período t, s = 1,2,...,12, que é dada por:

$$\Sigma_{f,g,t,s} = 1'_A\left(P'_{s(i),d(i)}\right)^{|t-s|}\begin{bmatrix} 0 & 0 \\ 0 & I_{(11x11)} \end{bmatrix} p_{i,min(t,s)}\, q_{it} u'_{is} - p_{it}^A q_{it} p_{is} u'_{is} \tag{3.43}$$

7. Covariância entre os estados do cônjuge e a variável de vigência entre os instantes t e s

Este resultado é obtido utilizando a definição de covariância e a informação do estado inicial do titular e cônjuge, portanto tem-se que:

$$\Sigma_{f,I,t,s}^{(sx,id)} = E\{[I_{it}f_{it} - E(I_{it}f_{it})][I_{is} - E(I_{is})]'|e_{i0} = 1_{e_0}, f_{i0} = 1_{f_0}\}$$

$$\Sigma_{f,I,t,s} = E\{[I_{it}f_{it} - p_{it}^A q_{it}][I_{is} - p_{is}^A]'|e_{i0} = 1_{e_0}, f_{i0} = 1_{f_0}\}$$

Ao desenvolver esta expressão chega-se a fórmula final para qualquer período t, s = 1,2,...,12, que é dada por:

$$\Sigma_{f,I,t,s} = 1'_A\left(P'_{s(i),d(i)}\right)^{|t-s|}\begin{bmatrix}0 & 0\\ 0 & I_{(11x11)}\end{bmatrix} p_{i,\min(t,s)}q_{it} - p_{it}^{A}q_{it}p_{is}^{A} \qquad (3.44)$$

8. Covariância entre os estados do filho entre os instantes t e s

Este resultado é obtido utilizando a definição de covariância e a informação do estado inicial do titular e dos filhos, portanto tem-se que:

$$\Sigma_{g,g,t,s}^{(sx,id)} = E\left\{[I_{it}g_{it} - E(I_{it}g_{it})][I_{is}g_{is} - E(I_{is}g_{is})]'|e_{i0} = 1_{e_0}, g_{i0} = 1_{g_0}\right\}$$

$$\Sigma_{h,h,t,s}^{(sx,id)} = E\left\{[I_{it}h_{it} - E(I_{it}h_{it})][I_{is}h_{is} - E(I_{is}h_{is})]'|e_{i0} = 1_{e_0}, h_{i0} = 1_{h_0}\right\}$$

$$\Sigma_{g,g,t,s}^{(sx,id)} = \Sigma_{h,h,t,s}^{(sx,id)} = \Sigma_{g,g,t,s} = \Sigma_{h,h,t,s}$$

$$\Sigma_{g,g,t,s} = E\left\{[I_{it}g_{it} - E(I_{it}g_{it})][I_{is}g_{is} - E(I_{is}g_{is})]'|e_{i0} = 1_{e_0}, g_{i0} = 1_{g_0}\right\}$$

Ao desenvolver esta expressão chega-se a fórmula final para qualquer período t, s = 1,2,...,12, que é dada por:

$$\Sigma_{g,g,t,s} = 1'_A\left(P'_{sx,id}\right)^{|t-s|}\begin{bmatrix}0 & 0\\ 0 & I_{(11x11)}\end{bmatrix} p_{i,min(t,s)}\,(U')^{|s-t|}\,diag\left(u_{i,min(t,s)}\right) - p_{it}^{A}u_{it}p_{is}^{A}u'_{is} \qquad (3.45)$$

9. Covariância entre os estados dos filhos entre os instantes t e s

Este resultado é obtido utilizando a definição de covariância e a informação do estado inicial do titular e dos filhos, portanto tem-se que:

$$\Sigma_{g,h,t,s}^{(sx,id)} = E\left\{[I_{it}g_{it} - E(I_{it}g_{it})][I_{is}h_{is} - E(I_{is}h_{is})]'|e_{i0} = 1_{e_0}, g_{i0} = 1_{g_0}, h_{i0} = 1_{h_0}\right\}$$

