Universidade Federal do Paraná
Setor de Ciências Exatas
Departamento de Estatística
Programa de Especialização em *Data Science* e *Big Data*

Delvo Sabino Santiago

Convolução entre as distribuições Poisson e Gamma para o cálculo da perda agregada máxima

Curitiba
2020

Delvo Sabino Santiago

Convolução entre as distribuições Poisson e Gamma para o cálculo da perda agregada máxima

Monografia apresentada ao Programa de Especialização em Data Science e Big Data da Universidade Federal do Paraná como requisito parcial para a obtenção do grau de especialista.

Orientador: Prof. Wagner Hugo Bonat

Curitiba

2020

Especialização em Data Science e Big Data
Universidade Federal do Paraná
dsbd.leg.ufpr.br

Convolução entre as distribuições Poisson e Gamma para o cálculo da perda agregada máxima

Delvo Sabino Santiago[1]
Prof. Wagner Hugo Bonat[2]

Resumo

Convolução é um termo estatístico que designa a função soma entre variáveis aleatórias. A convolução entre variáveis aleatórias com distribuições de Poisson e Gamma costuma ser chamada de distribuição Tweedie e tem ampla aplicação em vários campos da ciência, como em meteorologia, agricultura, saúde, investimentos e seguros, dentre outros. Neste artigo, focou-se, especificamente, na estimação da perda agregada máxima em um portfólio de riscos, representado pela carteira do seguro de automóvel. A perda agregada máxima representa o montante de indenizações suportável por uma cessionária de riscos dentro de um intervalo de tempo previamente fixado. Neste contexto a convolução entre variáveis aleatórias aparece de forma natural, uma vez que o custo do segurado é uma função do número (frequência) de sinistros modelado pela distribuição de Poisson e a severidade (custo) de cada sinistro modelado pela distribuição Gamma. Trabalhou-se com um banco de dados, disponibilizado pela Superintendência de Seguros Privados - SUSEP, cobrindo os anos de 2015 a 2019, compreendendo 16,4 milhões de registros. A solução está baseada em algoritmos desenvolvidos nas linguagens de programação SAS e Python, simulando múltiplos cenários, funções dos parâmetros das distribuições de probabilidade de Poisson e Gamma, havendo sido os referidos parâmetros estimados a partir do método da máxima verossimilhança. Na aplicação em foco, trabalhou-se com as seguintes variáveis atuariais: exposição, número de sinistros, indenizações incorridas, importâncias seguradas expostas, além das frequências de sinistros, severidade, prêmios de risco e carregamento de oscilação de risco. O resultado final é uma distribuição de probabilidade aproximada para a variável perda agregada máxima. Tal distribuição permite obter um intervalo de confiança para esta quantidade que é fundamental dentro do contexto de segurados e que é obtido como uma função não trivial dos parâmetros dos modelos Poisson e Gamma.

Palavras-chave: Seguros, Automóvel, Convolução, Poisson, Gamma, Sinistros, Tweedie.

Abstract

Convolution is a statistical term that designates the sum function between random variables. The convolution between random variables with Poisson and Gamma distributions is often called the Tweedie distribution and has wide application in several fields of science, such as meteorology, agriculture, health, investments and insurance, among others. In this article, we specifically focused on estimating the maximum aggregate loss in a risk portfolio, represented by the auto insurance portfolio. The maximum aggregate loss represents the amount of indemnities bearable by a risk assignee within a previously fixed time interval. In this context, the convolution between random variables appears naturally, since the insured cost is a function of the number (frequency) of claims modeled by the Poisson distribution and the severity (cost) of each claim modeled by the Gamma distribution. We worked with a data bank, made available by the Superintendency of Private Insurance - SUSEP, covering the years 2015 to 2019, comprising 16.4 million records. The solution is based on algorithms developed in the programming languages SAS and Python, simulating multiple scenarios, functions of the parameters of the Poisson and Gamma probability distributions, and these parameters were estimated using the maximum likelihood method. In the application in focus, we worked with the following financial variables: exposure, number of claims, indemnities incurred, insured amounts exposed, in addition to the frequency of claims, severity, risk premiums and risk fluctuation loading. The final result is an approximate probability distribution for the variable maximum aggregate loss. Such distribution allows obtaining a confidence interval for this quantity, which is fundamental within the unsafe context and which is obtained as a non-trivial function of the parameters of the Poisson and Gamma models.

