

CLÁUDIO DE NARDI QUEIROZ

**REDES BAYESIANAS NO GERENCIAMENTO E MENSURAÇÃO DE
RISCOS OPERACIONAIS**

Dissertação apresentada à Escola
Politécnica da Universidade de São Paulo
para obtenção do título de Mestre em
Engenharia

SÃO PAULO
2008

CLÁUDIO DE NARDI QUEIROZ

**REDES BAYESIANAS NO GERENCIAMENTO E MENSURAÇÃO DE
RISCOS OPERACIONAIS**

Dissertação apresentada à Escola
Politécnica da Universidade de São Paulo
para obtenção do título de Mestre em
Engenharia

Área de concentração: Engenharia
Elétrica - Sistemas Eletrônicos

Orientador: Professor Livre-Docente
Flávio Almeida de Magalhães Cipparrone

SÃO PAULO
2008

DEDICATÓRIA

Dedico este trabalho à Renata, minha esposa, com amor, admiração e gratidão por sua compreensão, carinho, presença e apoio, especialmente ao longo do período de elaboração deste trabalho.

Dedico a minha mãe, pelo exemplo, amor, carinho, esforço e determinação despendidos na minha vida.

AGRADECIMENTOS

Ao professor Flávio Almeida de Magalhães Cipparrone, pela orientação e oportunidade de amadurecimento e crescimento intelectual propiciado.

Agradeço ao Afonso de Campos Pinto, pelo apoio, orientação, conhecimento transmitido e incentivo em todo o decorrer deste trabalho.

Ao Antonio Carlos Avila Otte pela compreensão, confiança e incentivo sempre demonstrados.

Ao Marco Antônio Alexandre da Silva pelas lições de Risco Operacional.

Aos meus amigos Daniel Jugend, Paulo Roberto Godói de Oliveira e Rudini Menezes Sampaio pelas colaborações a este trabalho.

Aos colaboradores da MAPS Soluções e Serviços e a todos que contribuíram, direta ou indiretamente, na execução desta dissertação.

O que prevemos raramente ocorre; o que
menos esperamos geralmente acontece.

Benjamin Disraeli

RESUMO

A aplicação de Redes Bayesianas como modelo causal em Risco Operacional é extremamente atrativa do ponto de vista do gerenciamento dos riscos e do cálculo do capital regulatório do primeiro pilar do Novo Acordo da Basileia. Com as Redes é possível obter uma estimativa do VAR operacional utilizando-se não somente os dados históricos de perdas, mas também variáveis explicativas e conhecimento especialista através da possibilidade de inclusão de informações subjetivas.

Palavras-chave: Redes Bayesianas. Risco Operacional. Modelo Causal. Abordagem de Distribuição de Perdas. Abordagem de mensuração avançada.

ABSTRACT

The application of Bayesian Networks as causal model in Operational Risk is very attractive from the point of view of risk management and the calculation of regulatory capital under the first pillar of the New Basel Accord. It is possible to obtain with the networks an estimate of operational VAR based not only on the historical loss data but also in explanatory variables and expert knowledge through the possibility of inclusion of subjective information.

Keywords: Bayesian Networks. Operational Risk. Causal Model. Loss Distribution Approach. Advanced Measurement Approach.

LISTA DE ILUSTRAÇÕES

Figura 1 – Grafo direcionado e não direcionado.....	23
Figura 2 – Grafo direcionado cíclico e acíclico.....	23
Figura 3 - Grafo $G = (V, E)$ de uma RB com $V = \{X_1, X_2, X_3, X_4\}$ e $E = \{\{X_1, X_2\}, \{X_2, X_4\}, \{X_3, X_4\}\}$:	24
Figura 4 – Grafo da RB para o problema de um carro ligar.	26
Figura 5 - Estruturas básicas de d-separação.....	27
Figura 6 - Grafo para análise de d-separações.....	28
Figura 7 – Exemplo de <i>Markov Blanket</i>	30
Figura 8 – RB após a inferência com a variável X_2 evidenciada com $X_2 = x_{21}$	33
Figura 9 - RB com três variáveis para aprendizado de parâmetros.	39
Figura 10 – RB hipotética: X_1 possui 2 estados e X_2 possui 3 estados.	44
Figura 11 - LDA em uma linha de negócio / tipo de evento de perda.....	52
Figura 12 – RB mínima para cálculo de capital regulatório de RO.....	70
Figura 13 - RB de perdas por falha no processamento de transações bancárias.	71
Figura 14 – Variáveis com 20 estados discretos cada.....	73
Figura 15 - RB de perdas por falhas no processamento de transações bancárias com as variáveis discretizadas.	74
Figura 16 - Exemplo de probabilidades condicionais do nó <i>Severidade</i> da Figura 15.	75
Figura 17 – RB após inferência com informações sobre valores atuais dos ICRs... ..	77
Figura 18 – Rede da Figura 15 com ICR <i>Número médio de transações</i> instanciado no estado >5000 para verificação das novas distribuições de frequência e severidade condicionais ao valor do ICR.	79
Figura 19 – RB do estudo de caso 1	80
Figura 20 –RB após inferência sem conhecimento dos valores atuais dos ICRs do estudo de caso 1.....	84
Figura 21 –RB após inferência com conhecimento dos valores atuais dos ICRs do estudo de caso 1.....	84
Figura 22 – Distribuições de frequência estimada e empírica do estudo de caso 1..	86
Figura 23 – Distribuições de severidade estimada e empírica do estudo de caso 1.	87

Figura 24– Q-Q Plot da distribuição estimada dos dados empíricos do estudo de caso 1.....	87
Figura 25 –RB de perdas por o evento de perda por roubos em ATMs com as variáveis de frequência e severidade rediscritizadas do estudo de caso 1.....	90
Figura 26 - RB de perdas trabalhistas (estudo de caso 2).	91
Figura 27 –Gráfico de evolução temporal das perdas agregadas mensais médias do evento de perdas trabalhistas do estudo de caso 2.	94
Figura 28 – RB de perdas trabalhistas após inferência sem conhecimento dos valores atuais dos ICRs.	97
Figura 29 – RB após inferência com conhecimento dos valores dos ICRs <i>Número médio de demissões e Salário anual médio</i>	97
Figura 30 – Distribuições de frequência estimada pelo LDA e empírica do estudo de caso 2.....	99
Figura 31 – Distribuições de severidade estimada e empírica do estudo de caso 2.	99
Figura 32 – Q-Q Plot da distribuição estimada dos dados empíricos do estudo de caso 2.....	100
Figura 33 – Software LDA: distribuições	103
Figura 34 – Software LDA: SMC para cálculo de VAR e das perdas esperada e inesperada.	104
Figura 35 – Software LDA: QQ-Plot	104
Figura 36 – Software LDA: histograma com função densidade de probabilidade da distribuição paramétrica estimada.....	105
Figura 37– Software LDA: informações dos dados empíricos e da distribuição teórica.	105
Figura 38 – Software LDA: tela de exportação de dados (arquivo texto e Excel©).	105
Figura 39 – Software LDA: tela de importação de dados (arquivo texto e Excel©).	105
Figura 40 – Software RB: definição da estrutura.....	106
Figura 41 – Software RB desenvolvido: inferência.....	107
Figura 42 – Software de Redes Bayesianas: probabilidades condicionais.	107
Figura 43– Software LDA desenvolvido: discretização.	108
Figura 44 – Software LDA desenvolvido: SMC.	108

- Figura 45 – Comparação das probabilidades condicionais ‘aprendidas’ pelo software desenvolvido para estudo (direita) e o *Hugin Lite*© (esquerda) para o nó *Carro liga*..... 115
- Figura 46 – Comparação das probabilidades condicionais ‘aprendidas’ pelo software desenvolvido e o *Hugin Lite*© para o nó *Medidor combustível*..... 115
- Figura 47 – Comparação das probabilidades condicionais ‘aprendidas’ pelo software desenvolvido e o *Hugin Lite*© para o nó *Bicos limpos*. 115
- Figura 48 – Comparação das probabilidades condicionais ‘aprendidas’ pelo software desenvolvido e o *Hugin Lite*© (esquerda) para o nó *Combustível*. 115

LISTA DE TABELAS

Tabela 3.1 – Probabilidades de X_3 condicionais aos estados de seus pais	33
Tabela 3.2 – Probabilidades de X_1 (incondicionais, pois X_1 não possui pais).	34
Tabela 3.3 – Probabilidades de X_2 (incondicionais, pois X_2 não possui pais).	34
Tabela 3.4 - Casos observados utilizados no aprendizado de parâmetros.	40
Tabela 3.5 - Distribuição de $P(X_1 pa(X_1)) = P(X_1)$ aprendida do exemplo.	42
Tabela 3.6 - Distribuição de $P(X_2 pa(X_2)) = P(X_2)$ aprendida do exemplo.	42
Tabela 3.7 - Distribuição de $P(X_3 pa(X_3)) = P(X_3 X_1, X_2)$ aprendida do exemplo. .	42
Tabela 4.1 – Valores dos percentuais da abordagem AS aplicados a cada linha de negócio.....	49
Tabela 4.2 - Eventos de perda de um evento hipotético	54
Tabela 4.3 - Série de freqüência dos eventos de perda da Tabela 4.2.....	54
Tabela 4.4 - Série de severidade dos eventos de perda da Tabela 4.2	54
Tabela 4.5 - Percentual de implantação para bancos e conglomerados.....	64
Tabela 4.6 - Percentual de implantação para as demais instituições	64
Tabela 6.1 – Dez primeiras linhas do banco de dados histórico simulado (com 100.000 linhas) do estudo de caso 1.....	82
Tabela 6.2 – Dez primeiras linhas do banco de dados histórico simulado com os estados de cada variável do estudo de caso 1.	83
Tabela 6.3 – Resultados obtidos com a RB do estudo de caso 1	85
Tabela 6.4 – Resultados obtidos com o LDA aplicado à base de dados simulados do estudo de caso 1	87
Tabela 6.5 – Estudo de caso 1: valores de perda esperada real, sem considerar valores reais de ICRs na RB e obtida pela LDA.....	88
Tabela 6.6 – Estudo de caso 1: resultados obtidos com a RB da Figura 25, com as variáveis discretizadas.	90
Tabela 6.7 – Estudo de caso 2: dez primeiras linhas do banco de dados histórico simulado com os valores de cada variável.....	93
Tabela 6.8 – Estudo de caso 2: dez últimas linhas do banco de dados histórico simulado.....	94
Tabela 6.9 – Estudo de caso 2: dez primeiras linhas do banco de dados histórico. .	95

Tabela 6.10 – Estudo de caso 2: dez últimas linhas do banco de dados histórico simulado.....	96
Tabela 6.11 – Resultados obtidos com a RB do caso de estudo 2	98
Tabela 6.12 – Resultados obtidos com o LDA aplicado à base de dados simulados do caso 1.....	100
Tabela 6.12 – Resultados consolidados do estudo de caso 2.....	100
Tabela 7.1 – Validação: resultados para Normal com $\mu = 100$ e $\sigma = 30$	109
Tabela 7.2 – Validação: resultados para Log-normal com $\mu = 15$ e $\sigma = 2$	110
Tabela 7.3 – Validação: resultados para Exponencial com $\lambda = 0.43$	110
Tabela 7.4 – Validação: resultados para Gumbel com $\alpha = 5.77$ e $\beta = 6.43$	110
Tabela 7.5 – Validação: resultados para Pareto com $\theta = 5$ e $\alpha = 4$	110
Tabela 7.6 – Validação: resultados para Rayleigh com $\beta = 4.81$	110
Tabela 7.7 – Validação: resultados para Weibull com $\alpha = 10$ e $\beta = 20$	111
Tabela 7.8 – Validação: resultados para Normal Inversa com $\mu = 2.22$ e $\lambda = 7.89$.	111
Tabela 7.9 – Validação: resultados para Gama com $\alpha = 21$ e $\beta = 16$	111
Tabela 7.10 – Validação: resultados para Beta com $\alpha = 3.74$, $\beta = 8.77$, $min = 0$, $max = 10$	111
Tabela 7.11 – Validação: resultados para Poisson com $\lambda = 1.5$	112
Tabela 7.12 – Validação: resultados para Binomial com $N = 5$ e $p = 0.88$	112
Tabela 7.13 – Validação: resultados para Geométrica com $p = 0.65$	112
Tabela 7.14 – Validação: resultados para Hipergeométrica com $N = 2$, $D = 7$, $M = 9$	112
Tabela 7.15 – Validação: resultados para Binomial Negativa com $s = 7$, $p = 0.85$.	113
Tabela 7.16 – Distribuição de <i>Combustível</i>	114
Tabela 7.17 – Distribuição de <i>Bicos limpos</i>	114
Tabela 7.18 – Distribuição de P(Carro liga Combustível, Bicos limpos).....	114
Tabela 7.19 – Distribuição de P(Medidor combustível Combustível)	114
Tabela 7.20 – Resultado da inferência sem evidências	116
Tabela 7.21 – Resultado da inferência com a evidência <i>Combustível = Sim</i>	116

LISTA DE ABREVIATURAS E SIGLAS

BCB	Banco Central do Brasil
BIS	Bank for International Settlements
ICR	Indicador chave de Risco
LDA	Loss Distribution Approach
MCMC	Markov Chain - Monte Carlo
PE	Perda esperada
PI	Perda inesperada
PRE	Patrimônio de referência exigido
RB	Rede Bayesiana
RBG	Rede Bayesiana Gaussiana
RO	Risco Operacional
SMC	Simulação de Monte Carlo
VAR	Value at Risk

SUMÁRIO

1	INTRODUÇÃO	16
2	REVISÃO BIBLIOGRÁFICA	18
3	REDES BAYESIANAS	22
3.1	GRAFOS.....	22
3.2	DEFINIÇÃO DE REDES BAYESIANAS	23
3.3	VARIÁVEIS DISCRETAS X CONTÍNUAS.....	24
3.4	D-SEPARAÇÃO	27
3.5	<i>MARKOV BLANKET</i> DE UMA VARIÁVEL.....	29
3.6	CAUSALIDADE.....	30
3.7	INFERÊNCIA.....	32
3.7.1	<i>Inferência com variáveis discretas</i>	32
3.7.2	<i>Inferência com variáveis contínuas</i>	37
3.8	APRENDIZADO DE PARÂMETROS PARA VARIÁVEIS DISCRETAS.....	38
3.8.1	<i>Método de contagem de frequências relativas</i>	38
3.8.2	<i>Método das distribuições a priori de Dirichlet</i>	42
4	RISCO OPERACIONAL.....	45
4.1	DEFINIÇÃO	45
4.2	ABORDAGENS BÁSICAS.....	47
4.2.1	<i>Abordagem do Indicador básico (BIA)</i>	47
4.2.2	<i>Abordagem Padronizada (SA)</i>	48
4.2.3	<i>Abordagem Padronizada Alternativa (ASA)</i>	49
4.3	ABORDAGEM DE MENSURAÇÃO AVANÇADA (AMA)	50
4.3.1	<i>Loss Distribution Approach (LDA)</i>	51
4.4	RISCO OPERACIONAL NO BRASIL	62
4.4.1	<i>Cronograma de implantação</i>	62
4.4.2	<i>Metodologia do Indicador Básico (BIA)</i>	63
4.4.3	<i>Metodologia Padronizada Alternativa (ASA)</i>	65
4.4.4	<i>Metodologia Padronizada Alternativa Simplificada (ASA₂)</i>	66
5	REDES BAYESIANAS EM RISCOS OPERACIONAIS	67

5.1	INTRODUÇÃO	67
5.2	DEFINIÇÃO DAS VARIÁVEIS DO MODELO E DA ESTRUTURA DA REDE	69
5.3	DISCRETIZAÇÃO DAS VARIÁVEIS	72
5.4	OBTENÇÃO DOS PARÂMETROS DA RB	74
5.5	INFERÊNCIA.....	76
5.6	OBTENÇÃO DA DISTRIBUIÇÃO AGREGADA E CÁLCULO DO VAR OPERACIONAL	77
5.7	SIMULAÇÕES COM A RB DE RISCO OPERACIONAL	78
6	ESTUDO DE CASOS	80
6.1	CASO 1: DISTRIBUIÇÕES CONSTANTES NO TEMPO	80
6.1.1	<i>Simulação da base histórica.....</i>	<i>81</i>
6.1.2	<i>Discretização das variáveis</i>	<i>82</i>
6.1.3	<i>Aprendizado de parâmetros</i>	<i>83</i>
6.1.4	<i>Inferência na RB.....</i>	<i>84</i>
6.1.5	<i>Cálculo do VAR Operacional com a RB</i>	<i>85</i>
6.1.6	<i>Aplicação da LDA</i>	<i>86</i>
6.1.7	<i>Análise e comparação dos resultados.....</i>	<i>88</i>
6.2	CASO 2: DISTRIBUIÇÕES QUE SE ALTERAM NO TEMPO.....	91
6.2.1	<i>Simulação da base histórica.....</i>	<i>92</i>
6.2.2	<i>Discretização das variáveis</i>	<i>95</i>
6.2.3	<i>Aprendizado de parâmetros</i>	<i>96</i>
6.2.4	<i>Inferência na RB.....</i>	<i>96</i>
6.2.5	<i>Cálculo do VAR Operacional com a RB</i>	<i>98</i>
6.2.6	<i>Aplicação da LDA</i>	<i>98</i>
6.2.7	<i>Análise e comparação dos resultados.....</i>	<i>100</i>
7	SOFTWARES DESENVOLVIDOS	102
7.1	SOFTWARE LDA.....	102
7.2	REDES BAYESIANAS.....	106
7.3	VALIDAÇÕES DOS SOFTWARES DESENVOLVIDOS	109
7.3.1	<i>Validação do software LDA desenvolvido.....</i>	<i>109</i>
7.3.2	<i>Validação do software de Redes Bayesianas desenvolvido.....</i>	<i>113</i>
8	CONCLUSÕES E TRABALHOS FUTUROS	117
9	REFERÊNCIAS BIBLIOGRÁFICAS	120

1 Introdução

As falências e grandes perdas financeiras de bancos, como no caso do banco inglês *Barings* em 1995 e do francês *Société Générale* em 2008, demonstram a importância de um controle adequado dos riscos operacionais das instituições financeiras.

Metodologias de cálculo aplicadas a Risco Operacional (RO) constituem um tema relativamente novo e ainda com poucas pesquisas publicadas. O debate sobre este assunto tem crescido bastante desde a publicação do Novo Acordo da Basileia, também conhecido como Basileia II, pelo *Bank for International Settlements* (BIS) em junho de 2004, indicando melhores práticas de gerenciamento dos riscos (mercado, crédito e operacional) dos bancos, como descrito em BASEL COMMITTEE ON BANKING SUPERVISION (2004).

No contexto da *Advanced Measurement Approach* (AMA) do Basileia II, as instituições são encorajadas a utilizar metodologias avançadas de cálculo do capital regulatório - dinheiro reservado para fazer frente aos riscos inerentes dos processos de negócio.

Na AMA modelos não causais são freqüentemente utilizados como, por exemplo, a *Loss Distribution Approach* (LDA). Na LDA procura-se a distribuição das perdas a partir das distribuições paramétricas de duas variáveis aleatórias supostas independentes: freqüência e severidade das perdas.

Redes Bayesianas (RB) aparecem como uma solução atrativa de modelagem causal de Riscos Operacionais, permitindo fácil visualização do comportamento das perdas em função de suas causas e a incorporação de conhecimento subjetivo de especialistas.

O objetivo deste trabalho é estudar todos os passos para utilização de RBs como modelo causal de RO no contexto da AMA do Basileia II, verificando a

adequação do modelo para utilizar todos os insumos citados no Novo Acordo da Basileia, bem como as diferentes formas de modelagem, cuidados a serem tomados e limitações. É objetivo também comparar os resultados obtidos com a abordagem de distribuição de perdas – LDA.

Técnicas de aprendizado dos parâmetros de Redes Bayesianas a partir de dados históricos e informações subjetivas de especialistas são descritas, assim como a Simulação de Monte Carlo (SMC) aplicada para obter a distribuição agregada das perdas e, conseqüentemente, o *Value at Risk* (VAR) operacional a partir das suas distribuições de freqüência e severidade. Estas não são supostas independentes como no modelo tradicional LDA.

No capítulo 2 uma revisão bibliográfica do tema é feita. A seguir, um capítulo é dedicado à definição conceitual do modelo de Redes Bayesianas. O quarto capítulo aborda o tema Risco Operacional, descrevendo as possíveis abordagens de cálculo segundo o Novo Acordo da Basileia e o panorama brasileiro. A seguir estuda-se a aplicação de Redes Bayesianas em Risco Operacional, detalhando-se como a modelagem é feita. No capítulo seguinte dois estudos de casos são feitos a partir de variáveis simuladas e os resultados obtidos com as redes são comparados com os obtidos do modelo LDA.

Para este trabalho, dois softwares foram desenvolvidos: um para estudo do modelo LDA e outro para o estudo de Redes Bayesianas. Eles são apresentados no capítulo 7.

2 Revisão Bibliográfica

Métodos quantitativos em RO constituem um tema muito recente e a literatura atual ainda é muito escassa. Não existe, até o momento, uma convergência sobre quais modelos melhor se aplicam em cada situação. Alguns autores escreveram sobre o uso de RBs na modelagem de RO, mas nenhum deles detalhou o processo completo e as possibilidades de uso desta modelagem.

Neil; Fenton e Tailor (2004) sugerem duas maneiras de utilização de RBs em RO. Na primeira utiliza-se a premissa de independência entre frequência e severidade de perdas e as variáveis envolvidas na RB são as mesmas do modelo LDA: frequência de perdas, severidade de perdas e perda total (agregada). Esta abordagem exclui variáveis explicativas (causas) das perdas. As informações subjetivas utilizadas são parâmetros de distribuições estatísticas de frequência e severidade, que são pouco intuitivas e de difícil percepção, mesmo para um especialista. Para obter a tabela de probabilidades condicionais do nó de perdas agregadas os autores sugerem usar uma SMC. A segunda abordagem sugerida considera dependência entre frequência e severidade através de uma causa comum, que o autor denomina *efetividade do processo*. Esta efetividade, por sua vez, pode ter outras causas, como indicadores de risco. Nesta segunda abordagem existe o problema de a variável *efetividade do processo* não ser observável, o que obriga o uso de informações subjetivas. Além disso, os indicadores de risco necessariamente explicam simultaneamente frequência e severidade das perdas.

Antonucci; Piatti e Zaffalon (2007) descrevem o uso de Redes Credais, uma generalização de RBs para probabilidades imprecisas, no gerenciamento e mensuração de riscos operacionais. Redes Credais permitem a especificação de intervalos de probabilidades ao invés de valores pontuais de probabilidades, como no caso de RBs, e também o uso de evidências suaves, isto é, evidências imprecisas sobre os estados das variáveis de estudo.

Adusei-Poku (2005) estuda o uso de RBs para o risco operacional em transações de câmbio, utilizando probabilidades condicionais subjetivas e objetivas (oriundas de dados históricos) e construindo duas redes: uma para a frequência e outra para a severidade de perdas. O autor detalha uma metodologia para obtenção de probabilidades condicionais e distribuições paramétricas subjetivas (*probability elicitation*) através de questionários, preparação de documentos e entrevistas com especialistas.

Assuda (2003) sugere o uso de inferência bayesiana através de distribuições conjugadas para obter distribuições *a posteriori* de frequência e severidade, com o uso de dados recentes como *verossimilhança* e de distribuições *a priori* obtidas através de conhecimento de especialistas, indicadores de risco e dados externos. Com as distribuições *a posteriori* de frequência e severidade, a SMC normalmente utilizada na LDA é utilizada pra obter a distribuição agregada de perdas operacionais.

Scandizzo (2005) descreve uma metodologia para o mapeamento do risco operacional com o objetivo de identificar os riscos inerentes nos diferentes passos de um processo de negócio, selecionando indicadores chaves de risco (ICRs), apresentando dois exemplos de aplicação da metodologia e mostrando como um conjunto relevante de ICRs pode ser identificado.

Shah (2003) descreve o uso do modelo de lógica difusa (*fuzzy logic*) como modelo causal em riscos operacionais, para uso gerencial e para a sua mensuração, como o cálculo do valor em risco operacional (VAR – *Value at Risk*). Lógica difusa fornece mais informação para ajudar no gerenciamento de riscos operacionais do que para cálculo do VAR, uma vez que se baseia somente em regras difusas subjetivas. Os cálculos são feitos somente com conhecimento subjetivo, sem a utilização de dados de perdas.

Alexander (2002) mostra as vantagens do uso de RBs e diagramas de influência no gerenciamento e mensuração de alguns riscos operacionais, como riscos humanos no processamento de transações. As redes são modeladas de forma a gerar diretamente a distribuição agregada de perdas, perdendo-se os

conceitos de frequência e severidade de pedras. Um exemplo de rede com variável contínua é exibido. Não são abordadas as formas de se obter os parâmetros (probabilidades condicionais) da rede.

Perera (2000) sugere o uso de redes neurais artificiais para alguns tipos de riscos operacionais. Uma rede para cada tipo de evento de perda ou linha de negócio é utilizada. Dados suficientes devem existir para treinar a rede neural, e a quantidade pode ser grande, o que não é a realidade de muitos eventos de perda em RO.

Coleman (2000) descreve algumas modelagens em risco operacional. Entre as não causais, destaca a abordagem *Loss Distribution Approach* (LDA) e a teoria dos Valores Extremos (EVT). Como modelagem causal sugere um modelo econométrico de dados de perdas através de modelos preditivos lineares e análise discriminante para obter uma classificação de instituições com relação ao seu risco operacional.

Cruz (2002) em seu consagrado livro de RO mostra como deve ser feita a modelagem de dados de RO, bem como diversas ferramentas para cálculo do VAR operacional. Expõe algumas distribuições estatísticas para modelagem de frequência e severidade de perdas operacionais, testes de hipótese de aderência dos dados, teoria dos valores extremos (GEV – *Generalized Extreme Value* e GPD – *Generalized Pareto Distribution*). Como modelo causal linear sugere o uso de regressão linear múltipla e aponta seus problemas. Como modelos causais são apontados, mas muito pouco detalhados: redes neurais artificiais, lógica *fuzzy* e Redes Bayesianas.

Baud; Frachot e Roncalli (2002) mostram uma metodologia para utilizar base de dados externos em RO, considerando que dados externos geralmente são viesados, uma vez que normalmente são coletados por um consórcio e por isso possivelmente representam as maiores perdas, por serem de domínio público. Um tratamento estatístico que considera que as perdas são coletadas a partir de um valor limiar é apresentado.

A seguir a definição e os conceitos envolvidos em Redes Bayesianas são apresentados, iniciando com uma breve descrição sobre grafos, que representam a estrutura de uma RB. Os algoritmos envolvidos nos cálculos serão descritos, alguns deles detalhados. Como será visto, RBs possuem características desejáveis para modelagem de Riscos Operacionais.

3 Redes Bayesianas

Rede Bayesiana é um modelo gráfico para representar os relacionamentos probabilísticos entre um grande número de variáveis e para realizar inferência probabilística com estas variáveis. Segundo Jordan (1998), modelos gráficos são um casamento entre teoria das probabilidades e teoria dos grafos, que provém uma ferramenta natural para tratar dois problemas que ocorrem em toda matemática aplicada e engenharia: incerteza e complexidade.

Este capítulo é dedicado a explicar os conceitos envolvidos nas RBs. Inicialmente uma breve descrição de grafos, que constituem a estrutura de uma RB, é feita. A seguir RBs são formalmente definidas. O uso de variáveis contínuas ou discretas impactam diretamente a forma como uma RB é construída. Este fato é explorado na seqüência. Os importantes conceitos de *d-separação* e *Markov blanket*, relacionados com relações de independência condicional entre variáveis, são a seguir descritos. Depois, uma breve descrição do conceito de causalidade e as vantagens do uso de RBs causais são exploradas. Por fim, algoritmos de inferência são introduzidos e os de aprendizado de parâmetros detalhados.

3.1 Grafos

Um grafo G é um par ordenado $G = (V, E)$, onde V é um conjunto não vazio cujos elementos são chamados vértices ou nós e E é um conjunto de pares dos vértices de V , cujos elementos são denominados arestas. Usualmente V (e conseqüentemente E) é finito, isto é, $V = \{X_1, X_2, \dots, X_n\}$. Dois vértices ligados por uma aresta são ditos adjacentes. Duas arestas são adjacentes se compartilham um mesmo vértice.

Se as arestas são constituídas de pares ordenados de vértices dizemos que o grafo é direcionado.

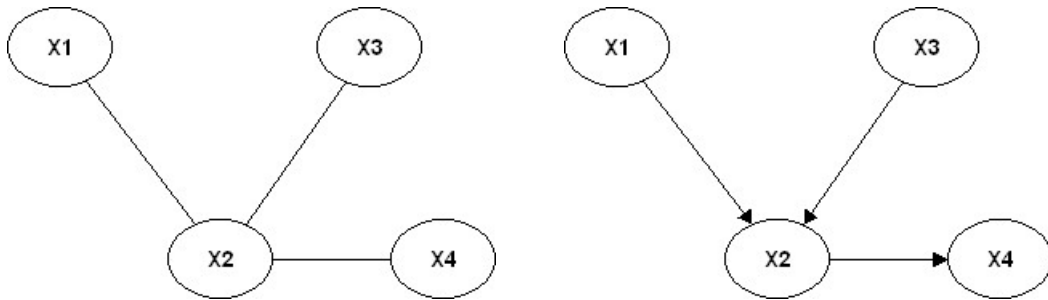


Figura 1 – Grafo direcionado e não direcionado. (a) O grafo da esquerda é não-direcionado. (b) O grafo da direita é direcionado.

