

**Estudo comparativo de métodos de avaliação de grandes quantis
de perda acumulada no modelo de seguro de Cramér-Lundberg.**

Armando Antonio Mendes

DISSERTAÇÃO APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DO TÍTULO
DE
MESTRE EM CIÊNCIAS

Área de Concentração: Estatística
Orientador: Prof. Dr. Vladimir Belitsky

São Paulo, agosto de 2008

Crença

Enquanto você não se comprometer, sempre haverá hesitação, oportunidade de recuar, ineficácia.

A respeito de todos os atos de iniciativa e criação há uma verdade elementar - ignorá-la mata inúmeras ideias e planos esplendidos.

No momento em que você se compromete, a providência também age.

Multidões de coisas que de outra forma jamais aconteceriam, acontecem para ajudar.

Da decisão brota uma torrente de acontecimentos, pondo a seu favor acidentes, encontros e ajuda material que você não esperava nem sonharia encontrar:

‘Tudo o que você pode fazer ou sonha que pode comece. A ousadia contém gênio, poder e magia.’

Do livro the Scottish Himalayan expedition de W. N. Murray.

As duas últimas linhas são atribuídas a Goethe, o poeta, filósofo e cientista alemão, um dos maiores espíritos dos últimos séculos.

Dedicatória

*À Deus,
a minha esposa Angela,
a minhas filhas Renata e Carla.*

Agradecimentos

*Algumas pessoas entram e saem de nossa vida silenciosamente;
outras ficam um pouco e deixam marcas em nossos corações;
daí em diante, jamais seremos os mesmos.*

(Anônimo)

Em virtude disso , meus sinceros agradecimentos a algumas pessoas que tiveram forte influência e ou participação na consecução desse projeto. Portanto agradeço:

Primeiramente *in memoriam* ao Prof. Dr. Flávio Wagner Rodrigues (falecido em 2004) que nos ensinou apreciar os trabalhos de Feller e plantou em mim a primeira semente no ramo das probabilidades.

Em segundo ao Prof. Dr. Wagner de Souza Borges por resgatar antigos projetos, há anos adormecidos.

Em terceiro ao Prof. Dr. Carlos Alberto de Bragança Pereira por nos ensinar a importância dos fundamentos e nos alertar constantemente que em probabilidade a intuição não costuma ser boa conselheira.

Em quarto aos colegas e amigos de mestrado Josivon, Fernando e Dalton três corações de ouro.

E finalmente ao último, mas não menos importante, Prof. Dr. Vladimir Belitsky, meu orientador, a quem estou profundamente em débito, por me conduzir eficientemente como aprendiz de ciências.

Resumo

A perda acumulada num período, de uma empresa de seguros, pode ser estimada a partir do modelo de Cramér-Lundberg que modela os instantes de chegadas de sinistros e suas severidades. Quando se tenta avaliar os valores extremamente grandes da perda acumulada estimada, surgem problemas técnicos que podem afetar a precisão da avaliação. Tais problemas são identificadas no presente trabalho, e alguns métodos matemáticos aptos a resolver ou amenizar tais problemas são sugeridos. Ao combinar os métodos, pode-se construir diversas abordagens para a estimativa de valores extremos da perda acumulada. A comparação dos resultados da aplicação de algumas destas abordagens para um problema real está feita no presente trabalho.

Palavras-chave: Modelo de Cramér-Lundberg, modelos de perda, perda acumulada, risco, valor-sobre-risco, convolução de distribuições probabilísticas, teoria de valores extremos.

THE UNIVERSITY OF CHICAGO

Abstract

We consider the collective risk model for aggregate loss of a insurance company or alike. Our concern is with the estimation of high quantiles of the distribution of the aggregate loss when the frequency distribution and the single-loss distribution of the model are inferred from data. We identify the causes for the inaccuracy of estimation and we present a number of methods applicable to enhance the accuracy. Combination of different methods gives rise to different estimation approaches. We execute several of them for a real data set and compare their results.

Keywords: Collective risk model, aggregate loss, extreme values, value-at-risk, convolution of probability distributions, extreme value theory.

Sumário

1	Introdução	3
2	Modelo de Cramér-Lundberg para companhias de Seguro	21
2.1	Definição do Modelo	21
2.2	Características do Modelo	24
2.2.1	Número de Ocorrências no Processo $N(t)$ -Poisson	24
2.2.2	Numero de Ocorrências no Processo $N(t)$ -Binomial Negativa	24
2.2.3	Distribuição da Severidade X_i	25
2.2.4	Total de Perdas Agregadas do Processo	26
3	Embasamentos teóricos	29
3.1	Valor em Risco	29
3.2	Proteção contra eventos severos	30
3.3	Operacionalizando o Modelo	35
3.3.1	Número de ocorrências $N(t)$	35
3.3.2	Severidade	38
3.3.3	Obtenção distribuição de Perdas- Método Analítico	38
3.3.4	Obtenção distribuição de Perdas- Métodos Computacionais	41

3.4	Teoria dos Valores Extremos	46
3.4.1	Teoria necessária para justificar o método POT	46
3.4.2	Família de distribuições de Pareto Generalizadas	49
3.4.3	Interligação entre as EVD's e GPD's	51
3.4.4	O Método POT- Peaks Over Threshold- Aspectos de Implementação. .	52
4	Aplicação do método para um caso real	55
4.1	Distribuição do número de Ocorrências	55
4.1.1	Ocorrências Mensais	55
4.1.2	Dados agrupados	56
4.1.3	Histograma	56
4.1.4	Principais Estatísticas	57
4.2	Distribuição da Severidade	58
4.2.1	Severidade distribuída no período-1	58
4.2.2	Severidade distribuída no período-2	59
4.2.3	Principais estatísticas da Severidade	60
4.2.4	Histograma englobando todo o período de estudo	61
4.3	Resumo Principais Estatísticas-Ocorrências e Severidade	62
4.4	Primeira Abordagem	62
4.4.1	Determinação dos momentos	62
4.4.2	Resultados Primeira Abordagem	63
4.5	Segunda Abordagem	63
4.5.1	Distribuição de Poisson	63
4.5.2	Distribuição Binomial Negativa	64

<i>SUMÁRIO</i>	xv
4.5.3 Computação	64
4.5.4 Aproximação para a Cauda	66
4.5.5 Resumo Resultados Segunda Abordagem	70
4.6 Terceira Abordagem	70
4.6.1 Parâmetros Estimados	71
4.6.2 Resultados Terceira Abordagem	71
5 Conclusões e Comentários	73
5.1 Conclusões	73
5.2 Novos Horizontes	75
Referências Bibliográficas	77

Capítulo 1

Introdução

O *problema* prático *principal*, ao qual aplica-se o presente trabalho, surgiu na atividade da seguradora AIG do UNIBANCO.

Para simplificar a formulação e preservar a sigilidade de dados, assume-se que a seguradora conceda cobertura a somente um tipo de seguro, digamos seguro de vida.

É frequente acontecer, que a família de um assegurado descubra muito tempo após sua morte que ele detinha uma apólice de seguro de vida e, portanto, venha a reclamar o pagamento do prêmio com certa defasagem.

Esta reclamação chama-se *sinistro ocorrido mas não avisado*, cujo nome destaca a sua característica principal: o *instante de ocorrência* está muito defasado em relação ao *instante de contabilização*.

Até aqui, não se vê nada de especial. O problema surge devido as exigências do Banco Central, que determina que para garantir a cobertura dos sinistros deste tipo, a seguradora precisa criar uma reserva de capital R ou seja um colchão.

Para contemplar esta exigência em termos mais precisos, determina-se um *horizonte* de tempo T , o qual, no caso deste problema, será um mês ($T = 1/12$ – pois a unidade de tempo é um ano). Então, o regulamento obriga que hoje – no primeiro dia de mês – a seguradora reserve um valor R que seja suficiente, com uma dada probabilidade α , para cobrir todos os sinistros ocorridos mas não avisados que serão contabilizados no decorrer do mês.

O valor de α – determinado pelo regulamento – é algo muito próximo a 1; nos trabalharemos com valores entre 0.95 e 0.999, normalmente aplicados ao problema descrito agora e em outras situações semelhantes ao problema aqui apresentado, entre as quais destacamos o problema de prevenção de risco operacional.

Notamos que a proximidade de α a 1 traz desafios matemáticos e computacionais à solução do problema e necessita de métodos estatísticos avançados que preocupam-se com a qualidade de estimativas para as caudas de distribuições de probabilidade. Os métodos vêm das áreas de estatística desenvolvidas recentemente. A função de distribuição da variável aleatória de perdas total S não pode ser calculada de maneira precisa devido as seguintes razões:

1. as distribuições das variáveis aleatórias X'_i e N não são precisamente conhecidas, pois só tem-se as amostras delas;
2. o cálculo preciso da distribuição da soma $X_1 + \dots + X_n$ é um problema computacionalmente pesado quando n é grande;
3. o fato da variável contadora não ser limitada exige o cálculo das distribuições de $X_1 + \dots + X_n$ para todo $n \in \mathbb{N}$, o que é computacionalmente inalcançável, a menos em casos particulares quando X é normal, exponencial ou gama o que nem sempre é o caso.

Existem alternativas diversas de lidar com cada um dos problemas (1)-(3) acima descritos. As combinações de soluções diferentes para cada um dos problemas proporciona **abordagens** diferentes de solução do *problema principal*. O objetivo do presente trabalho é expor em detalhes algumas das abordagens possíveis e comparar as eficiências das mesmas.

Este é o pilar do **caráter inovador** de nosso trabalho, pois a juventude dos métodos aplicados faz com que as diferenças em suas respostas não são plenamente expostas e analisadas – o nosso trabalho é uma contribuição para tais exposições e análise.

Para apresentar nossas abordagens ao problema principal e expor os problemas teóricos que surgem na comparação dos resultados das abordagens, estabeleceremos algumas notações. Denotaremos por S a soma das severidades dos sinistros que chegarão no intervalo de tempo $[0, T]$, quer dizer dos ocorridos mas não avisados cujos instantes de contabilização estarão em $[0, T]$. Em outras palavras S representa a perda acumulada no período considerado.

Tais instantes serão denotados por t_1, t_2, \dots, t_N em ordem crescente, mas estes não figurarão no modelo probabilístico do problema; o que será utilizado é N – o número deles, ou seja a frequência observada nos intervalos de tempo.. Por esta razão, aliás, as severidades dos sinistros serão denotadas por X_1, \dots, X_N e não por X_{t_1}, \dots, X_{t_N} . O *modelo probabilístico* para a soma de todos os sinistros a ocorrer, chamado de *perda acumulada*, será dado por :

$$S = \begin{cases} \sum_{i=1}^N X_i, & \text{para } N > 0, \\ 0 & \text{se } N = 0 \end{cases} \quad (1.1)$$

onde:

- N é a variável aleatória discreta positiva chamada *contadora*,
- X_1, X_2, \dots são variáveis aleatórias independentes e identicamente distribuídas.

Este é o modelo clássico (veja KLUGMAN, PANJER, WILLMOT [17]), que será aceito por nos aqui. Segundo o modelo, a perda acumulada S é uma variável aleatória; chamaremos sua função de distribuição por F_S . Com isto, podemos dar a resposta teórica ao problema principal de – lembrando – achar o valor do colchão R :

$$\begin{aligned} R &= \alpha\text{-quantil da distribuição } F_S \\ &= F_S^{-1}(\alpha) \text{ (re-escrevendo o quantil em termos precisos da Estatística)} \\ &= \text{VaR}_\alpha(S) \text{ (a notação para } F_S^{-1}(\alpha) \\ &\quad \text{que denomina-se } \textit{valor em risco ao nível } \alpha \text{ de } S \end{aligned} \quad (1.2)$$

Para aproveitar da resposta fornecido pelo modelo, é precisa modelá-lo aos dados. Os dados fornecidos por AIG do UNIBANCO expõem os sinistros ocorridos mas não avisados que foram contabilizados em $K = 28$ meses entre os anos de 2004 e 2007. Chamaremos por n_k o número das contabilizações no mês k , e por $x_{k1}, x_{k2}, \dots, x_{k,n_k}$ os valores dos sinistros contabilizados no correspondente mês.

Apresentamos na figura (1.1) parte dos dados coletados. Para uma visão completa dos dados coletados vide capítulo 3 onde fazemos uma análise exploratória completa dos dados disponíveis.

Sinal		+																							
Valor	Mês	Ano																							
		1			2			3			4		5		6										
Ocorr		2005	2006	2007	2005	2006	2007	2005	2006	2007	2005	2006	2005	2006	2005	2006	2005	2006	2007						
1	50	50	300	200	400	300	50	50	900	50	100	200	50	100	200	50	100	200							
2	2.600	50	1.000	400	600	500	200	100	1.000	50	200	200	100	200	200	100	200	200							
3	2.800	100	1.200	800	800	800	200	100	1.700	100	200	1.100	100	500	200										
4	3.200	100	1.400	800	2.600	800	200	200	11.800	100	300	1.400	100	500	200										
5	4.700	500	2.000	1.600	3.200	1.000	200	400	12.900	200	300	2.000	700	800	300										
6	4.800	700	2.700	1.800	13.300	3.600	500	500	16.400	200	500	2.100	700	1.000	600										
7	5.000	1.100	3.200	2.300	17.400	35.600	600	1.200	23.500	200	1.100	4.700	1.000	1.600	800										
8	6.000	1.700	3.600	2.900	19.900	42.100	800	1.600	70.200	200	1.500	4.800	1.800	2.700	900										
9	6.300	4.800	4.200	7.700	36.400	110.600	1.500	1.700	501.100	300	1.700	6.100	1.900	6.300	1.100										
10	7.200	7.600	9.300	17.600	50.000	434.200	3.700	1.900	1.064.600	400	1.900	7.900	7.400	8.400	2.100										
11	18.700	7.900	33.500	56.700	142.300	737.500	4.000	2.500	3.536.200	600	1.900	11.000	12.800	10.500	2.300										
12	33.200	21.900	119.400	108.800	510.900	2.704.400	8.200	2.500	5.177.100	800	3.000	11.200	29.500	38.300	4.600										
13	87.000	136.600	445.100	480.200	603.200	5.649.300	14.100	2.500		900	6.000	27.800	37.300	99.800	14.500										
14	656.200	151.000	759.300	770.200	2.352.900		30.400	2.600		1.200	8.800	34.100	74.800	555.400	27.700										
15	718.400	556.000	3.573.800	2.437.700	4.812.500		123.900	3.500		1.300	14.700	100.400	151.000	935.100	75.600										
16	2.718.000	758.700	4.924.400	4.406.300			523.500	4.200		1.400	37.900	528.600	372.100	2.295.700	89.400										
17	4.999.700	2.756.200					985.300	6.500		1.600	111.100	1.003.200	933.500	5.063.300	592.000										
18		4.592.600					2.379.500	14.000		3.600	685.600	2.530.400	3.029.300		874.100										
19							4.263.700	25.800		4.100	729.800	5.201.500	5.249.600		2.663.300										
20								34.500		8.600	2.879.500								5.415.900						
21								141.200		11.600	5.865.600														
22								157.200		18.300															
23								470.400		25.100															
24								951.800		55.100															
25								2.808.200		507.100															
26								4.235.500		792.300															
27										1.857.700															
28										4.983.700															

Figura 1.1: Dados Parciais AIG do UNIBANCO - maiores detalhes capítulo 3

Na primeira etapa do ajuste do modelo, usamos os dados para construir \hat{F}_X – a estimativa de F_X , a função de distribuição de severidade de qualquer sinistro, ou, em termos formais, a de cada uma das variáveis aleatórias X_i . Devido aos pressupostos do modelo, cada observação de severidade é um valor retirado da distribuição F_X de maneira independente das demais retiradas. Isto sugere naturalmente tomar todas as observações disponíveis e construir a função de distribuição amostral com base nelas. Simplificando a notação, chamaremos x_1, \dots, x_n todas as observações da severidade ($n = 535$).

Denotamos \hat{F}_X a função de distribuição amostral gerada por esta amostra. Esta é nossa aproximação para F_X . Por enquanto, pois mais tarde virão suas versões melhoradas. Para uma ideia inicial da característica desta distribuição amostral segue o histograma, figura (1.2), e a função de distribuição, figura (1.3):

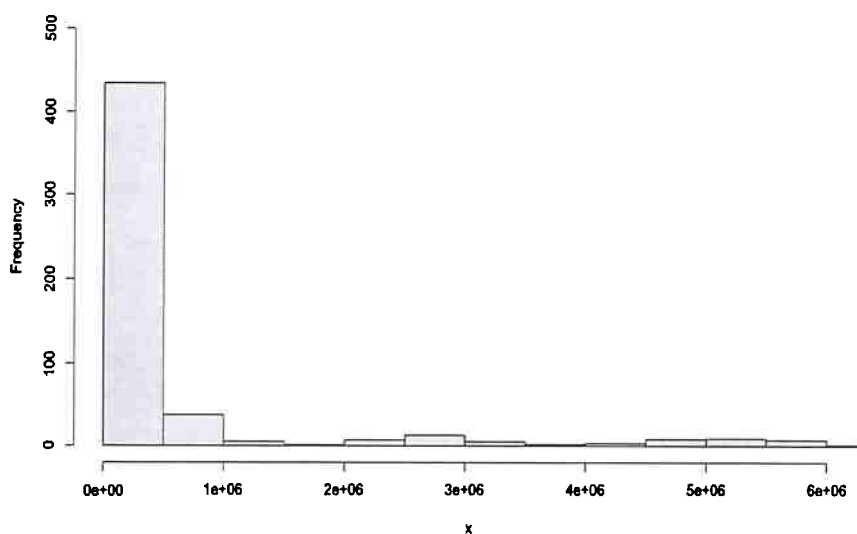


Figura 1.2: Nesta figura temos o histograma correspondente à amostra x_1, \dots, x_n das severidades dos sinistros observados

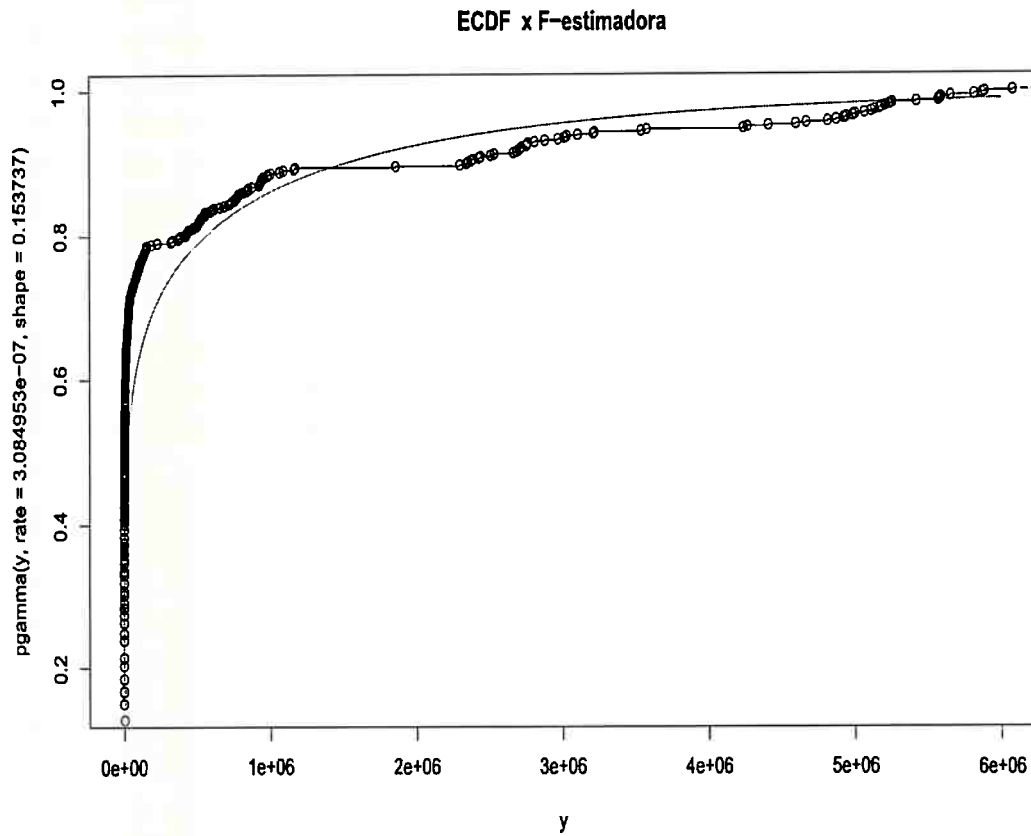


Figura 1.3: Aqui a figura mostra a função de distribuição amostral do histograma. Esta tem a forma de função-escada; junto com esta, há uma linha suave que representa a aproximação de sua cauda pela função F_{cauda} .