Keywords: *Insurance, Automobile, Convolution, Poisson, Gamma, Claims, Tweedie.*

1 Introdução

Quando um empregador oferta um seguro de automóvel aos seus empregados, ele se preocupa com a frequência dos sinistros, desconhecendo a aleatoriedade com que estes registros ocorrem, além de se preocupar com a gravidade do custo de cada sinistro o que caracteriza tamanho aleatório de cada sinistro. Mas eles estão espe-

[1]Aluno do programa de Especialização em Data Science & Big Data, delvo.santiagoo@gmail.com.
[2]Professor do Departamento de Estatística - DEST/UFPR.

cialmente preocupados com os sinistros agregados ou como está descrito neste artigo a perda agregada máxima, que é a soma total de todos os sinistros da carteira exposta em um determinado intervalo de tempo t. Esta é a soma de um número aleatório de variáveis aleatórias e, como tal, é extremamente desafiadora de analisar uma vez que sua distribuição de probabilidade é composta por dois componentes i) o número ou frequência de sinistros e ii) a severidade ou custo de cada sinistro.

Se for assumido que a frequência segue um processo de Poisson e que as severidades são independentes e seguem a distribuição de probabilidade Gamma, a distribuição de probabilidade resultante é a chamada distribuição *Compound Poisson* que é um caso particular do modelo Tweedie.[1].

O objetivo deste trabalho é estimar a Perda Agregada Máxima de uma companhia seguradora, visando mitigar a probabilidade de ruína de uma companhia seguradora durante sua continuidade de negócios. A técnica analítica de suporte foi desenvolvida por Byron[2] e apresentada por Thomas N. Herzog sendo conhecida como *"The Convolution Method"*[3].

No tópico "Conjunto de Dados"descreve-se a base de dados, detalhando as variáveis que a compõem, especificando os respectivos formatos.

Em "Metodologia"apresenta-se a conceituação básica da perda agregada máxima, acompanhada da correspondente formulação matemática, que embasa sua apuração e estimação. O viés prático de apuração da perda agregada é exibida em "Cálculo da Perda Agregada Máxima". Finalmente, no tópico "Resultados"são inscritas as conclusões, assim como as observações cabíveis.

2 Conjunto de dados

O conjunto de dados utilizado neste trabalho corresponde ao mercado brasileiro de seguros de automóveis, perfazendo um total de 14,992 milhões de expostos.

Os dados foram publicados no site da Susep[4] e extraídos utilizando técnicas de *Web Scraping* com o software *Python*. Os dados são semestrais entre 2015 até o primeiro semestre de 2019. A estrutura da base de dados e as variáveis disponíveis são apresentadas na Tabela 1.

Variável	Descrição	Tamanho	Formato
Quantitativas:			
EXPOSICAO	Quantidade de veículos expostos	8	BEST19.
PREMIO	Soma dos valores de prêmios, ponderados pela exposição de cada apólice	8	BEST20.
F_ROUBO_FURTO	Quantidade de sinistros da cobertura roubo/furto	8	BEST3.
I_ROUBO_FURTO	Total de indenizações de sinistros da cobertura roubo/furto	8	BEST7.
F_COLISAO_P	Quantidade de sinistros da cobertura colisão (parcial)	8	BEST4.
I_COLISAO_P	Total de indenizações de sinistros da cobertura colisão (parcial)	8	BEST7.
F_COLISAO_T	Quantidade de sinistros da cobertura colisão (total)	8	BEST3.
I_COLISAO_T	Total de indenizações de sinistros da cobertura colisão (perda total)	8	BEST7.
F_INCENDIO	Quantidade de sinistros da cobertura incêndio	8	BEST1.
I_INCENDIO	Total de indenizações de sinistros da cobertura incêndio	8	BEST6.
F_OUTROS	Quantidade de sinistros de outras coberturas, como assistência 24 hs, etc	8	BEST3.
I_OUTROS	Total de indenizações de sinistros de outras coberturas, como assistência 24 hs, etc	8	BEST6.
ANO_MODELO	Ano e Modelo dos Veículos	8	6.
Qualitativas:			
DT_REF	Semestrais / 2015 - 2019	6	
FAIXA_ETARIA	1-1825, 2-2635, 3-2645, 4-4655, 5->55	13	
GENERO	Masculino e Feminino	9	$30.

Tabela 1: Descrição das variáveis que compõem a base de dados disponibilizada pela SUSEP.

3 Metodologia

Nesta seção, serão apresentadas as ideias fundamentais para a estimação da perda agregada máxima, resultante da convolução das distribuições Poisson e Gamma, bem como, da estimação da incerteza associada utilizando técnicas de simulação estocástica.