Em um grafo direcionado se existe uma aresta de X_i para X_j , dizemos que X_i é pai de X_j (e X_j é filho de X_i). Se existe um caminho direcionado de X_i para X_j dizemos que X_i é ancestral de X_j (e X_j é descendente de X_i). Se um grafo direcionado não possuir ciclos direcionados, isto é, se para qualquer vértice $X_i \in V$ não existe um caminho direcionado que começa e termina em X_i então dizemos que é um grafo acíclico direcionado - *DAG (Directed Acyclic Graph)*.

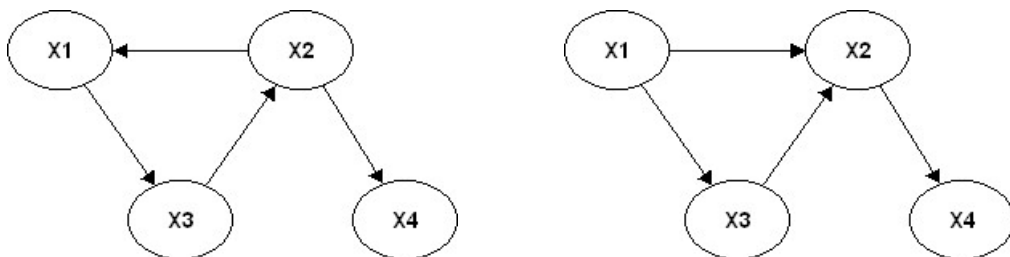


Figura 2 – Grafo direcionado cíclico e acíclico. (a) O grafo da esquerda é direcionado cíclico. (b) O da direita é direcionado acíclico – DAG (apesar de possuir um ciclo este não é direcionado)

3.2 Definição de Redes Bayesianas

Seja P a distribuição conjunta de probabilidades das variáveis aleatórias de algum conjunto V . Seja E um conjunto de pares ordenados dos elementos de V (arestas). Seja $G = (V, E)$ um grafo acíclico direcionado (DAG). Segundo Neapolitan (2004), (G, P) é uma Rede Bayesiana (RB) se para toda variável $X_i \in V$, X_i é condicionalmente independente do conjunto de todos os seus não descendentes dado o conjunto de todos os seus pais (condição de *Markov*).

Na Figura 3, temos um grafo $G = (V, E)$ com $V = \{X_1, X_2, X_3, X_4\}$ e $E = \{\{X_1, X_2\}, \{X_2, X_4\}, \{X_3, X_4\}\}$:

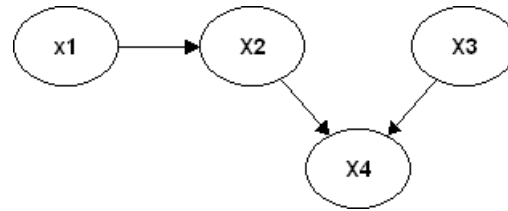


Figura 3 - Grafo $G = (V, E)$ de uma RB com $V = \{X_1, X_2, X_3, X_4\}$ e $E = \{\{X_1, X_2\}, \{X_2, X_4\}, \{X_3, X_4\}\}$:

Neste caso o grafo G , em conjunto com uma distribuição conjunta de probabilidades $P(X_1, X_2, X_3, X_4)$ será uma RB somente se:

- X_1 é independente de X_3 ;
- X_2 é independente de X_3 dado X_1 ;
- X_3 é independente de X_1 e X_2 ;
- X_4 é independente de X_1 , dados os valores de X_2 e X_3 .

Em uma RB, vale a *regra da cadeia*. Seja $V = \{X_1, \dots, X_n\}$ o universo de variáveis e $pa(X_i)$ o conjunto das variáveis que são pais de X_i . Então a distribuição conjunta P das variáveis da rede é igual ao produto das distribuições condicionais de todos os nós dados os valores dos seus pais:

$$P(V) = P(X_1, \dots, X_n) = \prod_i P(X_i | pa(X_i))$$

Vemos que pela regra da cadeia uma RB fornece uma representação mais compacta de $P(V)$.

3.3 Variáveis discretas x contínuas

A regra da cadeia vale tanto para distribuições condicionais discretas quanto para contínuas se a condição de Markov é satisfeita. Um resultado muito importante

é que, dado um grafo G em que os nós são variáveis aleatórias, para *quaisquer* distribuições discretas dos nós de G condicionais aos valores de seus pais, o produto destas distribuições resulta em uma distribuição conjunta P tal que, juntamente com o grafo G , sempre satisfazem a condição de *Markov*. Ou seja, (G, P) é sempre uma RB, independente de quais distribuições condicionais discretas foram escolhidas. O mesmo não é válido para distribuições condicionais contínuas de forma geral.

Uma RB fica, portanto, definida por um grafo e quaisquer distribuições condicionais discretas das variáveis. Este é o método mais utilizado para construir uma RB na prática. O uso mais comum de Redes Bayesianas é com variáveis discretas.

Jensen (2001) define uma RB supondo sempre variáveis discretas. Segundo ele uma RB consiste de:

- um conjunto de variáveis, cada uma com r_i estados discretos mutuamente exclusivos;
- um conjunto de arestas direcionadas entre variáveis;
- as variáveis junto com as arestas direcionadas formam um grafo acíclico direcionado (*DAG*);
- para cada variável X_i com pais X_1, X_2, \dots, X_n existe uma tabela de probabilidades condicionais $P(X_i / X_1, X_2, \dots, X_n)$.

Como exemplo, suponha a RB de variáveis discretas cujo grafo é representado na Figura 4. Esta rede representa o problema de saber se um carro vai ou não ligar de manhã supondo que existem duas possíveis causas para ele não ligar: a presença de combustível no tanque e se os bicos de injeção estão limpos ou sujos. Além disso, o medidor de combustível é consequência de haver ou não combustível no tanque.

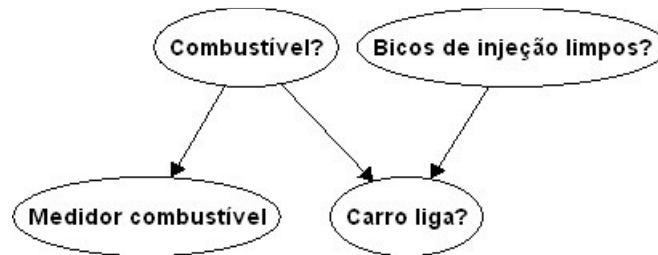


Figura 4 – Grafo da RB para o problema de um carro ligar.

Os possíveis valores que cada variável do exemplo podem assumir são:

- *Combustível?* (C): sim ou não
- *Bicos de injeção limpos?* (BIL): sim ou não
- *Medidor de combustível* (MC): vazio, meio tanque ou cheio
- *Carro Liga?* (CL): Sim ou não.

Nesta RB a distribuição de probabilidade conjunta de todas as variáveis pode ser calculada pela regra da cadeia:

$$P(V) = P(C, BIL, MC, CL) = P(C) \times P(BIL) \times P(MC|C) \times P(CL|C, BIL)$$

Ou seja, conhecendo-se as distribuições $P(C)$, $P(BIL)$, $P(MC|C)$ e $P(CL|C, BIL)$ podemos determinar $P(V)$. Como as variáveis são discretas, (G, P) satisfaz a condição de *Markov*.

Pelo fato de a rede do exemplo ser composta de variáveis discretas com estados mutuamente exclusivos, a distribuição conjunta de probabilidades $P(V)$ consiste de uma tabela com valores para todas as 24 possíveis combinações de estados das variáveis da rede. De forma análoga, as distribuições de cada variável X_i condicionais aos valores dos seus pais constituem tabelas com $q_i \times r_i$ probabilidades, onde q_i é o número de combinações dos estados dos pais de X_i e r_i é o número de estados de X_i .

3.4 D-Separação

A condição de *Markov* para Redes Bayesianas garante independência condicional de um vértice da rede em relação aos seus não descendentes dado o conjunto dos seus pais. A condição, entretanto, garante também outras independências condicionais além daquelas baseadas nos pais de um nó. Para isso, o conceito de d-separação em um *DAG* se faz necessário.

Seja G um *DAG* cujo conjunto de vértices é V . Dizemos que dois vértices distintos X e Y de V estão *d-separados* por um conjunto de vértices $A \subseteq V$ quando, para *todos* os caminhos (não direcionados) entre X e Y , qualquer uma das três afirmações seguintes for verdadeira:

- Existe um vértice $V \in A$ no caminho entre X e Y de forma que a conexão entre X e Y através de V é serial;
- Existe um vértice $V \in A$ no caminho entre X e Y de forma que a conexão entre X e Y através de V é divergente;
- Existe um vértice V , de modo que V e todos os seus descendentes não estão em A e a conexão de X e Y através de V é convergente.

A figura 5 ilustra os três tipos de conexão entre as variáveis:

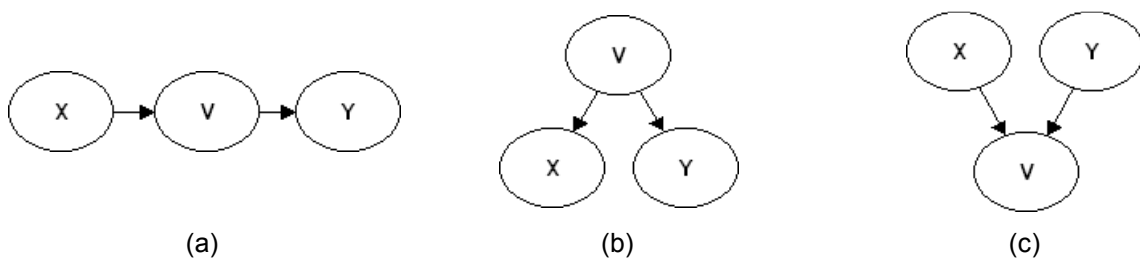


Figura 5 - Estruturas básicas de d-separação. Em (a) temos uma conexão serial: X e Y estão d-separados por $A=V$. Em (b) temos uma conexão divergente: X e Y estão d-separados por $A=V$. Em (c) temos uma conexão convergente: X e Y não estão d-separados por $A=V$ (estão d-separados por $A=\emptyset$)

A figura 6 exibe um *DAG* com sete variáveis. Analisando este *DAG* podemos verificar algumas d-separações. Por exemplo:

- X e R estão d-separados por $A = \{Y, Z\}$ porque o caminho $[X, Y, R]$ é d-separado por Y através de uma conexão serial e os caminhos $[X, Z, R]$ e $[X, Z, S, R]$ são bloqueados por Z por uma conexão serial;
- W e T não estão d-separados por Y porque apesar de o caminho $[W, Y, R, T]$ ser d-separado por Y (conexão serial) o caminho $[W, Y, X, Z, R, T]$ não é d-separado por Y (conexão convergente);
- Y e Z não estão d-separados pelo conjunto de todos os vizinhos de Y, $A = \{W, X, R\}$ porque o caminho $[Y, R, Z]$ não é d-separado por R (conexão convergente).

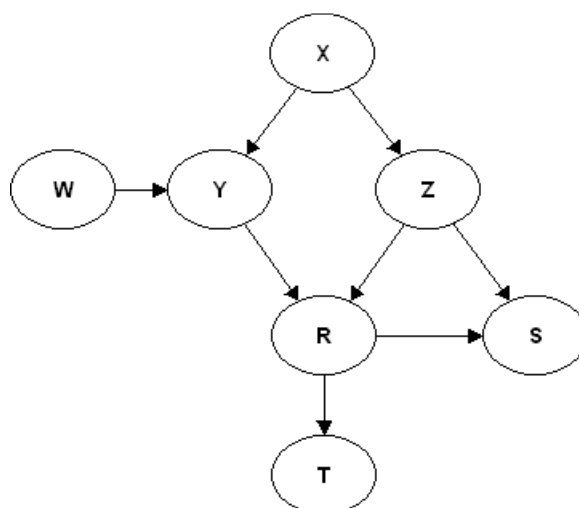


Figura 6 - Grafo para análise de d-separações.

A condição de *Markov* garante que se dois vértices quaisquer X e Y estão d-separados por um conjunto de vértices $A \subseteq V$ no grafo G de uma RB (G, P) , então X e Y são condicionalmente independentes em P dado A . Além disso, toda independência condicional conferida pela condição de *Markov* é identificada por uma d-separação em G .

Isto é, se (G, P) satisfaz a condição de *Markov*, então toda d-separação em G é uma independência condicional em P . Mais ainda, toda independência condicional que é comum a todas as distribuições conjuntas de probabilidade que satisfazem a condição de *Markov* com G é identificada por uma d-separação em G . Uma distribuição conjunta particular P que juntamente com o grafo G satisfaz a condição de *Markov* pode, porém, apresentar independências condicionais que não são identificadas por uma d-separação.

No caso da conexão serial mostrada na Figura 5a, vimos que X é d-separado de Y por V . Calculando a distribuição conjunta de probabilidades daquela RB utilizando a regra de cadeia verificamos que X é independente de Y dado V , ou seja:

$$P(X|V, Y) = \frac{P(X, Y, V)}{P(V, Y)} = \frac{P(X) \times P(V|X) \times P(Y|V)}{P(V, Y)} = \frac{P(X) \times P(V|X)}{P(V)} = P(X|V)$$

No caso da conexão divergente mostrada na Figura 5b, vimos que X é d-separado de Y por V . Utilizando a regra da cadeia temos:

$$P(X|V, Y) = \frac{P(X, Y, V)}{P(V, Y)} = \frac{P(X|V) \times P(V) \times P(Y|V)}{P(V, Y)} = P(X|V)$$

No caso da conexão convergente mostrada na Figura 5c, vimos que X não é d-separado de Y dado V , mas X é d-separado de Y dado $A=\emptyset$. Ou seja, X e Y são independentes quando V é desconhecido, isto é:

$$\begin{aligned} P(X|Y) &= \frac{P(X, Y)}{P(Y)} = \frac{\sum_V P(X, Y, V)}{P(Y)} = \frac{\sum_V P(X) \times P(V|X, Y) \times P(Y)}{P(Y)} \\ &= P(X) \times \sum_V P(V|X, Y) = P(X) \end{aligned}$$

3.5 *Markov Blanket* de uma variável

Uma RB pode ter um número grande de nós, e a probabilidade de um dado nó pode ser afetada pelo conhecimento do estado de um nó distante. Entretanto, o conhecimento dos estados de um conjunto fechado de nós pode blindar um nó do efeito de todos os outros.

Dizemos que se uma variável está instanciada quando o seu valor é conhecido. O *Markov blanket* de um nó X é o conjunto dos nós que são pais de X , filhos de X e dos nós que compartilham um filho com X . Se todas as variáveis do

Markov blanket de X estiverem instanciadas então X está d -separado de todas as outras variáveis da RB (conseqüentemente X é condicionalmente independente de todas as outras variáveis da rede dado seu o *Markov Blanket*).

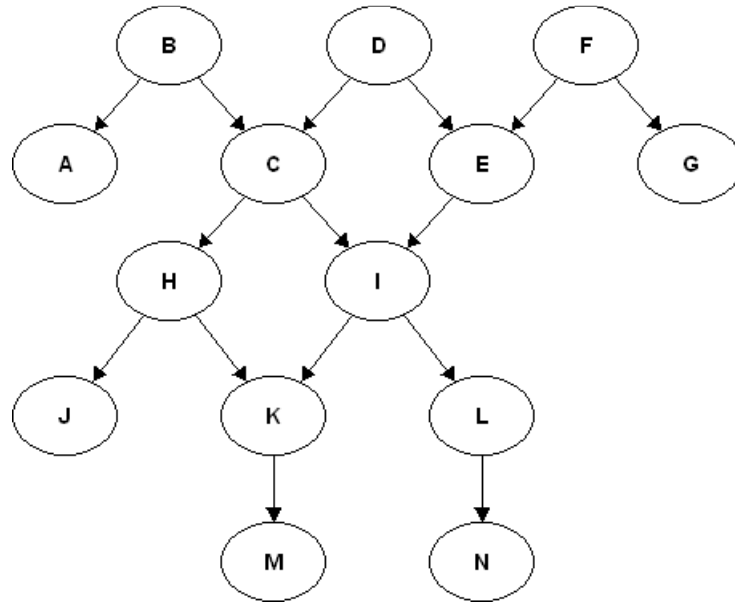


Figura 7 – Exemplo de *Markov Blanket*. O *Markov Blanket* de I é $MB = \{C, E, H, K, L\}$. I é independente de todas as demais variáveis dados os estados das variáveis do MB.

3.6 Causalidade

Segundo Cooper (1999), podemos dizer que X é causa de Y relativa a V se uma manipulação de X altera a distribuição de probabilidade de Y e não há nenhum subconjunto $W \subseteq V - \{X, Y\}$ tal que, se forem instanciadas as variáveis de W , uma manipulação de X não altera mais a distribuição de probabilidades de Y .

A estrutura de uma RB não precisa refletir necessariamente relações de causa e efeito. A única condição necessária é que as propriedades de *d-separação* da rede sejam satisfeitas. Existem, no entanto, boas razões para se trabalhar com redes causais:

- Um modelo perfeitamente causal é mínimo com respeito ao número de arestas.

- A manutenção de uma RB causal é mais fácil do que redes não causais, pois a adição e remoção de variáveis de uma rede causal são mais simples devido à lógica causal já existente na rede.

Uma aresta de um nó X para um nó Y em uma RB é dita causal se e somente se X é uma causa direta de Y . Um *DAG* é causal se todas as arestas que o compõem são causais.

Segundo Neapolitan (2004), se o *DAG* de uma RB for construído utilizando-se arestas causais então frequentemente teremos um *DAG* que satisfaz a condição de *Markov* com relação à distribuição de frequências relativas das variáveis. Ou seja, podemos construir Redes Bayesianas através de *DAGs* causais. Este fato, como será visto adiante, é importante no cálculo das distribuições condicionais de cada variável da rede, através de algoritmos de aprendizado de parâmetros.

Ao construir *DAGs* causais, entretanto, alguns cuidados devem ser tomados para que a condição de *Markov* seja sempre satisfeita com a distribuição conjunta (frequências relativas) das variáveis do grafo. Para isso deve-se garantir:

- Ausência de causas comuns ocultas;
- Ausência de viés de seleção;
- Ausência de laço de realimentação causal.

Deve-se ter a garantia da não existência de causas comuns ocultas. Por exemplo, se duas variáveis X e Y possuem uma causa comum C , então há uma dependência entre X e Y através de C . Se a causa comum for conhecida, então X e Y tornam-se independentes. Se C for uma causa oculta, ou seja, não estiver presente no grafo com arestas apontando para X e Y , haverá uma dependência entre elas que a condição de *Markov* identificará como independência.

Outra violação à condição de *Markov*, similar a não incluir uma causa comum, é o caso de viés de seleção, que pode ocorrer quando duas variáveis são causas comuns de outra não observada (oculta). Neste caso a condição de *Markov* identificará independência entre as duas variáveis observadas, que pode não ser

verdade para alguma população (onde o efeito é conhecido). Por exemplo, se supusermos que o uso da substância finasterida (F) e a apreensão sobre a falta de crescimento de cabelo (G) sejam causas de hipertensão (Y), e observarmos indivíduos hospitalizados para o tratamento de hipertensão, será verificada dependência entre F e G , que um grafo causal sem a variável G identificará como independência.

Por fim, deve-se sempre garantir que se uma variável X é causa de Y então Y não é causa de X , isto é, não existem laços de realimentação causais, de forma que o grafo associado é realmente direcionado acíclico.

Se por um engano uma aresta que liga uma variável X a outra Y for desenhada onde na verdade a influência causal de X em Y se dá através de outra variável, digamos Z , a condição de *Markov* não é violada. O *DAG* resultante não adicionará independências condicionais inexistentes. Na verdade o contrário ocorre, pois a Rede não será capaz de perceber a independência condicional entre X e Y dado Z .

3.7 Inferência

O processo de inferência em uma RB permite obter as distribuições de todas as variáveis da rede condicionais a determinado conjunto de variáveis. O processo de inferência permite atualizar nossas crenças a respeito das variáveis da rede à luz de novas evidências.

3.7.1 Inferência com variáveis discretas

Para uma RB de variáveis discretas podemos calcular a distribuição de qualquer variável X_i condicional a outras variáveis de um conjunto C por:

$$P(X_i|C) = \frac{P(X_i, C)}{P(C)} = \frac{\sum_{X_j \in V - \{X_i, C\}} P(V)}{\sum_{X_j \in V - \{X_i\}} P(V)} = \frac{\sum_{X_j \in V - \{X_i, C\}} \prod_{X_j \in V} P(X_j | pa(X_j))}{\sum_{X_j \in V - \{X_i\}} \prod_{X_j \in V} P(X_j | pa(X_j))}$$

A primeira igualdade deve-se ao teorema de Bayes de probabilidades condicionais, a segunda à regra de marginalização para variáveis com estados discretos mutuamente exclusivos e a terceira à regra da cadeia em uma RB. O denominador da equação de $P(X_i|C)$ é simplesmente uma constante de normalização, pois $\sum_i P(X_i|C) = 1$.

Evidência sobre uma variável é uma informação sobre o seu estado. O conjunto C é o conjunto das variáveis instanciadas (ou evidenciadas). Assim, realizar inferência em uma RB permite obter as distribuições *a posteriori* de todas as variáveis condicionais às evidências conhecidas.

Para exemplificar, suponha a RB de variáveis discretas cujo grafo aparece na Figura 8 e cujas probabilidades condicionais aparecem nas tabelas 3.1, 3.2 e 3.3. X_1 possui dois estados, x_{11} e x_{12} ; X_2 também possui dois estados, x_{21} e x_{22} e X_3 possui três estados, x_{31} , x_{32} e x_{33} .

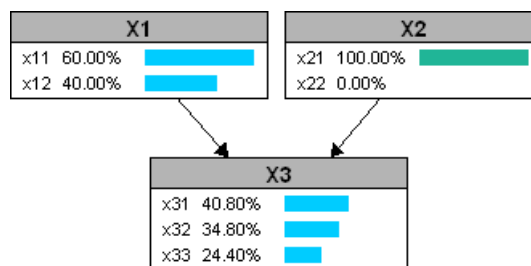


Figura 8 – RB após a inferência com a variável X_2 evidenciada com $X_2 = x_{21}$.

Tabela 3.1 – Probabilidades de X_3 condicionais aos estados de seus pais (X_1 e X_2)

X_1	X_2	$P(X_3 = x_{31} X_1, X_2)$	$P(X_3 = x_{32} X_1, X_2)$	$P(X_3 = x_{33} X_1, X_2)$
x_{11}	x_{21}	32%	46%	22%
x_{11}	x_{22}	60%	24%	16%
x_{12}	x_{21}	54%	18%	28%
x_{12}	x_{22}	8%	77%	15%

Tabela 3.2 – Probabilidades de X_1 (incondicionais, pois X_1 não possui pais).

$P(X_1 = x_{11})$	$P(X_1 = x_{12})$
60%	40%

Tabela 3.3 – Probabilidades de X_2 (incondicionais, pois X_2 não possui pais).

$P(X_2 = x_{21})$	$P(X_2 = x_{22})$
10%	90%

Com os dados fornecidos podemos, por exemplo, calcular a distribuição da variável X_3 condicional a $X_2 = x_{21}$, ou seja, determinar as probabilidades de todos os estados de X_3 condicionais a $X_2 = x_{21}$:

$$\begin{aligned}
 P(X_3 = x_{31} | X_2 = x_{21}) &= \frac{P(X_3 = x_{31}, X_2 = x_{21})}{P(X_2 = x_{21})} = \frac{\sum_{X_1} P(X_1, X_2 = x_{21}, X_3 = x_{31})}{P(X_2 = x_{21})} \\
 &= \frac{P(X_1 = x_{11}, X_2 = x_{21}, X_3 = x_{31}) + P(X_1 = x_{12}, X_2 = x_{21}, X_3 = x_{31})}{P(X_2 = x_{21})} \\
 &= \frac{P(X_1 = x_{11}) \times P(X_2 = x_{21}) \times P(X_3 = x_{31} | X_1 = x_{11}, X_2 = x_{21})}{P(X_2 = x_{21})} \\
 &+ \frac{P(X_1 = x_{12}) \times P(X_2 = x_{21}) \times P(X_3 = x_{31} | X_1 = x_{12}, X_2 = x_{21})}{P(X_2 = x_{21})} \\
 &= \frac{0,6 \times 0,1 \times 0,32}{0,1} + \frac{0,4 \times 0,1 \times 0,54}{0,1} = 40,80\%
 \end{aligned}$$

$$\begin{aligned}
 P(X_3 = x_{32} | X_2 = x_{21}) &= \frac{P(X_3 = x_{32}, X_2 = x_{21})}{P(X_2 = x_{21})} = \frac{\sum_{X_1} P(X_1, X_2 = x_{21}, X_3 = x_{32})}{P(X_2 = x_{21})} \\
 &= \frac{P(X_1 = x_{11}, X_2 = x_{21}, X_3 = x_{32}) + P(X_1 = x_{12}, X_2 = x_{21}, X_3 = x_{32})}{P(X_2 = x_{21})} \\
 &= \frac{P(X_1 = x_{11}) \times P(X_2 = x_{21}) \times P(X_3 = x_{32} | X_1 = x_{11}, X_2 = x_{21})}{P(X_2 = x_{21})} \\
 &+ \frac{P(X_1 = x_{12}) \times P(X_2 = x_{21}) \times P(X_3 = x_{32} | X_1 = x_{12}, X_2 = x_{21})}{P(X_2 = x_{21})} \\
 &= \frac{0,6 \times 0,1 \times 0,46}{0,1} + \frac{0,4 \times 0,1 \times 0,18}{0,1} = 34,80\%
 \end{aligned}$$

A probabilidade do último estado pode ser calculada em função dos demais, já que a soma deve sempre ser 1:

$$P(X_3 = x_{33}|X_2 = x_{21}) = 1 - P(X_3 = x_{31}|X_2 = x_{21}) - P(X_3 = x_{32}|X_2 = x_{21}) = 24,40\%$$

Como visto, a probabilidade conjunta $P(V) = P(X_1, \dots, X_n)$ é suficiente para os cálculos. Como, entretanto, a tabela de probabilidade conjunta cresce exponencialmente com o número de variáveis e a quantidade de estados das variáveis, métodos mais eficientes para o cálculo de $P(V)$ são necessários. Mesmo com a regra da cadeia, que garante uma representação mais compacta de $P(V)$, podem existir vértices com grande número de estados e com grande número de filhos. Segundo Neapolitan (2004) e Wu e Butz (2005), o problema de inferência em redes Bayesianas é NP - difícil.

3.7.1.1 Algoritmos exatos

Jensen (2001) desenvolveu um algoritmo de inferência que envolve a extração de um grafo triangulado não direcionado do *DAG* e a criação de uma árvore cujos vértices são cliques¹ deste grafo triangulado. Esta árvore é denominada *Junction Tree*. Probabilidades condicionais são então calculadas através da passagem de mensagens nesta *Junction Tree*. O algoritmo garante uma seqüência ótima de eliminação (marginalização) de variáveis, e é um dos métodos conhecidos mais eficientes.

Pearl (1986) desenvolveu um algoritmo de passagem de mensagens para inferência em redes Bayesianas cujo grafo é uma poli-árvore (que não contém ciclos). Dado um conjunto C de variáveis instanciadas o algoritmo determina $P(X|C)$ para todos os valores x de cada variável X da rede. Isso é feito através do envio de mensagens de cada variável instanciada para seus vizinhos, que por sua vez passam mensagens para os seus vizinhos, e assim por diante. O algoritmo não depende da ordem das mensagens originadas, o que significa que as evidências podem chegar a qualquer ordem.

¹ Clique é um conjunto completo que não é um subconjunto de outro conjunto completo. Um conjunto é completo se todos os nós são conectados dois a dois.

Os dois algoritmos mencionados acima são algoritmos *exatos*, pois permitem obter o valor preciso das probabilidades das variáveis, dadas as evidências disponíveis.

3.7.1.2 Algoritmos aproximados

Como mencionado, o problema de inferência em Redes Bayesianas é NP-difícil. Para redes muito grandes, os algoritmos exatos podem não ser eficientes em termos de espaço de armazenamento e tempo de processamento. Algoritmos aproximados foram desenvolvidos por pesquisadores para contornar este problema. Exemplos destes algoritmos são: *Logic Sampling* (ou *Forward Sampling*), *Likelihood Weighting*, *Gibbs Sampling* e *Metropolis-Hasting*.

O algoritmo *Logic Sampling* (ou *Forward Sampling*) é um método de amostragem simples para obtenção de probabilidades utilizando um gerador de números pseudo-aleatório. É lento quando existem evidências com baixa probabilidade de ocorrência.

O algoritmo *Likelihood Weighting* é semelhante ao *Logic Sampling*, utilizando amostragem para obtenção das probabilidades, mas contornando o problema de evidências com baixa probabilidade de ocorrência que torna o *Logic Sampling* lento.