A segunda etapa do ajuste do modelo é o ajuste da distribuição de N , o número de sinistros ocorridos mas não avisados por mês. O que carrega a informação sobre esta é a amostra n_1, \dots, n_K . Assumimos que estas são retiradas independentes da F_N . A inferência na F_N é paramétrica. Apresentamos a seguir, tabela (1.1), resumidamente as frequências das ocorrências de sinistros:

Ano	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
2004	-	-	-	-	-	-	-	-	-	-	-	23
2005	17	16	19	28	19	17	17	14	14	15	16	25
2006	18	15	26	21	19	20	20	27	17	23	27	21
2007	16	13	12	-	-	-	-	-	-	-	-	-

Tabela 1.1: Número de Ocorrências de Sinistros por mês.

Da mesma forma que fizemos para a distribuição da severidade segue um histograma, figura (1.4) e sua função densidade de probabilidade estimada, figura (1.5), para termos uma ideia inicial da característica desta distribuição amostral:

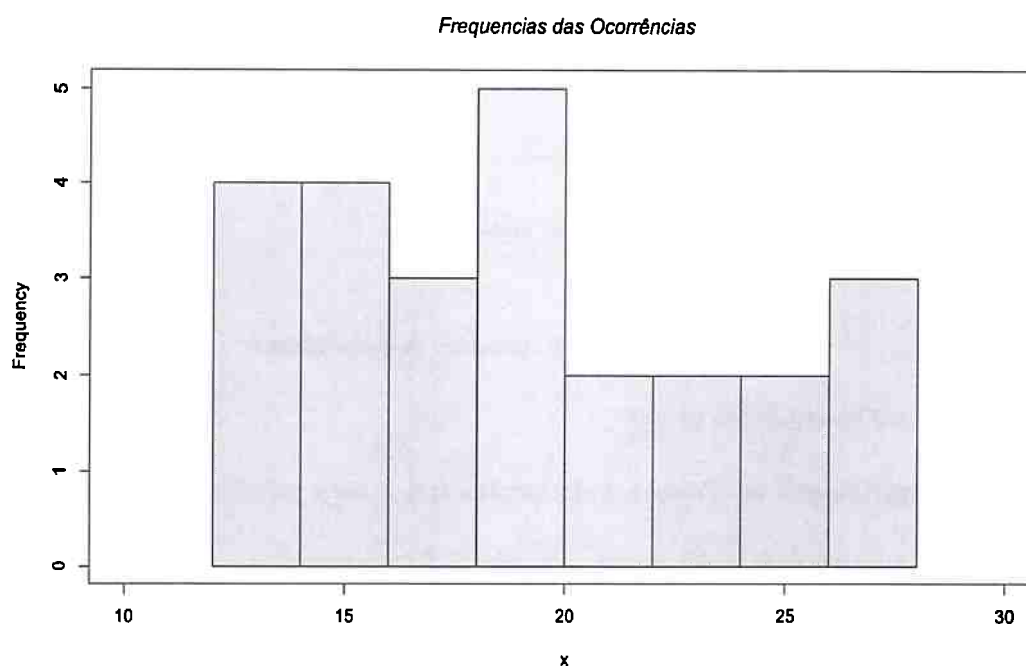


Figura 1.4: Histograma das frequências das ocorrências

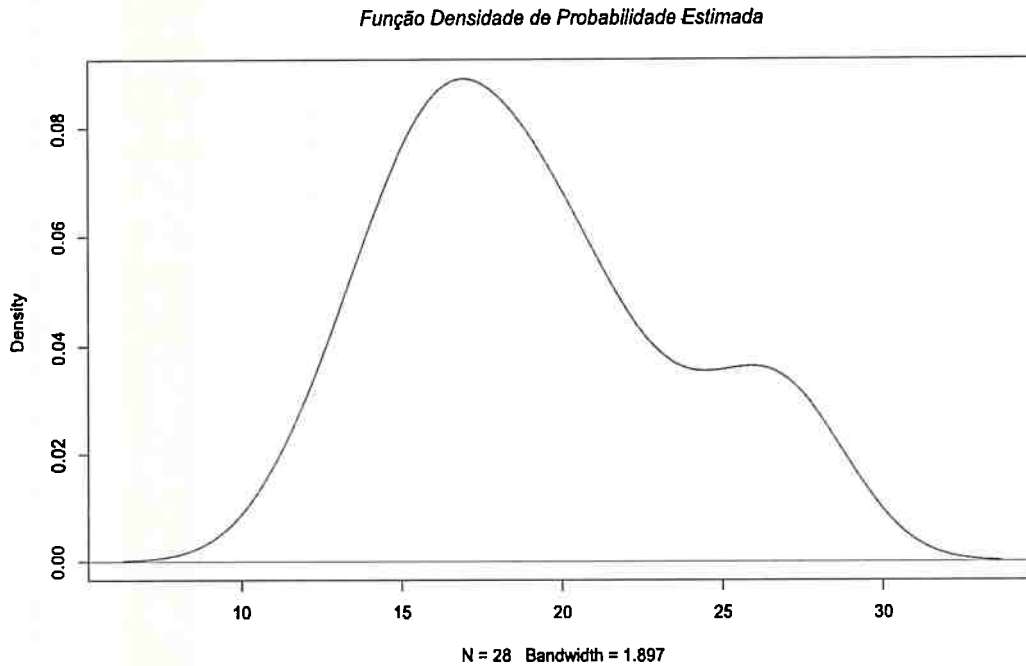


Figura 1.5: Função de densidade de probabilidade

Para F_N temos as seguintes opções:

- (1) F_N é a distribuição de Poisson cujo parâmetro λ será inferido a partir da amostra n_1, \dots, n_K ;
- (2) F_N é a distribuição Binomial Negativa, cujos parâmetros r e p serão inferidos a partir da amostra n_1, \dots, n_K e serão calculados pelo método dos momentos;

No capítulo 5 daremos o maior detalhamento às opções (1) e (2). Já para agora, escolhemos a opção (1).

A execução dela é simples: como a média é o estimador de máxima verossimilhança na distribuição de Poisson ela assume o valor de λ .

Assim que as duas etapas de ajuste descritas acima forem completadas, podemos executar a **primeira e segunda abordagem** da solução do problema. A idéia da abordagem é aproximar F_S por $F_{\hat{S}}$ onde

$$\hat{S} = \begin{cases} \sum_{i=1}^{\hat{N}} \hat{X}_i, & \text{para } \hat{N} > 0, \\ 0 & \text{se } \hat{N} = 0 \end{cases} \quad (1.3)$$

onde:

- \hat{N} é a variável aleatória discreta com função de distribuição \hat{F}_N ,
- $\hat{X}_1, \hat{X}_2, \dots$ são variáveis aleatórias independentes e identicamente distribuídas conforme \hat{F}_X .

E a resposta seria:

$$\text{VaR}_\alpha(\hat{S}) \approx F_{\hat{S}}^{-1}(\alpha). \quad (1.4)$$

Podemos mostrar esquematicamente o que foi feito até agora através da figura(1.6)onde fica evidente o comportamento da cauda da função distribuição resultante :

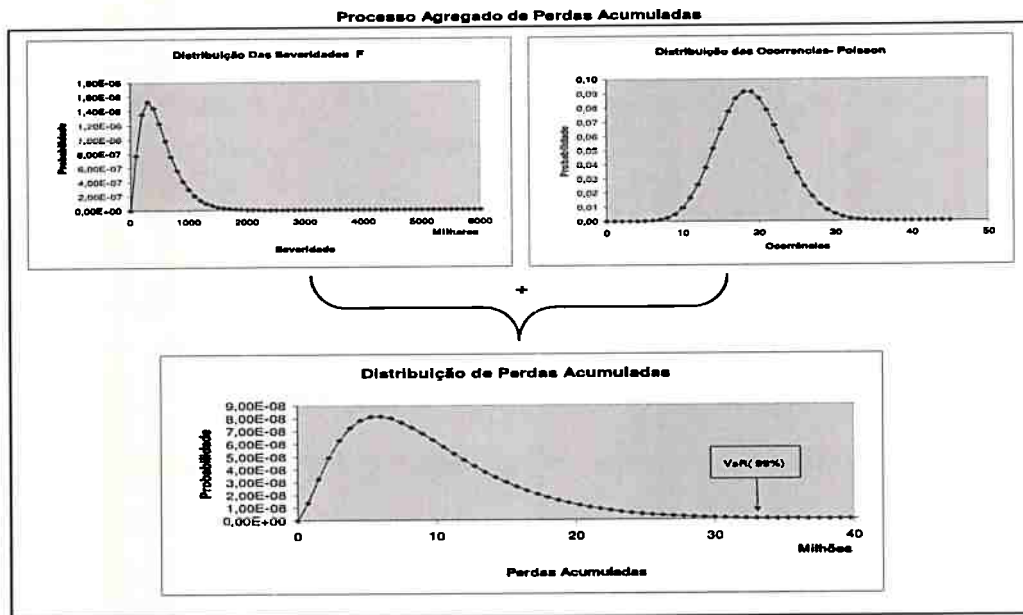


Figura 1.6: Processo Agregado das perdas acumuladas

Lembramos que existem problemas computacionais intrínsecos à construção de F_{ξ} . As razões deles e as soluções adotados no nosso trabalho serão reveladas e discutidas nos Capítulo 2.

Aquí só avisaremos que são quatro as soluções adotados ao problema de construirmos \hat{S} :

- Aproximação rápida pelo método dos momentos.
- Aproximação através de métodos analíticos.
- Aproximação através de simulações- métodos de Monte Carlo.
- Aproximação pelo método de Panjer.

Inicialmente obtivemos uma aproximação rápida, trabalhando com os momentos teóricos da função de perdas total, e valores para VaR próximos de 1, cujas previsões estão a seguir, na tabela (1.2). A este procedimento denominamos de **primeira abordagem**.

VaR	Capital a Reservar
95,0%	20 884 091
99,0%	28 418 951
99,9%	38 597 584

Tabela 1.2: Estimativa para a Reserva de Capital Primeira Abordagem.

Continuaremos a presente apresentação com a exposição do resultado obtido através de simulações. Os valores foram calculados para VaR próximos de 1 e estão na tabela (1.3).

VaR	Capital a Reservar
95,0%	20 446 862
99,0%	26 305 012
99,9%	33 545 101

Tabela 1.3: Estimativas de Reserva de Capital -Por Simulação

Os valores obtidos, resultantes da segunda abordagem, mostrados na tabela (1.3), foram obtidos através dos cálculos dos quantis elencados sobre a distribuição empírica resultante da simulação, cujos histograma e função densidade são mostrados nas figuras (1.7) e (1.8), respectivamente:

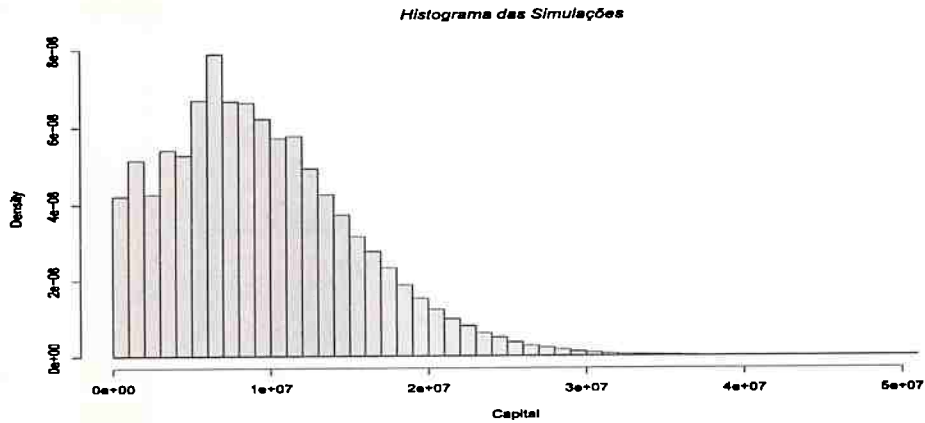


Figura 1.7: Histograma das Simulações 500 000 amostras

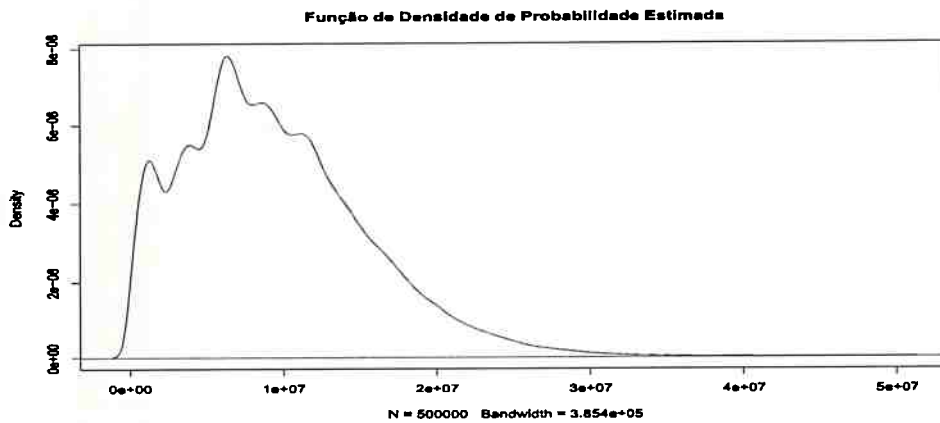


Figura 1.8: Função Densidade das Simulações 500 000 amostras

A partir desta, utilizando o método POT proveniente da Teoria de Valores Extremos, foi obtida \hat{F}_S^{cauda} – a aproximação da cauda direita da distribuição da perda acumulada S . Esta está em linha contínua na figura 1.9), apresentada a seguir.

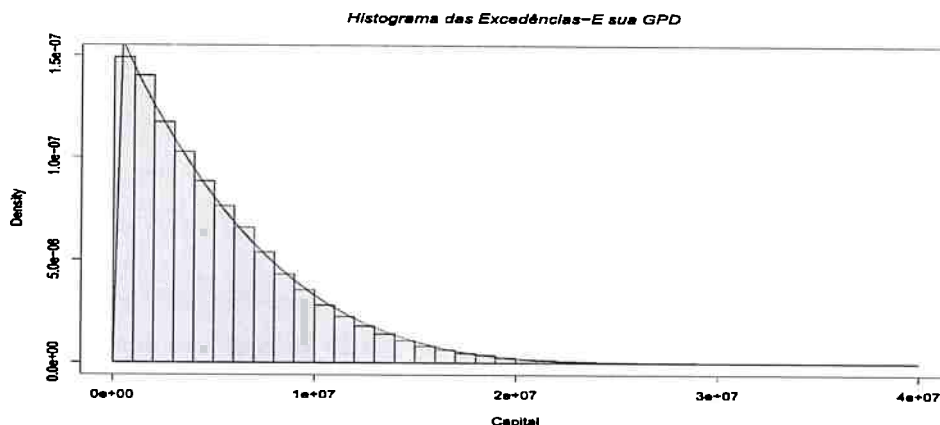


Figura 1.9: Temos aqui o histograma da cauda a direita da função de distribuição decorrente da segunda abordagem ao problema. Com uso do método POT proveniente da Teoria de Valores Extremos, foi obtida \hat{F}_S^{cauda} – a aproximação da cauda direita da distribuição da perda acumulada S . Esta está em linha contínua no desenho.

Acontece que a resposta (1.4) deve sub-estimar a verdadeira $\text{VaR}_\alpha(S)$ devido ao seguinte argumento: no procedimento de re-amostragem não toma-se em conta de que o valor do sinistro possa ultrapassar o maior valor da amostra de sinistros passados x_1, \dots, x_m .

Já que na realidade não é possível alegar que nenhum dos sinistros a ocorrer ultrapasse o maior sinistro observado até o presente, então a verdadeira distribuição de S tem cauda direita mais pesada de que a cauda da $F_{\hat{S}}$, quer dizer que o grafo de $F_S(x)$ fica abaixo do grafo $F_{\hat{S}}(x)$ para valores de x suficientemente grandes. Como reconstruir a verdadeira cauda direita de F_S a partir da cauda de $F_{\hat{S}}$ é o assunto discutido amplamente no âmbito da Teoria de Valores Extremos (TVE). Esta teoria fornece diversos métodos de reconstrução, um dos quais, o denominado POT *Peaks Over Threshold*, é utilizado por nos aqui.