Dentro do contexto atuarial, as principais técnicas de análise podem ser divididas em dois grandes grupos:

a) Perfeitas: Usa-se quando se presume conhecer a distribuição subjacente ao processo estocástico sob estudo, isto é, quando a função de probabilidade da variável aleatória de interesse é conhecida.
b) Imperfeitas: Quando a função de probabilidade é desconhecida, recorremos a métodos de ajustamento ou estimação, buscando desvendar a distribuição que melhor representa o vetor de dados observados.

O processo de seguros, ora em estudo, pode ser visto como um processo estocástico devido à natureza aleatória de seu comportamento, haja vista que, para que um risco seja considerado segurável, deve ser incerto, possível e futuro, com exceção para o seguro de vida em que a incerteza é caracterizada pela época da ocorrência do evento morte. Podemos citar como exemplos de eventos aleatórios em seguros, a colisão de um veículo ou um incêndio, roubo, entre outros.

Em um sistema estocástico de seguros há basicamente dois tipos de variáveis aleatórias: variáveis aleatórias de entrada e variáveis aleatórias de saída. As variáveis aleatórias de entrada podem ser representadas pelas distribuições de frequência de sinistros e pelas distribuições das consequências econômicas, enquanto a variável de saída é representada pela distribuição de perdas.

Sob a perspectiva do segurado, o pagamento dos prêmios ocorre de maneira mais previsível, enquanto os sinistros têm suas ocorrências regidas pela aleatoriedade. Pelo lado do segurador há a reserva inicial, que pode se mostrar suficiente ou não para o pagamento das indenizações associadas aos eventos cobertos incorridos. Ressaltamos que a natureza estocástica da saída do fluxo de caixa é dupla, ou seja, não é conhecido *a priori* o instante que ocorrerá o sinistro, nem tampouco o valor do montante proporcionado por cada sinistro, excluídos os seguros de vida e equivalentes, concebidos sob o regime da repartição simples ou dos capitais de cobertura, nos quais o montante a ser indenizado se caracteriza pelo capital segurado, de conhecimento antecipado.

O estudo do processo de sinistros normalmente é feito através da formulação de premissas, materializadas em funções estatísticas, seguida do ajustamento das mesmas aos dados observados, empregando técnicas de estimação. Dentro deste contexto os métodos de ajustamento ou estimação de parâmetros mais utilizados são os métodos de máxima verossimilhança e momentos.

O princípio da máxima verossimilhança diz que se deve usar como melhor estimativa para os parâmetros da distribuição aquelas que maximizam a verossimilhança do evento observado, ou seja, os valores que maximizam a probabilidade de a amostra observada ser obtida. Por outro lado, o método dos momentos apenas iguala os momentos teóricos ou populacionais com os momentos amostrais para obter os estimadores pela resolução de um sistema de equações. Neste trabalho optou-se pelo emprego do método da máxima verossimilhança.

O montante que uma seguradora irá despender com o pagamento de sinistros é uma função do número de sinistros ocorrido na carteira e da severidade de cada sinistro. O número de sinistros é uma das grandezas estocásticas contidas no modelo de risco coletivo, sendo seu entendimento fundamental para escolha da função de probabilidade, que descreverá o comportamento da variável "número de sinistros". Deste modo, um dos objetivos atuariais é a estimação da distribuição de probabilidade, associada à variável aleatória $k \rightarrow$ "número de sinistros"[5]:

$$P\left(K=k\right), \quad k = 0, 1, 2 \cdots .$$

Aceitando-se a suposição de que os sinistros acontecem independentemente em um determinado intervalo de tempo. A distribuição da variável aleatória K número de sinistros poderá ser aproximada pela distribuição de Poisson. Para expressarmos o processo gerador do número de sinistros como um processo de Poisson, é necessário que as seguintes condições sejam satisfeitas:

a) Os números de sinistros que ocorrem em quaisquer dois intervalos não superpostos são independentes (independência de incrementos);

b) A probabilidade de uma ocorrência durante um pequeno intervalo de tempo deve ser aproximadamente proporcional à amplitude do referido intervalo;

c) A probabilidade de ocorrência de 2 ou mais eventos em um reduzidíssimo intervalo de tempo deve ser de magnitude menor que o da ocorrência de 1 evento.