Ao desenvolver esta expressão chega-se a fórmula final para qualquer período t, s = 1,2,...,12, que é dada por:

$$\Sigma_{g,h,t,s} = 1'_A \left(P'_{s(i),d(i)}\right)^{|t-s|} \begin{bmatrix} 0 & 0 \\ 0 & I_{(11x11)} \end{bmatrix} p_{i,\min(t,s)} u_{it} u'_{it} - p_{it}^A u_{it} p_{is}^A u'_{is} \qquad (3.46)$$

10. Covariância entre os estados do filho e a variável de vigência entre os instantes t e s

Este resultado é obtido utilizando a definição de covariância e a informação do estado inicial do titular e dos filhos, portanto tem-se que:

$$\Sigma_{g,I,t,s}^{(sx,id)} = E\{[I_{it} g_{it} - E(I_{it} g_{it})][I_{is} - E(I_{is})]' | e_{i0} = 1_{e_0}, g_{i0} = 1_{g_0}\}$$
$$\Sigma_{h,I,t,s}^{(sx,id)} = E\{[I_{it} h_{it} - E(I_{it} h_{it})][I_{is} - E(I_{is})]' | e_{i0} = 1_{e_0}, h_{i0} = 1_{h_0}\}$$
$$\Sigma_{g,I,t,s}^{(sx,id)} = \Sigma_{h,I,t,s}^{(sx,id)} = \Sigma_{g,I,t,s} = \Sigma_{h,I,t,s}$$
$$\Sigma_{g,I,t,s} = E\{[I_{it} g_{it} - E(I_{it} g_{it})][I_{is} - E(I_{is})]' | e_{i0} = 1_{e_0}, g_{i0} = 1_{g_0}\}$$

Ao desenvolver esta expressão chega-se a fórmula final para qualquer período t, s = 1,2,...,12, que é dada por:

$$\Sigma_{g,I,t,s} = 1'_A \left(P'_{s(i),d(i)}\right)^{|t-s|} \begin{bmatrix} 0 & 0 \\ 0 & I_{(11x11)} \end{bmatrix} p_{i,\min(t,s)} u_{it} - p_{it}^A u_{it} p_{is}^A \qquad (3.47)$$

11. Covariância entre as variáveis de vigência entre os instantes t e s

Este resultado é obtido utilizando a definição de covariância e a informação do estado inicial do titular, portanto tem-se que:

$$\Sigma_{I,I,t,s}^{(sx,id)} = E\{[I_{it} - E(I_{it})][I_{is} - E(I_{is})]' | e_{i0} = 1_{e_0}\}$$
$$\Sigma_{I,I,t,s}^{(sx,id)} = \Sigma_{I,I,t,s}$$
$$\Sigma_{I,I,t,s} = E\{[I_{it} - E(I_{it})][I_{is} - E(I_{is})]' | e_{i0} = 1_{e_0}\}$$

Ao desenvolver esta expressão chega-se a fórmula final para qualquer período t, s = 1,2,...,12, que é dada por:

$$\Sigma_{I,I,t,s} = 1'_A \left(P'_{s(i),d(i)}\right)^{|t-s|} \begin{bmatrix} 0 & 0 \\ 0 & I_{(11x11)} \end{bmatrix} p_{i,\min(t,s)} - p^A_{it} p^A_{is} \qquad (3.48)$$

3.9
Modelo Paramétrico

O modelo paramétrico corresponde à utilização do modelo financeiro e do modelo estocástico pressupondo a aproximação da distribuição do total gasto com sinistros pela distribuição Normal. A média, variância e matriz de variância e covariância são calculadas como demonstrado na seção anterior.

$$T \sim N\big(E(T), V(T)\big) \qquad (3.49)$$

3.9.1
Cálculo do capital mínimo requerido utilizando o modelo paramétrico

Ao atribuirmos uma aproximação pela Normal para a distribuição do Total líquido de gasto com sinistros é possível obter o capital mínimo requerido calculando o quantil desta distribuição. O Capital Mínimo Requerido é obtido utilizando a fórmula abaixo:

$$CMR(\alpha) = \mu + z(\alpha) * \sigma \qquad (3.50)$$

Onde, $\mu = E(T)$ e $\sigma^2 = V(T)$. É necessário fazer uso da tabela da distribuição Normal para obter os valores críticos segundo os valores de alfa. O valor de alfa que será utilizado na ilustração será 0,001.