O método e a TVE serão revisados, na medida adequada à necessidade do presente trabalho, no Capítulo 3. O resultado – a função-aproximadora da cauda de F_S – está apresentada na Figura 1.9; denotamos esta função-aproximadora por $F_{\hat{S}}^{cauda}$. É ela que dá a resposta ao problema principal:

$$\hat{R} = \alpha\text{-quantil de } \hat{F}_{\hat{S}}^{cauda} \quad (1.5)$$

Os resultados obtidos com essa aproximação estão apresentados a seguir, tabela (1.4)

VaR	Capital a Reservar
95,0%	20 593 506
99,0%	26 032 302
99,9%	31 255 037

Tabela 1.4: Estimativas de Reserva de Capital

O procedimento acima descrito, o que resultou na resposta para (1.5), será denominado por **reamostrar e aproximar**, permitindo assim uma perda de rigor em nome da simplicidade da nomenclatura escolhida, pois nem a reamostragem nem a aproximação referidas por estes nome são procedimentos simplistas. O necessidade em dar o nome ao procedimento veio a tona pois há um outro procedimento, que gera uma resposta ao problema principal, e que será comparado com o procedimento descrito acima. Este outro será denominado por **aproximar e reamostrar**. O procedimento “aproximar e reamostrar” surge devido à aplicação da idéia da Teoria de Valores Extremos já naquilo que foi a primeira etapa da procedimento descrito acima. Seguindo esta idéia, o método POT deve ser aplicado já à estimativa da função da distribuição de X , pois o método permite determinar qual seria a distribuição dos sinistros mais severos que o maior da amostra x_1, \dots, x_n .

Na figura (1.10) fica mais claro como estamos querendo proceder.

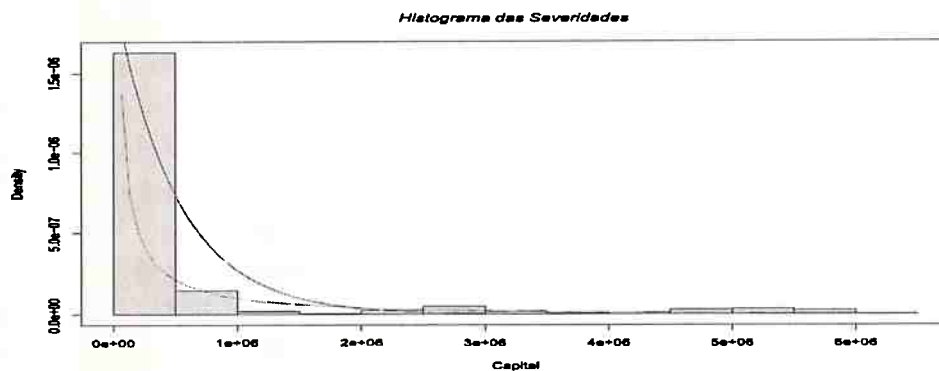


Figura 1.10: Histograma das Severidades com aproximações de distribuições Gama(mais a esquerda) e Exponencial

As funções contínuas mostrada na figura (1.10) denominaremos por \hat{F}_X^{cauda} . Chamaremos esta por **primeira etapa** da terceira abordagem. É óbvia a diferença entre ela e a primeira etapa da segunda abordagem. A **segunda etapa** da terceira abordagem estima a distribuição da variável aleatória contadora N pelos mesmos métodos que foi feito na segunda abordagem. Após as estimativas das distribuições de X e de N serem feitas, produzimos

$$\tilde{S} = \begin{cases} \sum_{i=1}^{\hat{N}} \tilde{X}_i, & \text{para } \hat{N} > 0, \\ 0 & \text{se } \hat{N} = 0 \end{cases} \quad (1.6)$$

onde:

- \hat{N} é a variável aleatória discreta com função de distribuição $F_{\hat{N}}$,
- $\tilde{X}_1, \tilde{X}_2, \dots$ são variáveis aleatórias independentes e identicamente distribuídas conforme \hat{F}_X^{cauda} .

E a resposta seria

$$\text{VaR}_\alpha(\tilde{S}) \equiv F_{\tilde{S}}^{-1}(\alpha). \quad (1.7)$$

Assim como na segunda abordagem, aqui também temos diversos problemas na construção de $F_{\tilde{S}}$. Os métodos aplicáveis para solucionar estes são:

- Métodos analíticos.
- Métodos de Monte Carlo.
- Método de Panjer.

Como pode-se ver, os procedimentos aqui adotados são basicamente os mesmos utilizados na segunda abordagem com as seguintes ressalvas :

1. A opção por métodos analíticos pode ser importante em função de resultados conhecidos para a convolução de variáveis aleatórias para algumas classes (especificamente para as distribuição exponencial e gama)

2. No método de Panjer há a necessidade de discretizar a função de distribuição obtida na aproximação da severidade, para torná-lo exequível do ponto de vista computacional.

Com o intuito de não complicar o presente argumento, neste momento, apresentamos o resultado da execução de (1.7) por métodos analíticos. Os resultados obtidos com esse procedimento estão apresentados a seguir, tabela (1.5).

VaR	Capital a Reservar
95,0%	20 820 042
99,0%	28 071 801
99,9%	37 726 789

Tabela 1.5: Estimativas de Reserva de Capital Terceira Abordagem

Os objetivos principais do nosso trabalho são: destacar a existência de abordagens diferentes ao problema descrito acima com o enfoque no que a especificidade do problema faz com que as abordagens possam ser separadas em três grupos, as que chamamos de primeira, segunda e terceira nos argumentos apresentados até agora; exibir a diferença nos resultados fornecidos por abordagens de grupos diferentes, mas indicando que a diferença relativa não é muito grande; exibir as sutilezas das abordagens mostrando as ramificações que, sendo seguidas, levam às execuções e resultados diferentes; descrever a teoria matemática que serve de amparo à construção dos procedimentos até aqui mencionados; sugerir alternativas que possam facilitar a vida dos gestores de risco; etc

Deve ser observado, que o problema acima descrito é um problema bastante genérico. Este surge em diversos ramos e situações. A situação pode se tornar mais complexa, pelo fato da falta de informações complementares e ou falta de base de dados confiáveis para evidenciar se os dados coletados pertencem a uma mesma ou várias linhas de negócios para permitir a separação de perdas por ramos, e mensurar adequadamente o risco operacional decorrente de atividades bancárias. A situação mais simples é de cobertura de sinistros de uma seguradora. O problema acima descrito foi formulado para os sinistros ocorridos mas não avisados, mas ele é o mesmo para qualquer caso que enquadra-se nas condições (1.) a (3.) descritas a seguir:

1. Sinistros (ou, em termos gerais “pedidos”) ocorrem ao tempo seguindo um processo estocástico.

2. Os valores de cobertura de sinistros (severidade) são realizações de uma distribuição probabilística.
3. A estratégia de provisão de sinistros é: fixar um período e fazer uma reserva da qual serão pagos os sinistros que chegarão neste período. Notamos que por causa da origem estocástica de sinistros, o valor acumulado dos que chegarão no período escolhido é uma variável aleatória. Por isso que a reserva, qualquer que seja seu tamanho, pode ser insuficiente para a provisão. Esta insuficiência está medida em probabilidade de sua confiança. A cada valor de confiança $\alpha \in (0, 1)$ corresponde o valor de reserva que garante a provisão com a probabilidade α ; este valor denota-se por VaR_α .

Notamos ainda que a estratégia alternativa a esta é não fazer reserva, mas cobrir os sinistros com o dinheiro que chega à seguradora devido aos prêmios pagos por segurados. Mas esta estratégia alternativa não é ótima, pois obriga a guardar prêmios no caixa da seguradora. É muito mais vantajoso aplicar este dinheiro para receber juros. Mas dinheiro aplicado fica “parado” por certo tempo no banco. Por isto, caso for adotada a aplicação de recursos, há necessidade da reserva R ou seja de um colchão. A estimativa de seu tamanho leva ao problema a ser abordado no nosso trabalho. Em outras palavras, nosso problema surge ou como um problema principal ou como um problema secundário na elaboração da estratégia de gestão de investimentos de qualquer seguradora.

Capítulo 2

Modelo de Cramér-Lundberg para companhias de Seguro

2.1 Definição do Modelo

Os fundamentos da moderna teoria do risco foram estabelecidos por Filip Lundberg com sua tese de 1903, (veja EMBRECHTS, KLUPPELBERG, MIKOSCH) [11] ,onde ele havia percebido que o processo de Poisson era o centro dos modelos atuariais de risco. Posteriormente com as contribuições de Harald Cramér aplicando os conceitos da teoria de processos estocásticos resultaram no modelo atualmente conhecido como Modelo de Cramér-Lundberg que apresenta a seguinte estrutura:

1. As ocorrências de sinistros(a partir de agora simplesmente sinistro), ocorrem em instantes aleatórios no tempo:

$$0 < T_1 < T_2 < T_3 < \dots \text{ e assim por diante.}$$

2. O numero de sinistros no intervalo $[0, t]$ é denominado

$$N(t) = \sup\{n \geq 1 : T_n \leq t\}, t \geq 0,$$

onde, por convenção , $\sup \phi = 0$, quer dizer que se $T_1 > t$, então $N(t) = 0$.

3. O tempo entre chegadas definidas como:

$$Y_1 = T_1, \quad Y_k = T_k - T_{k-1}, \quad k = 2, 3, \dots,$$

são variáveis aleatórias exponenciais independentes e identicamente distribuídas com média finita $E(Y_1) = 1/\lambda$.

4. A magnitude do sinistro(a partir de agora denominada severidade), $(X_k)_{k \in N}$ são variáveis aleatórias positivas, independentes e com distribuição de probabilidade F_X idênticas, possuindo média $\mu = E(X_1)$, e variância $\sigma^2 = Var(X_1) \leq \infty$.

5. As sequências (X_k) e (Y_k) são independentes entre si.

Como consequência dessas suposições podemos dizer que $N(t)$ é um processo homogêneo de Poisson com intensidade $\lambda > 0$. Portanto:

$$P(N(t) = k) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}, \quad k = 0, 1, 2, \dots \quad (2.1)$$

Temos também que a perda total(soma das severidades no período) desse modelo fica definida como:

$$S(t) = \begin{cases} \sum_{i=1}^{N(t)} X_i, & \text{para } N(t) > 0, \\ 0 & \text{se } N(t) = 0 \end{cases} \quad (2.2)$$

onde:

- $N(t)$ e X_1, X_2, \dots , como definidas acima.

¹mais adiante com as devidas justificativas esse resultado será obtido através de uma distribuição Binomial Negativa

Como as reclamações que chegam a seguradora se comportam como um processo de Poisson $N(t)$ de taxa $\lambda > 0$, e que X_i é a severidade da i -ésima reclamação temos que $S(t)$, $t > 0$, é um processo composto de Poisson.

Além disso após o uso da fórmula de probabilidade total, podemos finalmente apresentar a distribuição final de interesse que podemos chamar de distribuição total das perdas agregadas do processo como:

$$\begin{aligned} G_t(x) &= P(S(t) \leq x) \\ &= \sum_{i=0}^{\infty} \mathbb{P}[S(t) \leq x / N(t) = i] \mathbb{P}[N(t) = i] \\ &= \sum_{i=0}^{\infty} e^{-(\lambda t)} \frac{(\lambda t)^i}{i!} F^{*n}(x), \quad x \geq 0, \quad t \geq 0 \end{aligned} \quad (2.3)$$

onde $F^{*n}(x) = \mathbb{P}\left(\sum_{i=1}^n X_i \leq x\right)$ é a n -ésima convolução de F . E também que:

$$F^{*0}(x) = \begin{cases} 1 & \text{quando } x \geq 0; \\ 0 & \text{quando } x < 0. \end{cases}$$

$$\text{e } F^{*1}(x) = F(x).$$

Resultados estes que serão usados para calcular as convoluções.

Estudar o comportamento dessa distribuição faz parte dos objetivos deste trabalho uma vez que $F^{*n}(x)$ é a soma de distribuições independentes com determinadas características e verificar se essas características se mantêm ou se transformam em outras a medida que é agregada com a distribuição de probabilidades de Poisson se torna fundamental no processo de modelação que estamos interessados. Então passaremos a descrever os elementos necessários na avaliação desse modelo. Estaremos interessados em:

1. Estudar as características das variáveis $N(t)$, X_i e $S(t)$.

2. Analisar suas propriedades teóricas.
3. Baseando-se nas informações da seguradora AIG do UNIBANCO que forneceu dados de um período de três anos de $N(t)$ e de X_i produzir inferências para a distribuição das perdas totais agregadas do processo G_t .

2.2 Características do Modelo

2.2.1 Número de Ocorrências no Processo $N(t)$ -Poisson

Como definido na conceituação do modelo em (2.1), a variável aleatória *contadora* dos sinistros é a distribuição de Poisson:

$$\mathbb{P}(N(t) = k) = e^{-(\lambda t)} \frac{(\lambda t)^k}{k!}, \quad k = 0, 1, 2, \dots$$

Neste ponto temos um processo pontual de Poisson, abreviado como **PPP**. Com as seguintes propriedades:

$$\mathbb{E}[N(t)] = \lambda t \quad \text{e} \quad \text{Var}[N(t)] = \lambda t \quad \text{e também} \quad \hat{\lambda}t = \mathbb{E}[N(t)] \quad (2.4)$$

2.2.2 Numero de Ocorrências no Processo $N(t)$ -Binomial Negativa

Segundo KLUGMAN, PANJER, WILLMOT, [??], a distribuição binomial negativa tem sido extensivamente utilizada como uma alternativa para a distribuição de Poisson.

Como na distribuição de Poisson, seus valores de probabilidade são definidas para valores inteiros não negativos. Em função dela possuir dois parâmetros ela é mais flexível em forma do que a Poisson. A sua função é dada então por:

$$\mathbb{P}(N = k) = p_k = \binom{k-1}{r-1} p^r (1-p)^{k-r}, \quad (2.5)$$

para $k = r, r+1, \dots, r > 0$.

Da sua definição podemos calcular seus momentos, que resultam em:

$$\mathbb{E}[N(t)] = \frac{r}{p} \quad \text{e} \quad \text{Var}[N(t)] = \frac{r(1-p)}{p^2} \quad (2.6)$$

Como p é positivo, a variância da distribuição binomial negativa excede a média, em contraste com a distribuição de Poisson na qual a variância é igual a média.

Isto sugere que para alguns conjuntos de dados a distribuição binomial negativa pode ser melhor candidata na modelagem dos dados. É comum calcular relação entre a variância e a média da seguinte forma:

$$\frac{\text{Var}[N(t)]}{\mathbb{E}[N(t)]} = \frac{1-p}{p} = \frac{1}{p} - 1$$

e rescrevê-la como:

$$\text{Var}[N(t)] = \mathbb{E}[N(t)] \left(\frac{1}{p} - 1 \right) \quad (2.7)$$

onde $\left(\frac{1}{p} - 1 \right) \geq 1$

2.2.3 Distribuição da Severidade X_i

Neste momento não faremos nenhuma menção a respeito da característica da função de distribuição de severidade tomando apenas como premissas básicas que ela tenha seus primeiros momentos finitos, assim definidos:

$$\mathbb{E}[X_i] = \mu, \quad \text{e} \quad \text{Var}[X_i] = \sigma^2 \quad (2.8)$$

2.2.4 Total de Perdas Agregadas do Processo

Como definimos anteriormente $S(t)$ sendo :

$$S(t) = \begin{cases} \sum_{i=1}^{N(t)} X_i, & \text{para } N(t) > 0, \\ 0 & \text{se } N(t) = 0 \end{cases}$$

Temos então, independentemente do particular modelo considerado para $S(t)$, que responder uma questão de suma importância que é determinar a ordem de grandeza de $S(t)$, pois ela servirá de base para se calcular o prêmio do seguro que deverá cobrir as perdas acumuladas de $S(t)$. Podemos então avaliar os principais momentos de $S(t)$. Utilizando os momentos vistos em (2.4) e (2.8) :

$$\mathbb{E}[X_i] = \mu, \quad e \quad \text{Var}[X_i] = \sigma^2$$

$$\mathbb{E}N(t) = \lambda, \quad e \quad \text{Var}[N(t)] = \lambda$$

Podemos determinar a média e a variância de $S(t)$ através da esperança condicional resultando em:

$$\mathbb{E}[S(t)] = \mu\lambda \tag{2.9}$$

e

$$\text{Var}[S(t)] = \lambda\sigma^2 + \mu^2\lambda \tag{2.10}$$

Esses resultados derivam das propriedades da esperança condicional, como demonstraremos a seguir. Temos então:

$$\mathbb{E}[S(t)] = \sum_{i=0}^{\infty} \mathbb{E}[S(t)/N(t)] p_N(n)$$

$$\begin{aligned}
&= \sum_{i=1}^{\infty} \mathbb{E}[(X_1 + X_2 + \dots + X_N)/N(t)] p_N(n) \\
&= \sum_{i=1}^{\infty} \mathbb{E}[X_1 + X_2 + \dots + X_N] p_N(n) \\
&= \sum_{i=1}^{\infty} n\mu p_N(n) \\
&= \mu \sum_{i=1}^{\infty} n p_N(n) \\
&= \mu\lambda \quad \text{como apresentado em (2.9)}
\end{aligned}$$

Para a variância faremos:

$$\begin{aligned}
\text{Var}[S(t)] &= \mathbb{E}[(S(t) - \mu\lambda)^2] \\
&= \mathbb{E}[(S(t) - N\mu + N\mu - \lambda\mu)^2] \\
&= \mathbb{E}\left[\left(S(t) - N\mu + \mu(N - \lambda)\right)^2\right] \\
&= \mathbb{E}[(S(t) - N\mu)^2] + \mathbb{E}[\mu^2(N - \lambda)^2] + 2\mathbb{E}[\mu(S(t) - N\mu)(N - \lambda)]
\end{aligned}$$

Trabalhando individualmente com as parcelas dessa soma temos:

$$\begin{aligned}
\mathbb{E}[(S(t) - N\mu)^2] &= \sum_{i=0}^{\infty} \mathbb{E}[(S(t) - N\mu)^2 / N(t) = n] p_N(n) \\
&= \sum_{i=1}^{\infty} \mathbb{E}[(X_1 + X_2 + \dots + X_n - n\mu)^2 / N(t) = n] p_N(n) \\
&= \sum_{i=1}^{\infty} \sigma^2 n p_N(n) \\
&= \sigma^2 \sum_{i=1}^{\infty} n p_N(n) \\
&= \lambda\sigma^2
\end{aligned}$$

e

$$\begin{aligned}\mathbb{E}[\mu^2(N - \lambda)^2] &= \mu^2 \mathbb{E}[(N - \lambda)^2] \\ &= \mu^2 \lambda\end{aligned}$$

enquanto que:

$$\begin{aligned}\mathbb{E}[\mu(S(t) - N\mu)(N - \lambda)] &= \mu \sum_{i=0}^{\infty} \mathbb{E}[(S(t) - n\mu)(n - \lambda)/N(t) = n] P_N(n) \\ &= \mu \sum_{i=0}^{\infty} (n - \lambda) \mathbb{E}[(S(t) - n\mu)/N(t) = n] P_N(n) \\ &= 0\end{aligned}$$

porque

$$\mathbb{E}[(S(t) - n\mu)/N(t) = n] = \mathbb{E}[(X_1 + X_2 + \dots + X_n - n\mu) = 0]$$

O que juntando resulta em :

$$\text{Var}[S(t)] = \lambda\sigma^2 + \mu^2\lambda \quad \text{como apresentado em (2.10)}$$

Capítulo 3

Embasamentos teóricos

A aparente simplicidade do modelo esconde peculiaridades técnicas e operacionais, na avaliação do risco de perdas total, e o objetivo desse capítulo consiste em analisá-las.

