Projeto de Graduação

DEPARTAMENTO DE ENGENHARIA ELÉTRICA

02 de dezembro, 2011

SOMAS ALEATÓRIAS EM MODELOS DE RUÍNAS

Andréa Micheli Alzuguir

Luciana Schmid Blatter Moreira

PUC RIO

www.ele.puc-rio.br

Projeto de Graduação

DEPARTAMENTO DE ENGENHARIA ELÉTRICA

SOMAS ALEATÓRIAS EM MODELOS DE RUÍNAS

Alunas: Andréa Micheli Alzuguir

Luciana Schmid Blatter Moreira

Orientador: Cristiano Augusto Coelho Fernandes

Trabalho apresentado com requisito parcial à conclusão do curso de Engenharia Elétrica na Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, Brasil.

Agradecimentos

"É muito melhor arriscar coisas grandiosas, alcançar triunfos e glórias, mesmo expondo-se a derrota, do que formar fila com os pobres de espírito que nem gozam muito nem sofrem muito, porque vivem nessa penumbra cinzenta que não conhece vitória nem derrota."

Theodore Roosevelt

Resumo

Realizou-se esse trabalho com o objetivo de apresentar, estudar e modelar a Teoria das Ruínas aplicadas na área de seguros. Buscou-se capturar as características e peculiaridades das distribuições das variáveis envolvidas modelo principal através de estimações no software @Risk. Além disso, as deduções de alguns resultados foram realizadas através do software Maple. Em seqüência, unir-se-á os métodos de estimação de parâmetros aplicados aos dados referentes aos seguros de veículos o que foram descritos pela teoria com o que terá sido observado através dos resultados práticos obtidos.

Palavras-chave: Seguro; Teoria das Ruínas; Soma Aleatória; Momentos; Distribuições de Probabilidade.

DEPARTAMENTO
DE ENGENHARIA
ELÉTRICA

Random sums in ruin models

Abstract

This work was written in order to present, study and model the Ruin theory applied in insurance. The characteristics and peculiarities of the distributions of the variables involved in the main model were estimatimated in @ Risk software. Moreover, the results of some deductions were performed using the software Maple. Subsequently, using parameter estimation methods applied to vehicle insurance data will join what has been described by the theory with what has been observed through the practical results obtained.

Keywords: Insurance; Ruin theory; Moments; Random sums; Probability distributions.

Sumário

1. Introdução 1
   a. A história do seguro 1
   b. Teoria da Ruína 4
   c. Risco e incerteza 5
2. Metodologia 6
3. Conceitos Relevantes 7
   a. Teoria da probabilidade e variávia aleatórias 7
   b. Processos Estocásticos 8
4. Somas aleatórias 9
   a. Soma de variáveia aleatórias independentes 9
   b. Momentos de uma soma aleatória 10
5. Modelos Clássicos de Ruína 15
   a. O processo de risco 15
   b. Modelo de risco coletivo e o modelo de risco individual 16
6. Aplicações 17
   a. Estimações para seguros de veículos 17
   b. Construção da distribuição da soma aleatória 22
7. Conclusão 25
8. Bibliografia 26
9. Apêndice 27
   a. Método dos momentos 27
   b. Método da verossimilhança 27

Sumário de tabelas

- Figura 1- Relação da ruína ao longo do tempo 4
- Figura 2- Assimetria positiva (A), nula (B) e negativa (C) 13
- Figura 3- Classificações da curtose: leptocúrtica (a), mesocúrtica (b) e platicúrtica (c) 14
- Figura 4- Indenizações pagas pela seguradora no tempo 15
- Figura 5- Distribuição Poisson para a variável N 19
- Figura 6- Distribuição binomial negativa para a variável N 19
- Figura 7- Distribuição lognormal para a variável X 20
- Figura 8- Distribuição Pareto para a variável X 21
- Figura 9-Distribuição gama para a variável X 22
- Figura 10- Distribuição gama para soma aleatória 24

- Tabela 1 - Relação entre as taxas do PIB e os prêmios de seguros 2
- Tabela 2- Proporção dos prêmios em relação ao PIB 2
- Tabela 3 - Detalhamento dos prêmios de seguro 3
- Tabela 4- Distribuição de prêmios de seguros nos estados brasileiros. Fonte: SUSEP. 3
- Tabela 5 - Base de dados. Fonte: SUSEP 17

1. Introdução

Os seguros são instrumentos de proteção e reparação indispensáveis para o homem moderno, uma vez que este vive em busca de segurança e estabilidade. A implantação do seguro na sociedade está relacionada com o mutualismo (interação que traz benefício para ambas as partes envolvidas) e previdência (capacidade de manter reserva).

O mercado brasileiro de seguros movimentou mais de R$ 107 bilhões em 2009, número que corresponde a 3,1% do nosso Produto Interno Bruto (PIB). Cerca de 50% deste total retornou à sociedade em indenizações de seguros gerais, posicionando o Brasil na 17ª posição mundial deste segmento. Esses números são reflexos da retomada da economia mundial, mas também mostram claramente que a sociedade brasileira está cada vez mais consciente do papel positivo dos seguros.

O estudo da Ciência Atuarial aplicada à área de seguros tem conquistado um considerável espaço devido à uma crescente demanda mundial. Tendo isto em vista, este trabalho visa abordar diferentes estimações de distribuições probabilísticas que modelem de forma mais realística as variáveis aleatórias envolvidas no processo que define o valor das indenizações pagas por uma seguradora, bem como suas freqüências. A partir das conclusões obtidas com estas estimações, será possível a construção de um modelo que define o dispêndio total de uma seguradora. A ruína de uma seguradora, por definição, ocorrerá quando esse gasto for maior que seu montante inicial somado aos recebimentos mensais provenientes dos segurados.

a. A história do seguro

A busca pela estabilidade sempre esteve presente na vida do homem. Desde os primórdios o homem enfrenta provações através das mais diferentes situações que englobam desde variações climáticas a prejuízos financeiros. Os homens uniam-se em grupos através de associações que tinham como objetivo ressarcir membros que fossem atingidos por situações adversas. Com o passar do tempo, a cobertura do risco foi fortalecida, especialmente durante o mercantilismo, uma vez que se desenvolveram os chamados contratos de dinheiro e risco marítimo, dando início a gestão de risco. A proliferação das seguradoras ocorreu de forma mais intensa com a revolução industrial, uma vez que os avanços tecnológicos e as mudanças na vida cotidiana trouxeram riscos com possibilidade de perdas severas.

Objetivando o aumento da segurança e o controle do risco ao qual se é exposto, surge o seguro. Este pode ser apresentado como um contrato entre duas ou mais partes, onde uma, chamada segurador, assume um compromisso com o segurado, mediante o pagamento de um prêmio, pagando-lhe determinada quantia no caso de ocorrência do evento segurado.

O setor de seguros no Brasil passou por mudanças significativas desde a implementação do Real em 1994. A tabela a seguir apresenta a taxa de crescimento anual do PIB e a taxa de crescimento dos prêmios de seguros dos períodos indicados:

Período	Taxa de Crescimento do PIB (%)	Prêmios de Seguro	
		Nominal (%)	Real (%)
1996	2,2	17,4	7,1
1997	3,4	21,2	15,2
1998	0,0	5,5	3,7
1999	0,3	4,6	(4,0)
2000	4,3	13,3	6,9
2001	1,3	10,2	2,4
2002	2,7	8,9	(3,2)
2003	1,1	9,7	0,4
2004	5,7	14,1	6,0
2005	2,9	13,7	7,6
2006	3,7	11,9	8,5
2007	5,4	12,1	7,6
2008	5,1	15,1	9,2

Tabela 1 - Relação entre as taxas do PIB e os prêmios de seguros

Fonte: SUSESP, FENASEG e IBGE.

Atualmente, grandes valores são segurados pelas seguradas no Brasil. Tomando-se como base o setor de seguros de países desenvolvidos, acredita-se que o mercado nacional ainda não atingiu seu pleno potencial, como pode ser observado no gráfico a seguir:

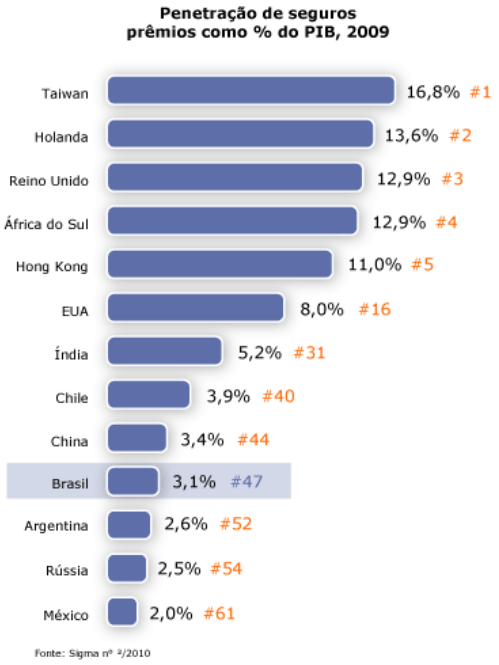

Tabela 2- Proporção dos prêmios em relação ao PIB

A receita de seguros no Brasil é proveniente basicamente da venda de apólices de seguros de automóveis, saúde e vida. A tabela a seguir detalha os prêmios de seguros nos períodos indicados:

Principais Segmentos	**Exercício Findo em 31 de dezembro**					
	2007	**Part. de Mercado**	**2008**	**Part. de Mercado**	**2009**	**Part. de Mercado**
Automóveis	13.535,12	27,7%	15.309,85	26,7%	17.293,50	28,5%
Outros ramos elementares	14.264,15	29,2%	17.028,67	29,7%	15.580,44	25,7%
Saúde	10.530,36	21,5%	12.792,88	22,3%	14.034,16	23,1%
Vida	10.586,45	21,6%	12.134,64	21,2%	13.810,86	22,7%
Total	**48.916,09**	**100,0%**	**57.266,03**	**100,0%**	**60.718,97**	**100,00%**

Tabela 3 - Detalhamento dos prêmios de seguro

Fonte: SUSEP e ANS.