3.10
Modelo Misto

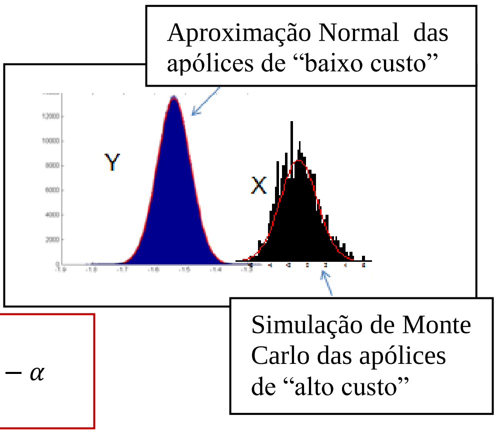
Um modelo deve considerar diversas situações futuras as quais a companhia poderá passar. Cada uma dessas situações é chamada de cenário. Esses cenários podem ser determinados a partir de um processo estocástico que captura a incerteza e a variabilidade dos riscos modelados, ou a partir de situações críticas determinísticas baseadas na experiência da companhia e em julgamentos subjetivos.

O modelo misto é a composição dos dois métodos o método paramétrico e o método não paramétrico. Portanto será utilizado o modelo financeiro, o modelo estocástico e a simulação paramétrica para obter através da convolução dos dois métodos a distribuição do total líquido do gasto com sinistro e o capital mínimo requerido calculando o quantil desta distribuição. Como foi visto anteriormente, a base de dados foi dividida em duas partes. A parte que poderia ser modelada seguindo o teorema central do limite, e a parte dos possíveis valores extremos. Os indivíduos que podem apresentar sinistros muito altos tiveram a vida de um ano simulada mês a mês. Assim, a variável T, total líquido gasto com sinistro, do ano pode ser contabilizada a cada simulação. Através da qual se tornou possível obter o gráfico do comportamento da freqüência absoluta desta variável. Como será observado no capítulo 4. Assim, ao obter o comportamento da variável total líquido gasto com sinistro das duas partes da base de dados pode-se então obter a distribuição final desta variável. Para ilustrar o método segue a figura abaixo:

Figura 5 – Modelo misto.

t^* é o Capital Mínimo Requerido

$$F(t^*) = P(X + Y < t^*)$$



t^* é obtido por uma busca não linear

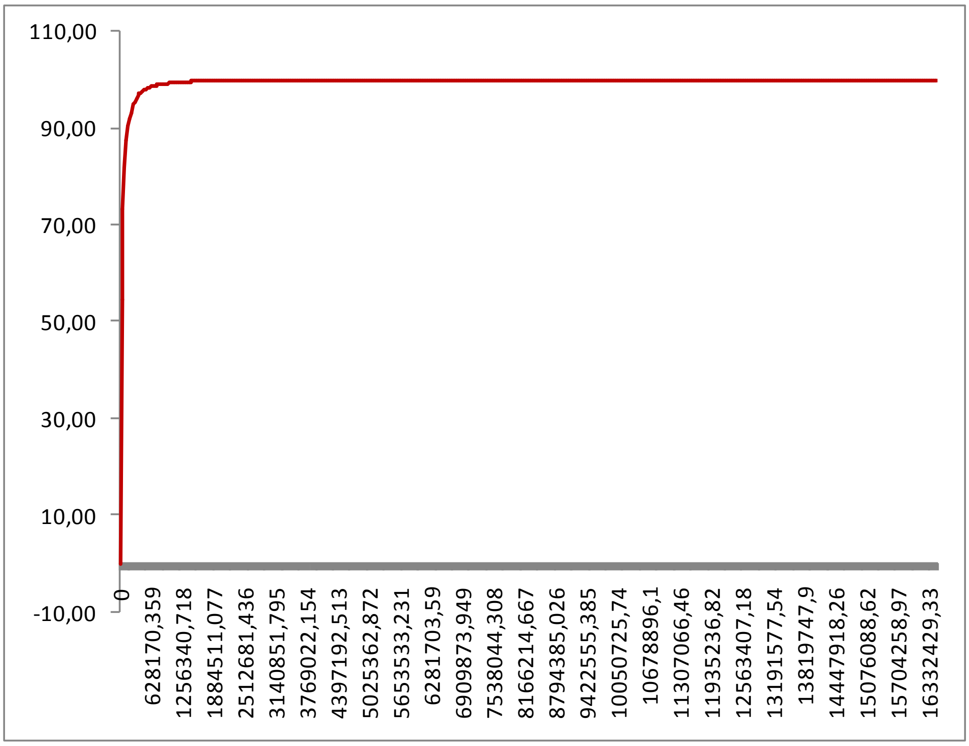
$$\boxed{F(t^*) = \frac{1}{N}\sum_{i=1}^{N}\left[\Phi_Y\left(\frac{t^* - x_i - \mu}{\sigma}\right)\right] = 1 - \alpha}$$