3.1 Valor em Risco

Valor em Risco (VaR, do inglês *Value at Risk*) pode ser considerado como uma medida de variação potencial máxima do valor de uma carteira ou portfólio em um determinado período t pré-fixado, com dada *probabilidade*.

Em outras palavras o VaR apresenta uma medida de quanto se pode perder, com probabilidade p , dentro de um horizonte t fixado.

Como uma medida estatística ela procura prever com um nível de acurácia alto a perda máxima do valor de uma carteira de seguros, por exemplo, decorrentes de suas operações em um determinado período (*holding period*).

Normalmente estaremos trabalhando com valores de VaR para probabilidades superiores a 99%.

Estabelecido o *holding period* para uma distribuição de probabilidade $F_x(\cdot)$, o VaR é o p – *quantil* da mesma e fica automaticamente determinado pela transformação inversa da função $F_x(\cdot)$ no nível p estabelecido.

A partir de agora trataremos o Valor em Risco como VaR_p . Visualmente o conceito fica claro na figura (3.1) a seguir:

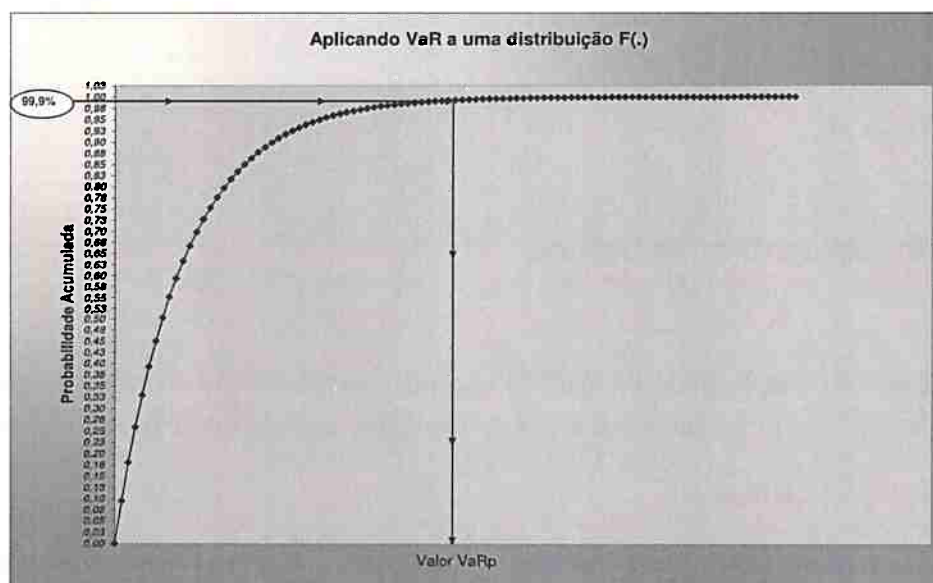


Figura 3.1: Valor em Risco para uma distribuição $F(\cdot)$ arbitrária

3.2 Proteção contra eventos severos

Neste caso trabalharemos com **VaR**(value-at-risk) com um nível de significância apropriado. Designamos por p o mais alto nível de confiança do VaR que queremos estimar. No caso de quisermos estimar somente um VaR, este p seria o nível deste VaR; por exemplo, se quisermos estimar $\text{VaR}_{95\%}$ então $p = 0,95$. Caso quisermos estimar **VaR**'es de um intervalo, então p seria igual ao extremo direito do intervalo; por exemplo, se quisermos estimar **VaR**'es de $\text{VaR}_{95\%}$ até $\text{VaR}_{97\%}$ então $p = 0,97$.

Analisando o gráfico da figura(3.2) fica bastante claro qual o objetivo que estamos perseguindo se quisermos por exemplo $\text{VaR}_{99\%}$.

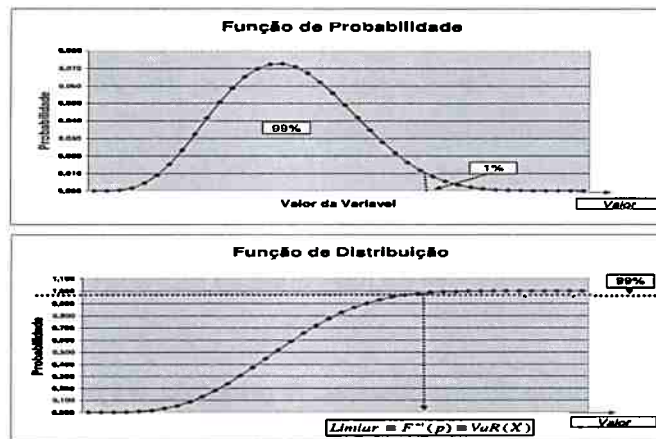


Figura 3.2: Função densidade de probabilidade e distribuição da variável aleatória contadora

Nosso objetivo é analisar as consequências na estimativa de **VaR** na expressão de G_t vista em (2.3) da parada de convoluções de ordem k_ε no extremo direito da função, onde k_ε é definido como o valor para o qual:

$$IP [N(t) = k_\varepsilon] = \varepsilon \quad (3.1)$$

onde $N(t)$ designa a variável aleatória Poisson de parâmetro λ , isto é

$$\mathbb{P} [N(t) = k] = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \dots \quad (3.2)$$

Esses argumentos funcionam somente quando

$$\varepsilon < 1 - p \quad (3.3)$$

É por isso que o parágrafo anterior demos uma atenção excessiva ao estabelecimento do valor de p .

Seja então $F(\cdot)$ uma função de distribuição. A princípio, ela é arbitrária. No caso que nos interessa, $F(\cdot)$ é a função de distribuição amostral construída com base no conjunto de valores de perdas observadas ou seja a distribuição de severidade.

Como já mencionamos anteriormente temos que $F^{*0}(\cdot)$ significa a função de distribuição que tem pulo de tamanho 1 no ponto 0. Isto porque ela corresponde à perda caso não ocorrer nenhum sinistro; está óbvio que não há perda neste caso.

Conforme nosso modelo (construído com base no Processo Pontual de Poisson) a função de distribuição da perda acumulada num ano é dada pela soma ponderada de convoluções de $F(\cdot)$; a saber:

$$\mathcal{C}(\cdot) = \sum_{k=0}^{\infty} \mathbb{P} [N(t) = k] F^{*k}(\cdot) \quad (3.4)$$

(a notação \mathcal{C} foi escolhida para assinalar que estamos falando da função Completa calculado em todo suporte da função ou seja $[0, \infty]$).

Nosso problema surge devido ao fato que não podemos somar infinitamente em (3.4). Para poder saber onde “parar” de somar, é precisa analisar a influência do erro devido à parada no valor de VaR. Para fazer esta análise, designamos por $k_\varepsilon - 1$ o tempo de parada, quer dizer,

o último membro da soma truncada, e designamos por $\mathcal{P}(\cdot)$ esta soma incompleta:

$$\mathcal{P}(\cdot) = \sum_{k=0}^{k_\epsilon-1} \mathbb{P}[N = k] F^{*k}(\cdot) \quad (3.5)$$

(a notação \mathcal{P} foi escolhida para assinalar que estamos falando da função *Parcial*).

Seja agora $\epsilon > 0$ um valor qualquer satisfazendo (3.3) e definiremos k_ϵ via:

$$\epsilon_k = \sum_{k=k_\epsilon}^{\infty} \mathbb{P}[N(t) = k] \quad (3.6)$$

Note a escolha de notações: k_ϵ e ϵ_k significam que um depende de outro. Deve-se notar que $\mathcal{P}(\cdot)$ não é uma função de distribuição. A razão está em que:

$$\sum_{k=0}^{k_\epsilon-1} \mathbb{P}[N(t) = k] = 1 - \epsilon_k < 1 \quad (3.7)$$

Isto faz com que \mathcal{P} torna-se uma soma ponderada de funções de distribuição com a soma dos pesos menor que 1. Daí que $\mathcal{P}(\cdot)$ não é uma função de distribuição. Para transformá-la numa função distribuição nos vamos reescalonar os pesos pelo mesmo fator para que a soma deles seja 1. É claro que o fator é $(1 - \epsilon_k)^{-1}$. Com os pesos reescalados, temos uma nova função a qual já é uma função de distribuição. Chamamos esta por $\mathcal{P}_r(\cdot)$ (*r* de reescalada). Eis a forma dela:

$$\mathcal{P}_r(\cdot) = \frac{1}{1 - \epsilon_k} \sum_{k=0}^{k_\epsilon-1} \mathbb{P}[N(t) = k] F^{*k}(\cdot) \quad (3.8)$$

Para nossa análise precisaremos de mais uma notação. Esta é:

$$\mathcal{A}(\cdot) = \sum_{k=k_\epsilon}^{\infty} \mathbb{P}[N(t) = k] F^{*k}(\cdot) \quad (3.9)$$

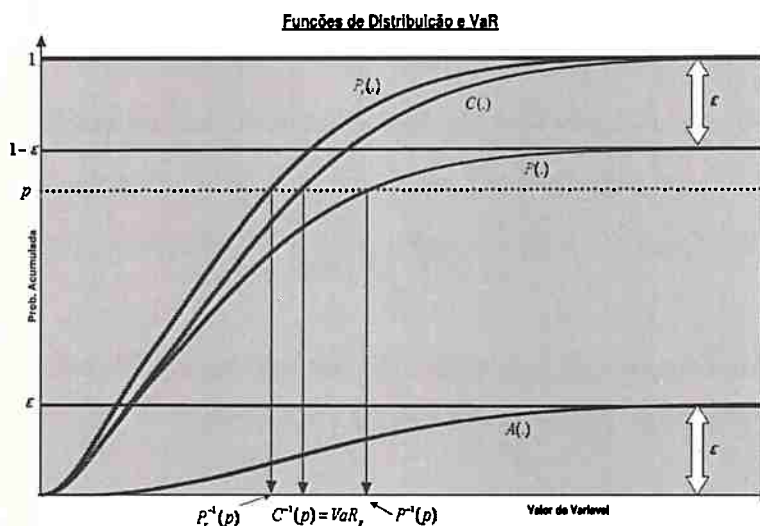


Figura 3.3: Procedimento para avaliar VaR.

Com esta notação, podemos re-escrever a fórmula (3.4) da seguinte maneira:

$$C(\cdot) = P(\cdot) + A(\cdot) \quad (3.10)$$

o que explica a escolha da notação A : esta lembra que a função é um Acréscimo à função P para completá-la até C .

Agora veja a figura(3.3). Nela desenhamos as funções $P(\cdot)$, $A(\cdot)$, $P_r(\cdot)$ e $C(\cdot)$. Marcamos também ε_k e p . Lembre-se que $p < 1 - \varepsilon_k$. Isto determina a posição de p no desenho. Notamos também um fato importante: a função $C(\cdot)$ está entre a função $P_r(\cdot)$ e $P(\cdot)$. Além disso chega-se a seguinte conclusão: VaR_p , o qual precisamos achar, é $C^{-1}(p)$ (p -quantil da função de distribuição $C(\cdot)$), e que este está entre $P_r^{-1}(p)$ e $P^{-1}(p)$. Note que a conclusão baseia-se no fato destacado acima.

A última conclusão gera o seguinte procedimento :

- 1 Fixar ε_k com base no valor de p .
- 2 Calcular as convoluções e simultaneamente acumular valor de

$$\sum_{k=0}^K \mathbb{P}[N = k] \quad (3.11)$$

- 3 Parar os cálculos do item 2 logo que a somatória (3.11) ultrapassar $1 - \varepsilon_k$.
- 4 Construir $\mathcal{P}(\cdot)$ e $\mathcal{P}_r(\cdot)$ e calcular $\mathcal{P}^{-1}(p)$ e $\mathcal{P}_r^{-1}(p)$.

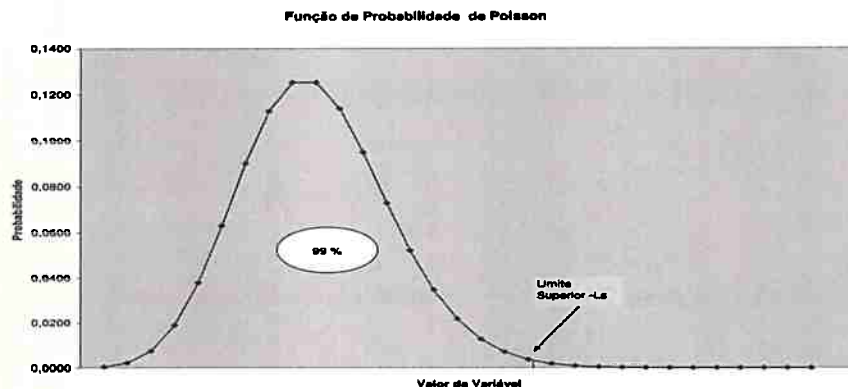
Caso estes dois valores não sejam muito distantes, poderemos emitir o resultado final: o verdadeiro VaR_p está entre estes valores. Caso contrário, é preciso continuar aumentando K da fórmula (3.11).

3.3 Operacionalizando o Modelo

Vamos agora estudar algumas características das duas variáveis aleatórias isoladamente e depois conjuntamente.

3.3.1 Número de ocorrências $N(t)$

É importante salientar que para a função de probabilidade de Poisson os valores de $\mathbb{P}(N = n)$ possuem valores significativos nas vizinhanças de λt considerado, e estabelecer limites para essa vizinhança é fundamental. Tendo em conta que estaremos sempre trabalhando com níveis de VaR em torno de 99% calcularemos esses limites como função de λt , o que nos ajudará a determinar a quantidade de convoluções necessárias para se atingir os níveis desejados quando formos determinar a distribuição agregada. A figura a seguir dá uma ideia clara do que desejamos:



Como podemos observar , no modelo de Poisson o ponto máximo de probabilidade ocorre para a variável assumindo o valor λt e vamos procurar a partir dele quantificar o número de convoluções necessárias para atingirmos os níveis de VaR desejados. Temos então resolver a equação abaixo uma vez λt fixado:

$$\sum_{k=0}^{Ls} e^{(-\lambda t)} \frac{(\lambda t)^k}{k!} \leq 0,99 \text{ onde } k = 1, 2, 3, \dots Ls \quad (3.12)$$

E também calcular o fator dado por:

$$Fator = \frac{Ls - \lambda t}{\sqrt{\lambda t}} \quad (3.13)$$

Os resultados dos cálculos dessa expressão estão apresentados na tabela a seguir.

Podemos então concluir que quase totalidade da probabilidade de interesse estará no intervalo de:

Tabela 3.1: Cálculo dos limites

λt	Limite Inferior	Limite Superior	Fator
1	0	4	3,00
2	0	6	2,83
3	0	8	2,89
4	0	9	2,50
5	0	11	2,68
6	0	12	2,45
10	2	18	2,53
20	9	31	2,46
30	17	43	2,37
50	33	67	2,40
70	50	90	2,39
100	76	124	2,40
200	167	234	2,40
300	259	341	2,37
400	353	447	2,35
500	448	553	2,37
600	543	658	2,37
700	638	762	2,34
800	734	870	2,37
900	830	971	2,37
1000	926	1074	2,34
Média	-	-	2,50

$$\lambda t - 2,5\sqrt{\lambda t} \leq \lambda t \leq \lambda t + 2,5\sqrt{\lambda t} \quad (3.14)$$

O que significa que para um valor de λt determinado serão necessárias um número de convoluções em torno de $\lambda t + 2,5\sqrt{\lambda t}$, para um Var de 99 %.

3.3.2 Severidade

Algumas distribuições são em geral mais usadas para a severidade. Entre as mais utilizadas temos:

- Distribuições de Pareto.
- Distribuição lognormal.
- Distribuição gama.
- Distribuição de perda Empírica.

Havendo a possibilidade de se trabalhar com um conjunto de informações disponíveis, que é o caso, podemos através de simulação obter a distribuição de perdas acumuladas no período considerado. No caso de optarmos por uma das outras distribuições na modelagem de dados, será necessário discretizá-la, para que possamos aplicar a metodologia discutida mais adiante.

3.3.3 Obtenção distribuição de Perdas- Método Analítico

Para algumas combinações de distribuições $N(t)$ e X_i a distribuição composta de perdas acumuladas pode ser avaliada através de resultados analíticos o que reduz consideravelmente os problemas computacionais. Vamos considerar o caso específico de quando a distribuição da severidade tem uma distribuição gama e em seguida uma distribuição exponencial. Lembrando que dado duas variáveis aleatórias com distribuições F_X e F_Y a convolução entre as duas, F_{X+Y} é dada por :

$$\begin{aligned}
 F_{X+Y} &= IP(X + Y \leq a) \\
 &= \iint_{X+Y \leq a} f_X(x) f_Y(y) dx dy \\
 &= \int_{-\infty}^{+\infty} \int_{-\infty}^{a-y} f_X(x) f_Y(y) dx dy
 \end{aligned}$$

$$\begin{aligned}
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{a-y} f_X(x) dx f_Y(y) dy \\
&= \int_{-\infty}^{+\infty} F_X(a-y) f_Y(y) dy
\end{aligned} \tag{3.15}$$

Para obtermos a a respectiva função de densidade basta diferenciarmos (3.15) e teremos:

$$\begin{aligned}
f_{X+Y}(a) &= \frac{d}{da} \int_{-\infty}^{+\infty} F_X(a-y) f_Y(y) dy \\
&= \int_{-\infty}^{+\infty} \frac{d}{da} F_X(a-y) f_Y(y) dy \\
&= \int_{-\infty}^{+\infty} f_X(a-y) f_Y(y) dy
\end{aligned} \tag{3.16}$$

Como a expressão da função de densidade de probabilidade da gama é dada por :

$$f(x) = \begin{cases} \frac{\lambda e^{-\lambda x} (\lambda x)^{t-1}}{\Gamma(t)}, & \text{para } x \geq 0, \\ 0 & \text{se } x < 0 \end{cases} \tag{3.17}$$

onde $\Gamma(t)$, chamada de função gama, é definida como $\Gamma(t) = \int_0^{\infty} e^{-y} y^{t-1} dy$

e aplicando-se o resultado de (3.16) sobre ela teremos:

$$\begin{aligned}
f_{X+Y}(a) &= \frac{1}{\Gamma(s)\Gamma(t)} \int_0^a \lambda e^{-\lambda(a-y)} [\lambda(a-y)]^{s-1} \lambda e^{-\lambda y} (\lambda y)^{t-1} dy \\
&= K e^{-\lambda a} \int_0^a (a-y)^{s-1} y^{t-1} dy
\end{aligned}$$

$$\begin{aligned}
&= Ke^{-\lambda a} a^{s+t-1} \int_0^1 (1-x)^{s-1} x^{t-1} dx \quad \text{fazendo } x = \frac{y}{a} \\
&= Ce^{-\lambda a} a^{s+t-1}
\end{aligned}$$

onde C é uma constante que não depende de a . E como a função calculada é uma função de densidade cuja integral resulta em 1, o valor de C é determinado e teremos finalmente:

$$f_{X+Y}(a) = \frac{\lambda e^{-\lambda a} (\lambda a)^{s+t-1}}{\Gamma(s+t)}$$

Em outras palavras se X e Y são variáveis aleatórias independentes com distribuição gama e respectivos parâmetros (s, λ) e (t, λ) , então $X + Y$ é uma variável aleatória gama com parâmetros $(s + t, \lambda)$.