De acordo com a SUSEP, que é o órgão responsável pelo controle e fiscalização do mercado de seguros, previdência privada aberta e capitalização, em 2009, a maior parcela de prêmios de seguros no Brasil (aproximadamente 65,9%) foi gerada na Região Sudeste, onde está concentrado o maior PIB do país. Além disso, em 2009 o Estado de São Paulo foi responsável por 47,1% do volume total de prêmios de seguros, seguido pelos Estados do Rio de Janeiro (10,4%), Minas Gerais (7,3%) e Rio Grande do Sul (7,1%). A tabela a seguir demonstra a distribuição de prêmios de seguros entre os estados brasileiros para os períodos indicados:

Principais estados	**Exercício Findo em 31 de dezembro**		
	2007 (%)	**2008 (%)**	**2009 (%)**
São Paulo	48,4	46,7	47,1
Rio de Janeiro	10,7	10,1	10,4
Minas Gerais	6,8	7,1	7,3
Rio Grande do Sul	7,0	7,0	7,1
Paraná	6,3	6,4	6,3
Outros	20,8	22,7	21,9
Total	**100,0**	**100,0**	**100,0**

Tabela 4- Distribuição de prêmios de seguros nos estados brasileiros. Fonte: SUSEP.

Segundo estimativas, é provável que este ano de 2011 entre na história do setor de seguros no Brasil. O mercado projeta prêmios acima de R$ 200 bilhões, com crescimento de 12%, em relação ao ano de 2010.

Em tempos de incerteza, a sociedade sente a necessidade de segurança, isso é observado tanto para o âmbito pessoal quanto para corporativo. Imerso nesse contexto, faz-se necessário o estudo de teorias e a modelagem de processos que respaldam as decisões tomadas por ambas as partes envolvidas em um seguro. Visto isso, segue o conceito de Teoria da Ruína, que na verdade se trata de um dos primeiros modelos que surgiram com o propósito de evitar e mostrar o ponto de ruína de uma seguradora.

b. Teoria da Ruína

A Teoria da Ruína é estudada na área de Ciências Atuariais. Uma vez que a teoria em questão é baseada na aplicação de diversos modelos probabilísticos e em processos estocásticos, ela também é conhecida como teoria do risco coletivo. Ela possui sua base pautada em modelos matemáticos que visam descrever os lucros e prejuízos de uma seguradora segundo sua vulnerabilidade e insolvência.

O desenvolvimento dessa teoria permite o cálculo de diversas medidas relacionadas a ruínas como: déficit na ocorrência da ruína, distribuição dos lucros da seguradora e a própria probabilidade de ruína, entre muitos outros. Sendo assim, esta área passou a atrair grande atenção nos últimos anos, além de ganhar diversas novas contribuições para seu desenvolvimento.

Existem diversas distribuições capazes de definir fielmente as variáveis envolvidas na ocorrência de um evento segurado e estimação das indenizações pagas, pois cada variável possui suas peculiaridades.

A equação que define o lucro de uma seguradora é apresentada a seguir:

$$U(t) = u + ct - S(t), \quad t \geq 0 \tag{1}$$

Onde:

- U(t) é o capital da seguradora no tempo t;
- O capital inicial é definido por u = U(0);
- O prêmio constante pago pelo segurado a cada período unitário de tempo é denominado por c;
- $S_t = \sum_{i=1}^{N(t)} X_i$ (2)

Onde X_i são as indenizações pagas pela seguradas e N(t) é o número de eventos de indenização (processo de contagem);

- S(t)= 0 se N(t)= 0.

Então, o lucro de uma seguradora é dado pela soma de seu montante inicial e dos prêmios pagos pelos segurados, subtraído das indenizações pagas pela seguradora. Portanto, uma seguradora será levada a ruína quando o valor de U(t)< 0 na equação (1), ou seja, quando as indenizações pagas pela seguradora forem superiores aos recebimentos da mesma, num determinado período. Para seguros de automóveis esse período é geralmente 1 ano.

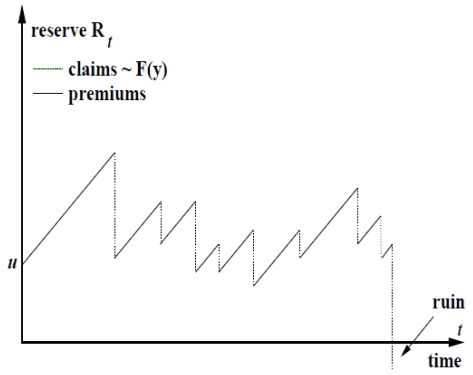

Figura 1- Relação da ruína ao longo do tempo

c. Risco e incerteza

O risco está relacionado aos nossos medos, ou seja, tudo aquilo que de alguma maneira representa algum tipo de perigo ou chance de perda devido a determinado grau de exposição. Dessa forma, define-se risco como sendo a probabilidade de resultados adversos ou uma estimativa para possíveis perdas. Esses resultados não esperados advêm das possibilidades de perdas, decorrentes das incertezas relacionadas aos eventos aleatórios envolvidos no processo.

Tendo em vista que se vive em um mundo cercado por incertezas, o gerenciamento de riscos permite a análise e identificação de alternativas viáveis na busca de um limite de exposição. Como conseqüência, surgem três conceitos básicos para esse sistema de apoio a decisão: risco, retorno e incerteza.

O retorno nada mais é que a remuneração, portanto um prêmio, pelo risco incorrido durante o exercício de uma atividade. O risco é a avaliação numérica das incertezas ligadas ao retorno. Resumindo, o risco é a medida de incerteza em relação ao retorno esperado.

Em seguros estes conceitos são vastamente aplicáveis uma vez que a essência de sua teoria é baseada na troca de riscos entre a seguradora e o cliente. Ou seja, o cliente busca se proteger de determinado evento, como por exemplo, colidir o automóvel, ou até mesmo assegurar qualquer possível problema de saúde. Em contra partida, a seguradora assume o risco da ocorrência do evento segurado. Desta forma, ela fornece ao cliente a proteção desejada, mediante o pagamento de um valor acordado previamente (o prêmio do seguro) resultante de estudos relacionados a modelagem, estocasticidade e probabilidade, estimado a partir da modelagem apropriada que caracteriza a dependência entre o evento adverso e seus fatores de risco.

Quantitativamente, o risco é representado pelo quantil extremo da distribuição de probabilidade das distribuições, o mesmo é descrito pela equação:

$$F_S(S) = \Pr(S_t \leq S^{(\alpha)}) = \alpha \tag{3}$$

Onde, α assume tipicamente os valores de 95% e 99% (como percentis extremos) e o $S^{(\alpha)}$ como sendo o quantil associado.

Essa medida de risco é conhecida como VaR, pois ilustra a perda "máxima" a ser incorrida. A probabilidade atribuída a esta perda máxima corresponde ao nível de significância do VaR, o qual é de 99% quando a probabilidade de perder um valor superior ao VaR é de 1%; de 95%, quando esta probabilidade risco é de 5%.

O VaR é calculado com dados do passado para projetar o futuro, e quando o futuro tem pouca correlação com o passado as previsões podem falhar. Finalmente, convém salientar que o VaR é sensível a bruscas movimentações, podendo oscilar bastante de um dia para o outro. Portanto, não se trata de uma medida que compreenda segurança e por isso, ela é utilizada, atualmente, junto a outras ferramentas na tomada de decisões e na análise de risco mais completa e eficiente.

2. Metodologia

Ao longo desse trabalho diferentes ferramentas serão utilizadas para promover a estruturação do modelo de somas aleatórias que representará a teoria apresentada. Inicialmente, a definição de alguns conceitos será relevante para o desenvolvimento do trabalho, uma vez que será necessário um maior respaldo para o entendimento da teoria em questão.

Além disso, será necessário deduzir algumas equações referentes aos momentos de uma soma aleatória e para isso usar-se-á o software Maple.

Com os resultados obtidos, selecionaremos as distribuições que mais se assemelham ao conjunto de dados baseados no setor de seguros automobilísticos. Vale ressaltar que os dados usados serão referentes aos sinistros ocorridos na Bahia entre os ano de 1998 e 2006. Para a variável aleatória que representa as indenizações, testaremos as distribuições de Pareto, lognormal e gama e para a variável aleatória que representa o número de ocorrências, testar-se-á a distribuição de Poisson e a binomial negativa.

Para implementar as estimações do modelo de somas aleatórias com suas distribuições já descritas será utilizado o *software* @Risk. Em seguida, far-se-á uma análise entre os dados observados na prática e teoricamente.

3. Conceitos Relevantes

Faz-se necessário a apresentação prévia de alguns conceitos básicos relacionados ao tema para o melhor entendimento dos temas apresentados neste trabalho.

a. Teoria da probabilidade e variáveis aleatórias

A teoria da probabilidade é pautada na ocorrência de eventos aleatórios através de estudos matemáticos desenvolvidos por seu principal fundador, Laplace. A probabilidade tem sua base desenvolvida por teoremas e axiomas, onde alguns destes teoremas são descritos abaixo:

- Soma das probabilidades de todos os eventos é igual 1;
- Para eventos aleatórios X e Y, a probabilidade de sua ocorrência simultânea será a soma das probabilidades de todos os eventos presentes tanto em X quanto em Y, no caso da interseção ser vazia, sua probabilidade é zero;
- Para estes mesmos eventos a probabilidade de ocorrência de um ou de outro será a soma dos eventos pertencentes a X e Y.

Logo, pode-se classificar um evento como aleatório se, ao ser realizado diversas vezes, sob as mesmas condições, seu resultado é impossível de ser previsto antecipadamente. O conjunto de todos os possíveis eventos de um experimento aleatório é denominado espaço amostral, Ω. Conseqüentemente, cada evento ou grupo de eventos será considerado um subconjunto de Ω.

É relevante também a definição do conceito de variável aleatória. Esta pode ser entendida como uma função que relaciona um evento qualquer pertencente a determinado espaço amostral, Ω, à somente um valor real. É representada por letras maiúsculas e suas ocorrências por letras minúsculas. Além disso, ao assumir valores reais, estas são ditas contínuas, ao passo que serão discretas caso pertençam a um conjunto finito ou infinito enumerável. Por conseguinte, pode-se classificá-la como uma variável quantitativa, cujo resultado é dependente de razões aleatórias.