Os algoritmos *Gibbs Sampling* e *Metropolis Hasting* utilizam *Markov Chain Monte Carlo* (MCMC), e iniciam a partir de uma configuração válida. Novas configurações são geradas através de sorteios dos estados das variáveis não evidenciadas, respeitando-se os estados das demais variáveis. A diferença entre os dois é que no caso de *Metropolis-Hasting* existe a probabilidade de rejeição de uma configuração, pois o algoritmo procura configurações mais prováveis que as anteriores. As distribuições de probabilidade das variáveis são obtidas das configurações geradas.

3.7.2 Inferência com variáveis contínuas

Como explicado em 3.3 *Variáveis discretas x contínuas*, quaisquer distribuições condicionais discretas das variáveis de uma RB resultam em uma distribuição conjunta $P(V)$ discreta que junto com o grafo G associado sempre satisfaz a condição de *Markov*. O mesmo não é válido, entretanto, para quaisquer distribuições condicionais contínuas.

Segundo Jensen (2001), devido à falta de desenvolvimento técnico, existem algumas restrições para o uso de variáveis contínuas. Primeiro, RBs podem somente tratar distribuições condicionais gaussianas: dada uma configuração dos pais, a distribuição do filho é uma combinação linear de distribuições normais. Uma segunda restrição é estrutural: não é permitido que uma variável contínua tenha filhos discretos.

Neapolitan (2004) detalha um algoritmo de inferência em RBs cujo grafo é uma poli-árvore, onde todas as variáveis da rede são contínuas. Supõe-se que, para uma variável X cujo conjunto de pais é $pa(X)$, X é combinação linear dos seus pais:

$$x = w_X + \sum_{Z \in pa(X)} b_{XZ} \times z,$$

onde W_X possui distribuição normal de média zero e variância $\sigma_{W_X}^2$, que representa a incerteza no valor de X dados valores dos seus pais e b_{XZ} é o peso que o pai Z de X possui sobre X . Variáveis sem pais são supostas normais com média e variância conhecidas. A rede descrita é denominada *Rede Bayesiana Gaussiana* (RBG).

3.8 Aprendizado de parâmetros para variáveis discretas

Em uma RB, o *DAG* é comumente chamado de estrutura e os valores das probabilidades de cada variável condicionais aos valores dos seus pais são denominados parâmetros. Métodos foram desenvolvidos para obter os parâmetros de uma RB a partir de dados existentes (banco de dados de casos). É possível obter os parâmetros de uma rede não só a partir de dados conhecidos, mas também incorporando conhecimento subjetivo de especialistas.

Existem duas situações distintas em que se utiliza aprendizado de parâmetros:

- Quando não se possui nenhuma probabilidade condicional;
- Para adaptar as probabilidades condicionais conhecidas para refletir melhor a experiência representada por casos passados acumulados.

3.8.1 Método de contagem de freqüências relativas

Um método simples mas eficiente de aprendizado de parâmetros é o de contagem de freqüências relativas, descrito em Jensen (2001).

Seja X_i uma variável da RB com r_i estados e pa_{ij} uma das q_i combinações dos estados dos pais de X_i . Seja N_{ij} o tamanho de amostra equivalente de $P(X_i | pa(X_i) = pa_{ij})$, que denota o grau de confiança na distribuição: quanto maior o valor de N_{ij} maior a confiança em $P(X_i | pa(X_i) = pa_{ij})$.

Inicialmente todas as probabilidades condicionais da rede são configuradas com valor $1/r_i$ (que reflete desconhecimento *a priori*), ou então elas podem ser inicialmente configuradas com probabilidades condicionais subjetivas de um especialista, relativas a um tamanho de amostra equivalente N_{ij} .

Para cada caso observado na base de dados históricos em que $X_i = x_{ik}$ e $pa(X_i) = pa_{ij}$ atualiza-se $P(X_i | pa_{ij})$ e N_{ij} através de:

$$P(X_i = x_{ik} | pa(X_i) = pa_{ij})' = \frac{P(X_i = x_{ik} | pa(X_i) = pa_{ij}) \times N_{ij} + 1}{N_{ij} + 1}$$

$$N'_{ij} = N_{ij} + 1$$

Segundo Jensen (2001) este método tende a superestimar a contagem de N_{ij} e com isso superestimar nossa certeza sobre a distribuição. Para resolver tal problema, sugere que seja utilizado *fading*: antes de somar 1 à N_{ij} para cada caso observado, multiplica-se inicialmente a contagem atual por uma constante $q \in (0,1)$. Desta forma, a influência de casos antigos decairá exponencialmente:

$$P(X_i = x_{ik} | pa(X_i) = pa_{ij})' = \frac{P(X_i = x_{ik} | pa(X_i) = pa_{ij}) \times N_{ij} \times q + 1}{N_{ij} \times q + 1}$$

$$N'_{ij} = N_{ij} \times q + 1$$

Como exemplo, suponha a RB da figura 9, com três variáveis: X_1 , X_2 e X_3 , cada uma com dois estados possíveis ($X_1 = \{x_{11}, x_{12}\}$, $X_2 = \{x_{21}, x_{22}\}$ e $X_3 = \{x_{31}, x_{32}\}$) e a tabela 3.4, com um conjunto de cinco casos observados para as variáveis. Suponha $q = 0,9$ e $N_{ij} = 5, \forall i, j$. Inicialmente todos os estados são considerados equiprováveis.

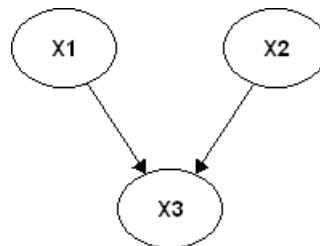


Figura 9 - RB com três variáveis para aprendizado de parâmetros

Tabela 3.4 - Casos observados utilizados no aprendizado de parâmetros.

caso	X_1	X_2	X_3
1	x_{11}	x_{21}	x_{32}
2	x_{12}	x_{21}	x_{31}
3	x_{11}	x_{22}	x_{31}
4	x_{11}	x_{22}	x_{31}
5	x_{11}	x_{22}	x_{31}

Para o caso 1 atualizamos as probabilidades condicionais:

$$P(X_1 = x_{11}) = \frac{0,5 \times 5 \times 0,9 + 1}{5 \times 0,9 + 1} = 0,59 \Rightarrow P(X_1 = x_{12}) = 1 - 0,59 = 0,41;$$

$$N_{11} = 5 \times 0,9 + 1 = 5,5;$$

$$P(X_2 = x_{21}) = \frac{0,5 \times 5 \times 0,9 + 1}{5 \times 0,9 + 1} = 0,59 \Rightarrow P(X_2 = x_{22}) = 1 - 0,59 = 0,41;$$

$$N_{21} = 5 \times 0,9 + 1 = 5,5;$$

$$P(X_3 = x_{31} | X_1 = x_{11}, X_2 = x_{21}) = \frac{0,5 \times 5 \times 0,9 + 0}{5 \times 0,9 + 1} = 0,41 \Rightarrow$$

$$\Rightarrow P(X_3 = x_{31} | X_1 = x_{11}, X_2 = x_{21}) = 1 - 0,41 = 0,59$$

$$N_{31} = 5 \times 0,9 + 1 = 5,5;$$

Para o caso 2:

$$P(X_1 = x_{11}) = \frac{0,59 \times 5,5 \times 0,9 + 0}{5,5 \times 0,9 + 1} = 0,49 \Rightarrow P(X_1 = x_{12}) = 1 - 0,49 = 0,51$$

$$N_{11} = 5,5 \times 0,9 + 1 = 5,95$$

$$P(X_2 = x_{21}) = \frac{0,59 \times 5,5 \times 0,9 + 1}{5,5 \times 0,9 + 1} = 0,66 \Rightarrow P(X_2 = x_{22}) = 1 - 0,66 = 0,34$$

$$N_{21} = 5,5 \times 0,9 + 1 = 5,95$$

$$P(X_3 = x_{31} | X_1 = x_{12}, X_2 = x_{21}) = \frac{0,5 \times 5 \times 0,9 + 1}{5 \times 0,9 + 1} = 0,59 \Rightarrow$$

$$\Rightarrow P(X_3 = x_{32} | X_1 = x_{12}, X_2 = x_{21}) = 1 - 0,59 = 0,41$$

$$N_{33} = 5 \times 0,9 + 1 = 5,5;$$

Para o caso 3:

$$P(X_1 = x_{11}) = \frac{0,49 \times 5,95 \times 0,9 + 1}{5,95 \times 0,9 + 1} = 0,57 \Rightarrow P(X_1 = x_{12}) = 1 - 0,57 = 0,43$$

$$N_{11} = 5,95 \times 0,9 + 1 = 6,36$$

$$P(X_2 = x_{21}) = \frac{0,66 \times 5,95 \times 0,9 + 0}{5,95 \times 0,9 + 1} = 0,56 \Rightarrow P(X_2 = x_{22}) = 1 - 0,56 = 0,44$$

$$N_{21} = 5,95 \times 0,9 + 1 = 6,36$$

$$P(X_3 = x_{31} | X_1 = x_{11}, X_2 = x_{22}) = \frac{0,5 \times 5 \times 0,9 + 1}{5 \times 0,9 + 1} = 0,59 \Rightarrow$$

$$\Rightarrow P(X_3 = x_{32} | X_1 = x_{11}, X_2 = x_{22}) = 1 - 0,59 = 0,41$$

$$N_{32} = 5 \times 0,9 + 1 = 5,5;$$

Para o caso 4:

$$P(X_1 = x_{11}) = \frac{0,57 \times 6,36 \times 0,9 + 1}{6,36 \times 0,9 + 1} = 0,64 \Rightarrow P(X_1 = x_{12}) = 1 - 0,64 = 0,36$$

$$N_{11} = 6,36 \times 0,9 + 1 = 6,72$$

$$P(X_2 = x_{21}) = \frac{0,56 \times 6,36 \times 0,9 + 0}{6,36 \times 0,9 + 1} = 0,47 \Rightarrow P(X_2 = x_{22}) = 1 - 0,47 = 0,53$$

$$N_{21} = 6,36 \times 0,9 + 1 = 6,72$$

$$P(X_3 = x_{31} | X_1 = x_{11}, X_2 = x_{22}) = \frac{0,59 \times 5,5 \times 0,9 + 1}{5,5 \times 0,9 + 1} = 0,66 \Rightarrow$$

$$\Rightarrow P(X_3 = x_{32} | X_1 = x_{11}, X_2 = x_{22}) = 1 - 0,66 = 0,34$$

$$N_{32} = 5,5 \times 0,9 + 1 = 5,95$$

Para o caso 5:

$$P(X_1 = x_{11}) = \frac{0,64 \times 6,72 \times 0,9 + 1}{6,72 \times 0,9 + 1} = 0,69 \Rightarrow P(X_1 = x_{12}) = 1 - 0,69 = 0,31$$

$$N_{11} = 6,72 \times 0,9 + 1 = 7,05$$

$$P(X_2 = x_{21}) = \frac{0,47 \times 6,72 \times 0,9 + 0}{6,72 \times 0,9 + 1} = 0,41 \Rightarrow P(X_2 = x_{22}) = 1 - 0,41 = 0,59$$

$$N_{21} = 6,72 \times 0,9 + 1 = 7,05$$

$$P(X_3 = x_{31} | X_1 = x_{11}, X_2 = x_{22}) = \frac{0,66 \times 5,95 \times 0,9 + 1}{5,95 \times 0,9 + 1} = 0,71 \Rightarrow$$

$$\Rightarrow P(X_3 = x_{32} | X_1 = x_{11}, X_2 = x_{22}) = 1 - 0,71 = 0,29$$

$$N_{32} = 5,95 \times 0,9 + 1 = 6,36$$

As tabelas de probabilidades condicionais obtidas através do método de contagem das frequências relativas aparecem nas tabelas 3.5, 3.6 e 3.7:

Tabela 3.5 - Distribuição de $P(X_1|pa(X_1)) = P(X_1)$ aprendida do exemplo.

$P(X_1 = x_{11})$	$P(X_1 = x_{12})$
0,69	0,31

Tabela 3.6 - Distribuição de $P(X_2|pa(X_2)) = P(X_2)$ aprendida do exemplo.

$P(X_2 = x_{21})$	$P(X_2 = x_{22})$
0,41	0,59

Tabela 3.7 - Distribuição de $P(X_3|pa(X_3)) = P(X_3|X_1, X_2)$ aprendida do exemplo.

X_1	X_2	$P(X_3 = x_{31} X_1, X_2)$	$P(X_3 = x_{32} X_1, X_2)$
x_{11}	x_{21}	0,41	0,59
x_{11}	x_{22}	0,71	0,29
x_{12}	x_{21}	0,59	0,41
x_{12}	x_{22}	0,50	0,50

3.8.2 Método das distribuições *a priori* de *Dirichlet*

Neapolitan (2004) argumenta que as funções densidade de probabilidade de *Dirichlet* fornecem um modo natural de quantificação das crenças *a priori* sobre frequências relativas, bem como uma forma de atualizar estas crenças à luz de evidência.

A função densidade de probabilidade de *Dirichlet* com parâmetros inteiros a_1, a_2, \dots, a_r com $M = \sum_{k=1}^r a_k$ é:

$$\rho(f_1, f_2, \dots, f_{r-1}) = \frac{\Gamma(M)}{\prod_{k=1}^r \Gamma(a_k)} \times f_1^{a_1-1} \times f_2^{a_2-1} \times \dots \times f_r^{a_r-1}$$

Variáveis aleatórias F_1, F_2, \dots, F_r que possuem esta função densidade de probabilidade possuem a distribuição *Dirichlet*. A função densidade de *Dirichlet* é denotada *Dir* ($f_1, f_2, \dots, f_{r-1}; a_1, a_2, \dots, a_r$). F_r é unicamente determinada pelo valor das $r-1$ variáveis anteriores: $f_r = 1 - \sum_{h=1}^{r-1} f_h$.

A função densidade de *Dirichlet* é uma generalização da função densidade Beta: as distribuições marginais de cada variável F_i são distribuições *Beta* (α, β) com $\alpha = a_i$ e $\beta = M$. O valor esperado de F_i é dado, portanto, por:

$$E[F_i] = \frac{a_i}{\sum_{k=1}^r a_k} = \frac{a_i}{M}$$

O método supõe, para cada variável da RB, uma distribuição de *Dirichlet* para cada combinação dos estados dos seus pais, que representa a distribuição conjunta das freqüências relativas de cada estado dada a combinação dos estados dos seus pais.

Seja q_i o número de combinações distintas dos estados dos pais de X_i , r_i o número de estados de X_i e supondo-se que estas combinações seguem uma ordenação $[pa_{i1}, pa_{i2}, \dots, pa_{iq_i}]$. Procura-se então, a partir de um conjunto de casos observados, para toda variável X_i e todas as combinações pa_{ij} dos estados dos pais de X_i os novos valores dos parâmetros $a_{ij1}, a_{ij2}, \dots, a_{ijr_i}$. Seja s_{ijk} o número de vezes na amostra em que a variável X_i encontrava-se no estado x_{ik} e seus pais encontravam-se no estado pa_{ij} . Os novos coeficientes são calculados por:

$$a'_{ijk} = a_{ijk} + s_{ijk}$$

A função de densidade de *Dirichlet* de X_i dada a combinação pa_{ij} de seus pais é atualizada através de:

$$F'_{ij} = Dir(f_{ij1}, f_{ij2}, \dots, f_{ij(r_i-1)}; a'_{ij1}, a'_{ij2}, \dots, a'_{ijr_i})$$

As probabilidades condicionais de qualquer vértice X_i da rede podem ser calculadas através de:

$$P(X_i = x_k | pa_{ij}) = E[F_{ijk}] = \frac{a_{ijk}}{\sum_{k=1}^{r_i} a_{ijk}}$$

Sendo N o *tamanho de amostra equivalente a priori*, que denota o grau de confiança nas probabilidades iniciais $P(X_i = x_{ik} | pa_{ij})$, os valores iniciais de a_{ijk} devem ser calculados através de

$$a_{ijk} = P(X_i = x_{ik} | pa_{ij}) \times P(pa_{ij}) \times N$$

Para expressar indiferença *a priori* Neapolitan (2004) recomenda o uso de $N = \max_i r_i$, $P(X_i = x_{ik} | pa_{ij}) = 1/r_i$ e $P(pa_{ij}) = 1/q_i$. Assim:

$$a_{ijk} = \frac{\max_i r_i}{r_i \times q_i}, k = 1, \dots, r_i$$

Como exemplo suponha a rede da Figura 10 abaixo, onde o nó X_1 possui dois estados (x_{11}, x_{12}) e o nó X_2 possui três estados (x_{21}, x_{22}, x_{23}).

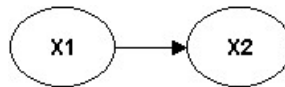


Figura 10 – RB hipotética: X_1 possui 2 estados e X_2 possui 3 estados.

Neste caso X_1 teria uma função densidade de *Dirichlet* $F_{11} = Dir(f_{111}; a_{111}, a_{112})$, sobre duas variáveis: f_{111} (frequência relativa de $X_1 = x_{11}$), f_{112} (frequência relativa de $X_1 = x_{12}$). X_2 teria duas funções densidade de *Dirichlet* (uma para cada combinação dos valores de seu pai, X_1): $F_{21} = Dir(f_{211}, f_{212}; a_{211}, a_{212}, a_{213})$, $F_{22} = Dir(f_{221}, f_{222}; a_{221}, a_{222}, a_{223})$. A variável aleatória f_{211} representa a frequência relativa de $X_2 = x_{21}$ dado que $X_1 = x_{11}$.

4 Risco Operacional

Segundo Cruz (2002), o termo *Risco Operacional* foi provavelmente mencionado pela primeira vez logo após a falência do banco *Barings*, onde um *trader* causou a quebra do banco escondendo posições nos mercados futuro e de derivativos na Ásia. Este evento trouxe consciência ao mercado de que alguns riscos, até então ignorados, podem afetar consideravelmente os resultados e podem não ser classificados como risco de mercado ou risco de crédito.

Mesmo com as aparentes evoluções no controle de Riscos Operacionais, o caso do banco *Société Générale* mostra a obrigação de cuidados extremos no gerenciamento de RO, uma vez que perdas inesperadas podem facilmente decretar a falência de uma instituição financeira. Neste caso as atividades fraudulentas de um *trader* resultaram em janeiro de 2008 na maior perda bancária já reportada – 4,9 bilhões de euros.

Este capítulo é dedicado a RO. Inicialmente a definição conceitual do *Basiléia II* é fornecida. A seguir, as abordagens básicas de cálculo do capital regulatório definidas pelo BIS são descritas. Os conceitos da abordagem de mensuração avançada são a seguir explicados. Por fim, o panorama brasileiro é colocado, com o cronograma de implantação definido pelo Banco Central do Brasil (BCB) e com as adequações feitas às metodologias básicas de cálculo.

4.1 Definição

Segundo o Comitê da Basiléia em Supervisão Bancária, subordinado ao *BIS* (*Bank for International Settlements*), Risco Operacional (RO) é definido como o risco de perda resultante de processos internos falhos ou inadequados, pessoas e

sistemas ou de eventos externos. Esta definição inclui risco legal² mas exclui risco estratégico e reputacional.

O *BIS* é uma organização internacional que encoraja e estimula cooperação internacional monetária e financeira e serve como um banco para os bancos centrais. Hoje 55 bancos centrais são membros do BIS, entre eles os bancos centrais brasileiro, europeu e americano. Em junho de 2004 o BIS lançou o documento *Basel II: International Convergence of Capital Measurement and Capital Standards: a Revised Framework*, conhecido como Novo acordo da Basileia, ou simplesmente Basileia II.

O *Basileia II* descreve um padrão mínimo para adequação de capital, procurando alinhar requerimentos de capital regulatório mínimo com os riscos que os bancos enfrentam e promover uma abordagem que permita uma visão de longo prazo, encorajando bancos a identificar os riscos que eles podem correr, hoje e no futuro, e desenvolver a habilidade de gerenciar esses riscos.

O capital regulatório mínimo é o dinheiro que os bancos devem reservar devido aos riscos inerentes às suas atividades. Seu cálculo compreende o primeiro de três pilares definidos no acordo da Basileia. O segundo pilar compreende o processo de regulação pelos bancos centrais, para garantir que os bancos sigam processos rigorosos, calculem seus riscos de uma forma robusta e confiável e tenham capital suficiente para proteger a ampla comunidade bancária contra implicações sistêmicas dos riscos que eles tomam para si. O terceiro pilar compreende disciplina de mercado, e funciona como uma alavanca para reforçar a confiabilidade e segurança do sistema bancário, através da melhor descoberta pelo mercado de exposições a riscos e nível de capital, de forma a facilitar a estimação do nível de solvência de um banco.

Para cálculo do capital regulatório mínimo o acordo da Basileia indica quatro abordagens: Indicador Básico (*BIA – Basic Indicator Approach*), Padronizada (*SA – Standardized Approach*), Padronizada Alternativa (*ASA – Alternative Standardized*

² Risco legal inclui, mas não é limitado a, exposições a multas, penalidades ou perdas punitivas resultantes de ações de supervisão, bem como acordos privados.

Approach) e a Mensuração Avançada (*AMA - Advanced Measurement Approach*). As abordagens na ordem em que foram citadas apresentam um aumento contínuo de sofisticação e sensibilidade ao risco.

Espera-se que as primeiras abordagens sejam extremamente punitivas para as instituições financeiras em termos de exigência de capital. Bancos que optarem por abordagens mais simples, além de não terem os benefícios de um melhor gerenciamento dos seus riscos operacionais irão desembolsar mais e com isso apresentarão desvantagens competitivas. O capital alocado tende a se reduzir na medida em que a instituição adota abordagens mais sofisticadas de mensuração.

Por outro lado, existem custos crescentes associados ao desenvolvimento de modelos, geração de base de dados e adequação aos critérios de qualificação na medida em que as abordagens tornam-se mais avançadas.

4.2 Abordagens básicas

As abordagens básicas são quantitativamente muito simples quando comparadas com a AMA. Elas tendem ser muito rigorosas para as instituições financeiras que optarem por adotá-las. São três as abordagens básicas descritas no Basileia II: Abordagem do Indicador Básico (BIA), Abordagem Padronizada (SA) e Abordagem Padronizada Alternativa (ASA).

4.2.1 Abordagem do Indicador básico (BIA)

Bancos utilizando o indicador básico devem reservar capital para risco operacional igual a 15% da receita bruta anual média dos últimos três anos:

$$P_{OPR}^{BIA} = \frac{1}{N} * \sum_{i=1}^3 \max(\alpha \times GI_i, 0)$$

, onde:

$\alpha = 15\%$;

GI_i : receita bruta anual no i -ésimo ano anterior;

N : número de anos dos últimos três em que a receita bruta anual foi positiva.

4.2.2 Abordagem Padronizada (SA)

Na abordagem padronizada, as atividades dos bancos são divididas em oito linhas de negócio e o capital a ser alocado é igual à média dos últimos três anos do somatório de um percentual da linha de negócio (que varia entre 12% e 18%) aplicado sobre a receita bruta da linha de negócio:

$$P_{OPR}^{SA} = \frac{1}{3} * \sum_{i=1}^3 \max\left(\sum_{j=1}^8 \beta_j \times GI_{ij}, 0\right)$$

, onde:

GI_{ij} : receita bruta anual da linha de negócio j , no ano i -ésimo ano anterior

β_j : percentual da linha de negócio j

Os valores de β_j aparecem na Tabela 4.1:

Tabela 4.1 – Valores dos percentuais da abordagem SA aplicados a cada linha de negócio

Linha de negócio	Fator β
Finanças corporativas (β_1)	18%
Negociação e vendas (β_2)	18%
Pagamentos e liquidações (β_3)	18%
Serviços de agente financeiro (β_4)	15%
Administração de ativos (β_5)	12%
Corretagem de varejo (β_6)	12%
Varejo (β_7)	12%
Comercial (β_8)	15%

4.2.3 Abordagem Padronizada Alternativa (ASA)

Na Abordagem Padronizada Alternativa, assim como na Padronizada, as atividades dos bancos são divididas em oito linhas de negócio. O capital a ser alocado é igual à média dos últimos três anos do somatório de percentuais das linhas de negócio (que variam entre 12% e 18%, iguais ao do SA) aplicados sobre um valor base, que depende da linha de negócio. Para seis linhas de negócio esse valor base é a receita bruta anual da linha de negócio. Para duas linhas de negócio (varejo e comercial) o valor base é igual ao somatório de empréstimos e adiantamentos no ano da linha de negócio:

$$P_{OPR}^{ASA} = \frac{1}{3} \times \sum_{i=1}^3 \max \left[\left(\sum_{j=1}^6 \beta_j \times GI_{ij} \right) + \left(\sum_{j=7}^8 \beta_j \times m \times LA_j \right), 0 \right]$$

, onde:

GI_{ij} : receita bruta anual da linha de negócio j , no ano i -ésimo ano anterior;

β_j : percentual da linha de negócio j , de acordo com a tabela 4.1;

LA_j : é a média total dos adiantamentos e empréstimos (*loans and advances*) em aberto sem risco ponderado e bruto de provisões dos três anos da linha de negócio j ;

$m = 3,5\%$ (fixo).

4.3 Abordagem de Mensuração Avançada (AMA)

Na Abordagem de Mensuração Avançada o requerimento de capital regulatório mínimo é igual à medida de risco gerada pelo sistema de mensuração interno do banco utilizando critérios qualitativos e quantitativos definidos pelo comitê da Basileia.

Um sistema de mensuração avançada deve estimar de modo aceitável perdas esperadas e inesperadas baseado no uso combinado de dados internos, dados externos relevantes, análise de cenários e fatores de controle internos (com fatores que refletem o ambiente de negócio). O sistema deve também ser capaz de suportar alocação de capital econômico para risco operacional através das linhas de negócio de maneira a criar incentivos na melhoria do gerenciamento de riscos operacionais das linhas de negócio do banco.

O comitê da Basileia não obriga nenhuma metodologia específica para o cálculo do capital econômico a ser alocado, mas afirma que a metodologia deve conseguir obter a máxima perda com 99,9% de confiança, no período de um ano, o *Value at Risk* (VAR) operacional. Deve também ser capaz de capturar eventos de perda com cauda pesada.

Na AMA aparecem os conceitos de VAR, perda esperada (*PE*) e perda inesperada (*PI*). A perda esperada é a média da distribuição de perdas, e a perda inesperada é a diferença entre o VAR e a perda esperada. Para um determinado horizonte de tempo (que a Basileia sugere que seja um ano) e um nível de confiança (α) a seguinte relação é válida:

$$VaR(\alpha) = PE + PI(\alpha)$$

Espera-se que o cálculo do capital regulatório seja substancialmente menor na AMA que nas demais abordagens vistas. A AMA é muito mais sofisticada que as demais, e segundo o BIS (2004) seu uso estará sujeito a requisitos quantitativos e qualitativos mínimos.

4.3.1 Loss Distribution Approach (LDA)

A abordagem de distribuição de perdas (LDA) foi citada pelo comitê da Basileia em 2001 no anexo seis de seu documento de consulta sobre riscos operacionais *Consultative Document – Operational Risk: Supporting Document to the New Basel Capital Accord*. Desde então ganhou força e tornou-se a metodologia mais empregada no contexto da AMA. A metodologia é bem detalhada em Cruz (2002).

A LDA supõe que os dados de perda constituem o mais objetivo indicador de risco disponível. Na LDA busca-se a distribuição estatística das perdas baseando-se somente na série histórica observada. Outras variáveis como as causas das perdas e os valores atuais destas causas não são consideradas.

A LDA utiliza técnicas atuárias padrões para modelar o comportamento das perdas operacionais através da estimação de frequência e severidade para produzir uma estimativa de perdas esperadas e inesperadas.

O primeiro passo da LDA é a separação da série de perdas em outras duas: a de frequência e a de severidade das perdas. A seguir, para cada uma das séries, estimam-se os parâmetros de distribuições estatísticas paramétricas. Para as duas séries, escolhem-se as distribuições que tiveram maior aderência aos dados, desde que uma aderência mínima tenha sido obtida, a partir de testes estatísticos de aderência.

Com as distribuições paramétricas de frequência e severidade realiza-se uma SMC para cálculo da distribuição agregada de perdas, já que de forma geral não

existe uma solução analítica para calculá-la. A partir da distribuição agregada é possível calcular o VAR operacional, a perda esperada e a perda inesperada.

O procedimento descrito acima é realizado para cada tipo de evento de perda em cada linha de negócio da instituição financeira. O comitê da Basileia recomenda a soma de cada VAR para o cálculo do capital regulatório. As instituições que provarem ao órgão regulador que mensuraram e contabilizaram as perdas esperadas nas suas práticas de negócio internas poderão alocar apenas a perda inesperada ao invés de todo o VAR.