Se K é uma variável aleatória com distribuição de Poisson, então as probabilidades correspondentes aos diferentes valores de K podem ser obtidas através da equação

$$P(K_t = k) = \frac{(\lambda t)^k \exp -\lambda t}{k!}, \tag{1}$$

onde λ é um número real positivo, conhecido como parâmetro de Poisson, sendo igual ao valor esperado de K, isto é $E(K) = \lambda$. Neste contexto o parâmetro da distribuição de Poisson, representa o número de sinistro por unidade de tempo. É importante lembrar que no caso da distribuição de Poisson o estimador de momentos para λ coincide com o estimador de máxima verossimilhança e corresponde a média amostral.

Outra importante distribuição que fará parte do nosso modelo, e que modelará a severidade é a distribuição Gamma. As principais distribuições do tipo contínuo para modelagem da severidade são a Exponencial, Pareto, Log-normal, dentre outras. Como estamos tratando de uma carteira de automóveis, optamos por modelar a severidade através da distribuição Gamma, devido a suas propriedades matemáticas e formato assimétrico, que costuma garantir boa aderência aos dados observados na prática atuarial para este ramo de seguro.

Sendo assim, neste artigo a severidade dos sinistros será modelada utilizando a distribuição Gamma, que por sua vez é descrita por dois parâmetros α e β. Para a variável aleatória X severidade do sinistro os estimadores de momentos da distribuição Gamma são dados pelas expressões abaixo, extraídas de Thom[6].

$$\hat{\alpha} = \frac{1 + \sqrt{1 + 4\frac{D}{3}}}{4D}, \quad \text{e} \tag{2}$$

$$\hat{\beta} = \frac{\bar{X}}{\hat{a}}, \tag{3}$$

em que

$$D = \ln(\bar{X}) - \frac{1}{n}\sum_{1}^{n} \ln(x_i) \quad \text{e} \quad \bar{X} \quad \text{é a média amostral.}$$

O montante de sinistros agregado é a soma dos sinistros individuais cobertos por um contrato de seguros, em um dado período. No modelo de risco coletivo, as perdas estarão representadas segundo duas componentes: a frequência de sinistro e a severidade. A distribuição da perda agregada S será a combinação destas duas variáveis, gerando a distribuição da perda agregada. A bibliografia denomina este processo como Poisson Composta (*Compound Poisson*).

4 Cálculo da Perda Agregada Máxima

Existem basicamente duas abordagens clássicas para o estudo da distribuição de perdas em um portfólio de seguros. São elas: modelo individual e modelo coletivo.

No modelo individual, a variável aleatória S, que representa o volume de indenizações do portfólio de risco, é dada pela seguinte equação:

$$S = \sum_{i=1}^{N} X_i \tag{4}$$

onde X_i corresponde ao valor indenizado por conta do risco de ordem i que varia de 1 a N, sendo N o total de itens cobertos. Qual seja N não é uma variável aleatória.

Já no modelo de risco coletivo, a perda no portfólio está representada pela seguinte equação:

$$S = \sum_{i=1}^{N} X_i \tag{5}$$

onde N é uma variável aleatória que indica a quantidade de sinistros ocorridos no portfólio, enquanto a variável aleatória X_i representa o valor da indenização associado ao $i$ésimo sinistro. De uma forma, mais precisa tem-se

$$S = \begin{Bmatrix} \sum_{i-0}^{n} X_i \, se \, N = n > 0 \\ 0 \, se \, N = 0 \end{Bmatrix} \tag{6}$$

onde,

- S representa o montante da perda agregada;
- N representa o número de sinistros ocorridos no portfólio;
- X_i representa o montante do $i$ésimo sinistro.

Para uma melhor compreensão do modelo é preciso conhecer duas suposições importantes:
1. $X_1 + X_2 + \ldots + X_N$ são variáveis aleatórias identicamente distribuídas;
2. As variáveis aleatórias $X_1 + X_2 + \ldots X_N$, são mutuamente independentes.

Uma terceira condição a ser considerada nas aplicações práticas é a homogeneidade do grupo representado pela referida soma, o que explica a necessidade da modelagem atuarial para devida segregação dos riscos, segundo à classe a que pertencem.

A média e a variância da variável aleatória S são dadas por

$$E(S) = E(N) \cdot E(X) \quad \text{e}$$

$$Var(S) = E(X)^2 Var(N) + Var(X) \cdot E(N).$$

Conforme podemos observar na equação E(S), o modelo de risco coletivo mostra que o processo estocástico gerador da distribuição de perdas é governado por duas distribuições de probabilidades, sendo elas: a distribuição de frequência (de onde se extrai o número de sinistros) e severidade (ou valor médio da indenização) dos sinistros.