Para determinar as probabilidades de eventos, é natural definir a distribuição de probabilidade, que é uma função usada para criar hipóteses sobre os eventos. A função distribuição possui diversas especificações. No caso de uma única dimensão esta é conhecida como função de distribuição cumulativa, podendo ser conhecida previamente ou não:

$$F_x(x) = \Pr[X \leq x], \quad -\infty < x < \infty \tag{4}$$

Esta distribuição pode ser definida como discreta, caso pertença a um conjunto contável e discreto, ou contínua. A obtenção da probabilidade de um evento descrito por uma distribuição de probabilidade é dada pela integração da função de densidade de probabilidade vinculada à mesma.

Para uma variável aleatória X com função distribuição cumulativa Fx (x), é interessante definir algumas características, como a média (ou expectativa), sua variância e seu k-ésimo momento em torno da origem, respectivamente:

$$E[X] = \int x \cdot dF(x) = \mu \tag{5}$$

$$Var[X] = E[(X) - E[X])^2] = \int (X - E[X])^2 \cdot dF(x) = \sigma^2 \tag{6}$$

$$E[X^k] = \int x^k \cdot dF(x) = \mu_k \tag{7}$$

Ainda associada à variável aleatória X, já definida anteriormente, se faz necessário definir a Função Geradora de Momentos (FGM) da mesma:

$$M(t) = E[e^{t \cdot x}] = \int e^{t \cdot x} dF(x) \tag{8}$$

Nota-se que a FGM só pode ser definida no intervalo de convergência da integral acima descrita. A FGM lida principalmente com variáveis aleatórias cujas funções de distribuições cumulativas têm caudas suficientemente bem comportadas.

Além disso, é correto afirmar que se duas variáveis aleatórias têm a mesma FGM, então suas distribuições devem ser idênticas, uma vez que existe uma relação unívoca entre a FGM e a função distribuição. Se derivarmos k vezes a equação de M(t) em relação a t, e finalmente usando t=0, obtêm-se:

$$E[X^k] = M^k(0), k = 1,2,.. \tag{9}$$

b. Processos Estocásticos

Processos estocásticos podem ser definidos como processos aleatórios que dependem do tempo, ou seja, trata-se de um conjunto de variáveis aleatórias, definidas em um espaço estado, indexadas por um parâmetro t pertencente a R. Com um olhar mais abrangente pode-se inferir qualquer evolução no tempo (probabilística ou determinística) analisável em termos de probabilidade como um processo estocástico. Este pode ser classificado em relação ao seu estado e tempo como discreto (cadeia), quando finito e enumerável e, contínuo, caso contrário.

O processo discreto tem uma importante propriedade chamada Markoviana, ou seja, para a previsão de valores futuros, apenas o valor atual do processo é relevante. Logo, no momento presente estará contida toda a informação importante para determinação da distribuição de probabilidade de amanhã. Resumindo, pode-se afirmar que toda a informação passada está expressa no valor atual da série.

Observa-se a aplicabilidade destes conceitos na área de seguros ao tratar-se tanto na probabilidade de ocorrência de eventos adversos como na precificação do risco indexado a cada tipo de situação passível de cobertura pela seguradora. Uma vez que não se consegue determinar o futuro com precisão, é necessária uma estimativa para a realização de uma análise quantitativa dos possíveis eventos para a obtenção de uma melhor interpretação do cenário.

4. Somas Aleatórias

a. Soma de variáveis aleatórias independentes

Define-se S como uma soma aleatória de variáveis aleatórias X_i (ver em equação (2)), onde $X_i \sim f(X_i)$, para todo i, i = 1, 2, ...,n. Usando alguns conceitos apresentados na sessão anterior, considere $S= X_1+ X_2$ na equação (8) da Função Geradora de Momentos:

$$M_s(t) = E[e^{t\cdot(X_1+X_2)}] = E[e^{t\cdot X_1} \cdot e^{t\cdot X_2}] \tag{10}$$

Pela propriedade já conhecida do valor esperado sob independência de X_1 e X_2, segue que:

$$E[e^{t\cdot X_1} \cdot e^{t\cdot X_2}] = E[e^{t\cdot X_1}] \cdot E[e^{t\cdot X_2}] \tag{11}$$

Assim segue que a FGM da soma aleatória S é um produto das FGMs de X_1 e X_2:

$$M_s(t) = M_{X_1}(t) \cdot M_{X_2}(t) \tag{12}$$

Se um conjunto de variáveis aleatórias $X_1,...,X_n$ forem independentes e com função distribuição cumulativa F_i (para i=1,..,n), a distribuição de sua soma será dada por $F_1*...*F_n$ e tem de ser calculada recursivamente.

Mas se o conjunto de variáveis aleatórias X_i for independente e tiver distribuição idêntica F, a distribuição da soma delas será F^{*n}. E se as mesmas variáveis tiverem uma FGM, M(t), comum a todas as variáveis aleatórias, por conseguinte, a FGM de suas somas será $M_x(t)^n$. Como se prova abaixo, para uma soma aleatória onde as variáveis aleatórias são independentes e identicamente distribuídas, $S=X_1+..+X_N$, tem-se que:

$$M_S(t) = E(e^{S\cdot t}) = E(e^{t(X_1+X_2+..+X_N)}) = E(e^{t\cdot X_1} \cdot e^{t\cdot X_2} ... \cdot e^{t\cdot X_N}) = (E(e^{t\cdot X}))^N = M_X^N(t) \tag{13}$$

Finalmente, diante de todas estas definições para variáveis aleatórias independentes e identicamente distribuídas, é natural definir o valor esperado e a variância destas variáveis tendo como base a soma aleatória S, descrita a seguir, onde, inicialmente, assume- se N constante:

$$S = \sum_{i=1}^{N} X_i \tag{14}$$

O valor esperado quando o N é fixo, por definição, é dado pela equação:

$$E(S) = \sum_{i=1}^{N} E(X_i) = N \cdot E(X) \tag{15}$$

A variância de uma soma aleatória é obtida pela expressão geral a seguir:

$$Var(S) = \sum_{i=1}^{N} Var(X_i) + 2 \cdot \sum_i \sum_j \operatorname{cov}(X_i, X_j) \tag{16}$$

Visto que as variáveis aleatórias em questão, X_i's e X_j's são independentes, pode-se particularizar a equação (16), uma vez que a covariância é zero para qualquer i diferente de j. A partir disso, obtem-se o seguinte resultado:

$$Var(S) = N \cdot Var(X) \tag{17}$$

b. Momentos de uma soma aleatória

Para aproximar o modelo da realidade, o número de termos N, que foi citado na sessão anterior como constante presente na soma aleatória S_t, será suposto agora como uma variável aleatória, N(t), com distribuição a ser definida. Assim temos:

$$S_t = \sum_{i=1}^{N(t)} X_i \tag{18}$$

A expressão geral para o momento de ordem m da soma aleatória pela Lei das Expectativas Iteradas pode ser obter como:

$$E(S_t^m) = E[E(S_t^m \mid N_t)] = E[N_t \cdot E(X_i^m)] = E(N_t) \cdot E(X_i^m)$$, com m= 1, 2, 3,.. (19)

Para o cálculo da variância de St, utiliza-se a expressão conhecida:

$$\begin{aligned} Var(S_t) &= E[Var(S_t \mid N_t)] + Var[E(S_t \mid N_t)] \\ &= E(Var \cdot \sum X_i) + Var[N_t \cdot E(X_i)] \\ &= E[N_t \cdot Var(X)] + Var[N_t \cdot E(X)] \end{aligned}$$, e assim, segue que:

$$Var(S_t) = Var(X) \cdot E(N_t) + [E(X)]^2 \cdot Var(N_t) \tag{20}$$

A função geradora de momentos da soma aleatória S é dada por:

$$M_S(t) = E(e^{t \cdot S}) = E[E(e^{t \cdot S} \mid N_t)] = E[M(t)^N]$$, onde M(t) é a função geradora de momentos de X_t (21)

Antes de prosseguir, é válido definir N como sendo uma variável com espaço amostral $\mathbb{N}$. Logo a sua função geradora de momento será dada por:

$$m(t) = \sum_{n=0}^{\infty} e^{t \cdot n} \cdot p_n \tag{22}$$

Então, desenvolvendo a equação (21) para seus momentos em torno da origem, temos:

$$M_s(t) = E[M(t)^N] = \sum_{n=0}^{\infty} p_n \cdot [M(t)]^n \tag{23}$$

Igualando (22) e (23):

$$\sum_{n=0}^{\infty} p_n \cdot [M(t)]^n = \sum_{n=0}^{\infty} e^{t \cdot n} \cdot p_n \therefore e^{t \cdot n} = M(t) \therefore t = \log M(t) \tag{24}$$

Ou seja, pode- se concluir que:

$$M_S(t)=m(\log M(t)). \tag{25}$$

A fim de encontrar os m momentos da soma aleatória S_t através da função $M_s(t)$, basta derivar a expressão anterior, avaliando-as em t=0. Além disso, vale ressaltar algumas igualdades relevantes, como: $E[S_t^m]=M_S^m(0)$, $E[N_t^m]=m^m(0)$ e $E[X_t^m]=M^m(0)$, onde m=1,2,3,.... A seguir apresentam-se os resultados para os momentos de uma soma aleatória em torno da origem:

1º momento:

$E[S_t] = \mu = Ms'(0) = m'(0).\ M'(0)$,

ou seja, $E[S]=E[N].E[X]$. (26)

2º momento:

$E[S_t^2] = Ms''(0) = m''(0).\ M'(0)^2 + m'(0).[M''(0)-M'(0)^2]$,

ou seja, $E[S_t^2]=E[N^2].(E[X])^2+E[N].Var[X]$. (27)

3º momento:

$E[S_t^3]=Ms'''(0)=m'''(0).M'(0)^3+3.m''(0).M''(0).M'(0)-3.m''(0).M'(0)^3+m'(0).[M'''(0)-3M''(0).M'(0)+2.M'(0)^3]$,

ou seja, $E[S_t^3]=E[N^3].(E[X])^3-3.E[N^2].E[X].Var[X]+E[N].(E[X^3]-3E[X^2].E[X]+2.E[X]^3)$. (28)

4º momento:

$E[S_t^4] = Ms''''(0) = m''''(0).\ (M'(0))^4 + 6.m'''(0).M''(0).(M'(0))^2 - 6.m'''(0).\ (M'(0))^4 + 3.m''(0).(M''(0))^2 - 18.\ m''(0).\ (M'(0))^2.M''(0) + 4.m''(0).M'(0).M'''(0) + 11.m''(0).(M'(0))^4 + m'(0).M''''(0) - 4.m'(0).M'''(0).M'(0) + 12.m'(0).M''(0).(M'(0))^2 - 3.m'(0).(M''(0))^2 - 6.m'(0).(M'(0))^4$,

ou seja,

$E[S_t^4]=E[N^4].(E[X])^4+6.E[N^3].E[X^2].E[X]^2-6.E[N^3].E[X]^4+3E[N^2].(E[X^2])^2-18.E[N^2].(E[X])^2.E[X^2]+4.E[N^2].E[X].E[X^3] + 11.\ E[N^2].E[X]^4 + E[N].E[X^4] - 4.E[N].E[X^3].E[X]+ +12.E[N].E[X]^2.E[X^2] - 3.E[N].(E[X^2])^2 - 6E[N].(E[X])^4$. (29)

A fim de deduzir os momentos da soma aleatória S_t em torno da média, torna-se necessário definir alguns momentos em torno da média, e a relação entre momentos em torno da média com momentos em torno da origem:

- Momento em torno da média

$$\mu_m = E[S - E(S)]^m, m = 1,2,3,... \quad (30)$$

- Momento em torno da origem

$$\mu'_m = E[S^m], m = 1,2,3,... \quad (31)$$

$$\mu = E[S] \quad (32)$$

- Relações entre os momentos em torno da média e os momentos em torno da origem:

$$\mu_2 = E[S - E(S)]^2 = \mu'_2 - \mu^2 \tag{33}$$

$$\mu_3 = E[S - E(S)]^3 = \mu'_3 - 3 \cdot \mu \cdot \mu'_3 + 2 \cdot \mu^3 \tag{34}$$

$$\mu_4 = E[S - E(S)]^4 = \mu'_4 - 4 \cdot \mu \cdot \mu'_3 + 6 \cdot \mu \cdot \mu'_2 - 3 \cdot \mu^4 \tag{35}$$

A partir destes resultados é possível então obter os momentos da soma aleatória em torno da média utilizando os momentos centrais de S. Logo, substituindo (30), (31) e (32) em (33), (34) e (35), obtêm-se:

$$E[S - \mu]^2 = E[S^2] - E[S]^2 \tag{36}$$

$$E[S - \mu]^3 = E[S^3] - 3 \cdot E[S] \cdot E[S^2] + 2 \cdot E[S]^3 \tag{37}$$

$$E[S - \mu]^4 = E[S^4] - 4 \cdot E[S] \cdot E[S^3] + 6 \cdot E[S^2] \cdot E[S]^2 - 3 \cdot E[S]^4 \tag{38}$$

Desenvolvendo as equações acima propostas (36), (37), (38) a partir das equações já deduzidas para os momentos de uma soma aleatória em torno da origem (26) a (29), chega-se aos seguintes resultados em função das características dos momentos das variáveis aleatórias X e N:

- Variância de S:

$$E[S - \mu]^2 = \mathrm{E[N^2].(E[X])^2 + E[N].Var[X] - E[N]^2.E[X]^2;} \tag{39}$$

- Terceiro momento de S em torno da média:

$$\begin{aligned} E[S - \mu]^3 &= \mathrm{E[N^3].(E[X])^3 - 3.E[N^2].E[X].Var[X] + E[N].(E[X^3] - 3E[X^2].E[X] + 2.E[X]^3)} \\ &\mathrm{- 3 \cdot (E[N] \cdot E[N^2].(E[X])^3} + E[X] \cdot \mathrm{E[N]^2.Var[X]) + 2 \cdot E[N]^3.E[X]^3;} \end{aligned} \tag{40}$$

- Quarto momento de S em torno da média:

$$\begin{aligned} E[S - \mu]^4 &= \mathrm{E[N^4].(E[X])^4 + 6.E[N^3].E[X^2].E[X]^2 - 6.E[N^3].E[X]^4} \\ &\mathrm{+ 3 \cdot E[N^2].(E[X^2])^2 - 18.E[N^2].(E[X])^2.E[X^2] + 4.E[N^2].E[X].E[X^3] + 11.E[N^2].E[X]^4} \\ &\mathrm{+ E[N].E[X^4] - 4.E[N].E[X^3].E[X] + 12.E[N].E[X]^2.E[X^2] - 3.E[N].(E[X^2])^2 - 6E[N].(E[X])^4} \\ &- 4 \cdot (E[N] \cdot E[X]) \cdot \mathrm{E[N^3].(E[X])^3 - 3.E[N^2].E[X].Var[X] + E[N].(E[X^3] - 3E[X^2].E[X] + 2.E[X]^3)} \\ &\mathrm{+ 6 \cdot (E[N^2].(E[X])^2 + E[N].Var[X])} \cdot E[N]^2 \cdot E[X]^2 - 3 \cdot E[N]^4 \cdot E[X]^4; \end{aligned} \tag{41}$$

A partir do terceiro e quarto meomento em torno da média, pode-se obter os coeficientes de assimetria e de curtose associados à variável aleatória S.

O coeficiente de assimetria é, por definição, o quanto sua densidade se desvia ou se afasta da posição simétrica. Uma densidade é simétrica em torno da média μ se e somente se, f(x-μ) = f(μ-x), para qualquer que seja x. Pode-se provar que se $f_x(x)$ é simétrica em torno de μ, então $E[x-\mu]^3=0$. É importante salientar que se ocorre simetria, então $E[x-\mu]^3=0$, mas a recíproca não é sempre verdadeira, ou seja, existe distribuição assimétrica onde $E[x-\mu]^3=0$. Ao possuir esta característica, o coeficiente será dito como assimétrico nulo. Os desvios positivos e negativos têm a mesma preponderância e as caudas da distribuição possuem o mesmo formato. O coeficiente pode ser classificado como assimétrico positivo, assimétrico negativo e assimétrico nulo. A sua expressão para a variável S é dada por:

$$A_s = \frac{E[S-\mu]^3}{E[[S-\mu]^2]^{\frac{3}{2}}} \tag{42}$$

Já numa distribuição assimétrica positiva, os desvios positivos são preponderantes em relação aos negativos. Nesse caso, a cauda à direita é mais alongada que a cauda à esquerda e $E[x-\mu]^3>0$.

Finalmente, na distribuição assimétrica negativa os desvios negativos são preponderantes em relação aos positivos. Nesse caso, a cauda à esquerda é mais alongada que a cauda à direita e $E[x-\mu]^3<0$.

A figura a seguir ilustra as distribuições assimétrica positiva (A), simétrica (B), assimétrica negativa (C) e μ_B é a média da distribuição simétrica (B):

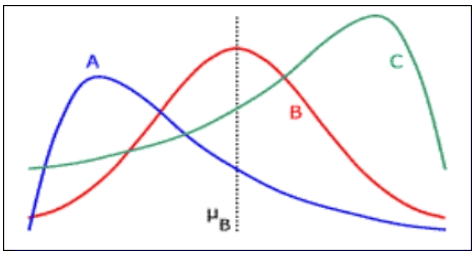

Figura 2- Assimetria positiva (A), nula (B) e negativa (C).

Pode-se obter este coeficiente de forma direta, uma vez que suas parcelas já foram desenvolvidas nas equações (39) e (40) com referência as equações (26) a (29). Com isso, tem-se:

$$A_s = \frac{[E[N^3].(E[X])^3 - 3.E[N^2].E[X].Var[X] + E[N].(E[X^3] - 3E[X^2].E[X] + 2.E[X]^3) - \\ - 3\cdot(E[N]\cdot E[N^2].(E[X])^3 + E[X]\cdot E[N]^2.Var[X]) + 2\cdot E[N]^3.E[X]^3]}{(E[N^2].(E[X])^2 + E[N].Var[X] - E[N]^2.E[X]^2)^{\frac{3}{2}}} \tag{43}$$

Por sua vez, o coeficiente de curtose captura o grau de achatamento de uma densidade em relação a distribuição normal. Elas podem ser classificadas como mesocúrtica, leptocúrtica e platicúrtica. Sua fórmula é apresentada a seguir:

$$K_s = \frac{E[S-\mu]^4}{E[[S-\mu]^2]^2} \tag{44}$$

Uma distribuição é classificada como platicúrtica se possuir pico mais arredondado com caudas mais curtas e mais magras, além de ter menor probabilidade que a normal de ter valores próximos à média e menor probabilidade que a normal de ter valores extremos.

Uma distribuição pode ser dita como leptocúrtica se tiver pico mais agudo com caudas mais longas e mais pesadas, além de possuir maior probabilidade que a normal de ter valores próximos à média e maior probabilidade que a normal de ter valores extremos.

Finalmente, se a distribuição é normal, então ela é definida como mesocúrtica. Ela será simétrica tendo seu coeficiente de curtose K_S=4.

A figura a seguir ilustra as distribuições leptocúrtica (a), mesocúrtica (b) e platicúrtica (c):

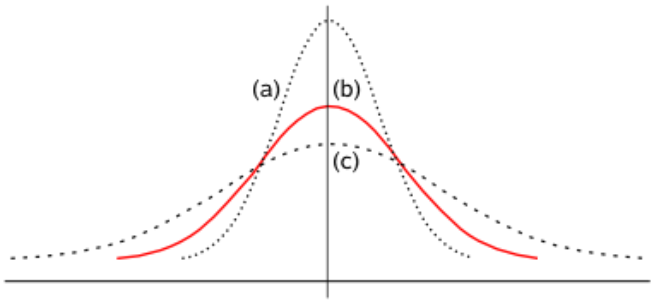

Figura 3-Classificações da curtose: leptocúrtica (a), mesocúrtica (b) e platicúrtica (c).