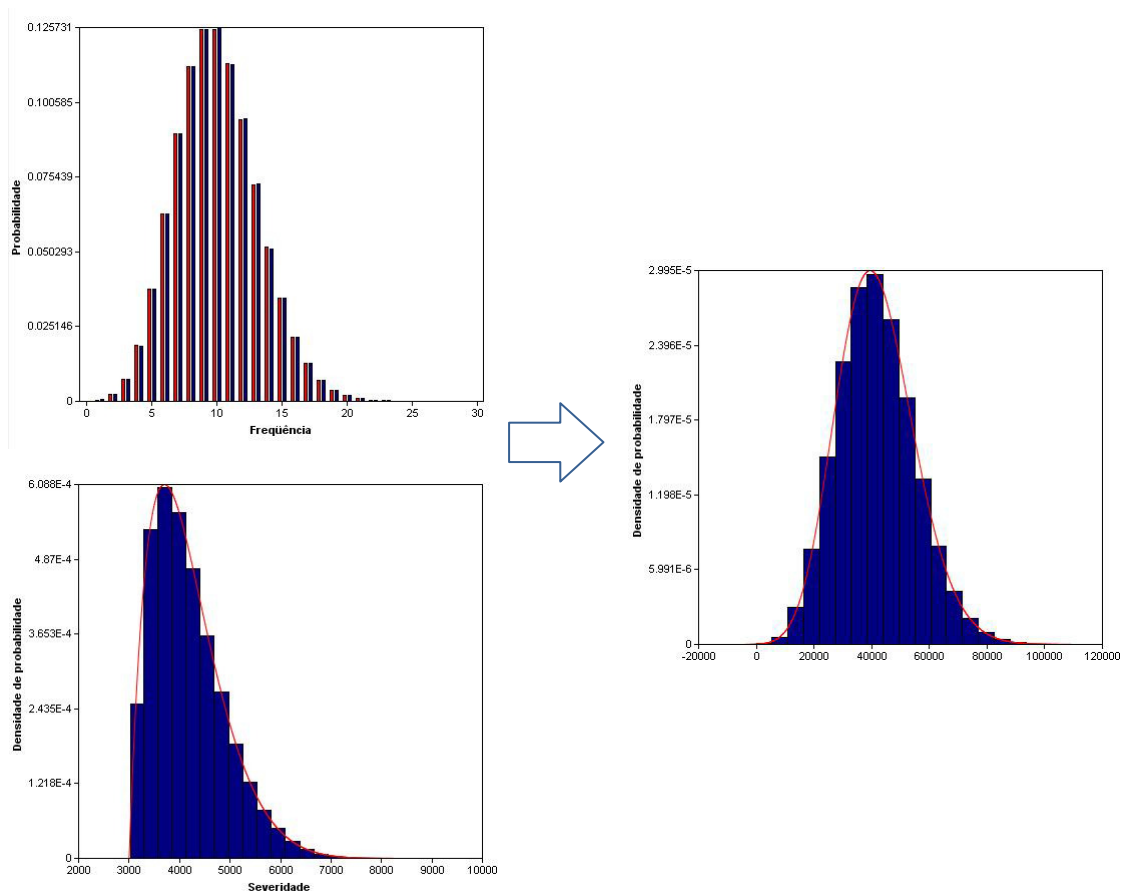


Figura 11 - LDA em uma linha de negócio / tipo de evento de perda: a distribuição de freqüência (cima, esquerda) é uma Poisson com $\lambda=10$ e a distribuição de severidade (baixo, esquerda) uma Beta com $\alpha=2$, $\beta=10$, $min=3000$, $max=10000$. A distribuição agregada possui média igual a 41.666,67 e $Var(99,9\%) = 99.800,00$ e é obtido através de uma SMC.

A Figura 11 exemplifica o uso do LDA para os dados históricos de um tipo evento de perda hipotético. No histograma de freqüência, as barras azuis

representam as probabilidades observadas na amostra histórica, e as vermelhas as probabilidades teóricas de uma distribuição Poisson ($\lambda = 10$). No gráfico de severidade vemos em azul o histograma dos dados históricos de severidade, e a linha vermelha é a função densidade de probabilidade de uma distribuição Beta com $\alpha=2$, $\beta=10$, $\min=3.000$ e $\max=10.000$. O gráfico à direita mostra o histograma da distribuição agregada obtida das distribuições paramétricas de frequência e severidade através de uma SMC.

4.3.1.1 Separação em frequência e severidade

Na LDA supõe-se que um evento de perda é função exclusiva de duas variáveis aleatórias independentes: frequência e severidade das perdas. A frequência é o número de vezes que um evento de perda ocorre em um determinado período de análise. É, portanto, uma variável discreta. A severidade corresponde ao valor individual de cada perda no período de análise. Poderia ser considerada uma variável discreta, já que as perdas são capturadas com precisão de centavos (segunda casa decimal), mas na LDA considera-se a severidade uma variável contínua. Seja P a variável aleatória das perdas totais, F a frequências das perdas e S a severidades das perdas. Então:

$$P = \sum_{i=1}^F S$$

Como exemplo suponha os dados de perdas da tabela 4.2. Vê-se que no dia 01/01/2008 duas perdas foram registradas. Esta é a frequência de perdas do dia. No dia 03/01/2008 não foram observadas perdas, sendo a frequência observada zero. As frequências de perdas relativas às perdas da Tabela 4.2 aparecem na Tabela 4.2. A série de severidade das perdas está na Tabela 4.4. Observe que no dia 03/01/2008 não houve perdas, de forma que um valor zero aparece na série de frequência (a série de severidade só possui valores positivos).

Tabela 4.2 - Eventos de perda de um evento hipotético

Data	Perda
01/01/2008	1121,68
01/01/2008	1002,85
02/01/2008	884,73
04/01/2008	937,58
04/01/2008	949,54
04/01/2008	974,59

Tabela 4.3 - Série de frequência dos eventos de perda da Tabela 4.2

Data	Frequência
01/01/2008	2
02/01/2008	1
03/01/2008	0
04/01/2008	3

Tabela 4.4 - Série de severidade dos eventos de perda da Tabela 4.2

Severidades
1121,68
1002,85
884,73
937,58
949,54
974,59

4.3.1.2 Estimação de parâmetros

Para estimar os parâmetros das distribuições os métodos mais utilizados são: momentos, máxima verossimilhança e mínimos quadrados.

As distribuições contínuas paramétricas de severidade mais utilizadas são: Normal, Log-Normal, Exponencial, Gumbel, Pareto, Rayleigh, Cauchy, Weibull, Wald (Normal Inversa), Gama e Beta.

As distribuições discretas paramétricas de frequência mais utilizadas são: Poisson, Binomial, Geométrica, Hipergeométrica, Binomial Negativa e Polya-Aeppli.

4.3.1.2.1 Método dos momentos

O método dos momentos é um dos métodos de estimação mais simples e mais antigos conhecido. Ele consiste no equacionamento de momentos amostrais com momentos populacionais não observáveis de forma que ao resolvermos estas equações obtemos estimadores para os parâmetros.

Seja $E[X_i^r]$ o r -ésimo momento (populacional) de uma variável aleatória X_i , cuja função densidade de probabilidade $f(x|\vec{p})$ possui k parâmetros $\vec{p} = \{p_1, p_2, \dots, p_k\}$. $E[X_i^r]$ depende de \vec{p} , isto é, $E[X_i^r] = E[X_i^r, \vec{p}]$. Seja:

$$m_r = \frac{1}{n} \times \sum_{i=1}^n x_i^r$$

o r -ésimo momento amostral de uma amostra observada de X_i de n valores $\{x_1, x_2, \dots, x_n\}$. Os valores dos parâmetros $\vec{p}' = \{p'_1, p'_2, \dots, p'_k\}$ que satisfazem o conjunto de equações

$$\begin{cases} E[X_i^1, \vec{p}'] = m_1 \\ E[X_i^2, \vec{p}'] = m_2 \\ \vdots \\ E[X_i^k, \vec{p}'] = m_k \end{cases}$$

são os estimadores dos parâmetros \vec{p} de $f(x|\vec{p})$ pelo método dos momentos.

Como exemplo, suponha a distribuição de *Poisson*, cuja função densidade de probabilidade depende de um único parâmetro, λ : $f(x|\lambda) = \lambda \times e^{-\lambda \times x}$. O conjunto de equações se resume a uma única neste caso:

$$E[X] = \frac{1}{n} \times \sum_{i=1}^n x_i = \bar{X}$$

$E[X]$ pode ser calculado pela função densidade de probabilidade de X :

$$E[X] = \int_{-\infty}^{+\infty} x \times f(x) \times dx = \frac{1}{\lambda}$$

Assim, o estimador de λ segundo o método dos momentos é:

$$\lambda' = \frac{1}{\bar{X}}$$

4.3.1.2.2 Método da máxima verossimilhança

Seja $\{x_1, x_2, \dots, x_n\}$ uma amostra de n valores de uma distribuição de parâmetros $\vec{p} = \{p_1, p_2, \dots, p_k\}$ e função densidade de probabilidade $f(x_i|\vec{p})$. A função de verossimilhança de X com relação à amostra é dada por:

$$L(\vec{p}) = \prod_{i=1}^n f(x_i|\vec{p})$$

Os estimadores $\vec{p}' = \{p'_1, p'_2, \dots, p'_k\}$ do conjunto de parâmetros \vec{p} de $f(x_i|\vec{p})$ são aqueles que maximizam a função de verossimilhança $L(\vec{p})$.

4.3.1.2.3 Método dos mínimos quadrados

Seja $\{x_1, x_2, \dots, x_n\}$ uma amostra de n valores de uma distribuição de parâmetros $\vec{p} = \{p_1, p_2, \dots, p_k\}$ e função distribuição acumulada $F(x_i|\vec{p})$. Os estimadores $\vec{p}' = \{p'_1, p'_2, \dots, p'_k\}$ do conjunto de parâmetros \vec{p} de $F(x_i|\vec{p})$ são aqueles que minimizam a função

$$M = \sum_{i=1}^n \left(F(x_i|\vec{p}') - \frac{2 \times i - 1}{2 \times n} \right)^2$$

4.3.1.3 Teste de aderência

Para verificar a aderência entre a distribuição paramétrica teórica estimada e a distribuição empírica (dos dados históricos), utiliza-se o teste do Qui-quadrado (χ^2) para as distribuições discretas e os testes de Anderson-Darling (AD) e Kolmogorov-Smirnov (KS) para as distribuições contínuas. Os testes de aderência são testes de hipótese cuja hipótese nula é que a distribuição empírica é igual à distribuição teórica paramétrica estimada.

Outras ferramentas utilizadas na verificação da aderência são os gráficos *QQ-Plot* e *PP-Plot* (verificação visual).

4.3.1.3.1 Teste do Qui quadrado (χ^2)

Essa forma de testar a aderência foi desenvolvida por Karl Pearson e baseia-se na estatística:

$$\chi_v^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

, onde:

χ_v^2 é a estatística do teste, com v graus de liberdade;

O_i é a frequência observada dos dados empíricos da classe i ;

E_i é frequência esperada, segundo a distribuição paramétrica teórica da classe i ;

k é número de classes ou valores considerados

Pearson mostrou que, se a hipótese nula for verdadeira e se todas as classes i forem tais que $E_i \geq 5$, a estatística acima definida terá aproximadamente distribuição χ^2 com $v=k-1-m$ graus de liberdade, sendo k o número de classes e m o número de parâmetros estimados independentemente a partir da amostra. Caso existam classes com $E_i < 5$, estas deverão ser fundidas às classes adjacentes.

O cálculo de frequências esperadas é feito através da expressão $E_i = n \times p_i$, onde p_i é a probabilidade, segundo a distribuição teórica, de se obter um valor da variável na classe considerada, e n é o número de elementos da amostra.

O teste é unilateral, sendo que a hipótese nula será rejeitada se $\chi_v^2 \geq \chi_{v,\alpha}^2$, sendo $\chi_{v,\alpha}^2$ o valor crítico para o nível de significância α . O nível de significância corresponde à probabilidade de cometermos o erro de rejeitar a hipótese nula quando ela é verdadeira.

Alternativamente, ao invés de trabalharmos com valores críticos podemos trabalhar com o conceito de *p-valor*. O *p-valor* corresponde à probabilidade de uma variável com distribuição qui-quadrado com v graus de liberdade ser maior que a estatística do teste de aderência:

$$p - valor = P(\chi^2 > \chi_v^2)$$

Ou seja, para qualquer nível de significância maior que o p-valor a hipótese nula é rejeitada. Quanto mais próximo de 1 o valor do *p-valor* menor a chance de rejeitarmos a hipótese nula.

4.3.1.3.2 Teste Kolmogorov-Smirnov

Kolmogorov e Smirnov desenvolveram um método para testar a aderência, em que a variável de teste é a maior diferença observada entre a função distribuição de probabilidades (acumulada) do modelo e a da amostra. A estatística do teste é simplesmente

$$KS = \max |F_e(x) - F_t(X)|$$

, onde:

$F_e(x)$ é a função distribuição acumulada empírica (das amostras);

$F_t(X)$ é a função distribuição acumulada teórica (da distribuição teórica).

A estatística *KS* do teste pode ser comparada como valores críticos tabelados (em função do tamanho amostral e do nível de significância). Se *KS* for maior que o valor crítico a hipótese nula é rejeitada. O p-valor do teste de Kolmogorov-Smirnov pode ser calculado por simulação numérica.

4.3.1.3.3 Teste Anderson Darling

O teste de aderência de Anderson-Darling é uma modificação do teste de Kolmogorov-Smirnov. Enquanto o teste de Kolmogorov-Smirnov consiste em obter a máxima discrepância entre as funções distribuição de probabilidade teórica e empírica, o de Anderson-Darling considera as discrepâncias em todo o domínio da função distribuição de probabilidade. Além disso, este teste dá mais peso às caudas da distribuição. A estatística do teste é dada por:

$$AD = n \times \int_{-\infty}^{+\infty} \frac{(F_e(x) - F_t(x))^2}{(F_t(x) \times (1 - F_t(x)))} \times f_t(x) \times dx$$

, onde:

$F_e(x)$ é a função distribuição acumulada empírica (das amostras);

$F_t(x)$ é a função distribuição acumulada teórica.

$f_t(x)$ é a função densidade de probabilidade teórica.

Para um conjunto de amostras crescentemente ordenado $\{x_1, x_2, \dots, x_n\}$ pode-se escrever AD como:

$$AD = -n - \sum_{i=1}^n \frac{(2 \times i - 1)}{n} \times [\ln(F_t(x_i)) + \ln(1 - F_t(x_{n+1-i}))]$$

O teste de Anderson-Darling é mais sensível que o teste de Kolmogorov-Smirnov, mas possui desvantagem de que valores críticos são dependentes da distribuição teórica do teste. O p-valor do teste pode ser calculado por simulação numérica.

4.3.1.4 Cálculo da distribuição agregada – Simulação de Monte Carlo

Métodos de Monte Carlo compõem uma classe de algoritmos computacionais para simulação do comportamento de vários sistemas físicos e matemáticos. Eles fornecem soluções aproximadas para uma variedade de problemas realizando experimentos de amostragem estatística em computadores e se distinguem de outros métodos de simulação por serem estocásticos (não determinísticos de alguma forma), geralmente utilizando números pseudo-aleatórios. O uso destes métodos ocorre em diversas áreas, sendo um exemplo clássico o seu uso no cálculo de integrais definidas, particularmente integrais multidimensionais com condições de contorno complicadas.

Na LDA, a partir das distribuições estatísticas de frequência e severidade pode-se obter a distribuição agregada de perdas para cálculo do VAR operacional. Não existe de modo geral uma solução analítica para a distribuição agregada de perdas. Uma SMC fornece uma forma simples de obtermos esta distribuição.

A SMC utilizada em risco operacional é simples e requer somente um bom gerador de números pseudo-aleatórios e poder computacional. O algoritmo consiste na repetição de um número n grande (por exemplo, $n=10.000$) de vezes dos três passos seguintes:

- Gera-se um número f segundo a distribuição de frequência das perdas;
- Geram-se f números $\{s_1, s_2, \dots, s_f\}$ segundo a distribuição de severidade das perdas;
- Calcula-se a amostra simulada da perda agregada: $p = \sum_{i=1}^f s_i$.

O número n de simulações deve ser suficientemente grande para que a distribuição agregada seja povoada com um número grande de pontos. Para calibrar o número de simulações, inicialmente calcula-se o VAR para um número inicial de simulações, por exemplo, 10.000. A seguir aumenta-se n e calcula-se novamente o VAR. Se a diferença entre os dois valores de VAR obtidos for superior a um determinado ε , aumenta-se ainda mais o valor de n e recalcula-se o VAR, aceitando-se o VAR obtido quando a diferença for inferior a ε .

4.3.1.5 Cálculo do VAR operacional

Ao final da SMC teremos n amostras simuladas de perdas, $\{p_1, p_2, \dots, p_n\}$, representando a sua distribuição agregada. Ordenado esta amostra de forma crescente, $\{p'_1, p'_2, \dots, p'_n\}$, podemos obter o VAR operacional diretamente através do ponto correspondente ao percentil desejado. Por exemplo, se $n=1.000$ o VAR com 99% de confiança é p_{991} (o valor de p_{990} é o divisor entre 89,9% e 99%, de forma conservadora utiliz-se p_{991});

4.4 Risco Operacional no Brasil

O BCB, como signatário do acordo da Basileia, publicou a resolução 3.380 em 29 de junho de 2006, dispondo sobre a implementação de estrutura de gerenciamento de RO para todas as instituições financeiras e demais instituições autorizadas a funcionar pelo BCB. Nesta resolução, define-se RO de maneira idêntica à definição do BIS e colocam-se exemplos de eventos de risco operacional.

O BCB divulgou, em 29 de agosto de 2007, a resolução 3490, dispondo sobre a apuração do patrimônio de referência exigido (PRE), que nada mais é que o capital que deve ser alocado para fazer frente aos riscos das instituições financeiras. O PRE é composto de algumas parcelas, entre elas a parcela P_{OPR} relativa aos riscos operacionais.

4.4.1 Cronograma de implantação

Com o objetivo de observar os critérios mais adequados aos níveis de riscos associados às operações conduzidas pelas instituições financeiras para fins de requerimento de capital regulamentar, o BCB divulgou em setembro de 2007 o comunicado 16.137, com um cronograma de implantação do Basileia II no Brasil. As recomendações do Basileia II foram adaptadas às condições, peculiaridades e estágio de desenvolvimento do mercado brasileiro. O cronograma divulgado para riscos operacionais foi:

- até o final de 2007: estabelecimento de parcela de requerimento de capital;
- até o final de 2009: divulgação dos pontos-chave para modelos internos de apuração de requerimento de capital;
- até o final de 2011: estabelecimento dos critérios de elegibilidade para adoção de modelos internos de apuração de requerimento de capital e

divulgação do processo de solicitação de autorização para uso de modelos internos de apuração de requerimento de capital;

- até o final de 2012: início do processo de autorização para uso de modelos internos de apuração de requerimento de capital para risco operacional.

Em 30 de abril de 2008, foi publicada a circular 3.383, sobre a apuração da parcela de capital regulamentar de RO (P_{OPR}). Nesta circular vemos algumas diferenças em relação às abordagens descritas no Basileia II. Três metodologias são indicadas para cálculo do P_{OPR} , ficando por conta da instituição a escolha de qual será utilizada, devendo ser comunicada a opção até 1º de junho de 2008:

- Abordagem do Indicador Básico (BIA)
- Abordagem Padronizada Alternativa (ASA)
- Abordagem Padronizada Alternativa Simplificada (ASA₂)

A apuração da parcela P_{OPR} será semestral considerando os três últimos períodos anuais (seis últimos semestres).

4.4.2 Metodologia do Indicador Básico (BIA)

O BCB utiliza uma base de cálculo da parcela de alocação de capital de RO para o indicador básico diferente da sugerida pelo novo acordo da Basileia. Ao invés da receita bruta anual, utiliza um Indicador de Exposição ao RO. Além disso, incorporou um percentual de implantação, para que as instituições financeiras pudessem alocar capital de forma progressiva, uma vez que elas nunca foram obrigadas a reservar dinheiro por conta dos seus riscos operacionais.

O Indicador de Exposição ao RO é a soma dos valores semestrais das receitas de intermediação financeira e das receitas com prestação de serviços, deduzidas as despesas de intermediação financeira, para cada período anual.

Devem ser excluídas as perdas ou ganhos provenientes da alienação de títulos e valores mobiliários e instrumentos derivativos não classificados na carteira de negociação. A carta-circular 3316 divulgada pelo BCB em 30 de abril de 2008 detalha a composição do indicador de exposição ao RO.

Assim, a parcela de capital alocado pela BIA é dada por:

$$P_{OPR}^{BIA} = Z \times \frac{1}{n} \times \sum_{i=1}^3 \max(\alpha \times IE_i, 0)$$

, onde:

$\alpha = 15\%$; IE_i : Indicador de exposição ao RO no período anual i ;

N: número de vezes, nos últimos três períodos anuais, em que IE foi positivo;

Z: percentual de implantação, de acordo com as Tabelas 4.5 e 4.6.

Tabela 4.5 - Percentual de implantação para bancos e conglomerados

Validade	Z
até dezembro de 2008	20%
até junho de 2009	50%
até dezembro de 2009	80%
a partir de janeiro de 2010	100%

Tabela 4.6 - Percentual de implantação para as demais instituições ³

Validade	Z
até dezembro de 2008	5%
até junho de 2009	20%
até dezembro de 2009	35%
até junho de 2010	50%
até dezembro de 2010	80%
a partir de janeiro de 2011	100%

³ Agências de fomento, sociedades de crédito imobiliário, associações de poupança e empréstimo, cooperativas de crédito, sociedades de arrendamento mercantil, sociedades de crédito, financiamento e investimento, sociedades corretoras de títulos e valores mobiliários, sociedades corretoras de câmbio, companhias hipotecárias e sociedades distribuidoras de títulos e valores mobiliários.

4.4.3 Metodologia Padronizada Alternativa (ASA)

No caso da ASA, da mesma forma que na BIA, o BCB utiliza como base de cálculo das linhas de negócio (exceto varejo e comercial) o indicador de exposição ao RO. Nas linhas de negócio varejo e comercial ao invés de utilizar a média total dos adiantamentos e empréstimos (como definido no Basileia II), utiliza um indicador alternativo de exposição ao RO. Além disso, assim como na BIA, incorpora o percentual de implantação Z.

O indicador alternativo de exposição ao RO é a média aritmética dos saldos semestrais das operações de crédito, de arrendamento mercantil e de outras operações com características de concessão de crédito e dos títulos e valores mobiliários não classificados na carteira de negociação, multiplicada pelo fator 0,035, para cada período anual:

$$P_{OPR}^{ASA} = Z \times \frac{1}{3} \times \sum_{i=1}^3 \max \left[\left(\sum_{j=1}^6 \beta_j \times IE_{i,j} \right) + \left(\sum_{j=7}^8 \beta_j \times IAE_{i,j} \right), 0 \right]$$

, onde:

$IAE_{i,j}$: Indicador alternativo de exposição ao RO, no período anual i , para a linha de negócio j ;

$IE_{i,j}$: Indicador de exposição ao Risco Operacional, no período anual i , para as linhas de negócio j ;

β_j : fator de ponderação aplicado à linha de negócio j , de acordo com a Tabela 4.1;

Z: percentual de implantação, de acordo com as Tabelas 4.5 e 4.6.

4.4.4 Metodologia Padronizada Alternativa Simplificada (ASA₂)

A metodologia ASA₂ não é prevista no Basileia II. Ela surge no Brasil pela dificuldade que algumas instituições financeiras têm de obter os dados de IE separados por linha de negócio. A parcela a ser alocada para RO na ASA₂ é dada pela seguinte fórmula:

$$P_{OPR}^{ASA2} = Z \times \frac{1}{3} \times \sum_{i=1}^3 \max\{[(IAE_i \times 0,15) + (IE_i \times 0,18)], 0\}$$

, onde:

IAE_i: Indicador alternativo de exposição ao RO, no período anual i, apurado de forma agregada para as linhas de negócio comercial e varejo;

IE_i: Indicador de exposição ao RO, no período anual i, apurado de forma agregada para as demais operações.

5 Redes Bayesianas em Riscos Operacionais

Este capítulo é dedicado ao uso de RBs na modelagem de Riscos Operacionais. Inicialmente uma introdução é feita, apontando os problemas da abordagem LDA e descrevendo as vantagens do uso do modelo de Redes Bayesianas. A seguir, descreve-se como estruturar uma RB de RO. Na seqüência, descreve-se a necessidade de discretização das variáveis, apontando cuidados necessários neste processo. A seguir detalha-se o processo de obtenção das probabilidades condicionais da rede, a partir de dados históricos de perdas e de informações subjetivas de especialistas. Com a estrutura e probabilidades condicionais é possível realizar inferências na RB, e também realizar análise de cenários, descritas ao final deste capítulo.

5.1 Introdução

A metodologia LDA descrita, apesar de estar no contexto da AMA apresenta alguns problemas:

- a estimativa do VAR operacional é baseada somente em dados históricos, que podem não mais refletir a situação presente;
- dados de perda não estão sempre disponíveis em quantidade suficiente para permitir uma estimativa razoável de exposição;
- supõe-se que a freqüência de um evento de perda é obrigatoriamente independente da sua severidade;
- variáveis explicativas dos eventos de perda não são levadas em conta (não é um modelo causal);
- não permite a inserção no modelo de conhecimento especialista para verificar, por exemplo, a influência de eventos raros no VAR operacional;

- não permite a análise de cenários (*what-if analysis*), muito útil no gerenciamento de riscos operacionais.

O Novo Acordo da Basileia descreve uma série de padrões quantitativos para o cálculo do capital regulatório mínimo, entre eles:

Qualquer sistema de mensuração de risco operacional deve possuir certas funcionalidades chaves para atingir o padrão sólido de supervisão iniciado nessa seção. Esses elementos devem incluir o uso de dados internos, dados externos relevantes, análise de cenários e fatores que refletem o ambiente do negócio e sistemas de controle internos (BIS, 2004, p.145, tradução nossa).

Segundo Yoon (2003), a modelagem da causalidade de perdas de RO não é apenas um arcabouço elegante para uma análise compreensiva, mas a base vital para o entendimento de como os riscos de perdas de RO aparecem dentro da estrutura e operações da organização. Ela também provê uma base na qual o gerenciamento pode intervir para conseguir a alteração desejada no perfil de risco.

A falta de dados e a complexidade de operações sugerem a inclusão de informações de especialistas. Yoon (2003) argumenta que informações qualitativas são melhores incorporadas através de métodos não lineares, como lógica difusa (*fuzzy logic*), redes neurais e redes Bayesianas.

Segundo Hoffman (2002), a lógica difusa não pode substituir métodos estatísticos robustos na *mensuração* de RO, no sentido do VAR. Redes neurais, por sua vez, são úteis na modelagem de relacionamentos complexos entre variáveis que seriam difíceis de serem modelados com métodos lineares, mas possui a desvantagem de necessitar de disponibilidade de dados.

Redes Bayesianas apresentam-se como uma solução atrativa para análise causal de riscos operacionais, tanto do ponto de vista qualitativo do gestor de riscos quanto do quantitativo do cálculo do capital regulatório, pois permitem incorporar todos os elementos citados no texto do BIS.

Segundo Cornalba e Giudici (2004), a abordagem estatística Bayesiana permite integrar, através do teorema de Bayes, diferentes fontes de informação para fornecer conhecimento unificado que permite o gerenciamento de RO e, ao mesmo tempo, determinar melhor o capital regulatório mínimo e mais sensível ao risco através do VAR operacional. As fontes de informação podem ser oriundas de dados de perdas coletados, auto-avaliação, dados externos e opinião de gestores de risco. Assim, é possível combinar dados históricos com expectativas futuras e opiniões e, ao mesmo tempo, pode-se com RBs considerar a correlação entre perdas de diferentes linhas de negócios e tipos de risco e avaliar o impacto de fatores causais.

5.2 Definição das variáveis do modelo e da estrutura da rede

Os relacionamentos entre variáveis em uma RB (definidos pelas arestas) não precisam necessariamente satisfazer uma relação de causalidade. Entretanto, se os relacionamentos forem relações causais, então as distribuições obtidas da rede serão as distribuições das frequências relativas das variáveis. Por isso, Redes Bayesianas constituem um modelo causal por natureza.

Segundo o Novo Acordo da Basileia, Indicadores Chave de Risco (ICRs) são estatísticas ou métricas que devem prover uma visão preditiva sobre a posição de risco de um banco. Esses indicadores devem ser revistos periodicamente para alertar os bancos sobre alterações que podem ser indicativas de aumento ou diminuição de risco. Os ICRs devem também refletir potenciais fontes de risco operacional. Tais indicadores podem ser, por exemplo, o número de transações falhas ou taxa de rotatividade de pessoal. A *Risk Management Association* (RMA) gerencia a *KRI Library and Services*, uma iniciativa projetada para instituições financeiras que estão interessadas em melhorar o gerenciamento de seus riscos operacionais.

Podem existir ICRs que são relativos à frequência de perdas, outros que são relativos à severidade das perdas e ICRs relativos a ambos (mistos). Na presença

de causas que afetam tanto frequência quanto severidade de perdas, frequência e severidade não são necessariamente independentes, como suposto no modelo LDA.

Para construção de uma RB para cálculo do capital regulatório mínimo de Riscos Operacionais necessita-se, então, dos ICRs, da frequência de perdas e da severidade de perdas. Deve-se tomar extremo cuidado no mapeamento dos ICRs. Conforme visto em 3.6 - *Causalidade*, causas comuns ocultas podem fazer com que a condição de *Markov* não seja satisfeita. Scandizzo (2005) descreve uma metodologia para o mapeamento de risco operacional com o objetivo de identificar os riscos inerentes nos diferentes passos de um processo de negócio.

Por outro lado, a existência na RB de ICRs que na realidade não são causas da frequência ou severidade das perdas não é necessariamente um problema: as distribuições reais das conseqüências condicionais às causas neste caso são eqüiprováveis. Entretanto, se as probabilidades condicionais forem obtidas de dados históricos e estes não forem abundantes, ICRs que não são causas verdadeiras das perdas podem temporariamente apresentar uma influência inexistente nas perdas no modelo.