Sob condições do Teorema Central do Limite, temos

$$S \sim N(E(S), Var(S)).$$

onde N(A,B) representa a distribuição de esperança matemática A e variância B.

Não custa lembrar que o objetivo central é a produção de estimativas para a variável aleatória S. A presença da aleatoriedade nos conduz à necessidade de estimar os parâmetros da variável S, quais sejam, sua esperança matemática e sua variância (ou seu desvio-padrão). Assim, note que tanto a $E(S)$ como a $Var(S)$ são funções dos parâmetros λ, α e β. Tais parâmetros serão estimados a partir dos dados observados e portanto carregam um nível de incerteza que deve ser levado em conta. Neste trabalho para incorporar a incerteza, adotamos método de simulação estocástica conforme descrito na próxima seção.

4.1 Simulação da Perda Agregada

Descrever o comportamento da distribuição de probabilidades da variável aleatória Perda Agregada em um modelo de risco coletivo é um dos exemplos de aplicação de simulação em processos estocásticos. Para tanto, devemos gerar observações aleatórias que representem a frequência de sinistros de carros, m, de acordo com o parâmetro estimado para a distribuição de Poisson. O passo seguinte é simular os m possíveis montantes de perdas relativas ao período a que se referem, usando-se os parâmetros estimados para a distribuição Gamma, eleita para representar a variável aleatória "severidade". Considerando que todos os parâmetros necessários para a simulação foram devidamente estimados, temos que:

i. Gerar m valores, $X_1, X_2, X_3, \ldots, X_m$ identicamente distribuídos, usando o algoritmo específico à distribuição escolhida;
ii. Calcular a soma $S = X_1 + X_2 + X_3 + \ldots + X_m$;
iii. Fazer um histograma dos valores encontrados de S.
iv. Todos os passos deste processo - simulação da frequência e severidade de *Claims* - são reproduzidos inúmeras vezes, até que seja encontrada uma amostra suficientemente robusta que nos permita fazer as necessárias inferências, utilizando-se os métodos estatísticos convencionais.

4.2 Aplicação em Tarifação

Por fim, a partir dos valores obtidos nas seções anteriores podemos chegar a tarifação do seguro. A primeira quantidade de interesse é o chamado PP = Prêmio Puro da Carteira, que é obtido pela seguinte fórmula:

$$PP = E(S) + Z_{1-\alpha} D(S).$$

Como por definição PR = Prêmio de Risco = $E(S)$

$$PP = PR + Z_{1-\alpha} D(S)$$

ou

$$PP = \mu.M_1 + Z_{1-\alpha} \sqrt{\mu.M_2}.$$

$$PPI = \frac{PP}{Expostos},$$

onde

$$D(S) = \sqrt{Var(S)}.$$

Importante novamente ressaltar que todas as quantidades apresentadas são funções dos parâmetros das distribuições Poisson e Gamma. Tais parâmetros são estimados baseados nos dados e portanto, esses estimadores são também variáveis aleatórias e que devem ter sua distribuição de probabilidade avaliada. Note que as operações realizadas com tais estimadores não são sempre triviais e daí vem a complexidade em definir a distribuição de probabilidade associadas a cada variável. É neste ponto que as técnicas de simulação são fundamentais, pois ao invés de buscar a distribuição de probabilidade teórica usando técnicas usuais de transformações de variáveis aleatórias, usamos as técnicas de simulação para obter realizações de tais variáveis aleatórias e obtemos a sua distribuição empírica que pode ser observada através de histogramas.

Às perdas catastróficas estão associadas probabilidades cada vez menores de ocorrência, distribuição uniforme no intervalo $[0, 1]$ utilizando-se algoritmo específico para a distribuição escolhida.

5 Resultados

Nesta seção nos descrevemos a aplicação dos métodos descritos nas seções anteriores aos dados referidos na seção "Conjunto de Dados". O resultado final obtido é a distribuição de probabilidade da variável perda agregada máxima. Optamos por aplicar o método de Monte Carlo considerando as 5 regiões padronizadas pela Superintendência de Seguros Privados - SUSEP. Para cada região foram geradas 1000 corridas, empregando a ferramenta SAS. No final das simulações foi realizado o teste de aderência de Kolmogorov-Smirnov sendo que em todos os casos a hipótese nula não foi rejeitada, configurando assim a viabilidade de emprego da convolução, antes descrita, para descrever a experiência de sinistros do portfólio de riscos.