Diante disso e trabalhando com o resultado de (3.15) podemos por indução, considerar que se $X_i, i = 1, \dots, n$ são variáveis aleatórias independentes gama com respectivos parâmetros $(t_i, \lambda), i = 1, \dots, n$, então $(\sum_{i=1}^n X_i)$ também é gama com parâmetros $(\sum_{i=1}^n t_i, \lambda)$.

Então, fazendo $t = \beta$, temos:

$$F^{*n}(x) = \Gamma(\beta t, \lambda) \quad (3.18)$$

Como a variável aleatória gama com parâmetro t igual a 1 é um caso particular e resulta na distribuição exponencial com parâmetro λ , nós podemos afirmar que se $X_i, i = 1, \dots, n$ são variáveis aleatórias independentes com distribuição exponencial então $(\sum_{i=1}^n X_i)$ resulta em gama com parâmetros (n, λ) .

Com isso exposto, podemos então rescrever a equação (2.3) com os resultados da convolução de distribuições gama (3.18) ou exponenciais da seguinte forma:

$$\begin{aligned}
G_i(x) &= P(S(t) \leq x) \\
&= \sum_{i=0}^{\infty} P[S(t) \leq x/N(t) = i]P[N(t) = i] \\
&= \sum_{i=0}^{\infty} e^{-(\lambda t)} \frac{(\lambda t)^i}{i!} \left(\Gamma(n\beta, \lambda) \right), \quad x \geq 0, \quad t \geq 0
\end{aligned} \tag{3.19}$$

3.3.4 Obtenção distribuição de Perdas- Métodos Computacionais

- Força Bruta- Esse método consiste em se elaborar um programa onde se executa as convoluções necessárias, cuja quantidade foi dimensionada em (3.14), sucessivamente, até o nível de VaR desejado. Esse processo pode se tornar bastante trabalhoso e custoso, dependendo do valor de λt , considerado pois a quantidade de operações necessárias crescem exponencialmente. A nossa experiência com este método não foi muito satisfatória.
- Simulação via Métodos de Monte Carlo- Um rotina para simulação via esse método pode ser implementada com razoável facilidade em linguagem R e os resultados são obtidos com uma redução de tempo de processamento bastante significativa, o que nos fez optar em grande parte do nosso trabalho por esse método.
- Transformação rápida de Fourier(FFT)- esse método combina as propriedades da função característica de variáveis aleatórias e rotinas computacionais desenvolvidas em varias linguagens, como também na linguagem R.
- Método Recursivo de Panjer- Neste método, as convoluções de variáveis aleatórias é feito de maneira recursiva e é baseado em propriedades recursivas próprias das distribuições discretas de Poisson e Binomial Negativa consideradas. Daremos um tratamento mais detalhado deste método a partir de agora.

Como já foi mencionado, as distribuições de Poisson e Binomial Negativa apresentam a seguinte característica:

$$p_n = \left(a + \frac{b}{n}\right)p_{n-1} \quad \text{para } n = 1, 2, 3, \dots \quad (3.20)$$

onde:

p_n é igual a probabilidade de exatamente n reclamações em um determinado intervalo de tempo e a e b são constantes calculadas para cada situação.

Desse modo no caso da distribuição de Poisson temos:

- 1) $p_n = e^{-\lambda} \frac{\lambda^n}{n!}$ para $n = 0, 1, 2, \dots$
- 2) $p_0 = e^{-\lambda}$
- 3) $\frac{p_n}{p_{n-1}} = \frac{\lambda}{n}$ ou $p_n = p_{n-1} \frac{\lambda}{n}$
- 4) $a = 0$ e $b = \lambda$

Para a distribuição Binomial negativa teremos:

- 1) $p_n = \binom{\alpha+n-1}{n} p^n (1-p)^\alpha$ para $n = 0, 1, 2, 3, \dots$
- 2) $p_0 = (1-p)^\alpha$
- 3) $\frac{p_n}{p_{n-1}} = \frac{(\alpha+n-1)p}{n}$
- 4) $a = p$ e $b = \frac{(\alpha-1)}{p}$

Uma vez estabelecida essas considerações aplicando o método recursivo de Panjer a a uma distribuição agregada temos:

$$\begin{aligned} G_t(x) &= P(S(t) \leq x) \\ &= \sum_{n=0}^{\infty} p_n F^{n*}(x), \quad x \geq 0, \quad t \geq 0 \\ &= p_0 \quad \quad \quad x = 0 \end{aligned}$$

e sua respectiva função de probabilidade:

$$\begin{aligned} g_t(x) &= P(S(t) = x) \\ &= \sum_{n=0}^{\infty} p_n f^{n*}(x), \quad x \geq 0, \quad t \geq 0 \\ &= p_0 \quad \quad \quad x = 0 \end{aligned}$$

fica assim definido:

$$g_r = P(S_N = r) = \frac{1}{1 - a f_0} \sum_{i=1}^r \left(a + \frac{bi}{r} \right) f_i g_{r-i},$$

onde:

$$\begin{aligned} S_N &= X_1 + X_2 + X_3 + \dots + X_N \\ p_k &= P(N = k) = \left(a + \frac{b}{r} \right) p_{k-1}, \quad k = 0, 1, 2, \dots \\ f_k &= P(X_i = k), \quad k = 1, 2, 3, \dots \\ f_0 &= P(X_i = 0) = 0 \\ g_0 &= P(S_N = 0) = p_0 = P(N = 0) \\ g_1 &= P(S_N = 1), \quad g_0, g_1, g_2, \dots, g_{r-1}, g_r \end{aligned}$$

O que pode ser escrito então da seguinte forma:

$$g_0 = P(S_N = 0) = p_0$$

$$\begin{aligned}
g_1 &= P(S_N = 1) = \frac{1}{1 - af_0} \sum_{i=1}^1 \left(a + \frac{bi}{1}\right) f_i g_{1-i}, \\
g_2 &= P(S_N = 2) = \frac{1}{1 - af_0} \sum_{i=1}^2 \left(a + \frac{bi}{2}\right) f_i g_{2-i}, \\
g_3 &= P(S_N = 3) = \frac{1}{1 - af_0} \sum_{i=1}^3 \left(a + \frac{bi}{3}\right) f_i g_{3-i}, \\
g_4 &= P(S_N = 4) = \frac{1}{1 - af_0} \sum_{i=1}^4 \left(a + \frac{bi}{4}\right) f_i g_{4-i}, \\
&\vdots \qquad \qquad \qquad \vdots \qquad \qquad \qquad \vdots \\
g_r &= P(S_N = r) = \frac{1}{1 - af_0} \sum_{i=1}^r \left(a + \frac{bi}{r}\right) f_i g_{r-i},
\end{aligned}$$

Para demonstrarmos a validade da fórmula acima vamos primeiramente recordar o conceito de funções geratrizes e suas propriedades:

– Função geratriz de probabilidades:

$$P(z) = P_N(z) = E(z^N) = \sum_{k=0}^{\infty} p_k z^k$$

– Função geratriz de momentos:

$$M_x(t) = E(e^{tX}) = \sum_{k=1}^{\infty} e^{X_j} p(x_j)$$

– E as seguintes propriedades:

$$M_X(t) = P_X(e^t)$$

$$P_X(z) = M_X(\ln z)$$

Considerando então que a distribuição primária do processo agregado pode ser calculada recursivamente faremos:

$$\begin{aligned} p_n &= \left(a + \frac{b}{n}\right)p_{n-1} \\ np_n &= a(n-1)p_{n-1} + (a+b)p_{n-1} \end{aligned}$$

Multiplicando cada lado da igualdade por $[P_X(z)]^{n-1}P'_X(z)$ e somando sobre n resulta em:

$$\begin{aligned} \sum_{n=1}^{\infty} np_n[P_X(z)]^{n-1}P'_X(z) &= a \sum_{n=1}^{\infty} (n-1)p_{n-1}[P_X(z)]^{n-1}P'_X(z) + \\ &+ (a+b) \sum_{n=1}^{\infty} p_{n-1}[P_X(z)]^{n-1}P'_X(z) \end{aligned}$$

Como $P_S(z) = \sum_{n=0}^{\infty} p_n[P_X(z)]^n$ a equação anterior é

$$P'_S(z) = a \sum_{n=1}^{\infty} np_n[P_X(z)]^n P'_X(z) + (a+b) \sum_{n=1}^{\infty} p_n[P_X(z)]^n P'_X(z)$$

Portanto

$$P'_S(z) = aP'_S(z)P_X(z) + (a+b)P_S(z)P'_X(z)$$

Como cada lado pode ser expresso em potencias de z e os coeficientes de z^{k-1} nesta expansão, obrigatoriamente são os mesmos nos dois lados da equação então fica:

$$\begin{aligned} kg_k &= a \sum_{j=0}^k (k-j)f_j g_{k-j} + (a+b) \sum_{j=0}^k j f_j g_{k-j} \\ &= akf_0 g_k + a \sum_{j=1}^k (k-j)f_j g_{k-j} + (a+b) \sum_{j=1}^k j f_j g_{k-j} \\ &= akf_0 g_k + ak \sum_{j=1}^k (f_j g_{k-j}) + b \sum_{j=1}^k j f_j g_{k-j} \end{aligned}$$

E finalmente,

$$g_k = a f_0 g_k + \sum_{j=1}^k \left(a + \frac{b_j}{k} \right) f_j g_{k-j}$$

Julgamos importante salientar que para a aplicação do método recursivo de Panjer é necessário se trabalhar com uma unidade monetária h conveniente(R\$ 1000,00 por exemplo), para que a variável discreta possua incrementos de h em h , seja no caso de trabalharmos com a distribuição empírica ou a contínua discretizada.

3.4 Teoria dos Valores Extremos

A teoria de valores extremos surgiu em 1928 com os trabalhos de Fisher e Tippett quando se estudava o comportamento do máximo de variáveis aleatórias independentes e identicamente distribuídas. A partir daí várias aplicações foram implementadas com sucesso em muitos campos tais como: ciências atuariais, engenharia, hidrologia, economia e finanças. Dentro dessa teoria ,o método POT- Peaks Over Threshold surgido na década de 80,, tornou-se uma das ferramentas estatísticas mais importantes na estimativa da probabilidade de eventos raros e, em particular, na estimativa de risco de mercado, risco de crédito e de risco operacional. Neste trabalho procuraremos abordar as principais ideias dessa ferramenta.

3.4.1 Teoria necessária para justificar o método POT

Considerando que X_1, X_2, \dots, X_n são variáveis aleatórias independentes e identicamente distribuídas oriundas de um processo estocástico $\{X_i\}_{i \in \mathbb{N}}$ com distribuição $F(\cdot)$ desconhecida temos interesse em obter a distribuição de $M_n = \max(X_1, \dots, X_n)$ e conseqüentemente poder avaliar:

$$\mathbb{P} [\max\{X_1, \dots, X_n\} \leq M]$$

Podemos determinar a distribuição de M_n facilmente fazendo:

$$\begin{aligned}
F_{M_n}(M) &= \mathbb{P}[\max\{X_1, \dots, X_n\} \leq M] \\
&= \mathbb{P}[X_1 \leq M, \dots, X_n \leq M] \\
&= \{\mathbb{P}[X \leq M]\}^n \\
&= \{F_X(M)\}^n
\end{aligned} \tag{3.21}$$

Uma vez que $F(\cdot)$ não é conhecida devemos trabalhar com resultados assintóticos para a distribuição de M_n .

Para isso vamos nos apoiar no teorema de Fisher e Tippett que segue abaixo.

Teorema Fisher e Tippett- Considere $\{X_n\}$ uma sequência de variáveis aleatórias independentes e identicamente distribuídas com uma função de distribuição $F(\cdot)$. Se existirem sequências numéricas c_n e $d_n \in \mathbb{R}$ e uma distribuição H não degenerada de forma que:

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{M_n - d_n}{c_n} \leq x\right) = H(x)$$

ou seja

$$\frac{M_n - d_n}{c_n} \xrightarrow{d} H \text{ quando } n \rightarrow \infty,$$

podemos afirmar então que no limite $H(x)$ só pode pertencer a uma das três funções abaixo:

a família de Fréchet constitui-se das distribuições $\Phi_\alpha(\cdot)$ com $\alpha > 0$, onde:

$$\Phi_\alpha(x) = \begin{cases} 0, & \text{se } x \leq 0, \\ \exp\{-x^{-\alpha}\}, & \text{se } x > 0; \end{cases}$$

a família de Weibull constitui-se das distribuições $\Psi_\alpha(\cdot)$ com $\alpha > 0$, onde:

$$\Psi_\alpha(x) = \begin{cases} \exp\{-[(-x)^\alpha]\}, & \text{se } x \leq 0, \\ 1, & \text{se } x > 0; \end{cases}$$

a distribuição de Gumbel é definida por:

$$\Lambda(x) = \exp\{-e^{-x}\}, \text{ para todo } x \in \mathfrak{R}.$$

Sabemos que o Teorema do Limite Central -TLC, uma das mais importantes ferramentas na teoria das probabilidades e estatística, estabelece que a distribuição normal é a única distribuição resultante da soma de variáveis aleatórias independentes e identicamente distribuídas (sob condições de regularidade). De modo análogo a teoria de valores extremos estabelece, para a distribuição do máximo no limite, as três únicas possibilidades. Uma breve comparação desses resultados é mostrada na tabela (3.2), a seguir:

Analogia	Teorema do Limite Central	Teoria dos Valores Extremos
Condições de X_n	iid com segundo momento finito	iid e df tem variação regular
Objeto de Estudo	S_n (Soma)	M_n (Máximo)
Distribuição Limite	Normal	EVD

Tabela 3.2: Comparativo TLC e EVT

Se introduzirmos os parâmetros de locação (b) e escala (a) as funções de distribuição de valores extremos (EVD's- *extreme value distributions*) ficam então:

$$\text{Fréchet} : \Phi_\alpha\left(\frac{x-b}{a}\right), x \in \mathfrak{R}, \text{ para } \alpha > 0, a > 0, b \in \mathfrak{R};$$

$$\text{Weibull} : \Psi_\alpha\left(\frac{x-b}{a}\right), x \in \mathfrak{R}, \text{ para } \alpha > 0, a > 0, b \in \mathfrak{R}; \quad (3.22)$$

$$\text{Gumbel} : \Lambda\left(\frac{x-b}{a}\right), x \in \mathfrak{R}, \text{ para } a > 0, b \in \mathfrak{R}.$$

Através de uma reparametrização, onde fazemos $\xi = \frac{1}{\alpha}$ podemos observar que as EVD's descritas acima podem ser membros de uma família única de distribuições. A essa família denominamos como distribuição de valores extremos generalizada (GEVD-Generalized extreme

value distribution), com a seguinte expressão:

$$H_{\xi}(x) = \begin{cases} \exp \left\{ -(1 + \xi x)^{-1/\xi} \right\}, & \text{se } \xi \neq 0, \\ \exp \{ -\exp(-x) \}, & \text{se } \xi = 0 \end{cases} \quad (3.23)$$

sendo que esta expressão vale para

$$\begin{aligned} x &> -\xi^{-1} && \text{quando } \xi > 0, \\ x &< -\xi^{-1} && \text{quando } \xi < 0, \\ x &\in \mathfrak{R} && \text{quando } \xi = 0, \\ &1 + \xi x > 0. && \end{aligned} \quad (3.24)$$

Podemos observar que se $\xi = 0$, H_{ξ} corresponde à distribuição de Gumbel. E sua expressão é suportada pelo resultado do cálculo onde temos que:

$$\lim_{c \rightarrow \infty} (1 - cx)^{\frac{1}{c}} = e^x$$

Para os casos de $\xi < 0$ e $\xi > 0$ temos respectivamente a correspondência com as distribuições de Weibull e Fréchet.

3.4.2 Família de distribuições de Pareto Generalizadas

A GPD padrão com valor $\xi > 0$, tem a seguinte expressão:

$$G_{\xi}(x) = \begin{cases} 0, & \text{para } x < 0, \\ 1 - (1 + \xi x)^{-1/\xi}, & \text{para } x \geq 0. \end{cases} \left(\begin{array}{l} \text{esta fórmula} \\ \text{vale quando } \xi > 0 \end{array} \right) \quad (3.25)$$

A GPD padrão com valor $\xi = 0$, tem a seguinte expressão:

$$G_0(x) = \begin{cases} 0, & \text{para } x < 0, \\ 1 - e^{-x}, & \text{para } x \geq 0. \end{cases} \left(\begin{array}{l} \text{esta função} \\ \text{corresponde a } \xi = 0 \end{array} \right) \quad (3.26)$$

A GPD padrão com valor $\xi < 0$, tem a seguinte expressão:

$$G_{\xi}(x) = \begin{cases} 0, & \text{para } x < 0, \\ 1 - (1 + \xi x)^{-1/\xi}, & \text{para } x \in [0, -\frac{1}{\xi}], \\ 1, & \text{para } x > -\frac{1}{\xi}. \end{cases} \left(\begin{array}{l} \text{esta fórmula} \\ \text{vale quando} \\ \xi < 0 \end{array} \right) \quad (3.27)$$

Para tornar essas funções mais flexíveis vamos dotá-las com dois novos fatores , um de localização(s) e outro de escala(β) através de uma transformação de variáveis. Seja então a variável aleatória $Y = \beta X + s$, onde β e s são duas constantes arbitrárias e $\beta > 0$. Considerando $F_X(\cdot)$ e $F_Y(\cdot)$ as respectivas funções de distribuição de X e Y segue então dessa transformação que:

$$\begin{aligned} F_Y(x) &= \mathbb{P}[Y \leq x] = \mathbb{P}[\beta X + s \leq x] = \mathbb{P}\left[X \leq \frac{x-s}{\beta}\right] = \\ &= F_X\left(\frac{x-s}{\beta}\right), \quad \forall x \in \mathfrak{R}. \end{aligned} \quad (3.28)$$

Dessa forma podemos reescrever as distribuições Padrão de Pareto Generalizadas que com essa transformação passam a ser chamadas agora de distribuições de Pareto Generalizadas, (abreviadas por *GPD's*). Neste caso a distinção entre as GPD's passa a ser feita por três parâmetros, ξ , β e s , sendo que ξ e s podem assumir qualquer valor real e β somente valores reais positivos.