Outras variáveis além de ICRs, frequência e severidade podem ser adicionadas caso exista interesse de gerenciamento mais ativo dos riscos operacionais. Se uma variável não for observável (não existir série histórica), então obrigatoriamente as probabilidades condicionais relativas a ela deverão ser informadas por um especialista.

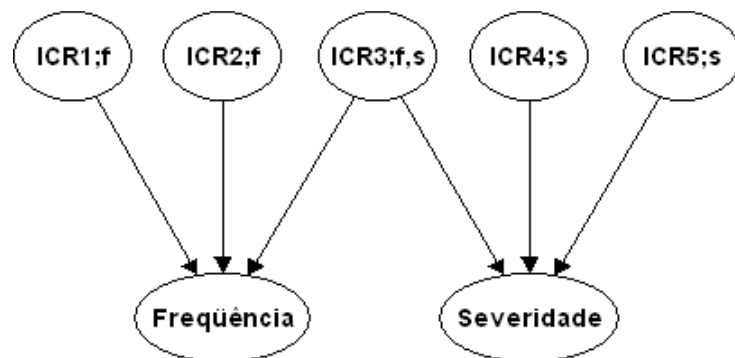


Figura 12 – RB mínima para cálculo de capital regulatório de RO: ICRs de frequência, ICRs de severidade, ICRs mistos, frequência e severidade de perdas. Neste exemplo frequência e severidade são independentes somente se o valor de ICR3;f,s é conhecido

Na existência de ICRs mistos, frequência e severidade não serão de forma geral independentes. A conexão entre frequência e severidade é feita através de conexões divergentes através dos ICRs mistos. Como visto em 3.4 – *D-Separação*, portanto, na existência de ICRs mistos, frequência e severidade só serão independentes se todos os valores dos ICRs mistos forem conhecidos. Na ausência de ICRs mistos, frequência e severidade serão sempre independentes, podendo ser modelados em Redes Bayesianas distintas.

Uma variável adicional, *perdas agregadas*, poderia ter sido colocada como filha das variáveis *frequência* e *severidade*. Como, entretanto, existe um bom algoritmo para gerar esta variável (uma SMC, descrita adiante) ela não será adicionada à Rede, uma vez esta adição obrigaria de forma desnecessária a informação de parâmetros que nem sempre são fáceis de serem obtidos nem de fácil percepção, mesmo para um especialista.



Figura 13 - RB de perdas por falha no processamento de transações bancárias.

Como exemplo, suponha a RB da Figura 13, que representa o evento de perdas por falhas no processamento de mensagens do sistema de pagamentos brasileiro (SPB). Nela vemos cinco variáveis: dois ICRs de frequência: *Taxa de erro do sistema* e *Número médio de transações*; um ICR de severidade: *Volume financeiro médio transacionado*; e as variáveis de interesse *Frequência de perdas* e *Severidade de perdas*.

5.3 Discretização das variáveis

Redes Bayesianas podem ser utilizadas com variáveis contínuas. Como visto, porém, existem grandes restrições ao seu uso. As distribuições condicionais têm que ser Gaussianas (dada uma configuração dos pais, a distribuição do filho é uma combinação linear de distribuições normais). Além disso, existe a restrição estrutural de não ser permitido que uma variável contínua tenha filhos discretos.

Redes Bayesianas são mais utilizadas com variáveis discretas. Variáveis contínuas podem ser *discretizadas*, isto é, seus valores podem ser agrupados em um número finito de estados discretos disjuntos. Por exemplo, supondo uma variável que possa assumir valores contínuos o intervalo $[a, b]$, pode-se dividir este intervalo em r estados discretos de largura fixa $(b-a) / r$.

Um cuidado na discretização deve ser tomado. Quanto maior o número de estados discretos de uma variável, maior será sua tabela de probabilidades condicionais. Um nó que tenha, por exemplo, 20 estados e possua dois pais, cada um com mais 20 estados vai apresentar uma tabela de probabilidades condicionais com $20^3 = 8000$ probabilidades condicionais. O tamanho da tabela é um problema por dois motivos:

- os algoritmos de inferência podem não ser eficientes o suficiente em termos de espaço de armazenamento e velocidade de processamento;
- o aprendizado das probabilidades condicionais a partir de dados históricos pode ser prejudicado, já que o número de observações históricas utilizadas pode ser insuficiente.

Se o número de casos históricos for pequeno e o número de probabilidades condicionais a serem obtidas grande, algumas combinações dos estados dos pais podem não ter sido observadas no passado. Neste caso, as probabilidades condicionais ou serão obrigatoriamente informadas por um especialista ou então serão configuradas distribuições uniformes, demonstrando falta de conhecimento *a priori*.

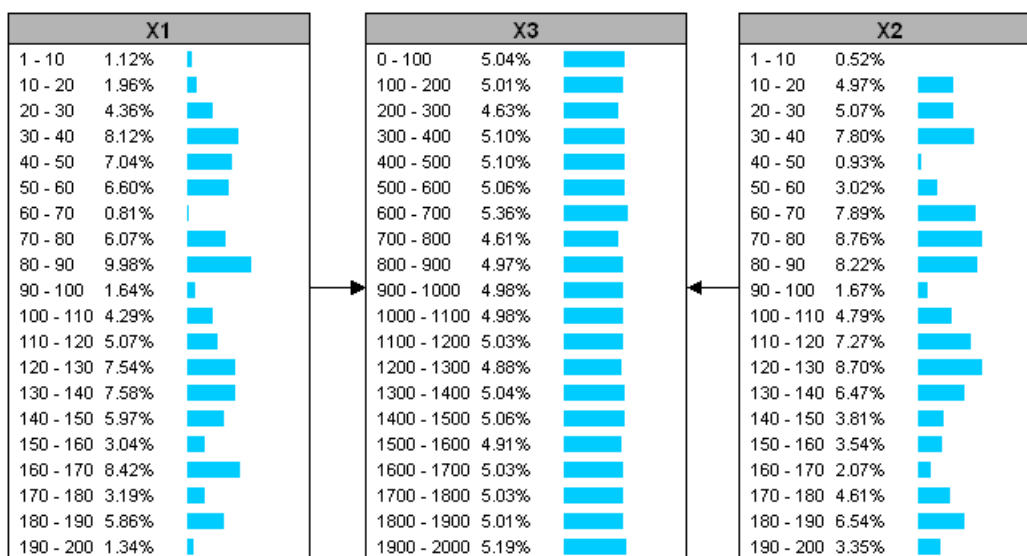


Figura 14 – Variáveis com 20 estados discretos cada. A tabela de probabilidades condicionais de X_3 (que possui X_1 e X_2 como pais) possui 8000 valores.

Por outro lado, se o número de estados discretos for muito pequeno, outros problemas aparecerão. Se um ICR for discretizado em classes de forma que, dentro de uma mesma classe existam valores do ICR que causem um intervalo grande de valores na frequência ou severidade das perdas, então a informação do estado atual deste ICR não trará informação sobre as perdas em um nível satisfatório. Se a frequência ou severidade das perdas for discretizada com um número pequeno de classes discretas, de forma análoga, a distribuição das perdas agregadas obtida através da SMC será prejudicada. Saber, por exemplo, que a probabilidade do número de perdas está entre zero e vinte fornece uma medida de VAR muito imprecisa.

A Figura 15 mostra a Rede Bayesiana da Figura 13 com as variáveis já discretizadas: a variável *Taxa de erro* com cinco estados disjuntos: *0% a 5%*, *5% a 10%*, *10% a 20%*, *20% a 50%* e *maior que 50%*; a variável *Número médio de transações* com cinco estados disjuntos: *1.000 a 2.000*, *2.000 a 3.000*, *3.000 a 4.000*, *4.000 a 5.000* e *maior que 5.000*; a variável *Volume financeiro médio* com cinco estados disjuntos: *0 a 10.000*, *10.000 a 20.000*, *20.000 a 50.000*, *50.000 a 100.000* e *maior que 100.000*; a variável *freqüência de perdas* com seis estados disjuntos: *0*, *1*, *2*, *3*, *4* e *maior que 4*; e a variável *severidade de perdas* com quatro estados disjuntos: *0 a 1.000*, *1.000 a 2.000*, *2.000 a 3.000* e *maior que 3.000*.

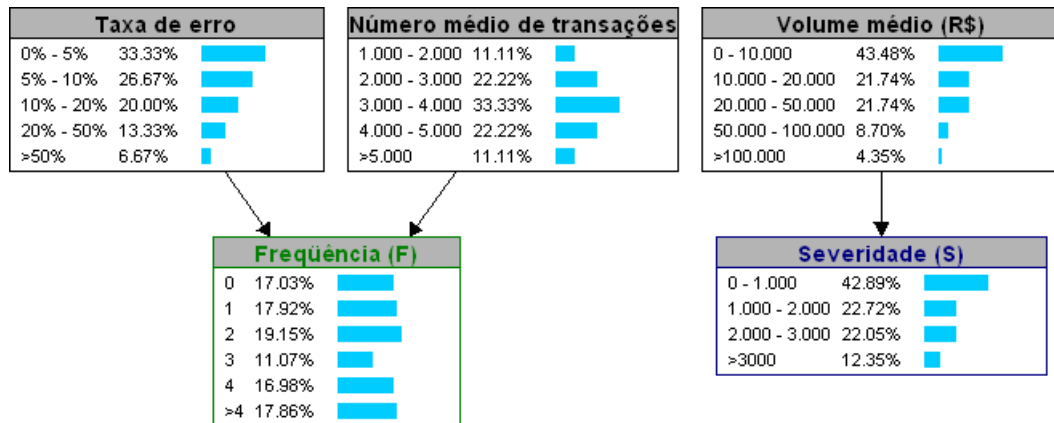


Figura 15 - RB de perdas por falhas no processamento de transações bancárias com as variáveis discretizadas.

5.4 Obtenção dos parâmetros da RB

O terceiro passo da modelagem causal com Redes Bayesianas consiste em obter, para cada variável, as distribuições de probabilidades de cada nó condicionais aos seus pais. Para vértices da rede que não possuem pais, são necessárias as probabilidades de cada um dos seus estados. Para vértices que possuem pais são necessárias as probabilidades de cada estado do vértice condicionais a cada uma das combinações dos estados dos seus pais.

Como exemplo, suponha a RB de perdas por falhas em transações bancárias da Figura 15. Para o nó *Taxa de erro*, que não possui pais, são necessárias as probabilidades incondicionais de seus estados, $P(0\% \text{ a } 5\%)$, $P(5\% \text{ a } 10\%)$, $P(10\% \text{ a } 20\%)$, $P(20\% \text{ a } 50\%)$ e $P(\text{maior que } 50\%)$. Para o nó *Frequência*, que possui dois pais (*Taxa de erro* e *Número médio de transações*) cada um com cinco estados, existem 25 combinações de estados possíveis de seus pais, sendo necessárias então $6 \times 25 = 150$ probabilidades condicionais. Para o nó *Severidade*, que possui quatro estados e um pai com cinco estados, são necessárias 20 probabilidades. A Figura 16 exhibe um exemplo de probabilidades condicionais do nó *Severidade*.

Probabilidades Condicionais				
Volume médio (R\$)	0 - 1.000	1.000 - 2.000	2.000 - 3.000	>3000
0 - 10.000	0.727272727	0.181818182	0.060606061	0.03030303
10.000 - 20.000	0.014117219	0.346757950	0.351990286	0.287134545
20.000 - 50.000	0.330338848	0.281422596	0.363609856	0.024628700
50.000 - 100.000	0.287036513	0.062995348	0.270510543	0.379457596
>100.000	0.295861972	0.139525641	0.345358410	0.219253978

Figura 16 - Exemplo de probabilidades condicionais do nó *Severidade* da Figura 15.

Para cada uma das q_i combinações dos estados dos pais pa_{ij} de um nó X_i o somatório das probabilidades de cada um dos r_i estados deve somar um, isto é:

$$\sum_{k=1}^{r_i} P(X_i = x_{ik} | pa_{ij}) = 1$$

Para obter os parâmetros existem três possibilidades:

- utilizar somente uma base dados históricos (*batch learning*);
- todas as probabilidades são informações subjetivas de especialistas;
- mesclar base de dados históricos com informações subjetivas de especialistas.

Como visto em 3.8, *Aprendizado de Parâmetros para variáveis discretas*, na utilização de base de dados históricos qualquer um dos dois algoritmos citados pode ser utilizado. Ambos permitem incluir opiniões de especialistas.

No caso da não utilização de base de dados históricos, todas as probabilidades devem ser informadas por especialistas. Na prática, dependendo do número de probabilidades não utilizar base histórica pode ser inviável.

O uso de informação subjetiva sobre as distribuições condicionais das variáveis da rede permite enriquecer as informações fornecidas pela base de dados histórica. O peso dado para as informações subjetivas em relação aos dados históricos pode ser configurado em ambos os algoritmos, sendo expresso como um tamanho de amostra equivalente, que indica o grau de confiança nas probabilidades condicionais subjetivas. É como se as informações do especialista tivessem sido

obtidas de uma amostra igual ao tamanho de amostra equivalente informado. Quanto maior esse tamanho de amostra equivalente, maior a influência dos dados subjetivos comparativamente à influência dos dados históricos.

O uso de informações subjetivas é também muito interessante em eventos de baixa probabilidade de ocorrência, onde existem poucas ocorrências históricas ou mesmo nenhuma. Este é um fato que tornam as Redes Bayesianas atrativas no gerenciamento de riscos operacionais, visto que em muitas situações os dados de perdas são escassos e especialistas têm como incorporar conhecimento ao modelo.

5.5 Inferência

Como visto em 3.7.1 – *Inferência com variáveis discretas*, o processo de inferência em uma RB permite obter as distribuições de todas as variáveis X_i da rede condicionais a determinado conjunto C de variáveis. No caso de uma RB de RO, para cálculo do capital regulatório mínimo, interessam as distribuições de frequência e severidade condicionais aos valores atuais dos ICRs.

Se a inferência for realizada sem utilizar o conhecimento dos valores atuais dos ICRs, as distribuições de frequência e severidade obtidas serão tais que consideram todos os possíveis valores dos ICRs e suas probabilidades de ocorrência. É como se todos os possíveis valores históricos de frequência e severidade fossem considerados, de forma análoga à LDA. Caso algum ICR seja desconhecido no momento do cálculo (e os demais conhecidos), ainda assim existem ganhos por se considerar os valores atuais dos ICRs conhecidos.

A Figura 17 mostra a RB de perdas por falhas no processamento de transações bancárias com os valores dos ICRs instanciados. As distribuições de frequência e severidade condicionais aos ICRs são exibidas.

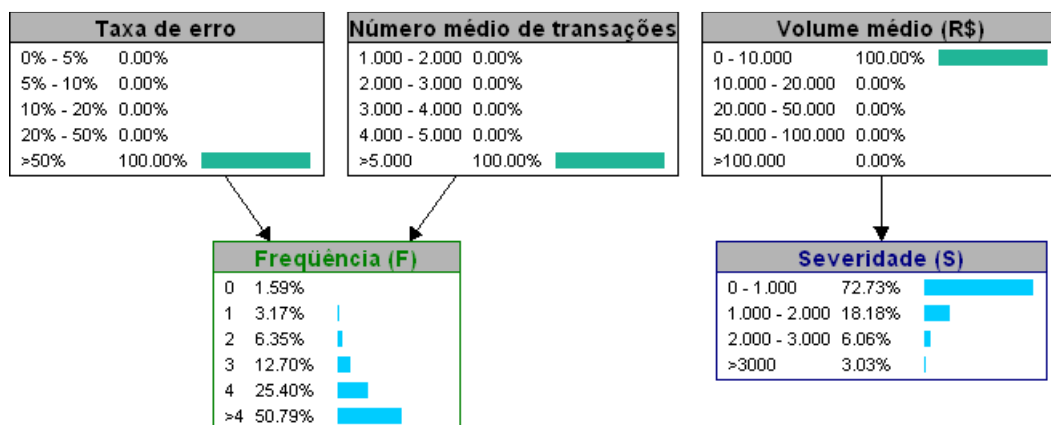


Figura 17 – RB após inferência com informações sobre valores atuais dos ICRs. As distribuições de frequência e severidade de perdas são condicionadas aos valores conhecidos dos ICRs.

5.6 Obtenção da distribuição agregada e cálculo do VAR operacional

Uma vez obtidas as distribuições de frequência e severidade de um determinado tipo de evento de perda em uma linha de negócio, pode-se obter a distribuição agregada de perdas e com ela calcular a perda esperada, a perda inesperada e o VAR operacional para qualquer nível de confiança desejado.

Para calcular o VAR operacional com nível de confiança α e um horizonte de tempo HT a partir das distribuições de frequência e severidade relativas ao horizonte de tempo HT realizam-se os seguintes passos:

1. Sorteia-se um número aleatório segundo a distribuição de frequência obtida na inferência da RB, $P(\text{Frequência})$. Seja f esse número;
2. Sorteiam-se f números aleatórios $\{s_1, s_2, \dots, s_f\}$ segundo a distribuição de severidade condicionada ao valor f da distribuição de frequência, $P(\text{Severidade} \mid \text{Frequência} = f)$;
3. Calcula-se o valor da perda agregada $P = \sum_{i=1}^f s_i$;
4. Repetem-se os passos de um a três por um número grande N de vezes (por exemplo, $N = 100.000$), obtendo-se uma série simulada de perdas agregadas $\{P_1, P_2, \dots, P_N\}$;
5. Ordena-se a série de perdas simuladas obtida no passo quatro;

6. Calcula-se $j = \text{int}(1 + N * \alpha)$, onde $\text{int}(x)$ representa a parte inteira do número real x . O VAR com nível de confiança α será o j -ésimo elemento da série obtida no passo quatro.

No passo dois, se não existirem ICRs mistos ou se todos os ICRs mistos forem conhecidos então $P(\text{Sev} | \text{Freq}) = P(\text{Sev})$, isto é, frequência e severidade são independentes. Este passo difere da SMC da LDA, que sempre utiliza $P(\text{Sev})$ ao invés de $P(\text{Sev} | \text{Freq})$.

O número N de simulações deve ser grande o suficiente de maneira que a diferença entre o VAR calculado com as N simulações e o VAR calculado com um número um pouco maior que N simulações não seja muito grande, ou seja, $|\text{VAR}(N) - \text{VAR}(N + \Delta N)| < \epsilon$.

5.7 Simulações com a RB de Risco Operacional

A RB permite análise de cenários. De forma gráfica e intuitiva, é possível observar como se alteram as distribuições de frequência e severidade das perdas (e conseqüentemente o VAR operacional) para quaisquer valores de suas causas, os ICRs. As ferramentas de RB disponíveis recalculam as novas distribuições automaticamente após a entrada de evidências na rede.

Por exemplo, na RB da Figura 18, existem 125 combinações de valores de ICRs. A Rede poderia ser utilizada para calcular os valores de VAR para qualquer uma das 125 combinações. Se o gestor de riscos operacionais sabe, por algum motivo, que existe uma boa chance de algum ICR aumentar ou diminuir, basta que ele instancie na RB o ICR com seu possível valor para verificar o valor de VAR correspondente.

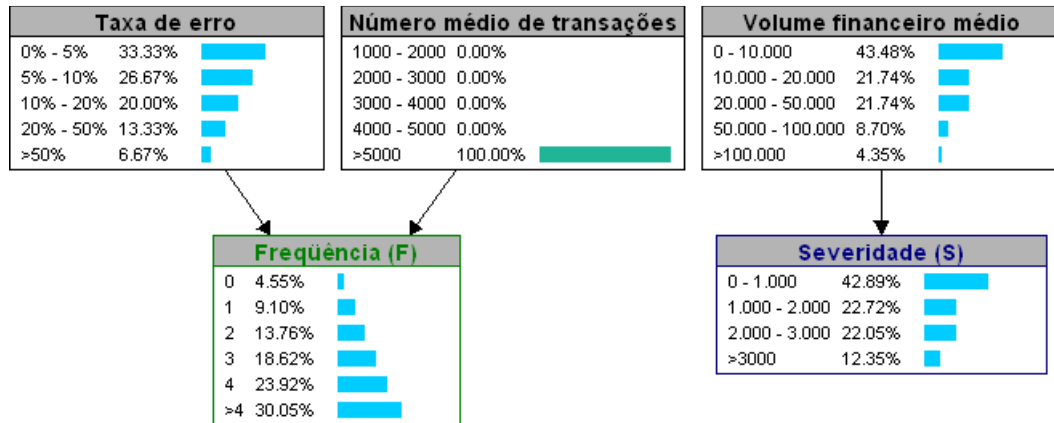


Figura 18 – Rede da Figura 15 com ICR *Número médio de transações* instanciado no estado >5000 para verificação das novas distribuições de frequência e severidade condicionais ao valor do ICR.

Neste ponto já sabemos como obter a estrutura e as probabilidades de cada variável condicionais aos seus pais de uma RB em RO. Sabemos também realizar inferências para obter a distribuição agregada de perdas, que por sua vez nos permite o cálculo da perda esperada, da perda inesperada e do VAR operacional. Com isso, podemos aplicar o conhecimento obtido em um ou mais casos onde são conhecidas as distribuições reais das variáveis (o que na prática não acontece), para ser possível comparar o modelo de RBs com a abordagem LDA. É o que será feito no próximo capítulo.

6 Estudo de casos

O objetivo deste capítulo é demonstrar a aplicação prática de RBs em RO. Para isso, dois estudos de caso serão realizados. O primeiro deles considera variáveis com distribuições conhecidas e constantes no tempo. O segundo também considera distribuições conhecidas, mas supõe que a distribuição de um ICR se altera no período de análise. Em ambos os casos, a partir das distribuições supostas das variáveis gera-se uma amostra simulando dados reais de perdas, e esta amostra é utilizada como dado de entrada na RB e na abordagem LDA. Como os dados foram gerados a partir de distribuições conhecidas, sabemos o resultado esperado, para comparar com os resultados obtidos pelos dois modelos.

6.1 Caso 1: distribuições constantes no tempo

Neste primeiro caso, suponha a RB de RO que aparece na Figura 19, que representa o evento de perda por roubos em ATMs (*Automated Teller Machine*, as máquinas disponibilizadas aos clientes de bancos para saques em dinheiro). O nó de frequência possui dois ICRs associados, o *Índice de desemprego* e o *Número de máquinas disponíveis* (em milhares). O nó de severidade também possui dois ICRs: o *Valor médio disponível nos ATMs* (em milhares) e o *Valor não segurado dos ATMs* (em milhares). Não existem ICRs mistos.

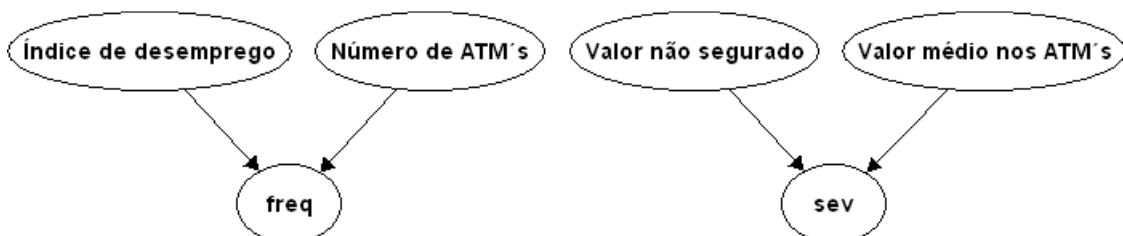


Figura 19 – RB do estudo de caso 1

Sendo *freq* e *sev* os nós de frequência e severidade, respectivamente, as seguintes distribuições constantes no tempo para as variáveis da rede foram supostas, para posterior comparação com o modelo LDA:

- *Índice de desemprego*: distribuição normal com $\mu = 5$ e $\sigma = 1$;
- *Número de ATMs*: distribuição beta com $\alpha = 2$, $\beta = 6$, $min = 0$ e $max = 10$;
- *Valor não segurado*: distribuição beta no intervalo com $\alpha = 10$ e $\beta = 3$, $min = 0$ e $max = 1000$;
- *Valor médio nos ATMs*: distribuição normal com $\mu = 1000$ e $\sigma = 200$;
- *freq*: distribuição Poisson com $\lambda = \text{Índice de desemprego} + \text{Número de ATMs}$;
- *sev*: distribuição normal com $\mu = \text{Valor não segurado} + \text{Valor médio nos ATMs}$ e $\sigma = \mu / 5$.

6.1.1 Simulação da base histórica

Para simular um banco de dados de observações históricas das variáveis da RB foram gerados dados que respeitam as distribuições das variáveis, seguindo-se os seguintes passos:

1. gera-se um número x_1 aleatório segundo a distribuição de *Índice de desemprego*;
2. gera-se um número x_2 aleatório segundo a distribuição de *Número de ATMs*;
3. gera-se um número x_3 aleatório segundo a distribuição de frequência, ou seja, um número aleatório segundo uma Poisson com $\lambda = x_1 + x_2$;
4. gera-se um número x_4 aleatório segundo a distribuição de *Valor não segurado*;
5. gera-se um número x_5 aleatório segundo a distribuição de *Valor médio nos ATMs*;

6. gera-se um número x_6 aleatório segundo a distribuição de severidade, ou seja, um número aleatório segundo uma Normal com $\mu = x_4 + x_5$ e $\sigma = \mu / 5$
7. repetem-se os passos de um a seis por N vezes, obtendo-se o banco de dados de desejado.

Para esta simulação utilizou-se $N = 100.000$. As dez primeiras linhas aparecem na tabela 6.1:

Tabela 6.1 – Dez primeiras linhas do banco de dados histórico simulado (com 100.000 linhas) do estudo de caso 1

Índice de desemprego	Número de ATMs	freq	Valor não segurado	Valor médio nos ATMs	sev
3,58	3,15	4	578,96	777,72	1378,89
4,05	1,51	5	861,21	927,36	2226,81
6,84	3,49	9	526,40	1168,10	1973,78
4,51	1,15	7	749,56	954,03	1717,10
4,39	3,94	11	917,36	1107,21	2287,13
6,13	0,65	6	674,31	1155,59	2313,20
4,09	1,15	4	830,95	881,33	1574,17
4,90	1,64	9	715,47	1132,68	1827,51
5,29	4,69	7	802,73	838,30	1842,87
4,18	2,12	1	755,83	996,70	1907,81

6.1.2 Discretização das variáveis

Para utilização da RB, todas as variáveis foram discretizadas em 10 estados de mesma largura. Os valores de *Índice de desemprego* encontram-se no intervalo $[0, 10[$, os de *Número de ATMs* em $[0, 10[$, os de frequência em $[0, 30[$, os de *Valor não segurado* em $[0, 1000[$, os de *Valor médio nos ATMs* em $[0, 2000[$ e os de severidade em $[0, 5000[$.

As 10 primeiras linhas da base de dados histórica simulada podem ser escritas em termos dos estados das variáveis ao invés dos seus valores. A Tabela 6.2 mostra os 10 primeiros registros:

Tabela 6.2 – Dez primeiras linhas do banco de dados histórico simulado com os estados de cada variável do estudo de caso 1.

Índice de desemprego	Número de ATMs	freq	Valor não segurado	Valor médio nos ATMs	sev
3 - 4	3 - 4	3 - 6	500 - 600	600 - 800	1000 - 1500
4 - 5	1 - 2	3 - 6	800 - 900	800 - 1000	2000 - 2500
6 - 7	3 - 4	9 - 12	500 - 600	1000 - 1200	1500 - 2000
4 - 5	1 - 2	6 - 9	700 - 800	800 - 1000	1500 - 2000
4 - 5	3 - 4	9 - 12	900 - 1000	1000 - 1200	2000 - 2500
6 - 7	0 - 1	6 - 9	600 - 700	1000 - 1200	2000 - 2500
4 - 5	1 - 2	3 - 6	800 - 900	800 - 1000	1500 - 2000
4 - 5	1 - 2	9 - 12	700 - 800	1000 - 1200	1500 - 2000
5 - 6	4 - 5	6 - 9	800 - 900	800 - 1000	1500 - 2000
4 - 5	2 - 3	0 - 3	700 - 800	800 - 1000	1500 - 2000

6.1.3 Aprendizado de parâmetros

A base de dados simulada foi utilizada para obtenção das probabilidades condicionais da RB. Neste estudo de caso o algoritmo utilizado foi o descrito em 3.8.1, *Método de contagem de frequências relativas*. Foram 2040 probabilidades obtidas: 1000 para frequência, 1000 para severidade e 10 para cada um dos ICRs.

6.1.4 Inferência na RB

Com a estrutura da RB e as probabilidades condicionais obtidas, é possível realizar qualquer inferência na RB. A Figura 20 ilustra as distribuições obtidas sem o conhecimento dos valores atuais dos ICRs (probabilidades não condicionais).