5.1 Região

A região 1 corresponde ao agrupamento de CEPs de risco que abrangem todo o estado de São Paulo (capital e interior).

Embora o cálculo tenha sido feito para todas as regiões do Brasil, os resultados entre regiões são muito similares, então trouxemos os gráficos da Região 1, a qual é a região mais representativa do artigo, região de São Paulo.

Demonstraremos a seguir com intervalo de confiança de $0,95$ os resultados da simulação que a modelagem alcançou diante dos paramentos do Gráfico de Distribuição de Probabilidade- Número de Sinistros, referente a região SP, da Região 1 (Figura 1).

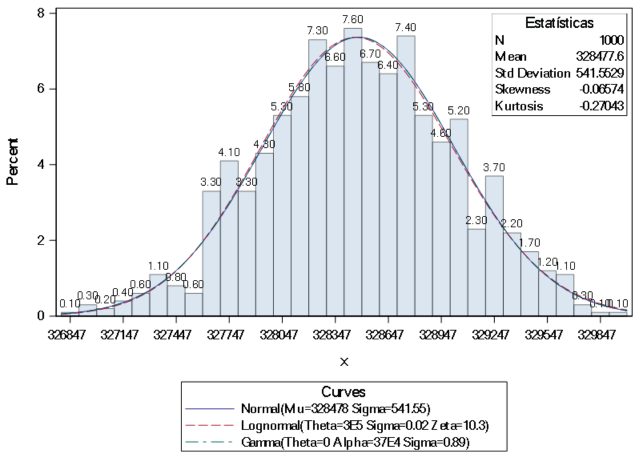

Figura 1: Região 1 - Gráfico de Distribuição de Probabilidade - Número de Sinistros, referente a região SP.

Na Figura 2, temos o Gráfico de Distribuição de Probabilidade - Severidade, referente a região SP, da Região 1:

Na abcissa do gráfico "Região 1 - Gráfico de Distribuição de Probabilidade- Número de Sinistros, referente a região SP", exibem-se as quantidades de sinistros resultantes da simulação, enquanto no eixo das ordenadas o percentual correspondente. A simetria acentuada do

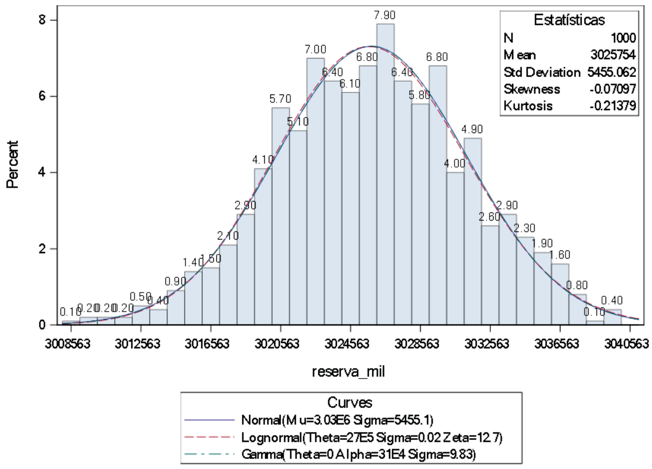

Figura 2: Região 1 - Gráfico de Distribuição de Probabilidade - Severidade, referente a região SP.

Na Tabela 2, temos o Resultado da Perda Agregada Máxima com IC, referente a região SP, da Região 1:

Sinistros Agregados / .000					
Min	Max	Std	5 - IC	Med	95 - IC
3.008.171,1	3.039.991,4	5.455,1	3.016.761,5	3.025.753,9	3.034.845,4

Prêmio de Risco			PPI	
Min	Med	Max	SES	%DIF
476,10	477,52	478,96	479,65	0,446%

Sinistros Agregados	Min	Sinistro Agregado Mínimo
	Max	Sinistro Agregado Máximo
	Std	Desvio Padrão
	5 - IC	Intervalo de Confiança de 5%
	Med	Média de Sinistros Agregados
	95 - IC	Intervalo de Confiança de 95%
Premio de Risco	Min	Premio de Risco Mínimo
	Med	Premio de Risco Médio
	Max	Premio de Risco Máximo
	SES	Premio de Risco Mercado
PPI	%DIF	Premio Puro Individual

Tabela 2: Região 1 - Resultado da Perda Agregada Máxima com IC, referente a região SP.

resultado induz à reflexão sobre o teorema do Limite Central, a partir do qual se demonstra à tendência de aproximação à distribuição Normal, na medida em que o tamanho da amostra tender para infinito.