Como se pode perceber, a interpretação da curtose não é muito simples. O maior ou menor pico, como usualmente é interpretado, não é o aspecto mais importante e sim a concentração de valores nas caudas.

É possível reescrever a fórmula (44), a partir das equações (39) e (41) dos segundo e quarto momentos em torno da média da soma aleatória. Resultando em:

$$
\begin{aligned}
E[S-\mu]^4 &= \mathrm{E[N^4].(E[X])^4} + 6.\mathrm{E[N^3].E[X^2].E[X]^2} - 6.\mathrm{E[N^3].E[X]^4} \\
&+ 3\cdot \mathrm{E[N^2].(E[X^2])^2} - 18.\mathrm{E[N^2].(E[X])^2.E[X^2]} + 4.\mathrm{E[N^2].E[X].E[X^3]} + 11.\mathrm{E[N^2].E[X]^4} \\
&+ \mathrm{E[N].E[X^4]} - 4.\mathrm{E[N].E[X^3].E[X]} + 12.\mathrm{E[N].E[X^2].E[X^2]} - 3.\mathrm{E[N].(E[X^2])^2} - 6\mathrm{E[N].(E[X])^4} \\
&- 4\cdot(E[N]\cdot E[X])\cdot \mathrm{E[N^3].(E[X])^3} - 3.\mathrm{E[N^2].E[X].Var[X]} + \mathrm{E[N].(E[X^3]} - 3\mathrm{E[X^2].E[X]} + 2.\mathrm{E[X]^3}) \\
&+ 6\cdot(\mathrm{E[N^2].(E[X])^2} + \mathrm{E[N].Var[X]})\cdot E[N]^2\cdot E[X]^2 - 3\cdot E[N]^4\cdot E[X]^4;
\end{aligned}
$$

$$
E[S-\mu]^2 = \mathrm{E[N^2].(E[X])^2} + \mathrm{E[N].Var[X]} - \mathrm{E[N]^2.E[X]^2} = Var(S);
$$

Sendo assim, pode-se obter este coeficiente de forma direta, como apresentado a seguir:

$$
K_s = \frac{E[S-\mu]^4}{[Var(S)]^2} \quad . \qquad (45)
$$

5. Modelos Clássicos de Ruína

a. O processo de risco

A alta probabilidade de ruína indica instabilidade de uma seguradora. A fim de evitar ocorrências adversas, o resseguro, o aumento dos prêmios e a captação de capital de giro são medidas consideráveis. No entanto, essa instabilidade não representa a probabilidade de quebra da seguradora num futuro próximo, na verdade, a ruína pode demorar a acontecer.

O processo de contagem descrito pela equação (1) é um processo estocástico contínuo no tempo N_t, $t \geq 0$, com $N_0=0$, cujas amostras formam uma função degrau, que é descontínua. A interpretação é que N_t denota o número de vezes que um determinado evento ocorre entre 0 e t, portanto as descontinuidades representam a ocorrência desses eventos e os pulos da função tem tamanho 1, o que significa que não há ocorrências simultâneas.

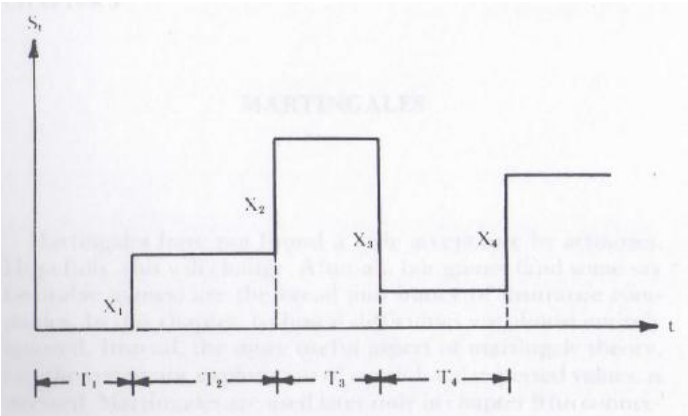

Figura 4- Indenizações pagas pela seguradora no tempo

Essas ocorrências podem ser representadas pelo número de indenizações, N(t), pagas por uma seguradora. Para modelagem de processos deste tipo, é comum o uso de distribuições do tipo binomial negativa ou Poisson. No entanto, tendo em vista que o processo de Poisson é uma escolha mais simples e popularmente usada para modelar processos de contagens, com parâmetro $\lambda>0$, o N(t) pode ser assumido como sendo um processo de Poisson. Então S_t é um processo composto de Poisson, uma vez que S_t é uma soma aleatória de parâmetro λ (valor esperado de ocorrências por unidade de tempo), onde sua média e variância são $E[S(t)]=\lambda.E(X).t$ e $Var[S(t)]=\lambda.E(X^2).t$, onde E(X) e $E(X^2)$ são os dois primeiros momentos centrais da variável aleatória X e o valor esperado e variância de N(t) equivalem a $\lambda.t$, casos particulares das equações (26) e (27).

A variável aleatória não negativa X_i representa o valor da indenização i, onde i varia de 1 até N(t). Logo, o somatório das indenizações pagas até o período de tempo t:

$$S(t)=X_1+X_2+...+X_{N(t)} \tag{46}$$

Para modelar as indenizações, assume- se que cada X_i deve ser uma variável aleatória definida em $\mathbb{R}^+$ positivo, independente e identicamente distribuído, sendo independentes do número de acionamentos N(t). Há algumas distribuições que podem ser usadas para caraterizar X são elas: lognormal, Pareto e gama. A modelagem por lognormal é mais usada quando a distribuição dos dados possui caudas mais pesadas, como no caso do seguro contra incêndios. Já a de Pareto se adequa melhor aos dados cuja probabilidade de ocorrência de valores de indenizações altas é considerável. O uso da distribuição gama é recomendado nos casos onde a cauda da FDC não é muito "pesada" caracterizando assim uma assimetria positiva. Vale ressaltar que esta última é comumente utilizada para a modelagem referente a seguros de veículos.

b. Modelo de risco coletivo e modelo de risco individual

No *modelo individual de risco*, o total de indenizações de um portfólio de contratos de seguros é a variável aleatória de relevância. Na verdade, é interessante estimar a probabilidade de certo capital ser suficiente para atender às ocorrências dos eventos segurados, ou seja, o Value-at-risk associado ao portfólio citado, sendo o quantil 95% de sua distribuição de probabilidade acumulada. O total de ocorrências é modelado como a soma de todas as ocorrências de todas as apólices individuais, que são independentes, vide equação (18). O Teorema do Limite Central, o qual envolve o ajuste de dois momentos, não é suficientemente preciso para a cauda direita de uma distribuição, para resolver este impasse existem dois métodos que são mais plausíveis para tal ajuste, são eles: Método da Aproximação pela gama Deslocada e o Método da Aproximação Normal.

O *modelo do risco coletivo* em Seguros modela um conjunto de ocorrências durante um determinado tempo de duração T. Ou seja, o risco coletivo não trata das apólices individualmente, mas sim de um grupo de apólices anônimas. É válido lembrar que as ocorrências dos eventos não desejados são variáveis aleatórias independentes e identicamente distribuídas, como previamente apresentado na equação (18), e aqui mostrado novamente a seguir:

$$S = \sum_{i=1}^{N(t)} X_i$$

Onde: X_i está associada a i-ésima apólice; N(t) é o somatório de das apólices individuais.

É comum assumir que a soma aleatória das ocorrências dos eventos segurados pode ser descrita pela distribuição de Poisson, com parâmetro λ e função de distribuição cumulativa de um grupo de ocorrências individuais.

Em termos de aplicabilidade, pode-se concluir que o *modelo de risco individual* não é o mais apropriado, uma vez que na maioria das vezes os dados fornecidos pela seguradora não são das apólices individuais, mas sim de um grupo de apólices. Visto isso, o *modelo de risco coletivo* é o mais indicado e por isso será utilizado neste trabalho.

6. Aplicações

a. Estimação para seguros de veículos

A estimação é a maneira mais concreta de aplicar toda a teoria vista até este momento, então, nessa seção, realiza-se uma aplicação do modelo de somas aleatórias para as indenizações de uma indústria de seguros de veículos.

Sendo assim, coletaram-se dados referentes ao setor de seguros de veículos nacional. Tendo em vista a extensão da base de dados, fornecida pela SUSEP, a qual engloba muitas categorias e grupos, realizou-se uma seleção dos dados. A subdivisão dos mesmos ocorreu para os anos de 1998 a 2006. As apólices escolhidas foram referentes às colisões de carros Fiat Palio 1.0 para pessoas do sexo masculino de idade entre 36 e 45 anos, residentes na Bahia. Seguem na próxima tabela as características explicitadas:

Cat.Tarifária	Região	Grupos	Ano	Sexo	Idade	Colisão	
						N° de Sinistros	Total de Sinistros
Todas	BA - Bahia	FIAT PALIO 1.0	1998	Masculino	Entre 36 e 45 anos	2	5.797,00
Todas	BA - Bahia	FIAT PALIO 1.0	1999	Masculino	Entre 36 e 45 anos	6	11.241,00
Todas	BA - Bahia	FIAT PALIO 1.0	2000	Masculino	Entre 36 e 45 anos	8	60.353,00
Todas	BA - Bahia	FIAT PALIO 1.0	2001	Masculino	Entre 36 e 45 anos	5	13.633,00
Todas	BA - Bahia	FIAT PALIO 1.0	2002	Masculino	Entre 36 e 45 anos	2	2.245,00
Todas	BA - Bahia	FIAT PALIO 1.0	2003	Masculino	Entre 36 e 45 anos	8	47.684,00
Todas	BA - Bahia	FIAT PALIO 1.0	2004	Masculino	Entre 36 e 45 anos	10	40.364,00
Todas	BA - Bahia	FIAT PALIO 1.0	2005	Masculino	Entre 36 e 45 anos	8	20.719,00
Todas	BA - Bahia	FIAT PALIO 1.0	2006	Masculino	Entre 36 e 45 anos	11	84.788,00

Tabela 5 - Base de dados. Fonte: SUSEP

Tendo posse desses dados e contrapondo-os com as distribuições sugeridas na seção 5.a, usou-se o *software* @Risk para estimar as distribuições das variáveis aleatórias N e X, descritas anteriormente como número de sinistros e indenizações, respectivamente.