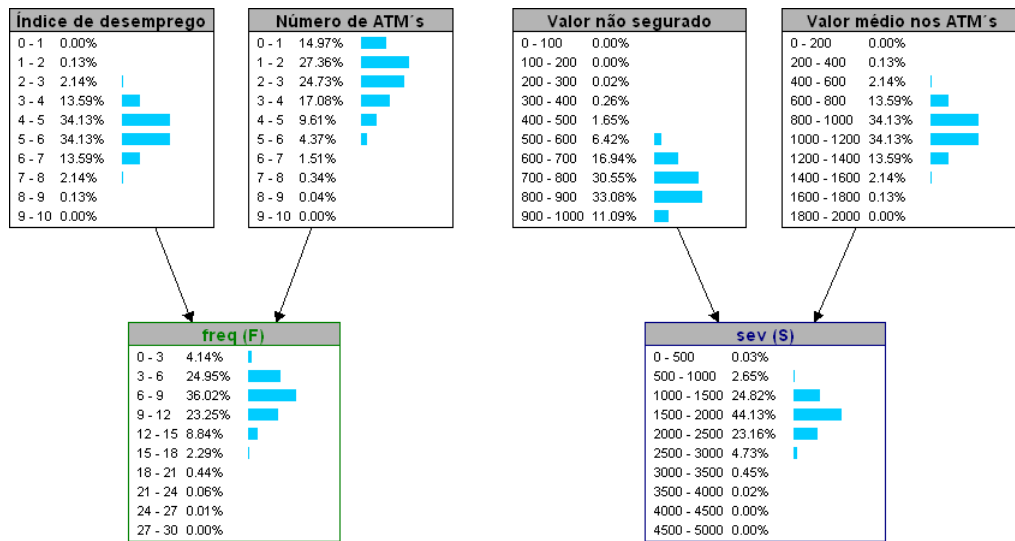


Figura 20 – RB após inferência sem conhecimento dos valores atuais dos ICRs do estudo de caso 1.

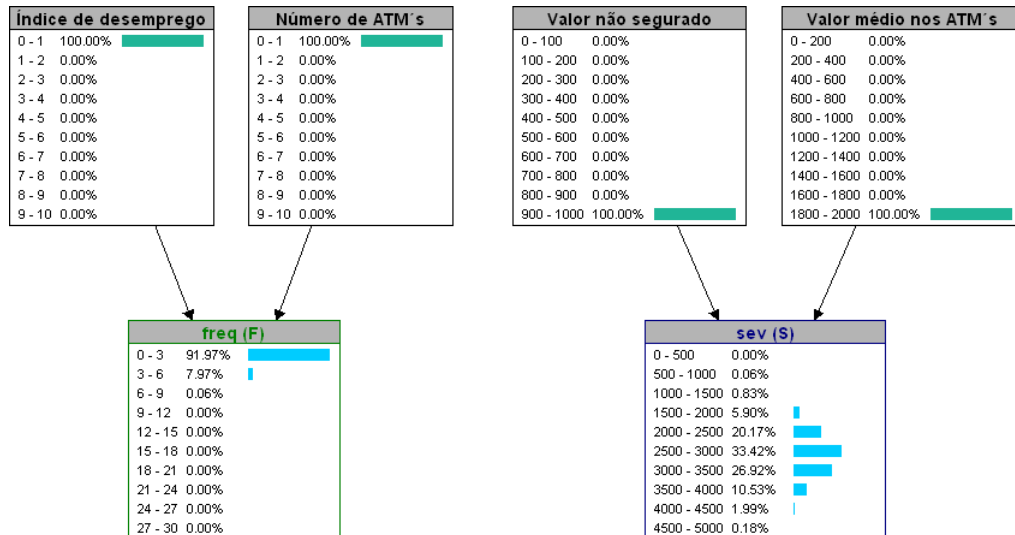


Figura 21 – RB após inferência com conhecimento dos valores atuais dos ICRs do estudo de caso 1.

Com a rede obtida é possível calcular as distribuições de frequência e severidade para todas as combinações de valores dos ICRs (no exemplo são 10.000 combinações possíveis), ou para os valores atuais dos ICRs. Como exemplo, a Figura 21 exibe as distribuições de frequência e severidade das perdas para o caso

em que *Índice de desemprego* = 0 – 1, *Número de ATMs* = 0 – 1, *Valor não segurado* = 900 – 1000 e *Valor médio nos ATMs* = 1800 – 2000:

6.1.5 Cálculo do VAR Operacional com a RB

Calculando-se o VAR operacional para os níveis de confiança 95% e 99,9%, e a perda esperada com N = 1.000.000 de simulações, são obtidos os valores da Tabela 6.3, para três situações:

1. não são conhecidos os valores atuais dos ICRs;
2. são conhecidos os valores atuais dos ICRs e eles são tais que as distribuições de frequência e severidade resultam nos menores valores de VAR e perda esperada (melhor caso);
3. são conhecidos os valores atuais dos ICRs e eles são tais que as distribuições de frequência e severidade resultam nos maiores valores de VAR e perda esperada (pior caso).

Tabela 6.3 – Resultados obtidos com a RB do estudo de caso 1

	Perda Esperada	VAR (95%)	VAR (99,9%)
ICRs desconhecidos	13.275,48	24.486,42	36.933,30
Melhor caso ⁴	310,95	918,91	1.857,04
Pior caso ⁵	53.732,29	75.028,11	87.368,77

A seguir calcularemos os valores de VAR, perda esperada e inesperada através da abordagem LDA, com os mesmos dados simulados utilizados nas RBs.

⁴ O Melhor caso corresponde ao menor valor de VAR e perda esperada, isto é, Índice de desemprego = 0 - 1, Número de ATMs = 0 - 1, Valor não segurado = 0 – 100 e Valor médio nos ATMs = 0 - 200

⁵ O pior caso corresponde ao maior valor de VAR e perda esperada, isto é, Índice de desemprego = 9 - 10, Número de ATMs = 9 - 10, Valor não segurado = 900 – 1000 e Valor médio nos ATMs = 1800 - 2000

6.1.6 Aplicação da LDA

Aplicando a abordagem *LDA* na base de dados históricos simulados de frequência e severidade das perdas as seguintes distribuições foram obtidas:

- Frequência: Binomial Negativa de parâmetros $s=18$, $p = 0.705918192$, obtida pelo método dos mínimos quadrados, com excelente aderência pelo teste do Qui-quadrado ($\chi^2 = 1.937695375$, p -valor = 0,982862709).
- Severidade: distribuição Beta de parâmetros $\alpha = 44,591620044$, $\beta = 668,852203498$; $min = -1.161,291824320$ e $max = 45.723,047380436$, obtida pelo método dos mínimos quadrados, com excelente aderência tanto pelo teste de Kolmogorov-Smirnov ($KS = 0,001674128$ e p -valor = 0,941802515) quanto pelo teste de Anderson-Darling ($AD = 0,200377082$ e p -valor = 0,990310085).

As distribuições de frequência e severidade, teóricas e empíricas aparecem nas Figuras 22 e 23. O QQ-Plot da distribuição de severidade aparece na Figura 24.

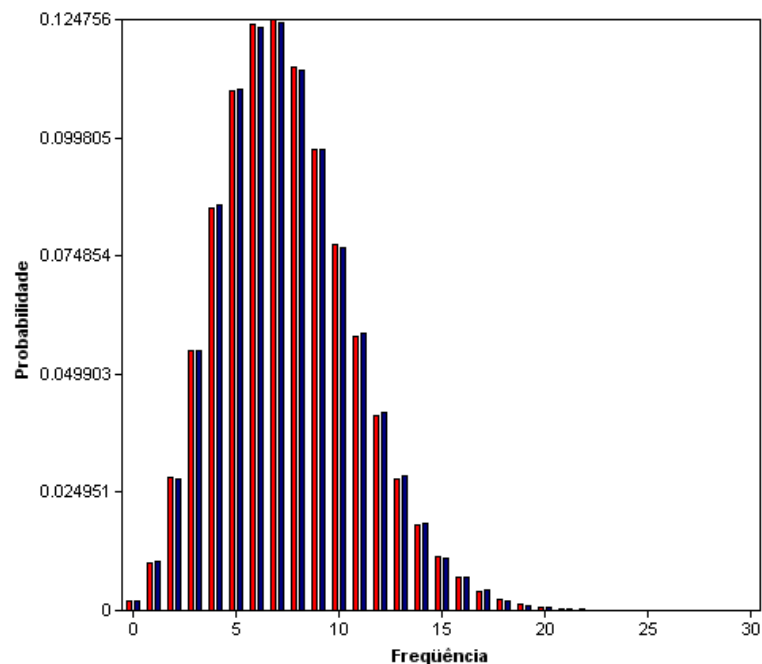


Figura 22 – Distribuições de frequência estimada e empírica do estudo de caso 1

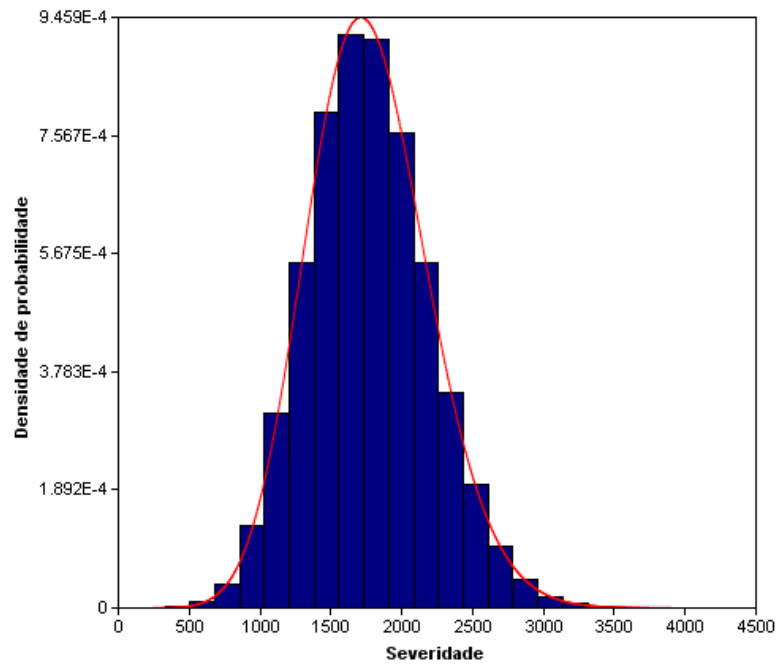


Figura 23 – Distribuições de severidade estimada e empírica do estudo de caso 1

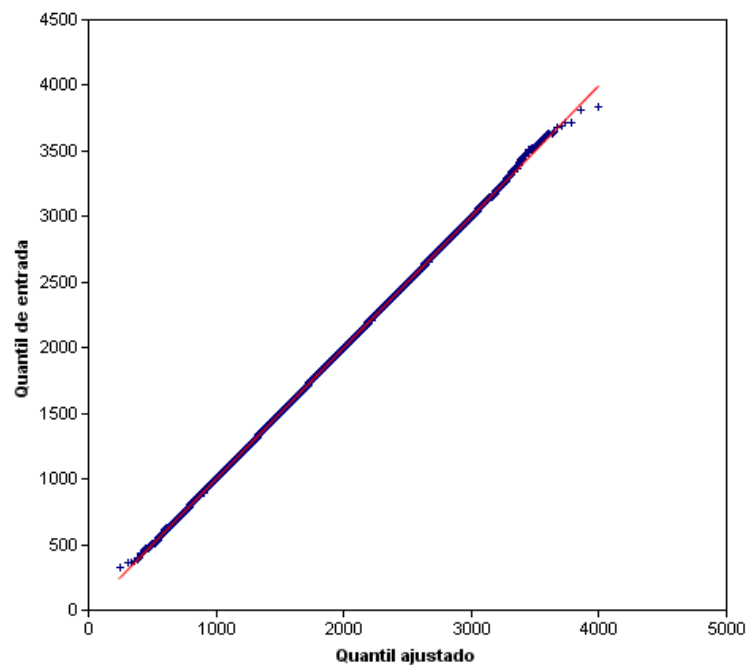


Figura 24– Q-Q Plot da distribuição estimada dos dados empíricos do estudo de caso 1.

Calculando-se a perda esperada, o VAR com 95% de confiança e o VAR com 99,9% de confiança obtém-se os dados da Tabela 6.4:

Tabela 6.4 – Resultados obtidos com o LDA aplicado à base de dados simulados do estudo de caso 1

	Perda Esperada	VAR (95%)	VAR (99,9%)
LDA	13.260,76	23.806,53	36.172,65

6.1.7 Análise e comparação dos resultados

Neste estudo de caso não existem ICRs mistos, de modo que frequência e severidade das perdas são independentes. Com isso podemos calcular o valor esperado das perdas por roubos em ATMs (considerando todos os possíveis valores dos ICRs) por:

$$\begin{aligned}
 E[\textit{perda}] &= E[\textit{freq}] \times E[\textit{sev}] = E[\lambda] \times E[\mu] \\
 &= E[\textit{Índice de desemprego} + \textit{Número de ATMs}] \\
 &\quad \times E[\textit{Valor não segurado} + \textit{Valor médio nos ATMs}] \\
 &= \left[5 + \left(\frac{2}{2+6} \times 10 \right) \right] \times \left[\left(\frac{10}{10+3} \times 1000 \right) + 1000 \right] = 13.269,23
 \end{aligned}$$

Comparando este resultado com o valor obtido pela RB quando os ICRs não estão instanciados (13.275, 48) as diferenças são pequenas. Da mesma maneira, vemos que as diferenças são pequenas ao comparar com o valor de perda esperada obtido com a LDA (13.260,76). A Tabela 6.5 ilustra as diferenças obtidas:

Tabela 6.5 – Estudo de caso 1: valores de perda esperada real, sem considerar valores auais de ICRs na RB e obtida pela LDA.

Valor Real	Redes Bayesianas	LDA
13.269,23	13.275,48	13.260,76

Calculando-se o valor esperado das perdas por roubos em ATMs considerando-se o pior caso, isto é, atribuindo *Índice de desemprego* = 9,5; *Número de ATMs* = 9,5; *Valor não segurado* = 950,0 e *Valor médio nos ATMs* = 1900,0 temos:

$$\begin{aligned}
 E[\textit{perda} \mid \textit{pior caso}] &= E[\textit{freq} \mid \textit{pior caso}] \times E[\textit{sev} \mid \textit{pior caso}] \\
 &= E[\lambda \mid \textit{pior caso}] \times E[\mu \mid \textit{pior caso}] \\
 &= E[\textit{Índice de desemprego} + \textit{Número de ATMs} \mid \textit{pior caso}] \\
 &\quad \times E[\textit{Valor não segurado} + \textit{Valor médio nos ATMs} \mid \textit{pior caso}] \\
 &= [9,5 + 9,5] \times [950,0 + 1900,0] = 54.150,00
 \end{aligned}$$

Comparando este resultado com o valor obtido pela RB quando os ICRs estão instanciados no pior caso (53.732,29) a diferença é pequena.

Calculando-se o valor esperado das perdas por roubos em ATMs considerando-se o melhor caso, isto é, atribuindo *Índice de desemprego* = 0,5; *Número de ATMs* = 0,5; *Valor não segurado* = 50,0 e *Valor médio nos ATMs* = 100,0, temos:

$$\begin{aligned}
 E[\textit{perda} \mid \textit{melhor caso}] &= E[\textit{freq} \mid \textit{melhor caso}] \times E[\textit{sev} \mid \textit{melhor caso}] \\
 &= E[\lambda \mid \textit{melhor caso}] \times E[\mu \mid \textit{melhor caso}] \\
 &= E[\textit{Índice de desemprego} + \textit{Número de ATMs} \mid \textit{melhor caso}] \\
 &\quad \times E[\textit{Valor não segurado} + \textit{Valor médio nos ATMs} \mid \textit{melhor caso}] \\
 &= [0,5 + 0,5] \times [50,0 + 100,0] = 150,0
 \end{aligned}$$

Vemos que a perda esperada calculada no melhor caso com as RBs foi diferente da real. Isto ocorreu devido à forma da discretização das variáveis. Ao agruparmos para a variável severidade os valores de 0 a 500 (cuja probabilidade de ocorrência obtida na RB de 100,0%) em um único estado a SMC utilizou o valor médio desta classe, o valor 250, com probabilidade de 100,00%, já que a SMC supõe distribuição equiprovável entre os valores da classe, por falta de informação. Na realidade, o valor esperado neste caso é 150,00.

O mesmo ocorreu com a variável frequência. Ao agruparmos os valores 0, 1 e 2 (de probabilidade de ocorrência obtida na RB de 91,97%) em um único estado a SMC utilizou o valor médio desta classe, o valor 1, com probabilidade de 91,97%. Na realidade, a probabilidade do valor 0 é de 36,79%, do valor 1 de 36,79% e do valor 2 de 18,39%, resultando em uma contribuição da classe [0, 3[para o valor esperado de 0,74. Ao agruparmos os três valores em uma única classe a contribuição da classe ao valor esperado calculada na SMC é 0,92. Da mesma maneira, para o segundo estado ([3, 6[), a probabilidade calculada pela RB foi de 7,97%, resultando em uma contribuição da classe ao valor esperado da variável obtida na SMC de 0,32. Na realidade este valor é 0,26.

O procedimento correto neste caso seria realizar novamente o processo de discretização para o estudo deste melhor caso. A RB fornece uma visualização clara de problemas de discretização, quando existem estados com probabilidade muito elevada e estados com probabilidade muito baixa, como é o caso exibido na Figura 21.

Ao discretizar novamente os nós de frequência e severidade e realizar novamente o processo de aprendizado de parâmetros temos a RB da Figura 25:

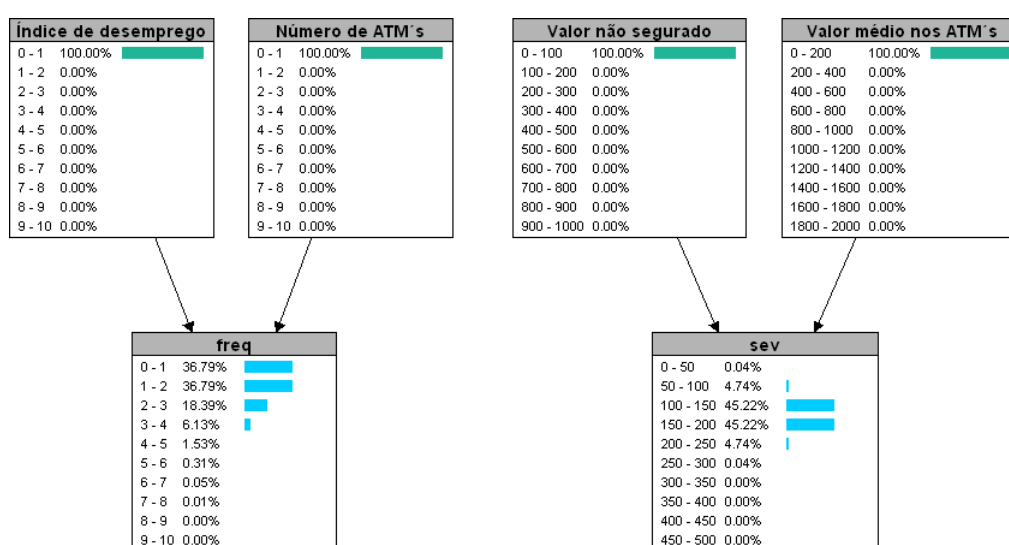


Figura 25 –RB de perdas por o evento de perda por roubos em ATMs com as variáveis de frequência e severidade rediscritizadas do estudo de caso 1.

Calculando o VAR, perda esperada e inesperada para esta nova RB temos os resultados da Tabela 6.6:

Tabela 6.6 – Estudo de caso 1: resultados obtidos com a RB da Figura 25, com as variáveis rediscritizadas.

	Perda Esperada	VAR (95%)	VAR (99,9%)
Melhor caso ⁶	150,15	448,91	828,91

O resultado agora obtido é muito mais satisfatório, uma vez que conhecemos o valor teórico da perda esperada. Na prática o valor real não é conhecido. Um extremo cuidado deve ser tomado, portanto, no processo de discretização das

⁶ O Melhor caso corresponde ao menor valor de VAR e perda esperada, isto é, Índice de desemprego = 0 - 1, Número de ATMs = 0 - 1, Valor não segurado = 0 - 100 e Valor médio nos ATMs = 0 - 200

variáveis de uma RB, e rediscritizações devem ser feitas quando alguma inferência resultar em estados com probabilidades muito elevadas, outros com probabilidades muito baixas e existe possibilidade de criar estados mais granulares.

RBs fornecem muito mais flexibilidade do que a abordagem tradicional LDA. Além disso, os valores calculados são muito mais precisos, desde que o mapeamento de ICRs esteja correto, uma vez que a LDA não utiliza informações do nível atual dos riscos operacionais, de forma que os valores calculados consideram todos os possíveis dados de perda históricos. Através dos dois casos extremos, vemos que a RB calcula um valor muito mais adequado aos níveis reais dos riscos. Se o melhor caso fosse a realidade, o valor de capital alocado pela LDA (perda inesperada com 99,9% de confiança) teria sido 3276% maior. Por outro lado, se o pior caso fosse a realidade, o valor de capital alocado pela LDA teria sido 32% menor.

6.2 Caso 2: distribuições que se alteram no tempo

Para este estudo de caso, suponha a RB da Figura 26, que mostra um evento de perdas trabalhistas. Para a frequência deste evento existe um ICR, o *Número médio mensal de demissões*. Para a severidade do evento existe também um único ICR, o *Salário anual médio* (em milhares).

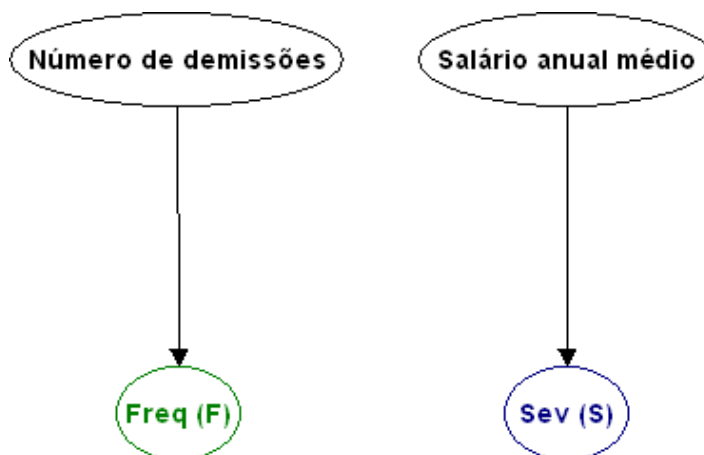


Figura 26 - RB de perdas trabalhistas (estudo de caso 2).

Para este estudo, assim como no anterior, serão supostas distribuições para todas as variáveis da RB. Ao contrário do anterior, porém, suporemos que a distribuição de uma das variáveis, *Número médio mensal de demissões*, será alterada no tempo, para refletir uma ação de mitigação deste risco com a diminuição do número médio de demissões e conseqüentemente da freqüência de perdas trabalhistas.

Os dados simulados serão correspondentes a um período de cinco anos, de janeiro de 2003 a dezembro de 2007. Nos três primeiros anos o *Número médio mensal de demissões* terá uma distribuição e a partir de janeiro de 2004 outra. A transição será abrupta. Sendo *Freq* e *Sev* os nós de freqüência e severidade, respectivamente, as seguintes distribuições para as variáveis da rede são supostas, para posterior comparação com o modelo LDA:

- *Número médio de demissões*: distribuição beta com $\alpha = 2$, $\beta = 2$, $min = 5$, $max = 10$ de janeiro de 2003 até dezembro de 2005. A partir de janeiro de 2006 até dezembro de 2007, distribuição beta com $\alpha = 2$, $\beta = 2$, $min = 2$, $max = 5$;
- *Salário anual médio*: distribuição beta com $\alpha = 3$, $\beta = 3$, $min = 900$, $max = 1100$;
- *Freq*: distribuição Poisson com $\lambda = \text{Número médio de demissões}$;
- *Sev*: distribuição normal com $\mu = \text{Salário anual médio}$ e $\sigma = \mu / 4$.

6.2.1 Simulação da base histórica

Para simular um banco de dados de observações históricas das variáveis da RB foram gerados dados de casos que respeitam as distribuições das variáveis, seguindo-se os seguintes passos:

1. gera-se um número x_1 aleatório segundo a distribuição de *Número médio de demissões*;

2. gera-se um número x_2 aleatório segundo a distribuição de freqüência, ou seja, um número aleatório segundo a Poisson com $\lambda = x_1$;
3. gera-se um número x_3 aleatório segundo a distribuição de *Salário anual médio*;
4. gera-se um número x_4 aleatório segundo a distribuição de *Valor médio nos ATMs*, ou seja, um número aleatório segundo a distribuição normal com $\mu = x_3$ e $\sigma = \mu / 4$;
5. repetem-se os passos de um a cinco por N vezes, obtendo-se o banco de dados desejado.

Para esta simulação utilizou-se $N = 1825$. As dez primeiras linhas aparecem na tabela 6.7.

Tabela 6.7 – Estudo de caso 2: dez primeiras linhas do banco de dados histórico simulado com os valores de cada variável.

Número médio de demissões	Freqüência	Salário anual médio	Severidade
7,20	3	956,59	1.280,92
6,12	6	996,68	447,34
8,57	11	1.040,09	1.195,87
8,32	14	1.077,73	378,03
5,77	5	993,58	663,43
6,85	8	1.023,59	865,74
7,72	7	986,24	1.472,86
8,30	12	1.040,87	1.120,26
9,30	13	1.051,48	683,63
6,61	8	1.029,30	1.039,68

As dez últimas linhas aparecem na tabela 6.8:

Tabela 6.8 – Estudo de caso 2: dez últimas linhas do banco de dados histórico simulado

Número médio de demissões	Frequência	Salário anual médio	Severidade
2.91	4	1.031,88	1.139,42
3.42	6	1.054,41	824,09
2.97	3	1.010,16	1.068,95
3.00	1	994,03	1.157,68
4.57	8	1.062,85	1.249,98
3.32	3	1.002,40	1.002,95
3.41	5	1.032,63	770,90
3.25	4	1.007,48	1.132,31
4.86	7	1.045,49	1.116,39
3.70	1	975,89	962,57

O gráfico de evolução temporal da média mensal das perdas agregadas obtido dos dados simulados aparece na Figura 27, evidenciando a alteração no comportamento das perdas a partir de janeiro de 2006.

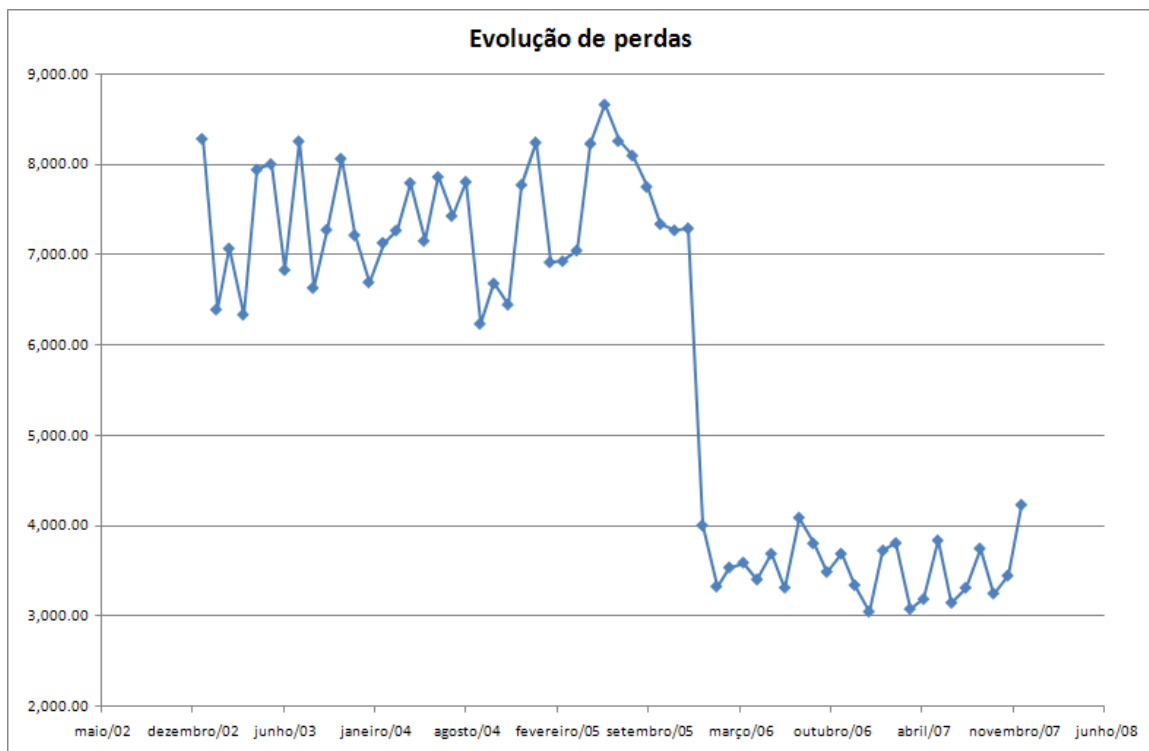


Figura 27 – Gráfico de evolução temporal das perdas agregadas mensais médias do evento de perdas trabalhistas do estudo de caso 2.

6.2.2 Discretização das variáveis

Para utilização da RB, todas as variáveis foram discretizadas em 5 estados de mesma largura. Os valores de *Número médio de demissões* encontram-se no intervalo $[0, 10[$, os de *Salário anual médio* em $[900, 1100[$, os de frequência em $[0, 20[$ e os de severidade em $[100, 2100[$.

As 10 primeiras linhas da base de dados histórica simulada podem ser escritas em termos dos estados das variáveis ao invés dos seus valores. A Tabela 6.9 mostra os 10 primeiros registros.