Na tabela associada a cada gráfico estão dispostas as seguintes informações consequentes das 1000 corridas de simulação:

MIN = corresponde ao valor mínimo obtido na simulação;

MAX = corresponde ao valor máximo obtido na simulação;

STD = desvio-padrão;

5 IC = limite inferior do intervalo de confiança para um nível de significância de 95%;

MED = mediana;

95 IC = limite inferior do intervalo de confiança para um nível de significância de 95%;

Na Tabela 3, temos o Resultado da Perda Agregada Máxima com IC, referente a região RS/PR/SC, da Região 2:

Sinistros Agregados / .000					
Min	Max	Std	5 - IC	Med	95 - IC
1.088.962,5	1.104.040,9	2.728,5	1.092.260,6	1.096.853,1	1.101.600,6

Prêmio de Risco			PPI	
Min	Med	Max	SES	%DIF
418,44	420,20	422,02	423,39	0,759%

Sinistros Agregados	Min	Sinistro Agregado Mínimo
	Max	Sinistro Agregado Máximo
	Std	Desvio Padrão
	5 - IC	Intervalo de Confiança de 5%
	Med	Média de Sinistros Agregados
	95 - IC	Intervalo de Confiança de 95%
Premio de Risco	Min	Premio de Risco Mínimo
	Med	Premio de Risco Médio
	Max	Premio de Risco Máximo
	SES	Premio de Risco Mercado
PPI	%DIF	Premio Puro Individual

Tabela 3: Região 2 - Resultado da Perda Agregada Máxima com IC, referente a região RS/PR/SC.

Na Tabela 4, temos o Resultado da Perda Agregada Máxima com IC, referente a região MG/RJ, da Região 3:

Sinistros Agregados / .000					
Min	Max	Std	5 - IC	Med	95 - IC
1.540.732,0	1.570.354,2	4.305,6	1.549.407,2	1.556.471,8	1.563.535,8

Prêmio de Risco			PPI	
Min	Med	Max	SES	%DIF
587,17	589,85	592,53	593,97	0,698%

Sinistros Agregados	Min	Sinistro Agregado Mínimo
	Max	Sinistro Agregado Máximo
	Std	Desvio Padrão
	5 - IC	Intervalo de Confiança de 5%
	Med	Média de Sinistros Agregados
	95 - IC	Intervalo de Confiança de 95%
Premio de Risco	Min	Premio de Risco Mínimo
	Med	Premio de Risco Médio
	Max	Premio de Risco Máximo
	SES	Premio de Risco Mercado
PPI	%DIF	Premio Puro Individual

Tabela 4: Região 3 - Resultado da Perda Agregada Máxima com IC, referente a região MG/RJ.

Na Tabela 5, temos o Resultado da Perda Agregada Máxima com IC, referente a região NO, da Região 4:

Sinistros Agregados / .000					
Min	Max	Std	5 - IC	Med	95 - IC
1.110.119,6	1.128.127,6	3.137,7	1.113.322,2	1.118.316,0	1.123.613,3

Prêmio de Risco			PPI	
Min	Med	Max	SES	%DIF
559,62	562,13	564,79	566,54	0,784%

Sinistros Agregados	Min	Sinistro Agregado Mínimo
	Max	Sinistro Agregado Máximo
	Std	Desvio Padrão
	5 - IC	Intervalo de Confiança de 5%
	Med	Média de Sinistros Agregados
	95 - IC	Intervalo de Confiança de 95%
Premio de Risco	Min	Premio de Risco Mínimo
	Med	Premio de Risco Médio
	Max	Premio de Risco Máximo
	SES	Premio de Risco Mercado
PPI	%DIF	Premio Puro Individual

Tabela 5: Região 4 - Resultado da Perda Agregada Máxima com IC, referente a região NO.

Na Tabela 6, temos o Resultado da Perda Agregada Máxima com IC, referente a região OUTROS, da Região 5:

Sinistros Agregados / .000					
Min	Max	Std	5 - IC	Med	95 - IC
831.758,4	847.171,9	2.292,9	836.549,0	839.940,5	843.902,8

Prêmio de Risco			PPI	
Min	Med	Max	SES	%DIF
590,05	592,45	595,24	597,56	0,863%

Sinistros Agregados	Min	Sinistro Agregado Mínimo
	Max	Sinistro Agregado Máximo
	Std	Desvio Padrão
	5 - IC	Intervalo de Confiança de 5%
	Med	Média de Sinistros Agregados
	95 - IC	Intervalo de Confiança de 95%
Premio de Risco	Min	Premio de Risco Mínimo
	Med	Premio de Risco Médio
	Max	Premio de Risco Máximo
	SES	Premio de Risco Mercado
PPI	%DIF	Premio Puro Individual

Tabela 6: Região 5 - Resultado da Perda Agregada Máxima com IC, referente a região OUTROS.