Para uma escolha mais apurada, o *software* @Risk elege a melhor distribuição dentre as candidatas apontadas usando o critério da menor estatística de aderência χ^2(chi quadrado). Ela se baseia no uso de hipóteses nulas que, particularizando para contexto deste trabalho, utiliza-se a hipótese nula assumindo determinada variável aleatória, y, como pertencente a determinada distribuição já postulada, $f_y(y)$. Ou seja:

$$\begin{aligned} H_0 &= y \in f_y(y) \\ H_a &= c.c. \end{aligned} \tag{47}$$

Como trabalhar-se-á com o valor esperado de y_i, o mesmo é defido pelo número de casos sob H_0 multiplicado pela probabilidade de ocorrência de cada caso, como pode ser observado na equação a seguir:

$$E[y_i] = n \cdot p(y_i) \tag{48}$$

Vale ainda ressaltar que no caso da variável aleatória y_i ser discreta $p(y_i)$ é obtido de forma direta. No caso contínuo (lognormal, gama, pareto, etc), i irá representar os intervalos, y_i será o número de observações nestes intervalos e $E[y_i]$ será o valor esperado y_i sob H_0 (hipótese nula). Dessa forma, as respectivas probabilidades poderão ser obtidas da seguinte forma:

$$p(y_i) = \int_{y\inf erior}^{y\sup erior} f(y_i)dy_i \tag{49}$$

O coeficiente χ^2 verifica se os valores observados nos dados se desviam do valor esperado sob uma certa distribuição postulada. O Chi- quadrado não é um teste paramétrico, ou seja, não depende de parâmetros populacionais, como a média e a variância. Ele pode ser calculado pela equação:

$$\chi^2 = \frac{\sum_i (y_i - E^{H_0}(y_i))^2}{E^{H_0}(y_i)} \tag{50}$$

Percebe-se que quando os dados observados são próximos do valor esperado sob hipótese nula, o valor de χ^2 é bem pequeno, o que mostra um bom encaixe entre os dados e a distribuição testada. O oposto é observado quando χ^2 assume valores altos. Um ponto fraco desta estatística é a definição dos intervalos a serem utilizados e isto acarreta em distintos resultados dependendo da especificação dos mesmos.

- Modelo para a variável N

Para a variável aleatória N foram testadas duas distribuições, Poisson e binomial negativa. Seguem os resultados obtidos:

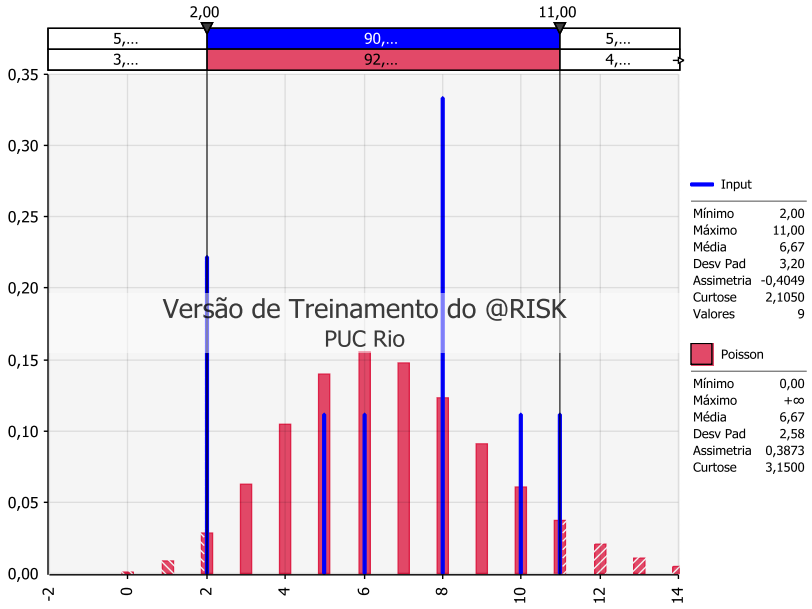

Figura 5- Distribuição Poisson para a variável N

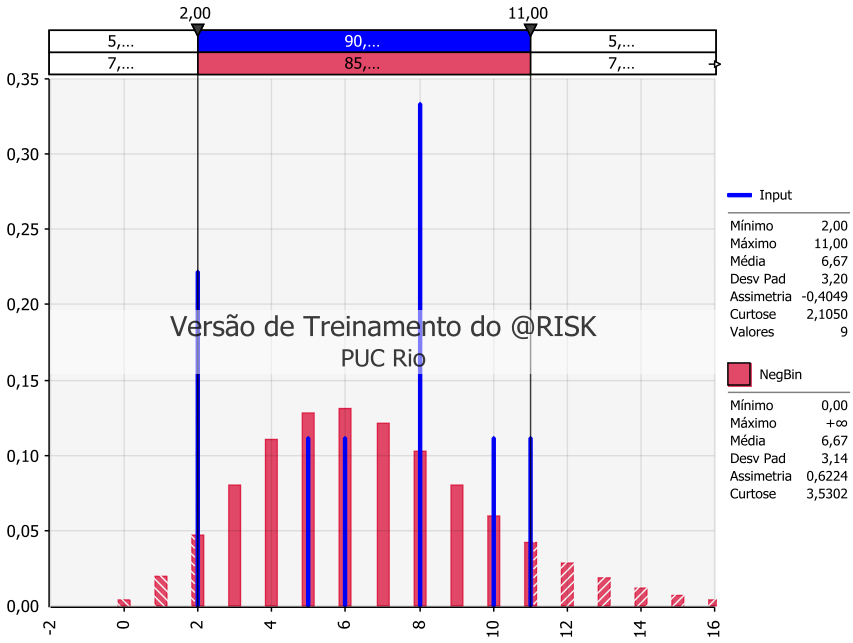

Figura 6- Distribuição binomial negativa para a variável N

Diante destes gráficos, pode-se observar que a melhor escolha para modelar a variável aleatória N, que representa o número de sinistros, é a distribuição Poisson. Seu parâmetro λ é 6,67, que corresponde a média das observações desta variável.

- Modelo para a variável X

Complementando esta fase de estimações já iniciada, é relevante testar a variável aleatória X, que contempla as indenizações. Como visto também na seção 5.a, as distribuições sugeridas são: gama, Pareto e lognormal. No entanto, como se está utilizando o modelo de risco coletivo e a base de dados em uso neste trabalho não fornece dados explícitos sobre o valor individual de cada indenização, não será possível uma estimação direta a partir da tabela 5.

Diante deste contexto, é razoável inferir alguns valores para as medidas estatísticas como média (μ) e variância (σ^2). Além disso, para tornar o modelo mais realista é importante a introdução de um fator de desvio que representa as franquias, este será denotado por x_0. A fim de se obter esses valores para estimação utilizou-se a tabela 5, dividindo-se o valor total de sinistros pelo seu correspondente número de ocorrências, gerando uma nova coluna de dados. A partir da mesma calcularam-se os seguintes valores para média e variância e um valor foi arbitrado para x_0, coerente com os demais valores:

$$\mu^*=4051{,}111;$$

$$\sigma^{2\,*}=5969596;$$

$$x_0^*= 1000.$$

O primeiro gráfico a ser apresentado é o que ilustra os resultados obtidos para a distribuição lognormal~(μ, σ). O mesmo segue, assim como suas características:

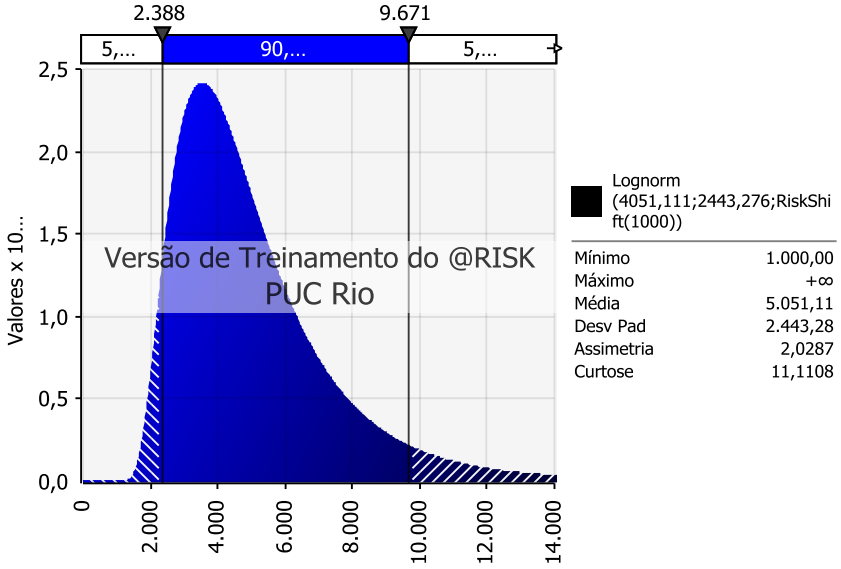

Figura 7- Distribuição lognormal para a variável X

Segue agora o resultado para a distribuição de Pareto:

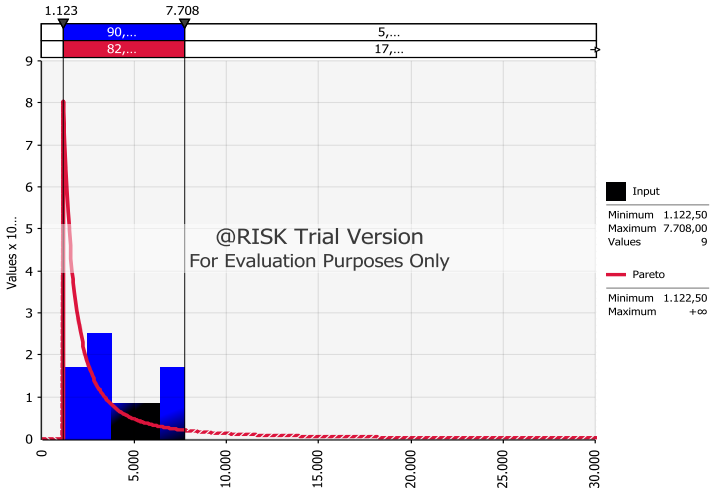

Figura 8- Distribuição Pareto para a variável X

A terceira distribuição sugerida pela teoria seria a distrubuição gama. Para fins de estimação da mesma, foi necessário o cálculo de mais alguns parâmetros (α, β, γ) a partir dos dados já fornecidos referentes à sua média, variância e desvio.