A GPD com $\xi > 0$ fica então com a seguinte expressão:

$$G_{\xi,\beta,s}(x) = \begin{cases} 1 - \left(1 + \xi \frac{x-s}{\beta}\right)^{-1/\xi}, & \text{para } x \geq s, \\ 0, & \text{para } x < s, \end{cases} \left(\begin{array}{l} \text{esta vale} \\ \text{quando } \xi > 0 \end{array} \right) \quad (3.29)$$

enquanto que a GPD com $\xi = 0$ é dada por:

$$G_{0,\beta,s}(x) = \begin{cases} 1 - e^{-\frac{x-s}{\beta}}, & \text{para } x \geq s, \\ 0, & \text{para } x < s, \end{cases} \left(\begin{array}{l} \text{esta} \\ \text{corresponde a } \xi = 0 \end{array} \right) \quad (3.30)$$

finalmente a GPD que corresponde a $\xi < 0$:

$$G_{\xi,\beta,s}(x) = \begin{cases} 0, & \text{para } x < s, \\ 1 - (1 + \xi \frac{x-s}{\beta})^{-1/\xi}, & \text{para } x \in [s, s - \frac{\beta}{\xi}], \\ 1, & \text{para } x > s - \frac{\beta}{\xi}, \end{cases} \left(\begin{array}{l} \text{esta fórmula} \\ \text{vale quando} \\ \xi < 0 \end{array} \right) \quad (3.31)$$

3.4.3 Interligação entre as EVD's e GPD's

Existe uma relação biunívoca entre as EVD's e as GPD's. A cada EVD corresponde um conjunto de GPD's e essa correspondência se dá pela seguinte fórmula:

$$H_{\xi}(x) = \exp\{-(1 - G_{\xi}(x))\} \quad \text{ou seja:}$$

Função de Distribuição EVD = $\exp\{-(\text{a função de sobrevivência de GPD})\}$

Podemos então aplicar a fórmula acima para os três tipos de funções que estamos analisando que resultam então em:

$$H_{\xi}(x) = \begin{cases} \exp\{-(1 - G_{\xi}(x))\}, & \text{para } x > -\xi^{-1} & \text{quando } \xi > 0, \\ \exp\{-(1 - G_{\xi}(x))\}, & \text{para } x \in [0, -\xi^{-1}] & \text{quando } \xi < 0, \\ \exp\{-(1 - G_0(x))\}, & \text{para } x > 0 & \text{quando } \xi = 0, \end{cases} \quad (3.32)$$

que simplificando também podem ser expressas como:

$$H_{\xi}(x) = \begin{cases} \exp\{(G_{\xi}(x) - 1)\}, & \text{para } x > -\xi^{-1} & \text{quando } \xi > 0, \\ \exp\{(G_{\xi}(x) - 1)\}, & \text{para } x \in [0, -\xi^{-1}] & \text{quando } \xi < 0, \\ \exp\{(G_0(x) - 1)\}, & \text{para } x > 0 & \text{quando } \xi = 0, \end{cases} \quad (3.33)$$

3.4.4 O Método POT- Peaks Over Threshold- Aspectos de Implementação.

O método POT consiste em uma abordagem específica para o problema de estimação da cauda. E a família de GPD's passa a ter uma importância fundamental neste método pois será com através delas que efetuaremos todo o processo de estimação. Como o próprio nome sugere, estamos interessados em estimar a cauda de uma função de distribuição $F_u(\cdot)$ acima de um determinado limiar u e portanto em obter uma distribuição $F_u(x)$ com a seguinte característica :

$$F_u(x) = \mathbb{P}(X - u \leq x | X > u) \quad x \in \mathfrak{R}$$

Que com as propriedades da fórmula da distribuição condicional pode ser reescrita com:

$$\begin{aligned} F_u(x) &= \mathbb{P}[X - u \leq x | X > u] = \frac{\mathbb{P}[X - u \leq x, X > u]}{\mathbb{P}[X > u]} \\ &= \begin{cases} 0, & \text{se } x < 0 \\ \frac{\mathbb{P}[u < X \leq u+x]}{\mathbb{P}[X > u]}, & \text{se } x \geq 0 \end{cases} = \begin{cases} 0, & \text{para } x < 0 \\ \frac{F(x+u) - F(u)}{1 - F(u)}, & \text{para } x \geq 0 \end{cases} \end{aligned}$$

Que resulta finalmente

$$F_u(x) = \begin{cases} 0, & \text{para } x < 0 \\ \frac{F(x+u) - F(u)}{1 - F(u)}, & \text{para } x \geq 0 \end{cases} \quad (3.34)$$

Ressaltamos ainda que, a função $F_u(\cdot)$ é denominada *função de distribuição de excessos* da distribuição $F(\cdot)$ *acima de um limiar u* , nome motivado pelo fato de a diferença $(X - u)$ ser usualmente chamada de *excesso* da variável aleatória X *acima do limiar u* , quando X ultrapassa u . Acrescentamos ainda, para completar a nomenclatura, que o valor da variável aleatória X que excede u chama-se *excedente de X acima de u* .

Para executarmos a metodologia do POT precisamos ainda definir o conceito de função da média de excessos. Temos então a função $e(\cdot)$ definida da seguinte forma:

$$e(u) := \mathbb{E}[X - u | X > u], \quad u < (\text{o valor máximo de } X), \quad (3.35)$$

chama-se *função da média de excessos de X*, ou, *função da média de excessos da função F(·)*, se esta última é a função de distribuição de X.

Temos aqui então a esperança matemática da variável aleatória $(X - u)$ condicionada à ocorrência do evento $\{X > u\}$, ou ainda, em linguagem matemática

$$\mathbb{E}[X - u | X > u] = \int_{-\infty}^{\infty} x dF_u(x), \quad (3.36)$$

onde $F_u(\cdot)$ designa a função da distribuição condicional de $(X - u)$ dado que ocorreu $\{X > u\}$. Vimos que esta função possui a forma:

$$F_u(x) = \mathbb{P}[X - u \leq x | X > u]$$

e que a sua respectiva função de densidade $f_u(x)$ é dada pela seguinte fórmula:

$$f_u(x) = \begin{cases} 0, & \text{se } x < 0, \\ \frac{f(x+u)}{1-F(u)}, & \text{se } x \geq 0. \end{cases} \quad (3.37)$$

E que a combinação de (3.36) e (3.37), resultam em:

$$e(u) = \int_0^{\infty} \frac{xf(x+u)}{1-F(u)} dx, \quad (3.38)$$

Temos ainda que apresentar a função da media amostral de excessos de uma amostra. Seja então uma amostra x_1, \dots, x_n , definimos $\hat{e}(u)$ como:

$$\begin{aligned}\hat{e}(u) &= \frac{1}{n} \sum_{i=1}^n (x_i - u) \\ &= \left(\frac{1}{n} \sum_{i=1}^n x_i \right) - u, \text{ para } u < \min\{x_1, \dots, x_n\},\end{aligned}\tag{3.39}$$

$$\hat{e}(u) = 0, \text{ para } u \geq \max\{x_1, \dots, x_n\}.\tag{3.40}$$

Para finalizarmos esta parte resta algumas considerações a respeito da identificação do valor u que seja o limiar ótimo de u que chamaremos de u^{opt} .

Com base nos dados da amostra devemos proceder da seguinte maneira:

1. Construir a função média de excessos para todo o intervalo de $\min(x_1, \dots, x_n) < u < \max(x_1, \dots, x_n)$, e construir o gráfico correspondente.
2. Fazendo a análise do gráfico procura-se regiões que apresentem comportamento linear crescente ou decrescente ou mesmo seja horizontal.
3. Percorrendo o gráfico da direita à esquerda devemos procurar a região que não seja muito irregular e que apresente comportamento aproximadamente linear. Declaramos então o valor do extremo esquerdo desta região como o valor ótimo do limiar.
4. Como estaremos trabalhando com dados da amostra e normalmente teremos valores discretos é importante ressaltar que pela própria definição da média de excessos consideramos sempre valores $> u$, então o valor de u não é contabilizado nos cálculos. Ordenando a amostra em ordem crescente e considerando que o valor de u^{opt} escolhido ocupe a k -ésima posição podemos então expressar a função da média amostral de excessos como:

$$\hat{e}(u) = \left(\frac{1}{n-k} \sum_{i=k+1}^n x_{(i)} \right) - u, \text{ para } u \in [x_{(k)}, x_{(k+1)}),\tag{3.41}$$

para qualquer k de 1 a $n-1$.

Capítulo 4

Aplicação do método para um caso real

No nosso objeto de estudo, temos que obter previsões para o volume de capital necessário para dar cobertura aos eventuais sinistros que ocorrem no período de um mês com um nível de probabilidade α especificado. Inicialmente vamos fazer uma análise exploratória dos dados coletados e em seguida conjuntamente com os conceitos teóricos apresentados no capítulo 2 obter essas previsões. Vamos recorrer ao que já foi apresentado e acrescentarmos o que for necessário para atingirmos este objetivo.

4.1 Distribuição do número de Ocorrências

Os dados referentes ao número de ocorrência dos sinistros estão apresentados a seguir.

4.1.1 Ocorrências Mensais

O número de ocorrências dos dados coletados foram distribuídos mensalmente, como mostra a tabela (4.1) abaixo e abrange um período de 28 meses entre os anos de 2004 e 2007.

Ano	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
2004	-	-	-	-	-	-	-	-	-	-	-	23
2005	17	16	19	28	19	17	17	14	14	15	16	25
2006	18	15	26	21	19	20	20	27	17	23	27	21
2007	16	13	12	-	-	-	-	-	-	-	-	-

Tabela 4.1: É a variável aleatória Contadora do Número de Ocorrências durante o mês.

4.1.2 Dados agrupados

A tabela (4.2) apresenta os dados agrupados e ordenados em ordem crescente de ocorrência.

Nr.Ocorr.	12	13	14	15	16	17	18	19	20	21	23	25	26	27	28
Freq.	1	1	2	2	3	4	1	3	2	2	2	1	1	2	1

Tabela 4.2: Frequência dos Sinistros durante o período.

4.1.3 Histograma

No histograma a seguir, figura (4.1) observamos classes com frequência variando de 2 a 5 e que também a média esta contida na classe com maior frequência.

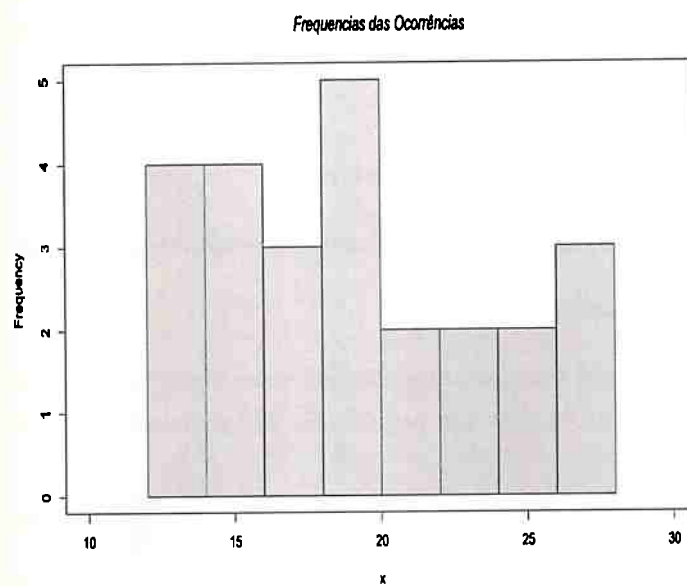


Figura 4.1: Histograma de Frequência do Número de Ocorrências

É a sua respectiva função de densidade de probabilidade que é apresentada na figura(4.2).

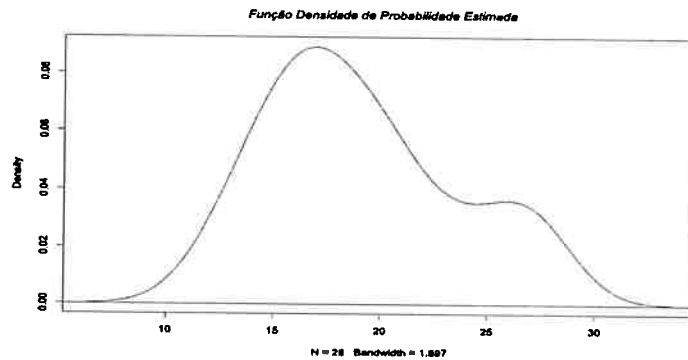


Figura 4.2: gráfico da função estimada da densidade de probabilidade

4.1.4 Principais Estatísticas

As principais estatísticas são apresentadas na tabela (4.3), e foram colocadas na tabela também os dois primeiros momentos, que serão usados para estimar os parâmetros das funções de distribuição que serão utilizadas no desenvolvimento do trabalho.

Descrição	Valor
Numero de meses observados	28
Media de ocorrências no mês	19
Moda de ocorrências no mês	17
Variância	20,25
Desvio padrão	4,50
Mínimo	12
Máximo	28
Primeiro Momento $E[N]$	19,00
Segundo Momento $E[N^2]$	381,25

Tabela 4.3: Principais Estatísticas Variável Contadora

4.2 Distribuição da Severidade

Os dados referentes aos valores dos sinistros estão apresentados a seguir.

4.2.1 Severidade distribuída no período-1

A figura a seguir(4.3), apresenta os sinistros ocorridos durante o período de janeiro a junho de 2005 a 2007.Os sinistros estão ordenados por ordem crescente de valor dentro do mês/ano.

Sinal		+														
Valor	Mês	Ano														
		1			2			3			4		5		6	
Ocorr		2005	2006	2007	2005	2006	2007	2005	2006	2007	2005	2006	2005	2006	2005	2006
1	50	50	300	200	400	300	50	50	900	50	100	200	50	100	200	200
2	2.600	50	1.000	400	600	500	200	100	1.000	50	200	200	100	200	500	200
3	2.800	100	1.200	800	800	800	200	100	1.700	100	200	1.100	100	500	200	200
4	3.200	100	1.400	800	2.600	800	200	200	11.800	100	300	1.400	100	500	200	200
5	4.700	500	2.000	1.600	3.200	1.000	200	400	12.900	200	300	2.000	700	800	300	300
6	4.800	700	2.700	1.800	13.300	3.600	500	500	16.400	200	500	2.100	700	1.000	600	600
7	5.000	1.100	3.200	2.300	17.400	35.600	600	1.200	23.500	200	1.100	4.700	1.000	1.600	800	800
8	6.000	1.700	3.600	2.900	19.900	42.100	800	1.600	70.200	200	1.500	4.800	1.800	2.700	900	900
9	6.300	4.800	4.200	7.700	36.400	110.600	1.500	1.700	501.100	300	1.700	5.100	1.900	6.300	1.100	1.100
10	7.200	7.600	9.300	17.600	50.000	434.200	3.700	1.900	1.064.600	400	1.900	7.900	7.400	8.400	2.100	2.100
11	18.700	7.900	33.500	56.700	142.300	737.500	4.000	2.500	3.536.200	600	1.900	11.000	12.800	10.500	2.300	2.300
12	33.200	21.900	119.400	108.800	510.900	2.704.400	8.200	2.500	5.177.100	800	3.000	11.200	29.500	38.300	4.600	4.600
13	87.000	136.600	445.100	480.200	603.200	5.649.300	14.100	2.500		900	6.000	27.800	37.300	99.800	14.500	14.500
14	656.200	151.000	759.300	770.200	2.352.900		30.400	2.600		1.200	8.800	34.100	74.800	555.400	27.700	27.700
15	718.400	556.000	3.573.800	2.437.700	4.812.500		123.900	3.500		1.300	14.700	100.400	151.000	935.100	75.600	75.600
16	2.718.000	758.700	4.924.400	4.406.300			523.500	4.200		1.400	37.900	528.600	372.100	2.295.700	89.400	89.400
17	4.999.700	2.756.200					985.300	6.500		1.600	111.100	1.003.200	933.500	5.063.300	592.000	592.000
18		4.592.600					2.379.500	14.000		3.600	685.600	2.530.400	3.029.300		874.100	874.100
19							4.263.700	25.800		4.100	729.800	5.201.500	5.249.600		2.663.300	2.663.300
20								34.500			8.600	2.879.500			5.415.900	5.415.900
21								141.200			11.600	5.865.600				
22								157.200			18.300					
23								470.400			25.100					
24								951.800			55.100					
25								2.808.200			507.100					
26								4.235.500			792.300					
27											1.857.700					
28											4.983.700					

Figura 4.3: Valores dos Sinistros por mês-primeira parte.

4.2. DISTRIBUIÇÃO DA SEVERIDADE

59

4.2.2 Severidade distribuída no período-2

A figura a seguir(4.4) apresenta os sinistros ocorridos durante o período de julho a dezembro de 2004 a 2006. Os sinistros estão ordenados por ordem crescente de valor dentro do mês/ano.

Valor	Mês		Ano											
	7		8		9		10		11		12			
Ocorr	2005	2006	2005	2006	2005	2006	2005	2006	2005	2006	2004	2005	2006	
1	300	50	50	700	100	50	100	50	50	50	100	100	50	
2	1.000	50	200	900	100	50	100	100	100	50	200	100	100	
3	1.000	50	600	2.100	900	200	200	600	200	100	500	200	300	
4	2.400	100	1.000	2.800	1.300	300	300	600	1.400	200	600	400	600	
5	3.700	200	2.000	2.800	1.400	400	800	700	2.300	300	800	400	1.100	
6	4.400	700	5.100	3.300	3.600	800	1.900	1.200	3.600	300	1.700	500	1.400	
7	7.000	1.100	7.700	3.300	13.200	1.600	2.100	3.000	3.800	400	2.700	600	1.700	
8	7.600	1.300	15.200	3.900	14.600	2.200	3.200	3.400	5.900	400	7.500	800	1.800	
9	8.500	2.900	56.400	5.000	40.900	2.700	11.000	4.300	7.800	1.300	7.500	1.200	2.700	
10	9.800	2.900	133.800	5.000	95.000	4.400	63.100	4.400	12.400	1.600	8.900	1.500	2.700	
11	10.800	3.400	429.300	5.800	334.100	4.700	67.700	4.800	74.400	1.900	12.500	2.600	3.800	
12	73.100	4.900	1.171.900	6.700	849.900	23.500	375.100	6.400	102.600	3.900	12.900	2.800	5.400	
13	229.600	5.600	2.339.700	6.700	2.709.600	88.900	936.700	6.800	505.400	8.700	13.700	3.700	7.100	
14	421.800	41.800	5.881.500	8.700	5.811.300	323.600	2.970.800	7.100	952.600	10.700	22.900	5.700	10.000	
15	981.900	44.800		8.700		780.900	5.108.200	8.000	2.752.400	13.500	24.100	6.400	10.800	
16	2.507.200	127.100		10.000		2.374.500		8.700	5.578.600	19.200	25.400	7.600	31.600	
17	4.938.400	540.200		10.300		6.072.700		22.600		21.400	32.400	8.600	155.600	
18		927.000		11.200				28.600		23.400	44.500	14.300	776.200	
19		3.218.600		11.500				124.200		29.500	187.900	17.800	810.600	
20		5.165.300		11.500				438.400		30.200	387.500	26.300	3.210.200	
21				17.700				849.100		47.000	1.163.800	62.900	5.132.000	
22				38.400				3.012.900		86.000	2.425.700	529.300		
23				110.800				5.232.400		154.100	4.664.000	923.600		
24				553.300						616.300		2.685.800		
25				1.089.100						839.200		4.870.800		
26				3.103.900						2.760.300				
27				5.582.300						5.564.100				

Figura 4.4: Valores dos Sinistros por mês-segunda parte.