3.10.1 Divisão da base de dados

Ao aplicar o Modelo Misto pretende-se explicar de forma mais minuciosa o comportamento da cauda da distribuição do total líquido do gasto com sinistro. Para tanto a base de dados deve ser dividida em duas partes. A base de dados tratada e unificada conforme a seção 3.1 será dividida em dois grupos: grupo das apólices de “baixo custo” e grupo das apólices de “alto custo”. Lembrando que as nomenclaturas correspondem ao possível custo futuro destas apólices para a seguradora. O grupo das apólices de “baixo custo” deve ser formado pelos indivíduos nos quais pode ser aplicado o Modelo Paramétrico, ou seja, o modelo em que para a soma das variáveis aleatória vale o Teorema Central do Limite. O grupo das apólices de “baixo custo” deve ser formado pelos indivíduos que apresentam valores mais discrepantes de capital segurado e que podem causar valores extremos, estes indivíduos serão modelados por simulação. Nesta seção pretende-se explicar como foi escolhida a forma de dividir a base de dados visando a aplicação do Modelo Misto.

Os dados provenientes de seguros e instituições financeiras, de uma forma geral, não apresentam um padrão de normalidade. Esses desvios são característicos de eventos extremos, com baixa probabilidade, mas que têm grande importância numa análise financeira. Esses eventos, concentrados nas caudas das distribuições embora ocorram com baixa probabilidade podem gerar grandes prejuízos. Desta forma, visando uma melhor compreensão do risco foi observado o comportamento da soma dos capitais segurados unitários por indivíduo. Sabe-se que um percentual do capital segurado será o valor pago por sinistro. Assim, a idéia é observar como se distribuem estes valores dentro da carteira. Como se pode constatar na figura abaixo e melhor na tabela onze a grande parte dos valores possíveis de sinistro estão concentrados até aproximadamente 90% dos casos, o que significa que até um possível valor de corte seria R$165.308,00.

Figura 6- Freqüência acumulada do valor da soma do capital segurado por indivíduo.



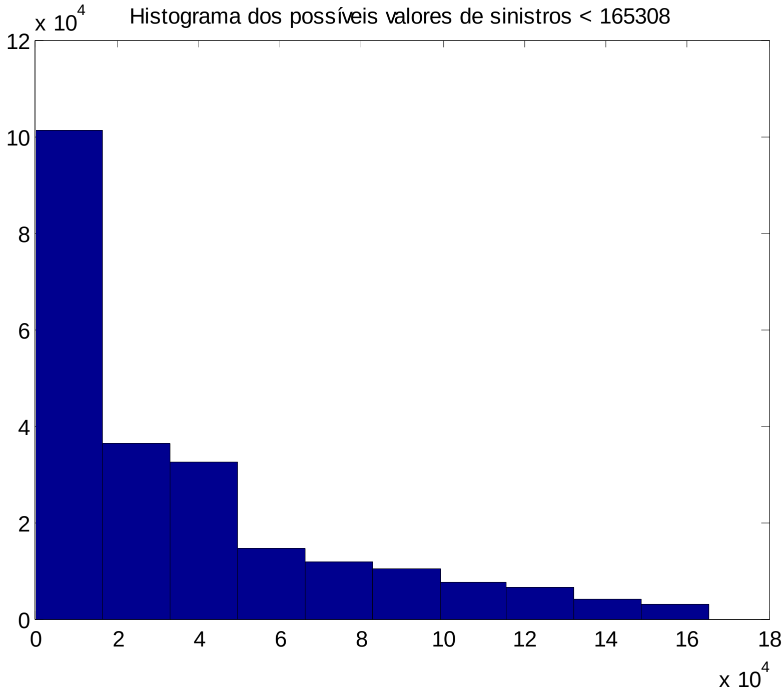
Pode-se observar no gráfico acima que a curva para a freqüência acumulada se tornar 100% ou bem próximo de 100% pode ser considerada como iniciando em aproximadamente 90%. Isto significa que os 10% de indivíduos acima da curva são os candidatos a possíveis valores extremos. Estes seriam os indivíduos cujo comportamento não permite a aplicação do Teorema Central do Limite. Desta forma, supõe-se que ao dividir a base em aproximadamente 90% e 10% pode-se se observar melhor o comportamento da cauda da distribuição do total líquido gasto com sinistro.