Como pode ser visto em Kaas, R.; Goovaerts, M.; Dhaene, J.; Denuit, M. Modern Actuarial Risk Theory (2001), páginas 35 e 36, o cálculo dos parâmetros α*, β* e x0* da distribuição de gama pelo método dos momentos, também descrito no apêndice, podem ser estimados da seguinte maneira:

$$\alpha^* = \frac{4}{\gamma^{2}*} \tag{51}$$

$$\beta^* = \frac{\gamma^* \cdot \sigma^*}{2} \tag{52}$$

$$x_o{}^* = \mu^* - \frac{2 \cdot \sigma^*}{\gamma^*} \tag{53}$$

Onde: μ* é a média da variável em questão; σ* é o desvio padrão da variável; γ* é a assimetria e x_0^* é o deslocamento da distribuição em relação à origem, representado aqui a franquia como já posto.

Atribuindo os valores já encontrados às variáveis μ* e σ^2* e o valor arbitrado a x_0*, pode-se resolver o sistema de equações proposto em (51), (52) e (53), donde, chega-se à estimação dos parâmetros mostrada a seguir:

α^*=1,559449;

β^*=1956,532;

γ^*=1,601565.

Visto que para a estimação da distribuição gama é necessário apenas os parâmetros α^*, β^* e x_0^*, obtem-se através do @Risk o seguinte resultado:

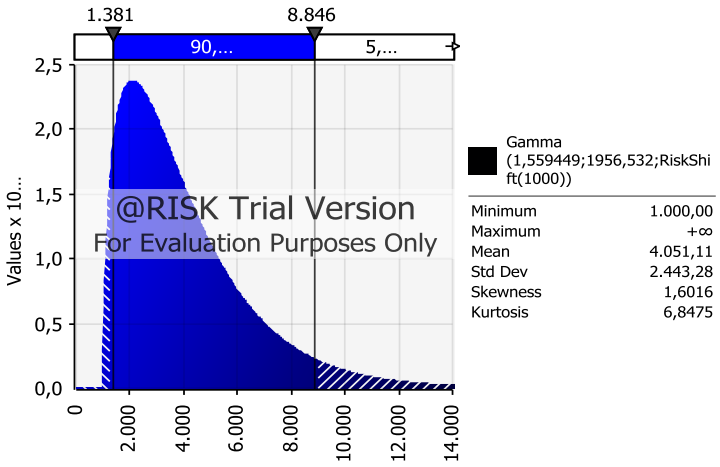

Figura 9-Distribuiçao gama para a variável X

A análise das três distribuições comprova o que foi afirmado pela teoria. A distribuição gama é, de fato, a que melhor descreve os dados, modelando assim de maneira mais fiel as indenizações em um modelo de ruínas. Pode-se observar que os dados utilizados para a construção da variável X estão bem ajustados e coerentes.

b. Construção da distribuição da soma aleatória

A escolha da distribuição que definirá a soma aleatória S será pautada no próprio princípio da ruína. A fim de evitar a quebra de uma seguradora, o número de ocorrências de eventos não desejados com uma indenização menor tem que ser mais freqüentes que aqueles cujas indenizações são muito altas. Essa definição sugere uma assimetria positiva, onde a cauda desta distribuição é alongada e fina. Portanto, fica claro que a distribuição que melhor se adéqua é a gama, tendo em vista também toda a teoria apresentada até este ponto.

A estimação da distribuição da soma aleatória pode ocorrer de duas maneiras distintas: a primeira é a partir de dados históricos, como os observados na tabela 5; a outra maneira possível é através da combinação dos momentos centrais das variáveis aleatórias X e N, que são variáveis aleatórias positivas e descrevem as indenizações e o número de sinistros, respectivamente. Ambos os métodos serão apresentados seguir.

De posse da tabela 5, onde a última coluna representa exatamente os valores de S, poderia-se obter uma estimação para a distribuição da soma aleatória. Porém, nota-se a ausência de resultado para a distribuição gama, devido à falha na convergência dos dados. Isto ocorre porque o software @Risk utiliza o método da verossimilhança, citado no apêndice, como base para suas estimações.

Esse impasse foi motivador para o cálculo dessa distribuição pelo método dos momentos, descrito também no apêndice. Como alternativa, a soma aleatória S pode ser também descrita pela combinação dos momentos centrais das variáveis aleatórias X e N, que são positivas e descrevem as indenizações e o número de sinistros, respectivamente. Visto o desenvolvimento e escolha das distribuições adequadas para estas variáveis (X, como gama, e N, Poisson) pode-se construir a distribuição da soma aleatória S.

A distribuição gama é dada pelos parâmetros α^*, β^* e x_0^*. Estes parâmetros são calculados a partir da média, variância e assimetria da variável em questão, como já visto nas equações (51), (52) e (53).

Para obter estes valores, faz- se necessária algumas definições que possibilitarão seu desenvolvimento para o caso particular da soma aleatória:

$$\mu_S * = E[S] \tag{54}$$

$$\sigma_S^2 * = E[S - \mu_S]^2 \tag{55}$$

$$\gamma * = \frac{E[S - \mu_S]^3}{(\sigma_S^2)^{\frac{3}{2}}} \tag{56}$$

Após a apresentação dessas equações, é relevante repetir as equações (26), (27) e (40) que representam suas parcelas:

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

$$E[S_t^2] = E[N^2] \cdot (E[X])^2 + E[N] \cdot Var[X]$$

$$E[S - \mu]^3 = \mathrm{E[N^3].(E[X])^3 - 3.E[N^2].E[X].Var[X] + E[N].(E[X^3] - 3E[X^2].E[X] + 2.E[X]^3)}$$
$$\mathrm{- 3 \cdot (E[N] \cdot E[N^2].(E[X])^3} + E[X] \cdot \mathrm{E[N]^2.Var[X]) + 2 \cdot E[N]^3.E[X]^3}$$

Visto que na seção 6.a foram realizadas estimações com as variáveis X e N, que resultaram em valores numéricos, pode-se aplicar os mesmos nas equações (54), (55), (56), obtendo assim os seguintes resultados da média, variância e assimetria de S:

μ_S^*=31869,33;

$\sigma^2_S{}^*$=796719651;

γ_S^*=0,82842638.

Já em posse destes valores, pode-se estimar diretamente α^*, β^* e x_0^*, parâmetros que definem a distribuição gama, a partir das equações (51), (52) e (53), chegando-se aos seguintes valores:

α^*=5,828438;

β^*=11691,67;

x_0^*= 36274,9.

A partir destes valores e fazendo uso do *software* @Risk estimou-se a distribuição gama para a soma aleatória S. O gráfico apresentado a seguir representa este resultado assim como fornece suas principais características estatísticas como média, variância, assimetria e curtose:

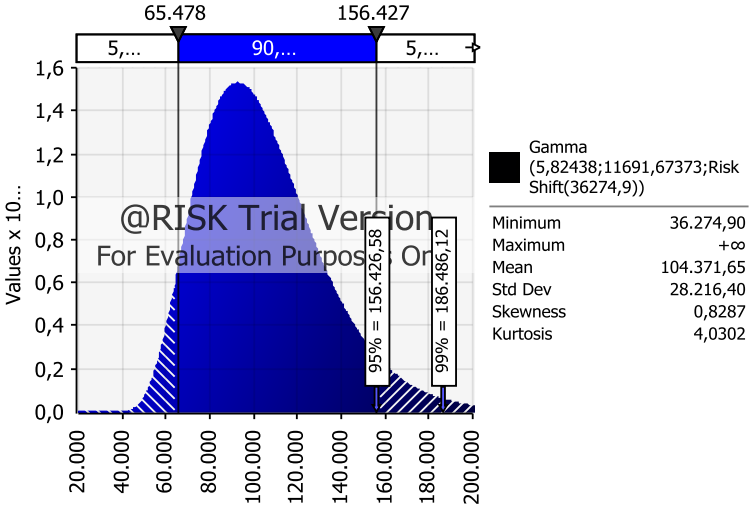

Figura 10-Distribuição gama para a variável S

Apesar da escassez de dados para a análise da teoria desenvolvida no decorer deste trabalho, conseguiu-se desenvolver estimações para as variáveis envolvidas no processo de ruína de uma seguradora. De fato, o gráfico obtido para a estimativa da soma aleatória satisfez as expectativas, uma vez que apresentou parâmetros coerentes com os dados trabalhados. No entanto, melhores resultados não foram possíveis de serem obtidos uma vez que a base de dados era limitada, pois a base de dados só engloba nove anos. Isto resultou numa estimação não fiel e pouco ilustrativa para a distribuição proposta, gama, à variável em questão, soma aleatória.

Pode-se notar pelo gráfico as marcações dos percentis referentes aos VaR de 95% e 99%. Como a distribuição gama apresenta uma nítida assimetria positiva, esses percentis apontam as ocorrências de eventos extremos, ou seja, de indenizações de maior valor, uma vez que o VaR ilustra a perda máxima incorrida. Dessa forma, pelo fato desses quantis representarem grandes perdas financeiras, sua ocorrência poderia levar ao aumento da probabilidade de ruína da uma seguradora.

Ressalta-se que diferentes tipos de técnicas de aproximação poderiam ter sido utilizados para estimar a distribuição da soma aletatória como, por exemplo, o método de Panjer. Este faz uso da recursão para o cálculo das probabilidades f(x). No entanto, por demandar intenso esforço computacional e matemático para a obtenção de seus resultados, foi feita neste trabalho, a opção pelo uso da distribuição gama na estimativa da distribuição que definiu a soma aleatória.