4.2.3 Principais estatísticas da Severidade

Nesta figura (4.5) são apresentadas as principais estatísticas calculadas dentro do ano/mês.

Sinal +		Mês											
Ano	Dados	1	2	3	4	5	6	7	8	9	10	11	12
2004	Nº Observações												23
	Mínimo												4.664.000
	Máximo												393.383
	Média												1,16E+12
	Variancia												1.079.205
	Desvio Padrão												
2005	Nº Observações	17	16	19	28	19	17	17	14	14	15	16	25
	Mínimo	50	200	50	50	200	100	300	50	100	100	50	100
	Máximo	4.999.700	4.406.300	4.263.700	4.983.700	5.201.500	5.063.300	4.938.400	5.881.500	5.811.300	5.108.200	5.578.600	4.870.800
	Média	545.321	518.500	438.976	295.600	498.879	530.600	541.676	717.461	705.429	636.087	625.222	366.960
	Variancia	1,77E+12	1,46E+12	1,19E+12	9,90E+11	1,67E+12	1,71E+12	1,68E+12	2,65E+12	2,69E+12	2,13E+12	2,24E+12	1,20E+12
	Desvio Padrão	1.329.506	1.208.184	1.091.413	995.003	1.291.598	1.306.610	1.296.301	1.627.617	1.640.823	1.460.848	1.497.137	1.093.738
2006	Nº Observações	18	15	26	21	19	20	20	27	17	23	27	21
	Mínimo	50	400	50	100	50	200	50	700	50	50	50	50
	Máximo	4.592.600	4.812.500	4.235.500	5.865.600	5.249.600	5.415.900	5.165.300	5.582.300	6.072.700	5.232.400	5.564.100	5.132.000
	Média	499.867	571.093	341.179	492.938	521.250	488.300	504.403	393.200	569.500	424.711	379.041	484.083
	Variancia	1,48E+12	1,75E+12	9,59E+11	1,93E+12	1,81E+12	1,73E+12	1,74E+12	1,47E+12	2,36E+12	1,51E+12	1,38E+12	1,65E+12
	Desvio Padrão	1.215.533	1.321.631	979.244	1.389.439	1.346.052	1.313.606	1.320.335	1.210.995	1.535.524	1.229.137	1.173.756	1.284.656
2007	Nº Observações	16	13	12									
	Mínimo	300	300	900									
	Máximo	4.924.400	5.649.300	5.177.100									
	Média	617.775	747.746	868.117									
	Variancia	2,11E+12	2,73E+12	2,88E+12									
	Desvio Padrão	1.453.859	1.650.946	1.696.362									

Figura 4.5: Principais estatísticas por ano/mês

4.2.4 Histograma englobando todo o período de estudo

No histograma a seguir, figura (4.6), os valores foram agrupados em classes e fica nítida a concentração de valores nas primeiras classes, sugerindo um comportamento exponencial para essa distribuição.

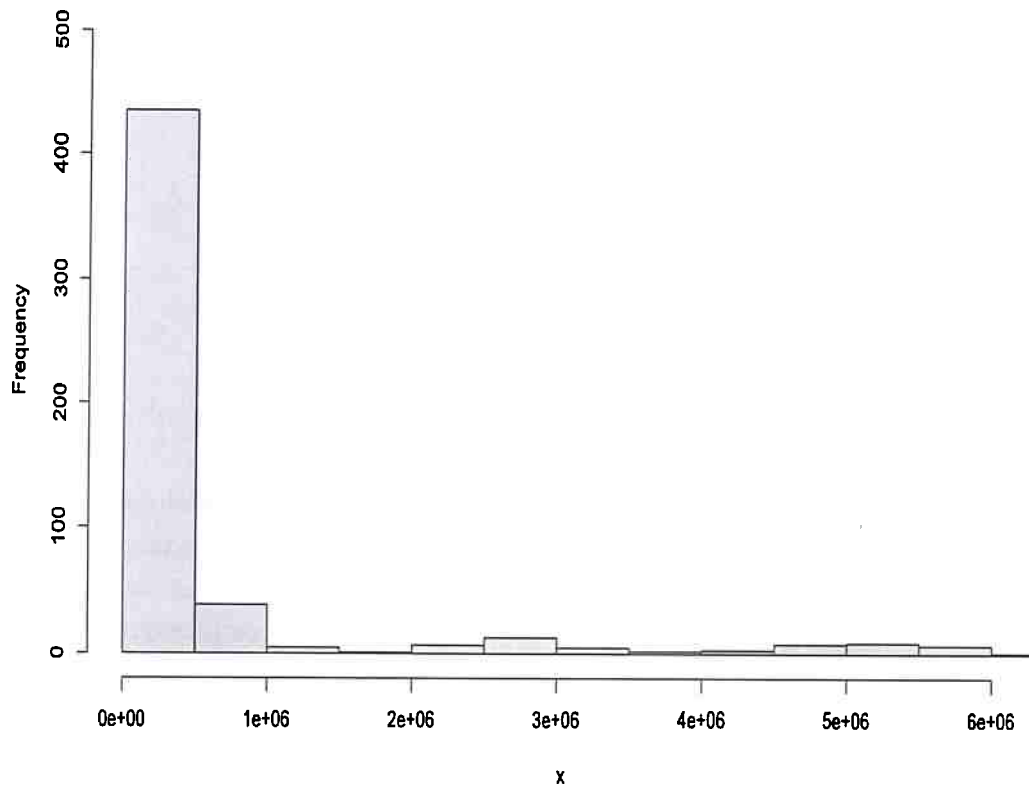


Figura 4.6: Histograma da Severidade

4.3 Resumo Principais Estatísticas-Ocorrências e Severidade

A tabela a seguir, tabela (4.4), apresenta as principais estatísticas considerando todo o período entre 2004 e 2007.

Descrição	Valor
Numero de Sinistros observados	535
Primeiro Quartil dos Sinistros	1 000
Mediana dos Sinistros	6 000
Media dos Sinistros	498 300
Terceiro Quartil	97 400
Moda de ocorrências no mês	17
Variância	1.615405e+12
Desvio padrão	1 270 986
Mínimo	50
Máximo	6 073 000
Primeiro Momento $E[X]$	498 300
Segundo Momento $E[X^2]$	1.86370789e+12

Tabela 4.4: Principais Estatísticas

4.4 Primeira Abordagem

Podemos facilmente obter estimativas para a reserva R mensal conhecendo-se apenas os primeiros momentos da distribuição da perda total. Para isso vamos inicialmente utilizar os dois momentos teóricos obtidos através das equações (2.9) e (2.10) vistas no capítulo 2, e os primeiros momentos obtidos na análise exploratória para se obter os primeiros momentos da distribuição de perdas total.

4.4.1 Determinação dos momentos

Os valores calculados estão na tabela (4.5) a seguir:

Estatísticas	Frequência	Severidade	Perda Total
Média	19	498 300	9 467 700
Variância	19	1,615405 e+10	3,5410457 e+13
Desvio Padrão	4,50	1 270 986	5 950 668,68

Tabela 4.5: Cálculo dos Momentos Perda Total

4.4.2 Resultados Primeira Abordagem

Com base nesses resultados, podemos ter uma primeira ideia do montante de capital a ser reservado por mês, para duas suposições quanto a característica da distribuição final, uma para distribuição normal e outra para distribuição gama. Resultados na tabela (4.6).

VaR	Distribuição Normal	Distribuição Gama
95,0%	19 255 679	20 884 091
99,0%	23 311 025	28 418 951
99,9%	27 856 648	38 597 584

Tabela 4.6: Estimativa para a Reserva de Capital.

No caso dos cálculos de VaR para a distribuição gama os parâmetros de escala e forma foram respectivamente 2.531381 e 2.673702e-07, obtidos através do método dos momentos.

4.5 Segunda Abordagem

Nesta segunda abordagem trabalharemos com a distribuição \hat{F}_X - a estimativa de F_X considerando a distribuição amostral empírica.

Para a distribuição contadora N ou seja, \hat{N} - a estimativa de N , teremos duas possibilidades: distribuição de Poisson e ou distribuição Binomial Negativa. Os procedimentos foram:

4.5.1 Distribuição de Poisson

O estimador de máxima verossimilhança neste caso é $\mathbb{E}[N]$ que no caso é igual a 19, ou seja temos em média a ocorrência de 19 solicitações ao mês de coberturas de seguro.

4.5.2 Distribuição Binomial Negativa

Como foi visto no capítulo 2, a distribuição Binomial Negativa possui dois parâmetros e eles podem ser estimados através de seus momentos, vide (4.1) onde temos de resolver o sistema :

$$E[N(t)] = \left(\frac{r}{p}\right) \quad \text{e} \quad \text{Var}[N(t)] = \left(\frac{r(1-p)}{p^2}\right)$$

As soluções encontradas foram : \hat{r} igual a 9,1975 e \hat{p} igual a 0,4841

4.5.3 Computação

Através da linguagem 'R' foram feitas as simulações/cálculos e construídos os gráficos com as premissas acima e obtivemos os seguintes resultados, tabela (4.7) para a obtenção da reserva R a ser feita no primeiro dia do mês:

VaR	Poisson	Binomial Neg.
95,0%	20 446 862	13 719 052
99,0%	26 305 012	19 109 470
99,9%	33 545 101	26 119 782

Tabela 4.7: Estimativas de Reserva de Capital

Valores esses obtidos através dos cálculos dos quantis elencados sobre a distribuição empírica resultante da simulação conforme figuras (4.7) e (4.8) a seguir:

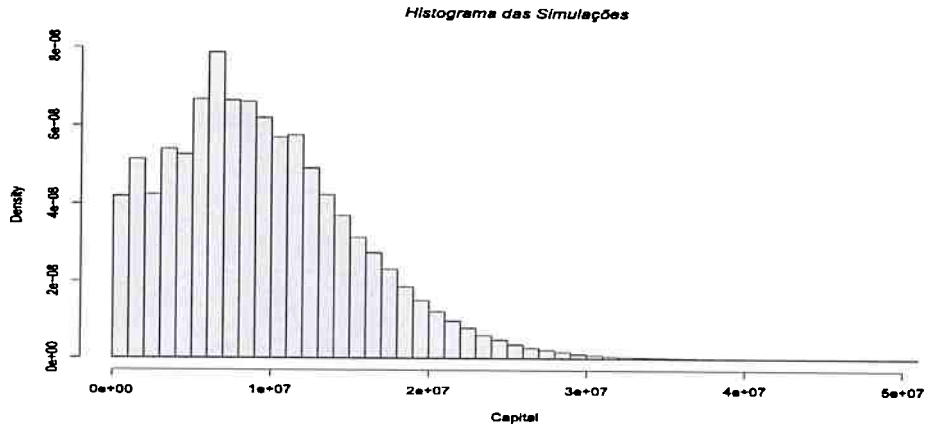


Figura 4.7: Histograma das Simulações 500 000 amostras

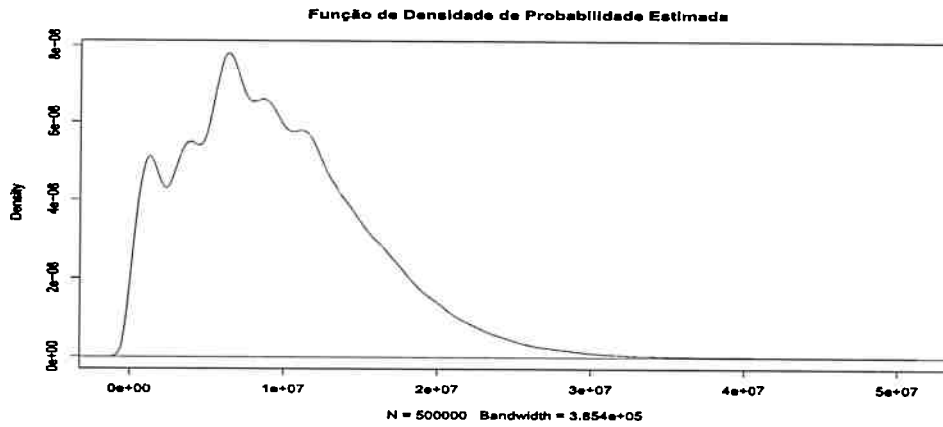


Figura 4.8: Função Densidade das Simulações 500 000 amostras

4.5.4 Aproximação para a Cauda

Com os dados obtidos faremos agora, através do método POT proveniente da Teoria dos Valores Extremos, a aproximação da cauda direita da distribuição da perda acumulada S . Então para \hat{F}_S^{cauda} temos os seguintes passos :

1. Construção da função média de excessos
2. Análise do limiar ótimo e sua escolha
3. Modelar com uma GPD
4. Obter os VaR's para os quantis elencados

Construção da função média de excessos

Pelas definições vistas em (3.18) os resultados da função média de excessos são mostrados na figura (4.9) a seguir:

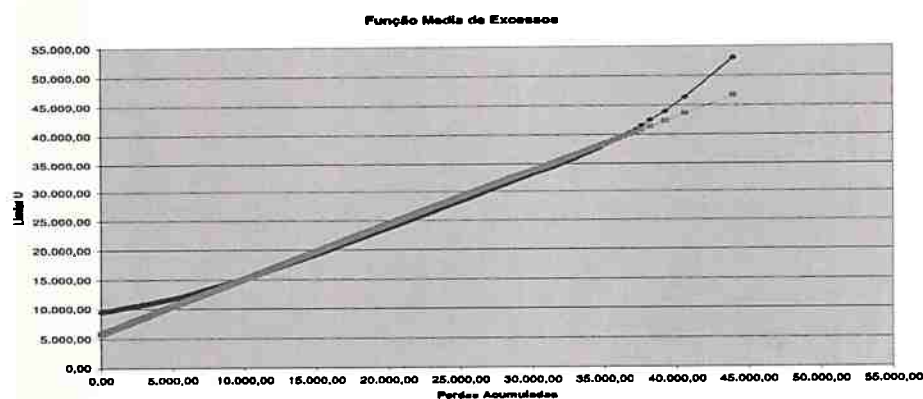


Figura 4.9: Função média de excessos acima de um limiar

Análise do limiar ótimo e sua escolha

Analisando o gráfico da função media de excessos percebemos um comportamento linear mais regular a partir de 10 500 000 e que se estende até 36 000 000. Escolhemos então o valor igual a 10 500 000 como o limiar ótimo para esse procedimento. Para maiores detalhes a respeito dessa escolha sugerimos BELITSKY e MOREIRA [2].

Modelar com uma GPD

Relembrando de (3.18) a estrutura da GPD a ser considerada é:

$$G_{\xi,\beta,0}(x) = \begin{cases} 0, & \text{para } x < 0, \\ 1 - (1 + \frac{\xi}{\beta}x)^{-1/\xi}, & \text{para } x \in [0, -\frac{1}{\xi}], \\ 1, & \text{para } x > -\frac{1}{\xi}. \end{cases} \left(\begin{array}{l} \text{esta fórmula} \\ \text{vale quando} \\ \xi < 0 \end{array} \right)$$

ou

$$G_{0,\beta,s}(x) = \begin{cases} 1 - e^{-\frac{x-s}{\beta}}, & \text{para } x \geq s, \\ 0, & \text{para } x < s, \end{cases} \left(\begin{array}{l} \text{esta} \\ \text{corresponde a } \xi = 0 \end{array} \right)$$

Os parâmetros foram calculados através do estimador de máxima verossimilhança e os valores encontrados foram:

$$\hat{\xi} = -0,206276 \quad \text{e} \quad \hat{\beta} = 6,05329e + 06 \quad \text{para a primeira equação, e}$$

$$\hat{\beta} = 1,99275157e - 07 \quad \text{para a segunda equação.}$$

O histograma da cauda e sua respectiva função estimadora \hat{F}_S^{cauda} estão na figura (4.10) a seguir:

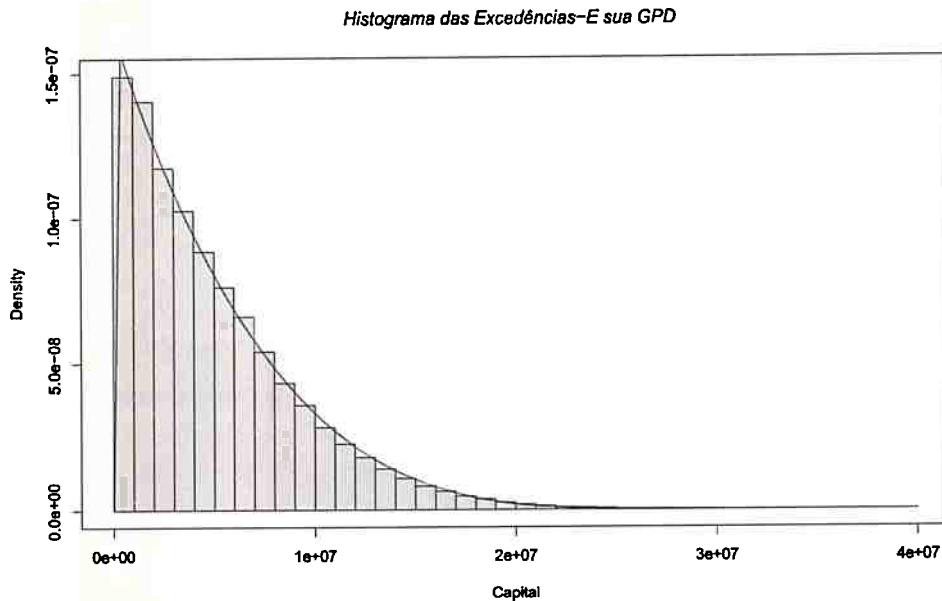


Figura 4.10: Temos aqui o histograma da cauda a direita da função de distribuição decorrente da segunda abordagem ao problema. Com uso do método POT proveniente da Teoria de Valores Extremos, foi obtida \hat{F}_S^{cauda} – a aproximação da cauda direita da distribuição da perda acumulada S . Esta está em linha contínua no desenho.