Tabela 11- Freqüência acumulada.

Bloco	Freqüência	% cumulativo
-	2	0,00%
33.061,60	137.590	54,62%
66.123,20	46.931	73,24%
99.184,79	22.011	81,98%
132.246,39	13.885	87,49%
165.307,99	7.083	90,31%
198.369,59	4.385	92,05%
231.431,18	3.047	93,26%
264.492,78	3.904	94,80%
297.554,38	1.756	95,50%
330.615,98	1.316	96,02%
363.677,58	1.840	96,75%
396.739,17	716	97,04%
429.800,77	845	97,37%
462.862,37	753	97,67%
495.923,97	593	97,91%

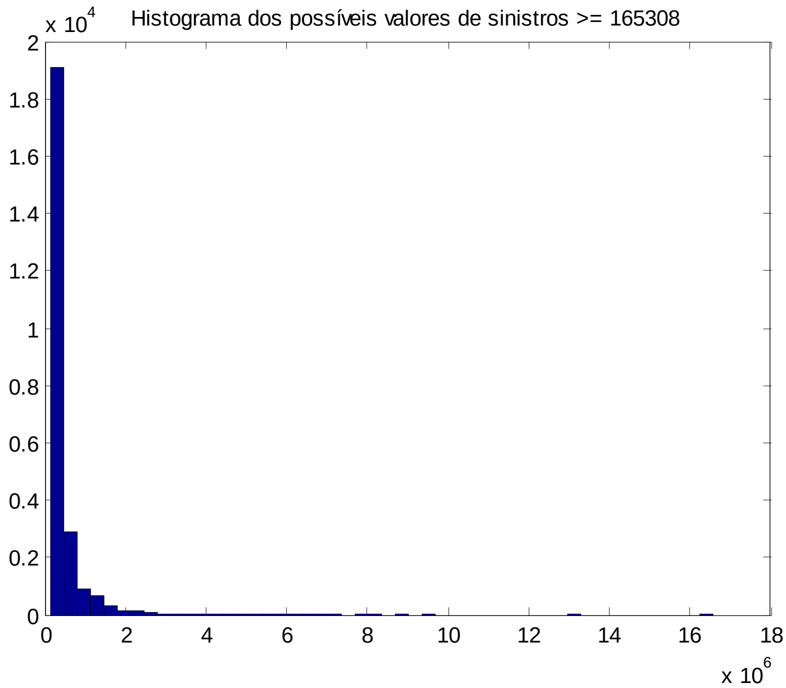
Na figura abaixo estão representados somente os valores menores que o ponto de corte. Há a possibilidade de ter a idéia de como a freqüência absoluta de capital segurado por indivíduo se distribui. Isto representa até aproximadamente 90% dos casos na base de dados.

Figura 7- Distribuição dos valores menores que o ponto de corte.