7. Conclusão

Com a constante necessidade do homem de se proteger contra a ocorrência de eventos não desejados e acrescido do fato do aumento de renda, consumo e expectativa de vida, a indústria de seguros vem apresentando um crescimento notório.

As pessoas passaram a valorizar a necessidade de planejar e proteger seu futuro, ainda mais em países emergentes, onde o mercado de seguros possui uma tímida penetração, uma vez que a cultura do seguro não é difundida nestes lugares.

Nesse trabalho desenvolveu-se uma análise de variáveis da teoria da ruína, a qual aponta o momento em que os lucros da seguradora passam a ser negativos. Diante de uma gama de dados disponíveis nesta área de seguros, escolheram-se os dados referentes às apólices de automóveis no estado da Bahia. Após análises críticas das características de cada distribuição em particular, pôde-se ressaltar aquelas que melhor se enquadraram apontando também o porquê das outras distribuições testadas não obterem sucesso.

Vale ressaltar que o modelo desenvolvido foi simplista, ou seja, presupostos heróicos não foram utilizados para realização do mesmo. Não cabe ao conteúdo desse trabalho a análise que define o ponto de ruína de uma seguradora, mas sim a análise das distribuições para as variáveis aleatórias envolvidas no processo através da simulação dos dados usados. Posteriormente, realizou-se a contraposição dos resultados com a teoria estudada e detalhada. Essas simulações colaboram para o desenvolvimento mais sólido do setor, uma vez que trazem um melhor entendimento dos eventos segurados passíveis de ocorrência.

A construção de um modelo que reflita mais fielmente a realidade deve considerar alguns graus de complexidade. Primeiramente, deve-se descobrir a dependência dentre as informações fornecidas ao modelo e como descrevê-las de forma coerente. Além disso, é relevante saber lidar com a dependência na gestão do risco inclusive pelo fato dos princípios dos seguros não serem freqüentemente válidos, uma vez que estes são suscetíveis a mudanças. A modelagem, portanto, desses modelos jamais refletiria a verdade absoluta observada na realidade.

Visto isso, existem diferentes formas para se obter uma aproximação para a modelagem dessa dependência e seus parâmetros. Pode- se alcançar então resultados mais precisos levando em consideração um termo referente ao erro; resultados assintóticos, representando uma tendência a determinado resultado; e o mais comumente utilizado, por se obter um melhor tratamento da informação, são os resultados obtidos pela modelagem estocástica. Supondo a inserção desses graus de complexidade para a construção de um modelo mais realístico algumas premissas devem ser adotadas, uma vez que os cenários são mais intricados, pois lidam com taxas de prêmios não constantes, premissas com independência super relaxadas para investimentos, dividendos e impostos incorporadas ao modelo.

A partir disso, seria uma solução ótima a seguradora investir parte de seus recebimentos em ações e outra parte no tesouro ou renda fixa. Esta estratégia a protege contra sua ruína, dado o rendimento financeiro que ela obterá e através da margem obtida, a seguradora conseguirá honrar a cobertura dos sinistros, sem que isto cause um impacto em seu capital de giro.

Apesar de diversos estudos serem conduzidos nesta área, a previsão da ocorrência dos eventos segurados bem como quantas vezes eles irão ocorrer é ineficaz, uma vez que são aleatórios. No entanto, a necessidade de estimar quantitativamente estes eventos levou à estruturação de alguns modelos que simulam estas ocorrências. Isso explica o notório desenvolvimento da matemática voltada para o setor de seguros.

8. Bibliografia

ALBRECHER, H. Dividends, Ruin and Bankruptcy. 5th International Conference on Statistical Modelling in Insurance & Finance, Maresias, April 10-15, 2011. (http://www.ime.usp.br/bcsmif/5th/index.php?inc=conf_programme; http://www.hec.unil.ch/people/halbrecher?vue=publications)

ALBRECHER, H. Solvency Modelling with Dependent Risks. 5th International Conference on Statistical Modelling in Insurance & Finance, Maresias, April 10-15, 2011. (http://www.ime.usp.br/bcsmif/5th/index.php?inc=conf_programme; http://www.hec.unil.ch/people/halbrecher?vue=publications)

GERBER, H. U. An Introduction to Mathematical Risk Theory. 1979, University of Pennsylvania, Philadelphia.

KAAS, R.; GOOVAERTS, M.; DHAENE, J.; DENUIT, M. Modern Actuarial Risk Theory. 2001, Kluwer Academic Publishers, Netherlands.

SPANOS, A. Probability Theory and Statistical Inference: Econometric Modelling with Observational Data. 1952, Cambridge University Press.

9. Apêndice – Métodos de estimação

a. Métodos dos momentos

Este método foi introduzido por Karl Pearson em 1902. O método dos momentos é pouco usado atualmente por ter sido substituído pelo método da verossimilhança, que será visto mais adiante, pois os estimadores deste possuem maior precisão e qualidade. Entretanto, a verossimilhança pode exigir recursos computacionais e técnicas de solução muito complexas de serem estimadas, tornando-se mais indicado o uso do método dos momentos, visto que este é mais rápido e de mais simples desenvolvimento. Sendo assim, as estimativas pelo método dos momentos podem servir como estimativas preliminares do método da máxima verossimilhança.

Sua descrição é simples e de fácil entendimento. Seja uma variável aleatória x, cuja população estatística é caracterizada por um conjunto de k parâmetros. Considere também uma amostra aleatória dessa população estatística de tamanho n.

Serão considerados dois momentos para este desenvolvimento. O primeiro seria o momento populacional estabelecido como função dos parâmetros:

$$\mu_k' = \mu_k'(\theta_1,\ldots,\theta_k) = E(y^k) \tag{57}$$

O outro é referente aos momentos amostrais:

$$m'_k = \frac{1}{n}\sum_{u\in 1:n} y_u^j,\ j \in 1:k \tag{58}$$

Este método de estimação se dará por meio da igualdade entre momentos populacionais com seus respectivos momentos amostrais. Com isso, chega-se a um sistema de equação resultante mostrado a seguir:

$$\mu_k'(\theta_1,\ldots,\theta_k) = m_k',\ k \in \{1,2,\ldots\} \tag{59}$$

A partir então da resolução deste conjunto de equações, obteremos o conjunto de estimadores referente ao método dos momentos.

b. Método da Verossimilhança

O método da verossimilhança é mais recomendável, pois possibilita realizar inferências com propriedades altamente desejáveis. Dentre os motivos para o seu uso pode-se citar:

- Ambiente consistente para tratamentos dos diferentes aspectos da inferência estatística: estimação e testes de hipótese;
- O conceito de verossimilhança é muito próximo ao de probabilidade, de modo que a idéia central da máxima verossimilhança possui forte apelo intuitivo;
- Pode ser utilizada como critério para seleção de modelos.

O estimador de máxima verossimilhança (EMV) é a estatística que maximiza a função de verossimilhança. É possível utilizar os recursos do cálculo diferencial e encontrar o máximo de $L(\theta)$ derivando-o e o igualando à zero: $dL(\theta)/d\theta=0$. Pode-se notar também que como o logaritmo natural de L é uma função crescente, então $\ln(L)=\ln(L(\theta))$ tem o mesmo máximo de $L(\theta)$, sendo esta conhecida como função de log-verossimilhança. Esta função é ainda mais usada do que a anterior uma vez que sua manipulação matemática é mais simples, pois torna os produtos e produtórios em somas e somatórios, respectivamente.

Além de possuir boas propriedades, caso certas condições denominadas condições de regularidade sejam satisfeitas, pode-se dizer que os estimadores possuem as seguintes características:

- Consistente;
- Assintoticamente normal;
- Assintoticamente não tendencioso;
- Assintoticamente eficiente;

É claro que, apesar de se tratar de um método amplamente utilizado especialmente por apresentar boas propriedades para sua aplicação na área estatística, algumas dificuldades referentes a seu uso devem ser aqui citadas:

- Em geral, necessita da especificação da população estatística para ser aplicado.
- Matemática não trivial.
- Com exceção de alguns casos, métodos numéricos sofisticados necessitam ser utilizados. Geralmente são métodos iterativos (necessitando, portanto, do estabelecimento de valores iniciais) e pouco triviais.
- Boas propriedades válidas para grandes amostras, podem não ser válidas para pequenas amostras.
- Estimadores de máxima verossimilhança nem sempre existem
- Pode ser sensível à escolha de valores iniciais. Em alguns casos, outros métodos, como o método dos momentos, podem ser utilizados para produzir valores iniciais.

Baseado nesse respaldo teórico o tratamento dado a estimação dos parâmetros da distribuição em questão, gama, pode ser realizado através do método da máxima verossimilhança. Por isso, particularizando para a busca dos parâmetros α e β referentes à distribuição gama, teremos a seguinte equação de estimação:

$$L(\alpha,\beta)=\frac{\beta^{-n}}{[\Gamma(\alpha)]^{n}}\cdot\frac{1}{\beta^{-n(\alpha-1)}}\cdot\prod_{i=1}^{n}y_i^{\alpha-1}\cdot e^{\frac{-y_i}{\beta}} \tag{60}$$

A partir da equação acima e ainda levando em consideração o já mencionado fato de que o trabalho com a log-verossimilhança além de possuir boas propriedades é de mais fácil tratamento matemático do que o método da verossimilhança puro, a equação a ser utilizada será obtida então a partir do operador logaritmo aplicado na equação antecedente como segue:

$$l(\alpha,\beta)=-n\cdot\log(\beta)-n\cdot\log(\Gamma(\alpha))+\sum_{i=1}^{n}[(\alpha-1)\cdot\log(y_i)-\frac{y_i}{\beta}] \tag{61}$$

De posse de ambas as equações, observa-se que como também já mencionado, tanto faz aplicarmos o método à (61) ou (62) que o resultado será o mesmo, ou seja, dl/dθ = dL/dθ = 0 resultarão na mesma resposta final.