PREMIO DE RISCO = valor a ser cobrado de cada unidade exposta para cobrir exclusivamente as indenizações incorridas;

PPI = prêmio de risco acrescido da margem de segurança estatística - analogamente, presta-se a cobrir os eventos esperados, com margem de segurança para proteção exclusiva da volatilidade de fundo aleatório.

Pelo ângulo da gestão estratégica, o segurador deve observar o limite superior do intervalo de confiança (95 IC) e avaliar a probabilidade de ocorrência de quantidades de sinistros acima deste patamar.

Num horizonte de 60 meses, por exemplo, considerando uma probabilidade de ocorrência de 5%, a expectativa de ocorrer sinistros acima do IC 95 é de 3 eventos. Estes parâmetros determinam a política de constituição da provisão técnica para fazer frente aos sinistros futuros.

Sob o ponto de vista prático, os gestores da seguradora observarão se os prêmios comerciais arrecadados no futuro serão suficientes para arcar com a carga de indenizações geradas nas simulações.

A partir desta análise os gestores decidirão entre aumentos de preços e aportes de capital.

Estabelecidas as políticas regionais de preços, análise adicional se fará agregando o conjunto de regiões para reavaliar as necessidades de capitalização da sociedade seguradora. Esta continuidade de análise está fora do escopo deste artigo.

6 Discussão

Os resultados atingidos são uma aproximação resultante do emprego do teorema do limite central. Com efeito, a soma de variáveis aleatórias independentes tende para a variável aleatória com distribuição normal, conforme a quantidade de parcelas aumenta.

Uma oportunidade para estudos futuros seria a agregação de fatores de risco, como Sexo, Idade e Perfil do Segurado, que deixaram de ser considerados por razões de certificação da base de dados e do objetivo deste estudo.

Objetivamente, a consideração da presença de fatores de risco e sua influência estratégica fica mais bem ajustada em estudos envolvendo técnicas clássicas de modelagem do prêmio, estando, portanto, fora do escopo deste artigo. Um efeito que foi devidamente considerado diz respeito à variância, como função do volume de unidades expostas.

Resumidamente, quando a variabilidade é mais elevada, um fator de penalização é imposto aos fatores que compõem a Perda Agregada Máxima e introduzido no modelo de modo a restabelecer a solvência da sociedade seguradora em um nível considerado aceitável. O processo de simulação afere, precisamente, a sensibilidade resultante da variação da massa exposta.

Agradecimentos

Agradeço ao Prof. Wagner Hugo Bonat pela dedicação e compromisso nesta aventura de pegar um trabalho tão complexo e acreditar que seria possível ser feito, foram duras horas de trabalho e dedicação para finalizar este artigo, sem compromisso teria sido impossível o feito. O modelo final é prático e será utilizado por muitos anos durante a minha vida profissional.

Agradeço ao Prof. Antônio Westenberger pela dedicação, apoio e paciência no processo de revisão conceitual deste artigo, por ter proporcionado seu vasto conhecimento sobre o tema, o qual me permitiu elaborar um trabalho fundamentado em bases sólidas.

Agradeço a minha família, minha esposa Leudina, meu filho Apolo e em especial, minha filha Mariana Ferreira Sabino, a qual me incentivou e auxiliou durante o desenvolvimento deste trabalho.

Referências

[1] D.A. Andersen and W.H. Bonat. *Double Generalized Linear Compound Poisson Models to insurance claims data*. Electronic Journal of Applied Statistical Analysis, 2017.

[2] B. J. T. Morgan. *Elements of Simulation*. Chapman & Hall, 1986.

[3] T. N. Herzog and G. Lord. *Applications of Monte Carlo Methods to Finance and Insurance*. Chapman & Hall, 1986.

[4] Susep. *Autoseg*. Susep, 2015-2019.

[5] C.D. Daykin. *Practical Risk Theory for Actuaries*. Chapman & Hall, 1994.

[6] H.C.S. Thom. *A note on the Gamma Distribution*. U.S.Weather Bureau, 1958.