Neste trabalho dois modelos serão comparados, o modelo paramétrico que será aplicado na base toda e o modelo misto que será aplicado na base dividida. Esta divisão deixará 90% dos indivíduos em uma base que será chamada de apólices de “baixo custo” e os 10% da cauda em uma base chamada apólices de “alto custo”. É importante lembrar que as duas juntas formam a base toda. Na base de apólices de “baixo custo” será utilizado o modelo paramétrico com a aproximação para a distribuição Normal. Já na base de apólices de “alto custo” será aplicado o modelo não paramétrico utilizando o método de simulação de Monte Carlo. A seguir pode-se observar que os indivíduos que formarão a base de apólices de “alto custo” apresentam as seguintes freqüências absolutas:

Figura 8- Distribuição dos valores maiores que o ponto de corte.



Ou seja, nesta análise há uma estrutura na cauda que possivelmente pode ser explicada, que não fica visível graficamente ao observar os dados todos juntos. Logo, espera-se que o modelo misto consiga captar estes comportamentos melhor que o modelo paramétrico.

3.10.2
Cálculo do capital mínimo requerido utilizando o modelo misto

Uma vez dividida a base de dados a parte da base de dados de apólices de “alto custo” será modelada por simulação e na parte das apólices de “baixo custo” será aplicado o modelo paramétrico. A simulação paramétrica, utilizando Monte Carlo para as apólices de “alto custo”, consiste em simular a vida dos indivíduos no período de um ano, mês a mês. Desta forma a variável total líquido gasto com sinistros pode ser contabilizada a cada simulação tornando possível obter, ao final de 10.000 anos simulados, a distribuição do total líquido gasto com sinistros para estes indivíduos.

Nesta etapa existem duas distribuições para o total líquido gasto com sinistros, a distribuição obtida pelo método paramétrico para as apólices “baixo custo” e a distribuição obtida por simulação de Monte Carlo para as apólices de “alto custo”. Para transformar as duas distribuições em uma única distribuição vamos supor Y sendo a distribuição obtida pelo método paramétrico, cujo total líquido gasto com sinistros pelo Teorema Central do Limite segue uma distribuição Normal, e X a distribuição obtida pelo método não paramétrico de simulação de Monte Carlo. A distribuição final pode ser obtida da seguinte forma:

$$F(t^*) = P(X + Y < t^*)$$

$$P(X + Y > t^*) = \sum_{i=1}^{N} P(X + Y > t *, X = x_i) =$$

$$= \sum_{i=1}^{N} P(X + Y > t * | X = x_i) P(X = x_i) = \sum_{i=1}^{N} P(x_i + Y > t *) P(X = x_i) =$$

$$\sum_{i=1}^{N} P(Y > t^* - x_i) P(X = x_i) = \sum_{i=1}^{N} \left[1 - \Phi\left(\frac{t^* - x_i - \mu}{\sigma}\right)\right] P(X = x_i)$$

Assim, dados os x_i valores da distribuição do total gasto líquido gasto com sinistro (T) obtidos por simulação, a esperança μ e o desvio padrão σ que foram achados no modelo paramétrico é possível achar o valor de t^*. O valor de t^* deve ser tal que respeite a probabilidade α do capital mínimo requerido para solvência ser maior que o valor t^* obtido. Implícito neste cálculo está o conceito do VaR. Assim, sabendo que:

$$F(t^*) = \frac{1}{N}\sum_{i=1}^{N} \Phi\left(\frac{t^* - x_i - \mu}{\sigma}\right) = 1 - \alpha \tag{3.51}$$

O valor de t^*, que corresponderá ao valor do capital mínimo requerido necessário para garantir a solvência da carteira no período de um ano, para cada α é determinado através de uma busca não linear baseada no método da bissecção como pode ser visto no apêndice 8.2. Através do cálculo da probabilidade podemos variar o t^* a ponto de encontrar o valor que satisfaça a condição acima. Achado o valor t^* para um alfa pré-determinado será feita uma análise de sensibilidade onde o valor do capital mínimo requerido será calculado variando os cenários com taxa de juros diferentes, o que dará a sensibilidade do modelo perante a variável macroeconômica. Em seguida a análise será feita ao simular diferentes cenários para a variável persistência, que influencia na probabilidade de fim da apólice no ano. Estas análises e seus resultados serão expostos no capitulo seguinte onde será feita uma ilustração do modelo proposto.