Tabela 6.9 – Estudo de caso 2: dez primeiras linhas do banco de dados histórico.

Número médio de demissões	Frequência	Salário anual médio	Severidade
6 - 8	0 - 4	940 - 980	900 - 1300
6 - 8	4 - 8	1020 - 1060	100 - 500
8 - 10	8 - 12	1060 - 1100	900 - 1300
8 - 10	12 - 16	980 - 1020	100 - 500
4 - 6	4 - 8	1020 - 1060	500 - 900
6 - 8	8 - 12	980 - 1020	500 - 900
6 - 8	4 - 8	1020 - 1060	1300 - 1700
8 - 10	12 - 16	1020 - 1060	900 - 1300
8 - 10	12 - 16	1020 - 1060	500 - 900
6 - 8	8 - 12	940 - 980	900 - 1300

A Tabela 6.10 mostra os 10 últimos registros.

Tabela 6.10 – Estudo de caso 2: dez últimas linhas do banco de dados histórico simulado.

Número médio de demissões	Freqüência	Salário anual médio	Severidade
2 - 4	4 - 8	1020 - 1060	900 - 1300
2 - 4	4 - 8	1020 - 1060	500 - 900
2 - 4	0 - 4	980 - 1020	500 - 900
2 - 4	0 - 4	980 - 1020	900 - 1300
4 - 6	8 - 12	1060 - 1100	1300 - 1700
2 - 4	0 - 4	980 - 1020	900 - 1300
2 - 4	4 - 8	1020 - 1060	1300 - 1700
2 - 4	4 - 8	980 - 1020	500 - 900
4 - 6	4 - 8	1020 - 1060	900 - 1300
2 - 4	0 - 4	940 - 980	900 - 1300

6.2.3 Aprendizado de parâmetros

A base de dados simulada foi utilizada para obtenção das probabilidades condicionais da RB. Neste estudo de caso o algoritmo utilizado foi o descrito em 3.8.1, *Método de contagem de freqüências relativas*. Foram 260 probabilidades obtidas: 125 para freqüência, 125 para severidade e 5 para cada um dos ICRs.

6.2.4 Inferência na RB

Com a estrutura da RB e as probabilidades condicionais obtidas, é possível realizar qualquer inferência na RB. A Figura 28 ilustra as distribuições obtidas sem o conhecimento dos valores dos ICRs (probabilidades não condicionais).

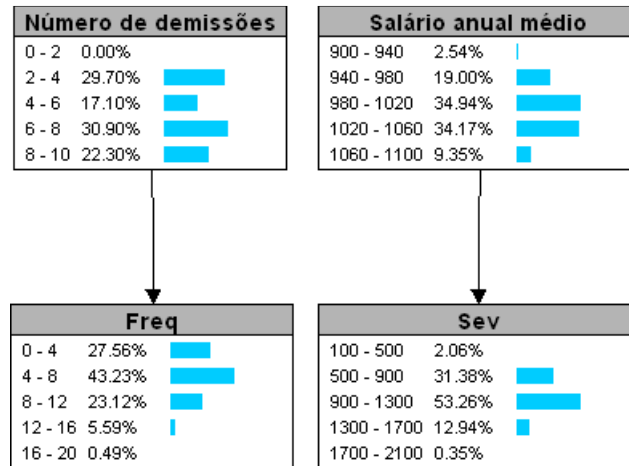


Figura 28 – RB de perdas trabalhistas após inferência sem conhecimento dos valores atuais dos ICRs.

Com a rede obtida é possível calcular as distribuições de frequência e severidade para todas as combinações de valores dos ICRs (no exemplo são 25 combinações possíveis), ou para os valores atuais dos ICRs. Como exemplo, a Figura 29 exibe as distribuições de frequência e severidade das perdas para o caso em que *Número de demissões* = 2 – 4 e *Salário anual médio* = 980 – 1020.

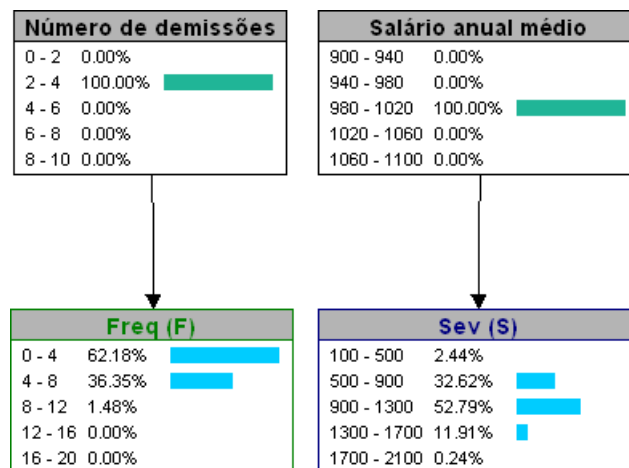


Figura 29 – RB após inferência com conhecimento dos valores dos ICRs *Número médio de demissões* e *Salário anual médio*.

6.2.5 Cálculo do VAR Operacional com a RB

Calculando-se a perda esperada e o VAR operacional para os níveis de confiança 95% e 99,9%, com $N = 1.000.000$ de simulações, são obtidos os valores da Tabela 6.11, para três situações:

1. não são conhecidos os valores dos ICRs;
2. somente o valor do ICR *Número de demissões* é conhecido, igual ao valor médio dos últimos dois anos (3,49).

Tabela 6.11 – Resultados obtidos com a RB do caso de estudo 2

	Perda Esperada	VAR (95%)	VAR (99,9%)
ICRs desconhecidos	5.903,63	12.631,72	19.203,04
ICR de frequência conhecido	3.113,61	7.379,98	11.860,57

6.2.6 Aplicação da LDA

Aplicando a abordagem *LDA* na base de dados históricos simulados de frequência e severidade das perdas as seguintes distribuições foram obtidas:

- Frequência: Polya-Aeppli de parâmetros $K = 3,881667447$ e $p = 0,664261244$ obtida pelo método dos mínimos quadrados, com estatística do teste do Qui-quadrado $\chi^2 = 12,712526788$ e p-valor = 0,122130899.
- Severidade: Beta de parâmetros $\alpha = 38,988947327$; $\beta = 58,869513401$, $min = -1.047,506749990$ e $max = 4.122,105131649$, com estatística do teste de Kolmogorov-Smirnov $KS = 0,007907606$ e p-valor = 0,521510663 e estatística do teste de Anderson-Darling $AD = 0,518817022$, p=valor = 0,727692541.

As distribuições de freqüência e severidade, teóricas e empíricas aparecem nas Figuras 30 e 31. O QQ-Plot da distribuição de severidade aparece na Figura 32.

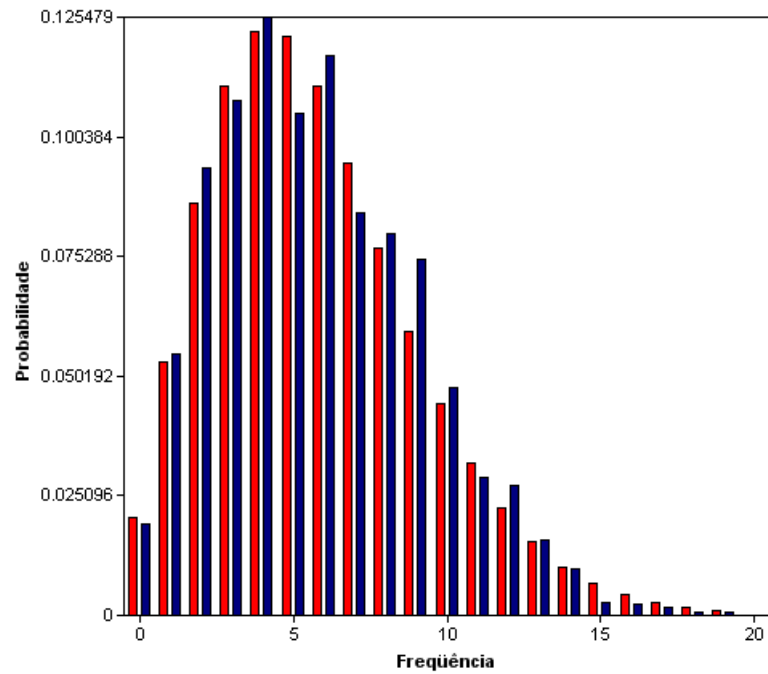


Figura 30 – Distribuições de freqüência estimada pelo LDA e empírica do estudo de caso 2.

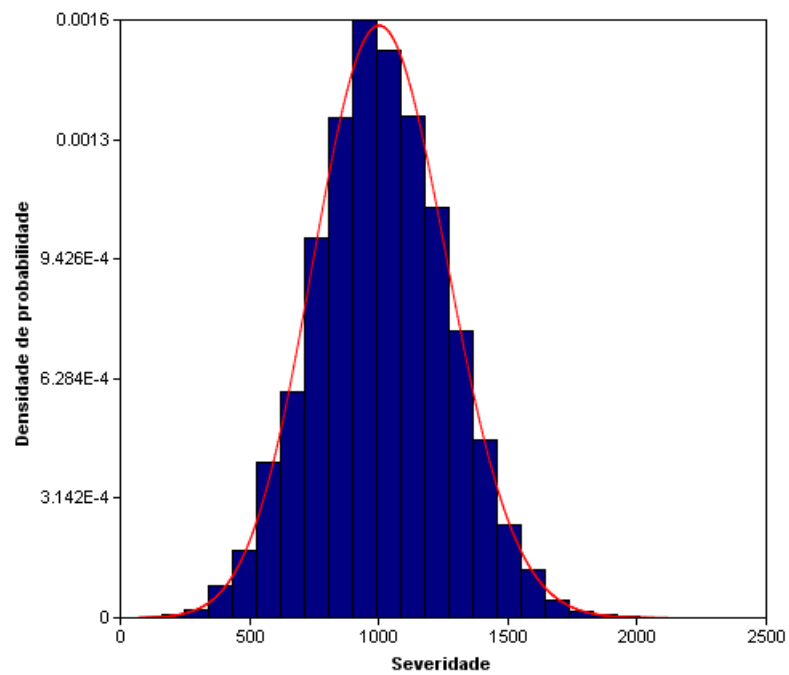


Figura 31 – Distribuições de severidade estimada e empírica do estudo de caso 2.

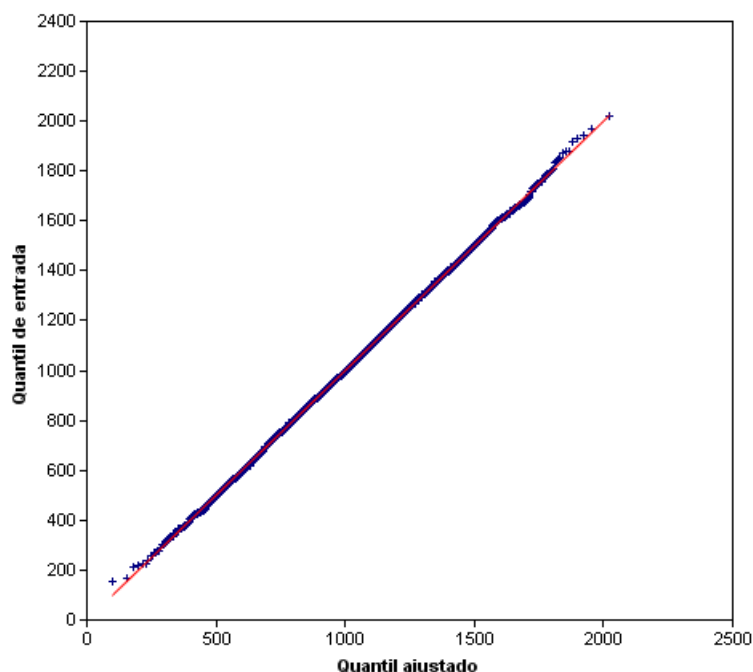


Figura 32 – Q-Q Plot da distribuição estimada dos dados empíricos do estudo de caso 2.

Calculando-se a perda esperada, o VAR com 95% de confiança e o VAR com 99,9% de confiança obtém-se os dados da Tabela 6.12:

Tabela 6.12 – Resultados obtidos com o LDA aplicado à base de dados simulados do caso 1

	Perda Esperada	VAR (95%)	VAR (99,9%)
LDA	5.912,80	12.431,16	20.833,87

6.2.7 Análise e comparação dos resultados

Os resultados obtidos com a RB e com a abordagem LDA aparecem consolidados na Tabela 6.12:

Tabela 6.12 – Resultados consolidados do estudo de caso 2

	Perda Esperada	VAR (95%)	VAR (99,9%)
RB - ICRs desconhecidos	5.903,63	12.631,72	19.203,04
RB - ICR de frequência conhecido	3.113,61	7.379,98	11.860,57
LDA	5.912,80	12.431,16	20.833,87

Podemos observar que os valores calculados pela LDA e pela RB sem informações de ICRs são muito próximos. Além disso, estes valores foram também muito próximos à perda média dos cinco anos calculada a partir da base simulada de perdas, cujo valor é 5.864,80.

No caso em que o ICR de frequência é conhecido, entretanto, vemos uma grande diferença de valores da RB com relação à LDA. O valor obtido com a RB utilizando a informação do valor atual do ICR de frequência é muito próximo à perda média dos últimos dois anos calculada a partir da base simulada, que é igual a 3.546,49.

Isto demonstra que a RB é capaz de captar alterações no comportamento das distribuições das variáveis da rede, desde que existam dados suficientes, após a alteração no comportamento das perdas, para realizar o aprendizado de parâmetros (ou existam dados mais antigos ainda com comportamento semelhante ao atual).

7 Softwares desenvolvidos

Para estudo do tema desta dissertação foram desenvolvidos dois softwares: um para a abordagem de distribuição de perdas LDA e outro de Redes Bayesianas. Optou-se por desenvolver os softwares ao invés de se utilizar softwares comerciais para um entendimento mais profundo de todos os conceitos e detalhes da LDA e de RBs, e também porque os softwares comerciais disponíveis de LDA e RBs não integram todas as funcionalidades necessárias para RO, como, por exemplo, a SMC para cálculo de VAR operacional.

Os dois softwares foram desenvolvidos utilizando linguagem C++. Eles são compostos de arquivos executáveis (um para cada software) e uma biblioteca de vínculo dinâmico (DLL), que possui todos os algoritmos utilizados em RBs e na LDA. A DLL foi desenvolvida em linguagem C++ ANSI. As interfaces gráficas foram desenvolvidas utilizando a biblioteca *Microsoft Foundation Classes* (MFC).

Todas as figuras deste trabalho foram obtidas dos softwares desenvolvidos. A seguir são listadas as funcionalidades e cada um dos softwares.

7.1 Software LDA

O software de LDA desenvolvido possui as seguintes funcionalidades:

- geração de números aleatórios segundo 11 distribuições contínuas⁷ e 6 distribuições discretas⁸;
- estimação de parâmetros das distribuições contínuas e discretas através de três métodos: momentos, máxima verossimilhança e mínimos quadrados;

⁷ Distribuições contínuas: Normal, Log-Normal, Exponencial, Gumbel, Pareto, Rayleigh, Cauchy, Weibull, Normal Inversa, Gama e Beta.

⁸ Distribuições discretas: Poisson, Binomial, Binomial Negativa, Geométrica, Hipergeométrica, Polya-Aeppli.

- teste de aderência: KS e AD para distribuições contínuas e χ^2 para distribuições discretas;
- importação / exportação de dados através de arquivos texto e planilha Excel©;
- gráficos: Histograma e QQ-Plot;
- cálculo de VAR operacional através de SMC a partir de uma distribuição discreta de frequência e uma distribuição contínua de severidade.

As figuras 33, 34, 35, 36, 37, 38 e 39 mostram as telas do software LDA desenvolvido.

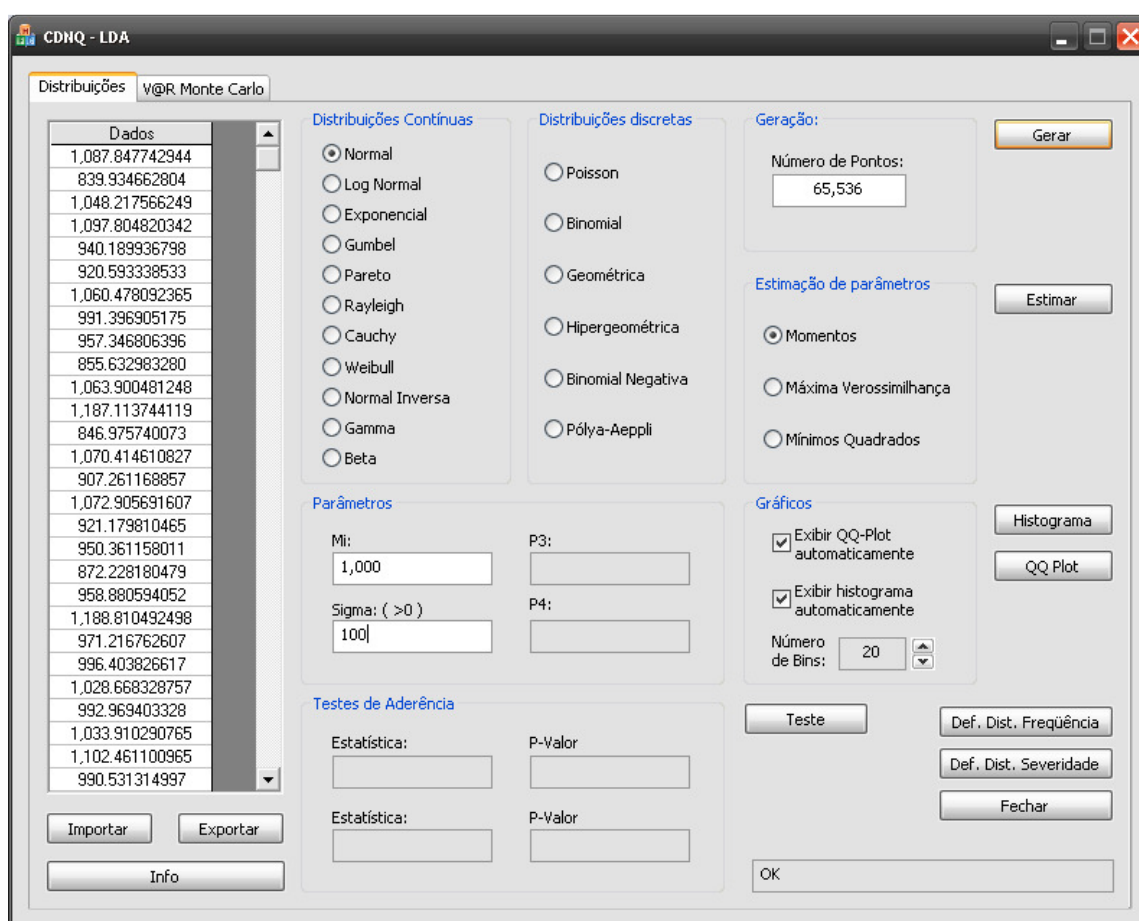


Figura 33 – Software LDA: distribuições

CDNQ - LDA

Distribuições V@R Monte Carlo

Distribuição de frequência

Poisson

Lambda:

Binomial

Geométrica

Hipergeométrica

Binomial Negativa

Pólya-Aeppli

P2:

P3:

Distribuição de severidade

Normal

Log Normal

Exponencial

Gumbel

Pareto

Rayleigh

Cauchy

Weibull

Normal Inversa

Gamma

Beta

Mi:

Sigma:

P3:

P4:

Número de rodadas:

Semente:

Horizonte de tempo:

Perda Esperada:

Nível de confiança	V@R	Perda não esperada
99.80 %	8,088,250.44	788,080.79
99.81 %	8,091,678.01	791,508.36
99.82 %	8,097,895.10	797,725.45
99.83 %	8,101,222.48	801,052.83
99.84 %	8,107,595.22	807,425.57
99.85 %	8,111,080.98	810,911.33
99.86 %	8,118,058.17	817,888.53
99.87 %	8,125,678.85	825,509.20
99.88 %	8,132,884.54	832,714.89
99.89 %	8,140,556.42	840,386.77
99.90 %	8,148,518.82	848,349.17
99.91 %	8,156,632.31	856,462.66
99.92 %	8,168,395.15	868,225.51
99.93 %	8,180,347.23	880,177.58
99.94 %	8,192,056.01	891,886.37
99.95 %	8,209,502.24	909,332.59
99.96 %	8,226,114.46	925,944.82
99.97 %	8,246,226.58	946,056.94
99.98 %	8,271,178.98	971,009.33
99.99 %	8,328,157.83	1,027,988.19

Calcular Fechar

Figura 34 – Software LDA: SMC para cálculo de VAR e das perdas esperada e inesperada.

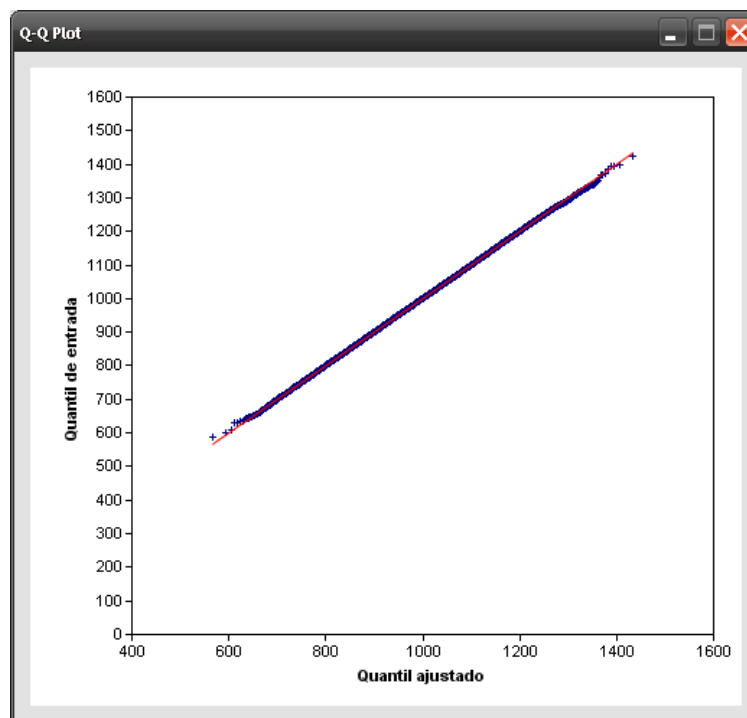


Figura 35 – Software LDA: QQ-Plot

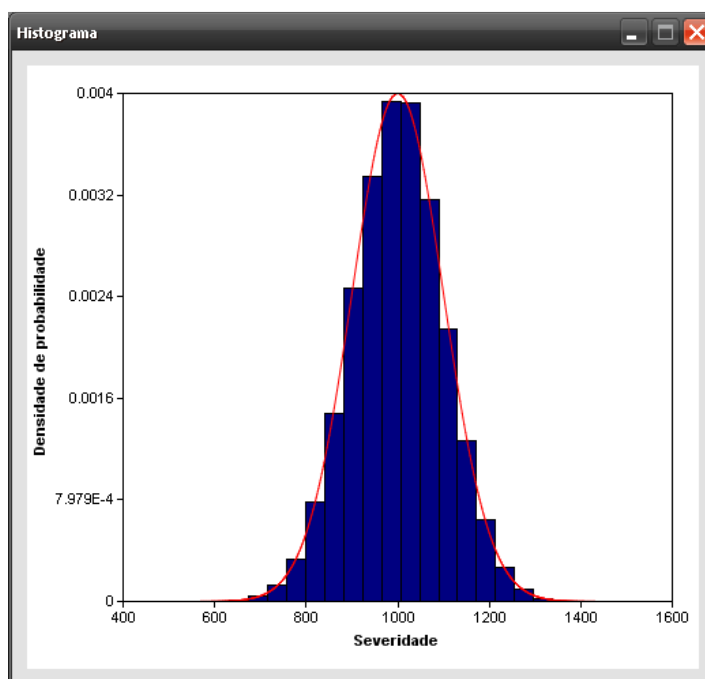


Figura 36 – Software LDA: histograma com função densidade de probabilidade da distribuição paramétrica estimada.

Estatística	Valor
μ_1	1000.4988882161
μ_2	1010949.1568363
μ_3	1031361648.7237
μ_4	1062044086428.2
Média	1000.4988882161
Variância	9951.1315146451
Assimetria	-0.0040806434253381
Curtose	-0.0020413576398117
Média teórica	1000
Variância teórica	10000

Figura 37– Software LDA: informações dos dados empíricos e da distribuição teórica.

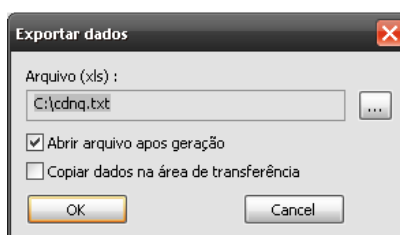


Figura 38 – Software LDA: tela de exportação de dados (arquivo texto e Excel©)

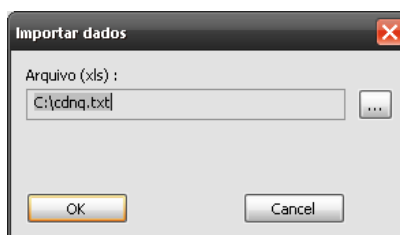


Figura 39 – Software LDA: tela de importação de dados (arquivo texto e Excel©)

7.2 Redes Bayesianas

O software de Redes Bayesianas desenvolvido possui as seguintes funcionalidades:

- aprendizado de parâmetros pelo método de contagem de freqüências relativas e pelo método das distribuições *a priori* de *Dirichlet*;
- inferência utilizando algoritmo exato *Junction Tree*;
- geração de base histórica de dados a partir das probabilidades condicionais existentes na Rede;
- cálculo do VAR Operacional utilizando nós de freqüência e severidade.

As figuras 40, 41, 42, 43 E 44 mostram as telas do software de RBs desenvolvido.

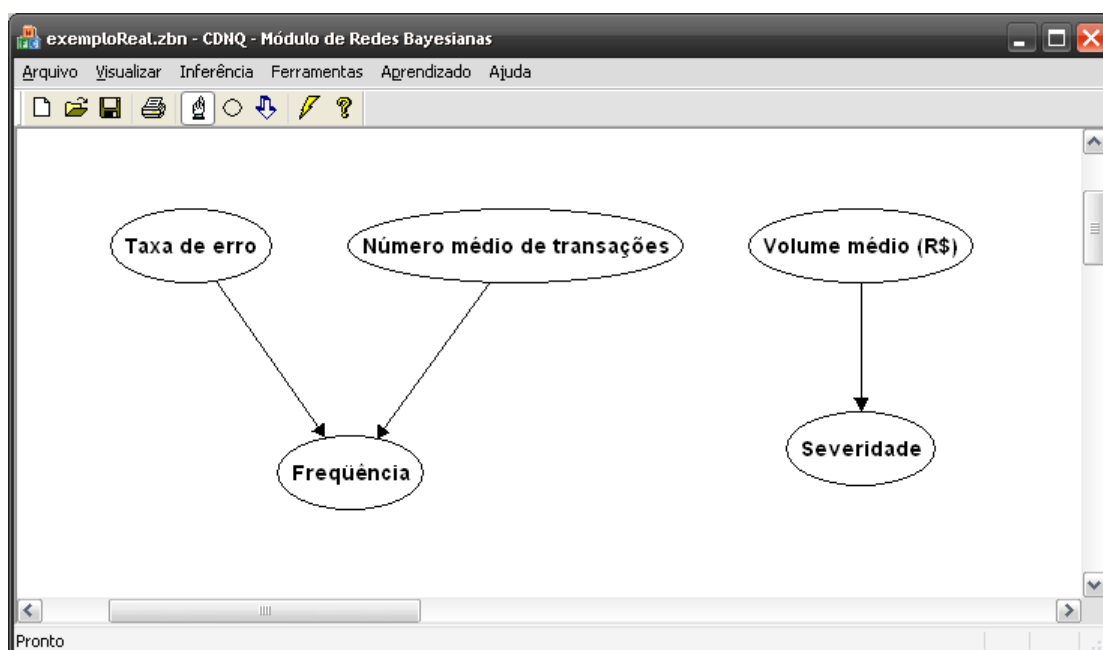


Figura 40 – Software RB: definição da estrutura.