Obter os VaR's para os quantis elencados

Para obtermos os VaR's desejados é necessário levarmos em conta da necessidade de combinarmos os resultados da probabilidade acumulada até o valor do limiar ótimo escolhido e ponderarmos pela sua função complementar sobre a GPD encontrada. Isso foi feito então com os seguintes valores:

- probabilidade acumulada até o limiar = 0,614125
- probabilidade complementar acima do limiar = 0,385875

Temos então os seguintes resultados que estão apresentados na tabela (4.8) a seguir:

VaR	$\xi < 0$	$\xi = 0$
95,0%	20 593 506	20 754 617
99,0%	26 032 302	28 831 078
99,9%	31 255 037	40 385 880

Tabela 4.8: Estimativas de Reserva de Capital Via GPD

Repetindo os mesmos procedimentos e considerando a variável contadora N com uma distribuição Binomial Negativa obtivemos:

$$\hat{\xi} = -0,04556154 \quad \text{e} \quad \hat{\beta} = 3,634783e + 06 \quad \text{para a primeira equação, e}$$

$$\hat{\beta} = 2,878563e - 07 \quad \text{para a segunda equação.}$$

Que resultaram no histograma a seguir (4.11):

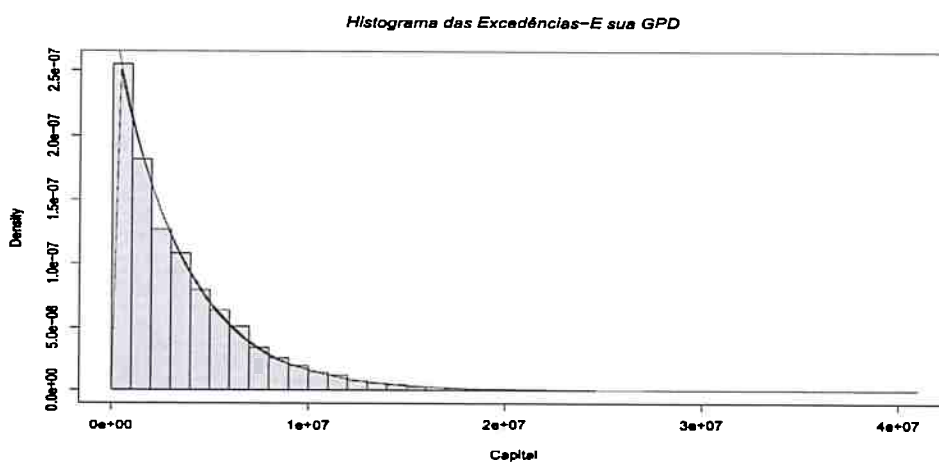


Figura 4.11: Temos aqui o histograma da cauda a direita da função de distribuição considerando a função contadora N com tendo distribuição Binomial Negativa

Como para esse caso temos que também ajustar o valor dos quantis para os cálculos de VaR, os seguintes valores foram considerados:

- probabilidade acumulada até o limiar = 0,87849
- probabilidade complementar acima do limiar = 0,12151

Valores estes que proporcionaram as seguintes estimativas para R que estão mostradas na tabela (4.9) a seguir:

VaR	$\xi < 0$	$\xi = 0$
95,0%	13 663 295	13 584 903
99,0%	19 080 275	19 176 018
99,9%	26 171 099	27 175 096

Tabela 4.9: Estimativas de Reserva de Capital Via GPD

4.5.5 Resumo Resultados Segunda Abordagem

Na tabela (4.10) a seguir mostramos os resultados obtidos com os vários critérios:

Descrição	\tilde{N}	ξ	VaR 95,0%	VaR 99,0%	VaR 99,9%
Simulação	Poisson	-	20 446 862	26 305 012	33 545 101
Simulação	Bin.Neg.	-	13 719 052	19 109 470	26 119 782
Cauda GPD	Poisson	$\xi < 0$	20 593 506	26 032 302	31 255 037
Cauda GPD	Poisson	$\xi = 0$	20 754 617	28 831 078	40 385 880
Cauda GPD	Bin.Neg.	$\xi < 0$	13 663 295	19 080 275	26 171 099
Cauda GPD	Bin.Neg.	$\xi = 0$	13 584 903	19 176 018	27 175 096

Tabela 4.10: Resumo Estimativas de Reserva de Capital

4.6 Terceira Abordagem

Nesta abordagem trabalharemos com a distribuição \hat{F}_X - a estimativa de F_X considerando duas aproximações sobre a distribuição amostral empírica. A primeira através de uma distribuição gama e outra através de uma distribuição exponencial e com essas suposições utilizar

a expressão (2.3) do método analítico visto no capítulo 2.

4.6.1 Parâmetros Estimados

Através do método dos momentos e também do método de máxima verossimilhança os seguintes estimadores foram obtidos:

$\hat{\alpha} = 0,1540249$ e $\hat{\beta} = 3,09073e - 07$ no caso da distribuição gama, e

$\hat{\beta} = 2.006643e - 06$ para a distribuição exponencial

Os resultados dessas aproximações estão apresentados na figura(4.12) a seguir:

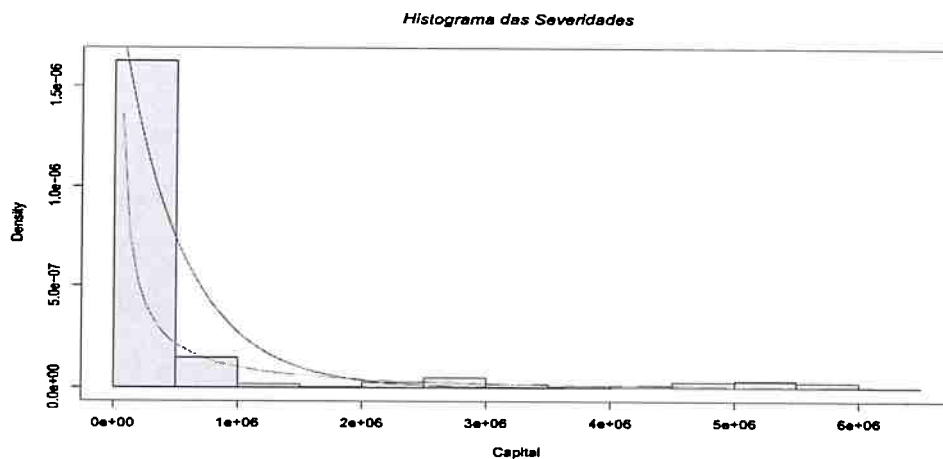


Figura 4.12: Histograma das Severidades com aproximações de distribuições Gama(mais a esquerda) e Exponencial

4.6.2 Resultados Terceira Abordagem

Foram obtidos então os seguintes resultados, tabela (4.11) para estimar a reserva R a ser feita no primeiro dia do mês:

VaR	Poisson+Gama	Poisson+ Exponencial.	Bin.Neg. + Gama	Bin.Neg.+Exponencial
95,0%	20 820 042	14 914 825	13 939 397	9 961 651
99,0%	28 071 801	17 671 362	20 515 392	12 887 220
99,9%	37 726 789	21 037 005	29 552 850	16 635 735

Tabela 4.11: Estimativas de Reserva de Capital Terceira Abordagem

Neste ponto também poderíamos fazer uma aproximação da cauda por uma GPD, como na segunda abordagem, no entanto não o faremos por considerar que a aproximação inicial para as severidades com distribuições gama e exponencial de uma certa forma já contempla isto por possuírem caudas infinitas.

Capítulo 5

Conclusões e Comentários

5.1 Conclusões

Reportando então todos os resultados em uma tabela única, tabela (5.1), podemos ter uma visão geral dos resultados obtidos e fazer algumas considerações.

Abordagem	Descrição	\tilde{N}	ξ	VaR 95,0%	VaR 99,0%	VaR 99,9%
Primeira	Normal (1)	Poisson	-	19 255 679	23 311 025	27 856 648
	Gama (2)	Poisson	-	20 884 091	28 418 951	38 597 584
Segunda	Simulação (3)	Poisson	-	20 446 862	26 305 012	33 545 101
	Simulação (4)	Bin.Neg.	-	13 719 052	19 109 470	26 119 782
	Cauda GPD (5)	Poisson	$\xi < 0$	20 593 506	26 032 302	31 255 037
	Cauda GPD (6)	Poisson	$\xi = 0$	20 754 617	28 831 078	40 385 880
	Cauda GPD (7)	Bin.Neg.	$\xi < 0$	13 663 295	19 080 275	26 171 099
	Cauda GPD (8)	Bin.Neg.	$\xi = 0$	13 584 903	19 176 018	27 175 096
Terceira	Severidade Gama (9)	Poisson	-	20 820 042	28 071 801	37 726 789
	Severidade Exponencial (10)	Poisson	-	14 914 825	17 671 362	21 037 005
	Severidade Gama (11)	Bin.Neg.	-	13 939 397	20 515 392	29 552 850
	Severidade Exponencial (12)	Bin.Neg.	-	9 961 651	12 887 220	16 635 735

Tabela 5.1: Resumo Global das duas Abordagens para Estimativas de Reserva de Capital

A partir da tabela (3.18) vamos construir um gráfico comparativo para os resultados de VaR igual 99,9%, figura (5.1), com o objetivo de compararmos ordem de grandeza facilmente. Relativamente as conclusões obtidas serão também válidas para os níveis de VaR inferiores.

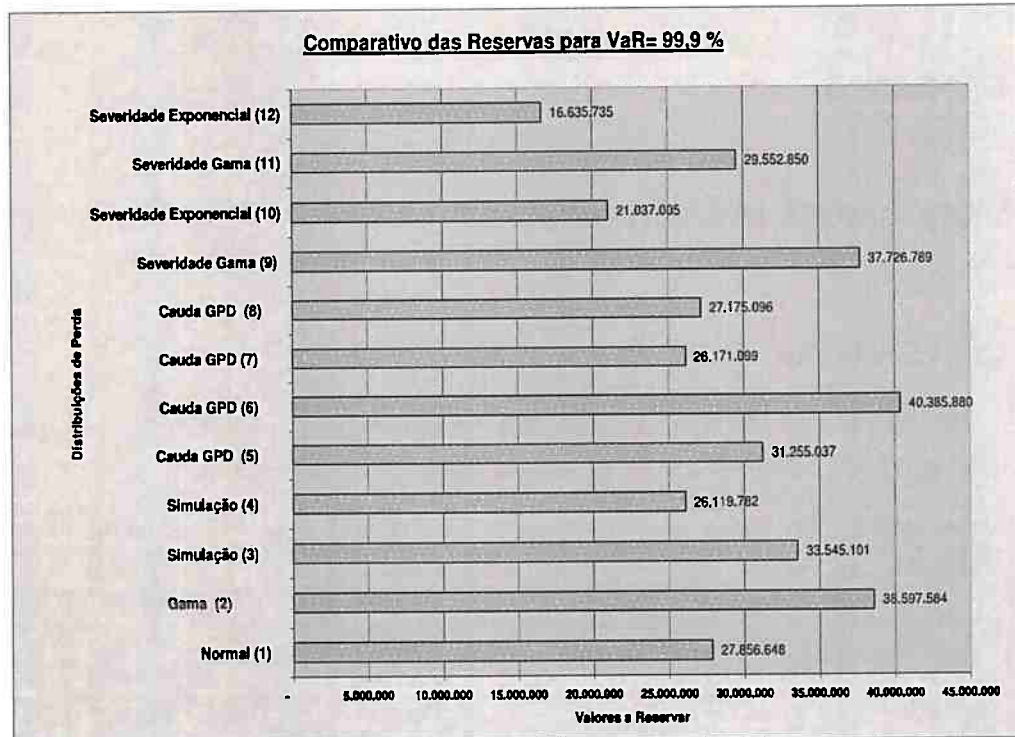


Figura 5.1: Gráfico comparativo de todos resultados obtidos para um nível de VaR de 99,9%

Trabalhando conjuntamente com a tabela e o gráfico concluímos o seguinte:

- Onde assumimos a variável contadora N tendo uma distribuição Binomial Negativa os valores da Reserva são bem menores apesar dela possuir variância maior que a distribuição de Poisson.

- Que os resultados das distribuições (2) e (9) estão próximos e que o cálculo de (2) é mais simples e o de (9) mais elaborado.
- O maior valor encontrado em (6) é fruto de uma aproximação da cauda por uma exponencial que possui cauda infinita enquanto que o valor encontrado em (5) é bem menor porque a sua cauda é finita.
- Que o resultado intermediário de (3) obtido por simulações pode ser uma alternativa interessante na escolha por uma delas.
- Que resultados próximos para algumas distribuições e distantes para outras não são conflitantes e podem definir intervalos de valores importantes para a decisão de qual opção a ser escolhida.

Naturalmente em todo trabalho de modelagem deve-se ter o cuidado de procurar os melhores ajustes possíveis, com as ferramentas possíveis porém devemos ter em mente que a validade das conclusões para esse conjunto de dados analisados não necessariamente serão válidos para outros em períodos. Por outro lado, uma vista rápida na amplitude amostral dos dados das severidades, caracteriza que o conjunto de dados analisado muito provavelmente fazem parte de linhas de atuação distintas e que um detalhamento melhor deles seria necessário. E desta forma o que foi exposto pode ser amplamente utilizado pois as propriedades do modelo permitem que se faça misturas de distribuições com taxas de ocorrências proporcionais a todas ocorrências o que pode ser determinado separando-se as severidades em grupos.

5.2 Novos Horizontes

Ainda há muito espaço para evoluir em termos metodológicos e teóricos nesse campo. Recentes artigos que tratam desse tema, especificamente o artigos de DUTTA e PERRY [10], e também EMBRECHTS, DEGEN e LAMBRIGGER [12], mostram o interesse em se buscar novas classes de distribuições para se modelar dados dessa natureza. A atual crise financeira mundial, particularmente no Estados Unidos com relação ao problema dos financiamentos imobiliários pode ser enquadrado no modelo que abordamos pois a interrupções (ocorrências) de pagamentos nos financiamentos com um impacto (severidade) nos negócios de uma instituição

financeira pode ser tratado como um caso de risco operacional e portanto devidamente avaliado e mesurado. O nosso trabalho aqui apresentado pode dar uma contribuição útil neste sentido.

Referências Bibliográficas

- [1] ASSAF, A.N. *Mercado Financeiro 6ª ed.* São Paulo: Atlas, 2005
- [2] BELITSKY, V.; MOREIRA, F. M. *Estimating Risk With Peaks-Over-Threshold Method - Short Course.* Maresias, March 25-30, 2007
- [3] BALAKRISHNAN, N.; NEVZOROV, V. B. *A Primer On Statistical Distributions.* New Jersey: John Wiley & Sons, 2003.
- [4] BEIRLANT, J.; GOEGEBEUR, Y.; SEGERS, J.; TEUGELS, J. *Statistics of Extremes - Theory and Applications.* Chichester: John Wiley & Sons, 2004.
- [5] BERNSTEIN, Peter L. *Desafio Aos Deuses: A Fascinante Historia do Risco 20ª ed.* Tradução Ivo Korylowski. Rio de Janeiro: Elsevier-Campus, 1997.
- [6] BOROWIAK, D. S. *Financial and Actuarial Statistics An introduction.* New York: Marcel Dekker Inc., 2003
- [7] CASTILLO, E. *Extreme Value Theory in Engineering* San Diego : Academic Press , 1988
- [8] COLES, S. *An Introduction to Statistical Modeling of Extreme Values.* London: Springer Verlag, 2001.
- [9] CRUZ, M.G. *Modeling Measuring and Hedging Operational Risk* Chichester: John Wiley & Sons, 2002.
- [10] DUTTA, K. ; PERRY, J. *A Tale of Tails: An Empirical Analysis of Loss Distribution Models for Estimating Operational Risk Capital* Working Papers Federal Reserve Bank of Boston, 2007

- [11] EMBRECHTS, P.; KLUPPELBERG, C.;MIKOSCH, T. *Modeling Extremal Events for Insurance and Finance*.Berlin: Springer (1997).
- [12] EMBRECHTS, P.;DEGEN, M.;LAMBRIGGER,D.D. *The Quantitative Modeling of Operational Risk: Between g-and-h and EVT* 2006
- [13] FELLER,W. *An Introduction to Probability Theory and its Applications vol 1.*, New York: John Wiley & Sons, 1950
- [14] FELLER,W. *An Introduction to Probability Theory and its Applications vol 2.* , New York: John Wiley & Sons, 1966
- [15] HOSKING,J.;WALLIS,J. *Parameter and quantile estimation for the generalized Pareto distribution*,Technometrics 29, 339-349, 1987
- [16] KASS,R.;GOOVAERTS M.;DHAENE, J.;DENUIT, M.*Modern Actuarial Risk Theory*. New York Kluwer Academic Publishers, 2002.
- [17] KLUGMAN, S. A.;PANJER,H.H.;WILLMOT, G. E. *Loss Models From Data Decisions*. New Jersey: John Wiley & Sons,2004.
- [18] LEADBETTER, M.R.; LINDGREN, G. E ROOTZÉN, H. *Extremes and Related Properties of Random Sequences and Processes*. Berlin: Springer-Verlag,1983.
- [19] MORGADO,A.C.;PITOMBEIRA,J.B.;CARVALHO,P.C.P. ;FERNANDEZ,P; *Análise Combinatória e Probabilidade*.Rio de Janeiro: Sociedade Brasileira de Matemática, 2004.
- [20] MOSCADELLI, M. *The modelling of operational risk: experience with the analysis of the data collected by the Basel Committee*.
- [21] PICKANDS, J. *Statistical inference using extreme order statistics. Ann. Statist.*, 3 1975, pp. 119-131.
- [22] REISS, R.D.;THOMAS, M. *Statistical analysis of extreme values: with applications to insurance, finance, hydrology and other fields*. (2nd edition) Birkhäuser Verlag, 2001.
- [23] RESNICK, S.I. *Adventures in Stochastic Processes*.Boston: Birkhäuser, 2005.

- [24] ROSS, S. *A First Course in Probability*. New Jersey: Prentice Hall, Inc Upper Saddle River: 2002.
- [25] ROSS, S. *Introduction to Probability and Statistics for Engineers and Scientists 3ª ed.* California: Elsevier Academic Press, 2004.
- [26] ROSS, S. *Introduction to Probability Models* San Diego : Academic Press, 2003.
- [27] ROSS, S. *Simulation 2nd ed* San Diego : Academic Press, 1997
- [28] SOLOMON, F. *Probability and Stochastic Processes*. New Jersey : Prentice-Hall, Inc., 1987.
- [29] STIGLER, S.M., Stuart. *Statistics on the Table*. London: Harvard University Press, 1999.
- [30] TAYLOR, H.M. ; KARLIN, S. *An Introduction To Stochastic Modeling*. London: Academic Press, Inc. Ltda, 1984.
- [31] TOLEDO, J. R. F. *Mercado de Capitais - Uma introdução* São Paulo: Thomson, 2006