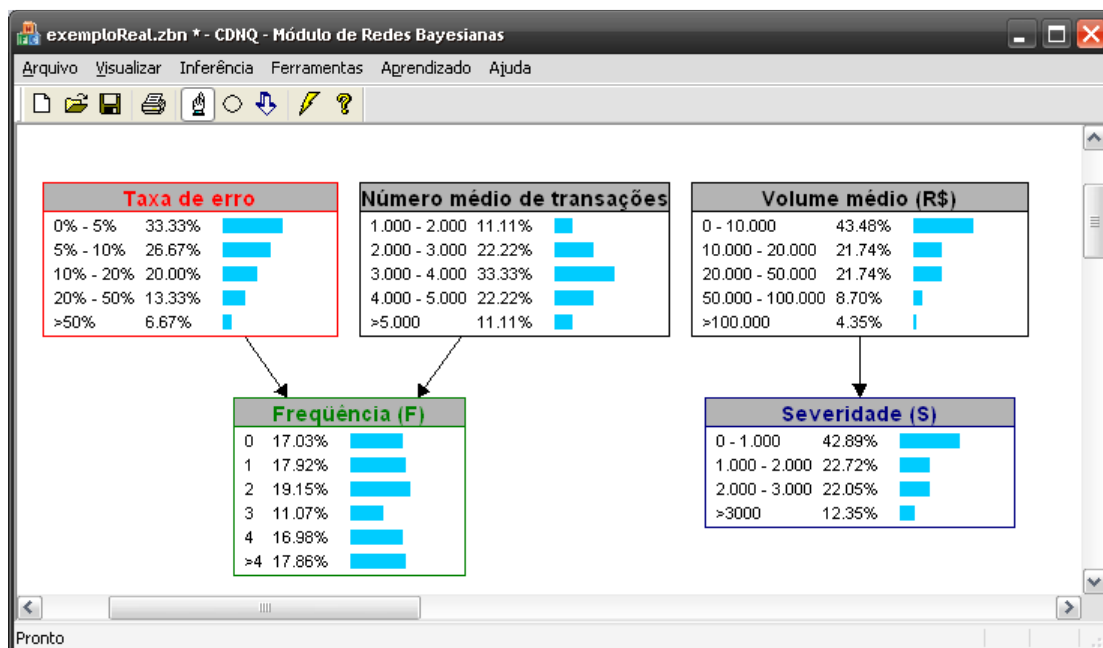


Figura 41 – Software RB desenvolvido: inferência



Figura 42 – Software de Redes Bayesianas: probabilidades condicionais.

Discretização

Nó de frequência: Frequência

Estado	Limite Inferior	Limite Superior
0	0	1
1	1	2
2	2	3
3	3	4
4	4	5
>4	5	1,000,000

O intervalo de cada classe é [Limite Inferior, Limite Superior]

Nó de severidade: Severidade

Estado	Limite Inferior	Limite Superior
0 - 1.000	0.00	1,000.00
1.000 - 2.000	1,000.00	2,000.00
2.000 - 3.000	2,000.00	3,000.00
>3000	3,000.00	1,000,000.00

O intervalo de cada classe é [Limite Inferior, Limite Superior]

OK Cancel

Figura 43– Software LDA desenvolvido: discretização.

Cálculo de V@R

Nível de confiança	V@R	Perda não esperada
99.82 %	2,296,289.30	328,981.29
99.83 %	2,298,360.59	331,052.58
99.84 %	2,300,526.03	333,218.02
99.85 %	2,302,770.48	335,462.48
99.86 %	2,304,787.01	337,479.01
99.87 %	2,307,347.02	340,039.01
99.88 %	2,310,052.84	342,744.84
99.89 %	2,312,450.41	345,142.40
99.90 %	2,315,703.05	348,395.04
99.91 %	2,319,988.03	352,680.03
99.92 %	2,323,752.24	356,444.23
99.93 %	2,327,916.68	360,608.67
99.94 %	2,333,372.64	366,064.63
99.95 %	2,339,526.80	372,218.79
99.96 %	2,346,196.51	378,888.50
99.97 %	2,356,519.29	389,211.29
99.98 %	2,369,269.85	401,961.84
99.99 %	2,390,069.29	422,761.29

Perda esperada: 1,967,308.01

Número de rodadas: 1000000 Horizonte de: 365 Semente: 1

Calcular Fechar

Figura 44 – Software LDA desenvolvido: SMC.

7.3 Validações dos softwares desenvolvidos

Os softwares desenvolvidos foram validados utilizando-se softwares comerciais disponíveis. Os resultados das validações dos softwares LDA e RB encontram-se a seguir.

7.3.1 Validação do software LDA desenvolvido

Para validação do software LDA foi utilizado outro software, comercial, o *BestFit*®, versão 4.5.5, da *Palisade Corporation*. Segundo o manual deste software, o método utilizado para estimação dos parâmetros das distribuições é o método da máxima verossimilhança.

Números aleatórios foram gerados, através do *BestFit*®, para cada uma das distribuições implementadas no software desenvolvido. A seguir estes números foram exportados do *BestFit*® para um arquivo texto e importados para o software de estudo, que foi utilizado para estimar os parâmetros das distribuições, utilizando três métodos: momentos, máxima verossimilhança e mínimos quadrados. As Tabelas 7.1 a 7.15 mostram os resultados obtidos.

Tabela 7.1 – Validação: resultados distribuição Normal com $\mu = 100$ e $\sigma = 30$

Normal	Momentos	Máxima Verossimilhança	Mínimos Quadrados	Palisade Bestfit ©
μ	100.074790394	100.074790394	99.995516053	100.074790394
σ	30.032556151	30.032556151	30.051603726	30.032556151
KS	0.004937811	0.004937811	0.003927061	0.004938
AD	0.419254115	0.419254115	0.393435259	0.4193

Tabela 7.2 – Validação: resultados distribuição Log-normal com $\mu = 15$ e $\sigma = 2$

Log-normal	Momentos	Máxima Verossimilhança	Mínimos Quadrados	Palisade Bestfit ©
μ	15.418043284	15.010147533	15.012212209	15.010147532
σ	1.741567598	1.997428599	1.999765433	1.997428599
KS	0.101139012	0.003465328	0.002901324	0.003465
AD	610.310688553	0.150845624	0.143676701	0.1508

Tabela 7.3 – Validação: resultados distribuição Exponencial com $\lambda = 0.43$

Exponencial	Momentos	Máxima Verossimilhança	Mínimos Quadrados	Palisade Bestfit ©
λ	0.426834015	0.426834015	0.423813866	0.426834015
KS	0.005376476	0.005376476	0.004187569	0.005376
AD	0.764681909	0.764681909	0.543818804	0.7647

Tabela 7.4 – Validação: resultados distribuição Gumbel com $\alpha = 5.77$ e $\beta = 6.43$

Gumbel	Momentos	Máxima Verossimilhança	Mínimos Quadrados	Palisade Bestfit ©
α	5.752872033	5.765578856	5.766344322	5.765579051
β	6.440800821	6.405230884	6.386981841	6.405231348
KS	0.005089863	0.004255310	0.004539763	0.004255
AD	0.819760370	0.626185425	0.643425300	0.6262

Tabela 7.5 – Validação: resultados para Pareto com $\theta = 5$ e $\alpha = 4$

Pareto	Momentos	Máxima Verossimilhança	Mínimos Quadrados	Palisade Bestfit ©
θ	4.956222863	4.991356329	5.004691784	4.991356329
α	3.994317963	4.000017169	4.000017169	4.000017169
KS	0.009380272	0.003607307	0.004073355	0.003607
AD	4.232533237	∞	∞	∞

Tabela 7.6 – Validação: resultados para Rayleigh com $\beta = 4.81$

Rayleigh	Momentos	Máxima Verossimilhança	Mínimos Quadrados	Palisade Bestfit ©
β	4.813598800	4.804699399	4.815039734	4.804699399
KS	0.004272029	0.004974123	0.004158658	0.004974
AD	0.650326930	0.798952628	0.647152390	0.7990

Tabela 7.7 – Validação: resultados para Weibull com $\alpha = 10$ e $\beta = 20$

Weibull	Momentos	Máxima Verossimilhança	Mínimos Quadrados	Palisade Bestfit ©
α	10.148135492	10.096380898	10.094265317	10.096380786
ϕ	19.988394262	19.990886402	19.985476506	19.990886445
KS	0.003986566	0.004442896	0.003464459	0.004443
AD	0.460376916	0.330944783	0.310186816	0.3309

Tabela 7.8 – Validação: resultados para Normal Inversa com $\mu = 2.22$ e $\lambda = 7.89$

Normal Inversa	Momentos	Máxima Verossimilhança	Mínimos Quadrados	Palisade Bestfit ©
μ	2.224606015	2.224606015	2.226141396	2.224606015
λ	7.928514984	7.908217814	7.877119397	7.908217814
KS	0.003238351	0.003174389	0.002628134	0.003174
AD	0.227294217	0.211727670	0.221390587	0.2117

Tabela 7.9 – Validação: resultados para Gama com $\alpha = 21$ e $\beta = 16$

Gama	Momentos	Máxima Verossimilhança	Mínimos Quadrados	Palisade Bestfit ©
α	21.146469342	21.104563960	20.872760801	21.104563960
β	15.886808799	15.918353767	16.105171518	15.918353767
KS	0.005405461	0.005280478	0.003853624	0.005280
AD	0.496919769	0.451628476	0.383634258	0.4516

Tabela 7.10 – Validação: resultados para Beta com $\alpha = 3.74$, $\beta = 8.77$, $min = 0$, $max = 10$.

Beta	Momentos	Máxima Verossimilhança	Mínimos Quadrados	Palisade Bestfit ©
α	3.576225424	3.869522819	3.772996568	3.869421021
β	8.785591578	9.735638803	10.493655147	9.735171397
min	0.071476307	-0.022055306	0.049667516	-0.022030089
max	10.099174520	10.507131690	11.114797896	10.506904871
KS	0.004497745	0.004877408	0.003299858	0.004877
AD	0.375593492	0.411651454	0.398171334	0.4117

Tabela 7.11 – Validação: resultados para Poisson com $\lambda = 1.5$

Poisson	Momentos	Máxima Verossimilhança	Mínimos Quadrados	Palisade Bestfit ©
λ	1.49155	1.49155	1.493088617	1.49155
χ^2	4.924073777	4.924073777	4.966943708	4.924
χ^2 p-valor	0.177442034	0.177442034	0.174233778	0.1774

Tabela 7.12 – Validação: resultados para Binomial com $N = 5$ e $p = 0.88$

Binomial	Momentos	Máxima Verossimilhança	Mínimos Quadrados	Palisade Bestfit ©
N	5	5	5	5
p	0.87869	0.87869	0.878694722	0.87869
χ^2	0.000344942	0.000344942	0.000508869	0.0003449
χ^2 p-valor	0.985182059	0.985182059	0.982002749	0.9852

Tabela 7.13 – Validação: resultados para Geométrica com $p = 0.65$

Geométrica	Momentos	Máxima Verossimilhança	Mínimos Quadrados	Palisade Bestfit ©
p	0.652315721	0.652315721	0.652728315	0.652315721
χ^2	0.009853835	0.009853835	0.000541151	0.009854
χ^2 p-valor	0.920926688	0.920926688	0.981440771	0.9209

Tabela 7.14 – Validação: resultados para Hipergeométrica com $N = 2$, $D = 7$, $M = 9$

Hipergeométrica	Momentos	Máxima Verossimilhança	Mínimos Quadrados	Palisade Bestfit ©
n	2	2	2	2
D	7	7	7	7
M	9	9	9	9
χ^2	4.226285714	4.226285714	4.226285714	4.226
χ^2 p-valor	0.039802447	0.039802447	0.039802447	0.0398

Tabela 7.15 – Validação: resultados para Binomial Negativa com $s = 7$, $p = 0.85$

Binomial Negativa	Momentos	Máxima Verossimilhança	Mínimos Quadrados	Palisade Bestfit ©
s	8	8	8	8
p	0.865814921	0.865814921	0.865974510	0.865814921
χ^2	0.085773220	0.085773220	0.037185517	0.08577
χ^2 p-valor	0.993488238	0.993488238	0.998114008	0.9935

As Tabelas 7.1 a 7.15 mostram que o software desenvolvido estima parâmetros das distribuições de maneira idêntica ao software comercial *Bestfit*©.

7.3.2 Validação do software de Redes Bayesianas desenvolvido

Para validação do software de RB foi utilizado outro software, comercial, o *Hugin Lite*, versão 6.9, da *Hugin Expert*. Dois algoritmos foram validados: o de aprendizado de parâmetros e o de inferência. Segundo o manual do fabricante, o *Hugin Lite* utiliza o método de contagem de freqüências relativas, também desenvolvido no software de estudo. Para realização de inferência na RB, o algoritmo exato *Junction Tree* é utilizado no software *Hugin Lite*, também o mesmo implementado no software de estudo.

7.3.2.1 Validação do algoritmo de aprendizado de parâmetros

A RB do problema do caro ligar, cujo grafo aparece na Figura 4 (3.3 – Variáveis discretas x contínuas) foi utilizada. O grafo da RB foi construído no *Hugin Lite* e as probabilidades condicionais das Tabelas 7.16, 7.17, 7.18 e 7.19 foram inicialmente alimentadas nos nós da rede.

Tabela 7.16 – Distribuição de *Combustível*

P(Combustível=Sim)	P(Combustível=Não)
0,85	0,15

Tabela 7.17 – Distribuição de *Bicos limpos*

P(Bicos limpos = Sim)	P(Bicos limpos = Não)
0,28	0,72

Tabela 7.18 – Distribuição de P(Carro liga | Combustível, Bicos limpos)

		Carro liga	
Combustível	Bicos Limpos	Sim	Não
Sim	Sim	0,90	0,10
Sim	Não	0,20	0,80
Não	Sim	0,00	1,00
Não	Não	0,00	1,00

Tabela 7.19 – Distribuição de P(Medidor combustível | Combustível)

		Medidor combustível		
Combustível	Vazio	Meio	Cheio	
Sim	0,00	0,70	0,30	
Não	1,00	0,00	0,00	

A seguir, utilizando a ferramenta *Generate Cases* do *Hugin Lite*®, um arquivo com 500 casos simulados foi gerado, respeitando as distribuições condicionais de cada nó da rede. A seguir, tanto no software desenvolvido quanto no *Hugin Lite*® e as probabilidades condicionais de todas as variáveis foram configuradas de forma que as distribuições fossem equívocas, refletindo total desconhecimento *a priori*. Os valores de N_{ij} (tamanho de amostra equivalente) foram configurados com o valor zero em ambos os softwares.

O arquivo gerado de casos simulados foi então utilizado como input para o aprendizado de parâmetros no software de estudo desenvolvido e também para o aprendizado de parâmetros no próprio *Hugin Lite*®. Os resultados de ambos foram idênticos, e aparecem nas Figuras 45, 46, 47 e 48.

Carro_Liga				Medidor_Combustivel				Bicos_Limpos		Combustivel	
Bicos_Limpos		Sim		Nao		Nao					
Combustivel		Sim	Nao	Sim	Nao						
Sim		0.855072	0	0.177852	0						
Nao		0.144928	1	0.822148	1						
Experience		138	16	298	48						

Probabilidades Condicionais			
Combustivel	Bicos_Limpos	Sim	Nao
Sim	Sim	0.855072464	0.144927536
Sim	Nao	0.177852349	0.822147651
Nao	Sim	0	1
Nao	Nao	0	1

Figura 45 – Comparação das probabilidades condicionais ‘aprendidas’ pelo software desenvolvido para estudo (direita) e o *Hugin Lite*© (esquerda) para o nó *Carro liga*.

Carro_Liga			Medidor_Combustivel		Bicos_Limpos		Combustivel	
Combustivel		Sim	Nao					
Vazio		0	1					
Meio		0.676606	0					
Cheio		0.323394	0					
Experience		436	64					

Probabilidades Condicionais			
Combustivel	Vazio	Meio	Cheio
Sim	0	0.676605505	0.323394495
Nao	1	0	0

Figura 46 – Comparação das probabilidades condicionais ‘aprendidas’ pelo software desenvolvido para estudo (direita) e o *Hugin Lite*© (esquerda) para o nó *Medidor combustivel*.

Carro_Liga		Medidor_Combustivel		Bicos_Limpos		Combustivel	
Sim		0.308					
Nao		0.692					
Experience		500					

Probabilidades Condicionais	
Sim	Nao
0.308	0.692

Figura 47 – Comparação das probabilidades condicionais ‘aprendidas’ pelo software desenvolvido para estudo (direita) e o *Hugin Lite*© (esquerda) para o nó *Bicos limpos*.

Carro_Liga		Medidor_Combustivel		Bicos_Limpos		Combustivel	
Sim		0.872					
Nao		0.128					
Experience		500					

Probabilidades Condicionais	
Sim	Nao
0.872	0.128

Figura 48 – Comparação das probabilidades condicionais ‘aprendidas’ pelo software desenvolvido para estudo (direita) e o *Hugin Lite*© (esquerda) para o nó *Combustivel*.

7.3.2.2 Validação do algoritmo de inferência

Com as probabilidades condicionais obtidas em 7.3.2.1 – *Validação do algoritmo de aprendizado de parâmetros*, duas inferências foram realizadas. A primeira foi feita sem evidências sobre valores de todas as variáveis. A segunda foi feita supondo valor conhecido (*Sim*) para a variável *Combustivel*. Nos dois casos os valores obtidos foram idênticos àqueles obtidos no *Hugin Lite*©, como mostrado nas Tabelas 7.20 e 7.21.

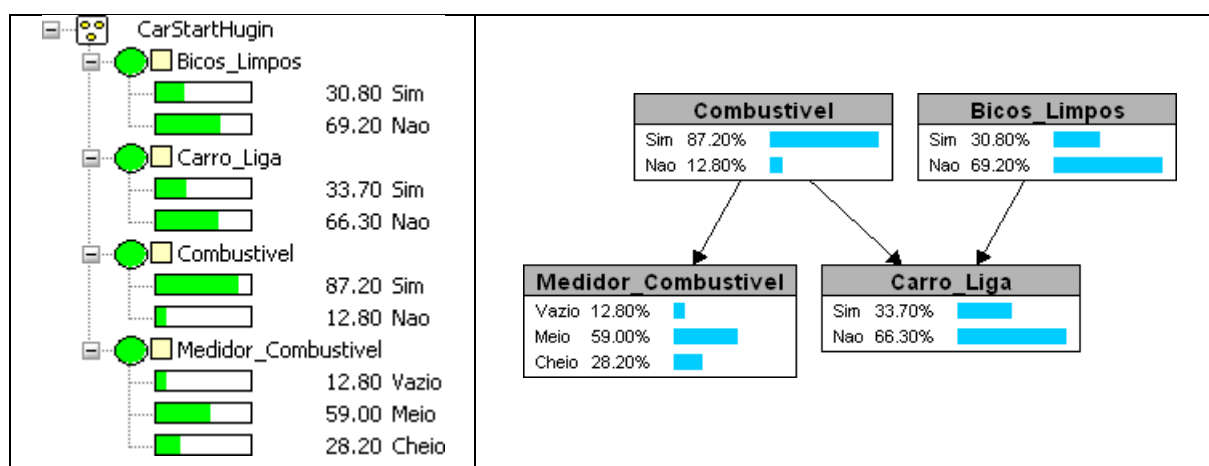
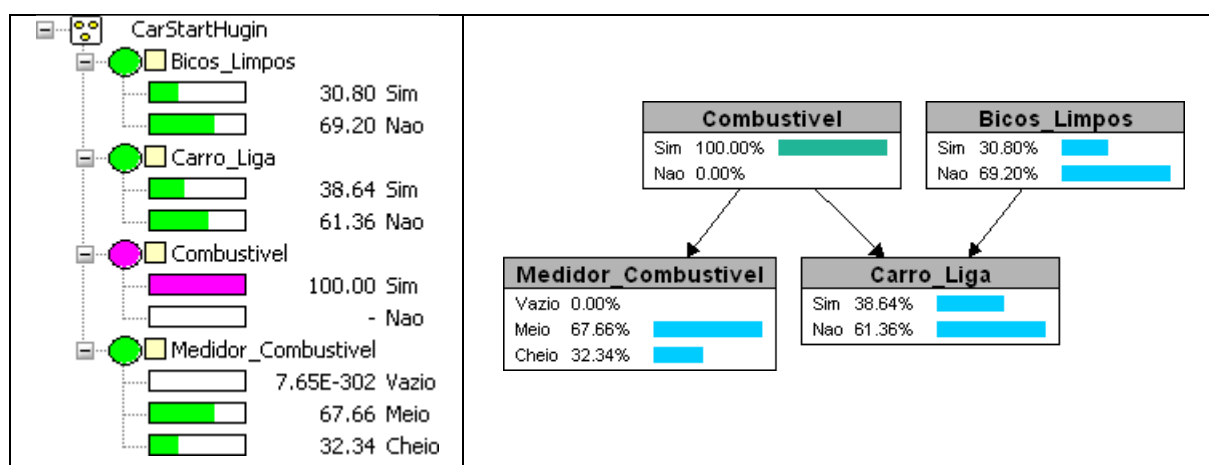


Tabela 7.20 – Resultado da inferência sem evidências

Tabela 7.21 – Resultado da inferência com a evidência *Combustível = Sim*

8 Conclusões e trabalhos futuros

Risco Operacional é ainda um tema muito novo para as instituições financeiras. As primeiras exigências regulatórias estão sendo impostas neste momento (2008), e o cronograma brasileiro definido pelo BCB prevê a aprovação e implantação de modelos até o final de 2012.

Como as abordagens básicas tendem a alocar um volume de capital regulatório mínimo muito grande, é bem provável que grande parte das instituições financeiras migre para modelos avançados de mensuração. Até por uma questão de sobrevivência, já que bancos que conseguirem alocar menos capital serão mais competitivos que aqueles que optarem por modelos mais simples. O capital regulatório mínimo deve ser substancialmente menor na AMA que nas demais abordagens vistas. O uso da AMA, entretanto, estará sujeito a requisitos quantitativos e qualitativos mínimos a serem definidos pelo BCB.

Esta dissertação procurou explorar o modelo de Redes Bayesianas aplicado na modelagem de Riscos Operacionais, como modelo de mensuração avançada para o cálculo de capital regulatório mínimo definido pelo BCB, seguindo as recomendações do Comitê da Basileia de Supervisão Bancária. O modelo representa uma alternativa à abordagem LDA, hoje a mais utilizada pelas instituições financeiras.

Através de exemplos e estudos de casos demonstrou-se o poder que RBs fornecem como modelagem causal de RO. Os principais conceitos teóricos de RBs, como definição da estrutura, discretização de variáveis, aprendizado de parâmetros, inferências e análise de cenários foram introduzidos para depois serem utilizados na prática nos estudos de casos. Todos os passos necessários para o uso de RBs em RO foram detalhados.

Cuidados e problemas no uso de RBs, como a forma de discretizar as variáveis e obter as probabilidades condicionais foram descritos e quando possível exemplificados.

O primeiro diferencial positivo de RBs observado foi a utilização de variáveis explicativas das perdas, os ICRs, ao invés de considerar somente dados de perdas passadas como faz a LDA. Como visto no primeiro estudo de caso, as diferenças de VAR calculado considerando-se o perfil atual de risco de um banco e todos os possíveis valores de perdas passadas podem ser muito grandes. Além disso, vimos no segundo estudo de caso que perdas dinâmicas, isto é, que têm seu comportamento variável no tempo, não são muito bem modeladas sem o uso de um modelo causal.

Um segundo diferencial positivo observado foi a possibilidade de incluir conhecimento subjetivo de especialistas no modelo para cálculo das perdas esperadas, inesperadas e do VAR operacional. No caso de eventos de baixa probabilidade de ocorrência, mesmo com poucos dados históricos (ou nenhum), RBs ainda permitem o cálculo com conhecimento subjetivo. Mais ainda, este conhecimento pode ser mesclado de forma clara com informações históricas existentes.

Outro ponto positivo do modelo é a forma gráfica com que RBs apresentam resultados, tornando o gerenciamento de RO mais amigável.

Apesar das vantagens citadas, RBs são mais complicadas de utilizar que a abordagem LDA. Envolvem conceitos estatísticos mais complexos e menos intuitivos que os da LDA. Uma questão que não foi abordada neste trabalho, e poderia ser alvo de trabalhos futuros, é a relação temporal entre causa e efeito de perdas. RBs constituem um modelo probabilístico que mensura relações de causa e efeito através de probabilidades condicionais. Relações temporais como *delay* de influência de um ICR na perda não são diretamente tratadas em uma RB. Podem existir inclusive cenários reais de *delay* variável, complicando ainda mais o problema. Influências que afetam a perda com *delay* zero ou fixo poderiam, entretanto, ser previamente tratadas deslocando as observações de acordo com o *delay* antes do aprendizado de parâmetros.

De forma geral, o modelo apresentou resultados bastante satisfatórios, vistos nos estudos de casos que aplicavam RBs para situações supostamente conhecidas. As grandes críticas do modelo LDA não se aplicam à modelagem com RBs, como a obrigatoriedade de dependência entre frequência e severidade das perdas, estimativa de VAR baseada somente em perdas históricas e ausência de variáveis explicativas.

Os dois softwares desenvolvidos neste trabalho foram de muita valia, ajudando a fixar muito bem os conceitos envolvidos, apesar da grande dificuldade enfrentada nos seus desenvolvimentos, principalmente no de RBs.

9 Referências Bibliográficas

ADUSEY-POKU, K. *Operational Risk Management – Implementing a Bayesian Network for Foreign Exchange and Money Market Settlement*. 2005. 152 p. Tese (Doutorado) - Faculty of Economics and Business Administration of the University of Göttingen. Göttingen, 2005.

ALEXANDER, C. *Bayesian Methods for Measuring Operational Risk*. Discussion Papers in Finance, ISMA Centre, University of Reading.

ANTONUCCI, A.; PIATTI, A.; ZAFFALON, M; *Credal Networks for Operational Risk Measurement and Management*. Lecture Notes in Computer Science, Itália, Volume 4693/2007. p. 604-611, setembro de 2007.

BASEL COMMITTEE ON BANKING SUPERVISION. *Sound Practices for the Management and Supervision of Operational Risk*. Fevereiro de 2003

BASEL COMMITTEE ON BANKING SUPERVISION. *International Convergence of Capital Measurement and Capital Standards – A Revised Framework*. Junho de 2004.

BASEL COMMITTEE ON BANKING SUPERVISION. *Observed range of practice in key elements of Advanced Measurement Approaches (AMA)*. Outubro de 2006.

BASEL COMMITTEE ON BANKING SUPERVISION. *Consultative Document - Operational Risk*. Janeiro de 2001.

BAUD, N.;FRACHOT, A.; RONCALLI, T. *Internal data, external data and consortium data for operational risk measurement: How to pool data properly?* França: Groupe de Recherche Op´erationnelle, Cr´edit Lyonnais, 2002.

CASTILLO, M.; MENDOZA, A. *Diseño de una Metodología para la Identificación y la Medición del Riesgo Operativo en Instituciones Financieras*. Revista de Ingeniería, Universidad de los Andes, Bogotá, Revista 19, p.45-52, 2004.

COLEMAN, R. *Using Modeling in Operational Risk Management*. Conference on Operational Risk in Retail Finance Services, Londres, 2000.

COOPER, G. F. *An Overview of the Representation and Discovery of Causal Relationships Using Bayesian Networks*. Computation, Causation and Discovery, AAAI Press, 1999, p. 3-62.

CORNALBA, S.; GIUDICI, P. *Statistical Models for operational risk management*. Physica A: Statistical Mechanics and its applications. P166-172, 2004.

CRUZ, M. *Modeling, Measuring and Hedging Operational Risk: A Quantitative Approach*. 1ª Edição. Wiley, 2002. p. 346.

GIUDICI, P. *Integration of Qualitative and Quantitative Operational Risk Data: A Bayesian Approach*. Operational Risk Modeling and Analysis: Theory and Practice, p. 131-138, 2004.

HOFFMAN, D. *Managing operational risk: 20 firmwide best practice strategies*. 1ª edição, Wiley, 2002, p. 540.

JENSEN, F. V. *Bayesian Networks and Decision Graphs*. 1ª Edição. Springer, 2001, p. 284.

JORDAN, M. I. *Learning in Graphical Models*. 1ª Edição. Springer, 1998. P. 644.

KING, J. L. *Operational Risk: Measurement and Modelling*. 1ª Edição. Wiley, 2001. p. 276.

NEAPOLITAN, R. E. *Learning Bayesian Networks*. 1ª Edição. Prentice Hall, 2004. p.674.

NEIL, M.; FENTON, N.; TAILOR, M. *Using Bayesian Networks to Model Expected and Unexpected Operational Losses*. Risk Analysis, Vol. 25, N^o. 4, p. 963-972, 2005

PEARL, J. *Causality: Models, Reasoning and Inference*. 1ª Edição. Cambridge University Press. 2000, p.384.

PEARL, J. *Fusion, Propagation and Structuring in Belief Networks*, Artificial Intelligence, Vol. 29, 1986.

PERERA, J. *Quantifying Operational Risk Using Neural Networks*. Julho de 2000. Disponível em <http://www.opriskandcompliance.com/public/showPage.html?page=269588>. Acesso em 05/08/2007.

PRESS, W.H.; FLANNERY, S.A.; TEUKOLSKY, W.T; VETTERLING, W.T. *Numerical Recipes: The Art of Scientific Computing*. 2ª Edição. Cambridge University Press, 1992, p 994.

SCANDIZZO, S. *Risk Mapping and Key Risk Indicators in Operational Risk Management*. Economic Notes by Banca Monte dei Paschi di Siena SpA, vol. 34, no. 2-2005, pp 231-256, 2005.

SHAH, S. *Measuring Operational Risks using Fuzzy Logic Modeling*. Setembro de 2003. Disponível em < <http://www.irmi.com/Expert/Articles/2003/Shah09.aspx>>. Acesso em 05/08/2007.

WU, D; BUTZ, C. *On the complexity of Probabilistic Inference in Singly Connected Bayesian Networks*. 10th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing, vol. 1, pp 581--590, 2005.

YASUDA, Y. *Application of Bayesian Inference to Operational Risk Management*. 2003. 44p. Tese (Doutorado) – Universidade de Tsukuba, Japão, janeiro de 2003.

YOON, Y. K. *Modelling Operational Risk in Financial Institutions Using Bayesian Networks*. 2003. 83p. Dissertação (Mestrado). Cass Business School. Londres, 2003.