

Editado por

Célia A. Zorzo Barcelos

Universidade Federal de Uberlândia - UFU
Uberlândia, MG, Brasil

Eliana X.L. de Andrade

Universidade Estadual Paulista - UNESP
São José do Rio Preto, SP, Brasil

Maurílio Boaventura

Universidade Estadual Paulista - UNESP
São José do Rio Preto, SP, Brasil



A Sociedade Brasileira de Matemática Aplicada e Computacional - SBMAC publica, desde as primeiras edições do evento, monografias dos cursos que são ministrados nos CNMAC.

Para a comemoração dos 25 anos da SBMAC, que ocorreu durante o XXVI CNMAC em 2003, foi criada a série **Notas em Matemática Aplicada** para publicar as monografias dos minicursos ministrados nos CNMAC, o que permaneceu até o XXXIII CNMAC em 2010.

A partir de 2011, a série passa a publicar, também, livros nas áreas de interesse da SBMAC. Os autores que submeterem textos à série Notas em Matemática Aplicada devem estar cientes de que poderão ser convidados a ministrarem minicursos nos eventos patrocinados pela SBMAC, em especial nos CNMAC, sobre assunto a que se refere o texto.

O livro deve ser preparado em **Latex (compatível com o Miktex versão 2.7)**, as figuras em eps e deve ter entre **80 e 150 páginas**. O texto deve ser redigido de forma clara, acompanhado de uma excelente revisão bibliográfica e de **exercícios de verificação de aprendizagem** ao final de cada capítulo.

Veja todos os títulos publicados nesta série na página
<http://www.sbmac.org.br/notas.php>



Sociedade Brasileira de Matemática Aplicada e Computacional

2012

INTRODUÇÃO À SIMULAÇÃO ESTOCÁSTICA PARA ATUÁRIA E FINANÇAS USANDO R

Hélio Côrtes Viera
lopes@mat.puc.rio.br

Departamento de Matemática
Pontifícia Universidade Católica do Rio de Janeiro

Alejandro C. Frery
acfrery@pesquisador.cnpq.br

Instituto de Computação
Universidade Federal de Alagoas

Luciano Vereda
lvereda@rdc.puc.rio.br

Instituto de Gestão de Riscos Financeiros e Atuariais
Pontifícia Universidade Católica do Rio de Janeiro



Sociedade Brasileira de Matemática Aplicada e Computacional

São Carlos - SP, Brasil
2012

Coordenação Editorial: Sandra Mara Cardoso Malta

Coordenação Editorial da Série: Eliana Xavier Linhares de Andrade

Editora: SBMAC

Capa: Matheus Botossi Trindade

Patrocínio: SBMAC

Copyright ©2012 by Hélio Côrtes Vieira, Alejandro C. Frery e Luciano Vereda.

Direitos reservados, 2012 pela SBMAC. A publicação nesta série não impede o autor de publicar parte ou a totalidade da obra por outra editora, em qualquer meio, desde que faça citação à edição original.

Catálogo elaborado pela Biblioteca do IBILCE/UNESP
Bibliotecária: Maria Luiza Fernandes Jardim Froner

Vieira, Hélio

Introdução à Simulação Estocástica para Atuária e Finanças usando **R**
- São Carlos, SP : SBMAC, 2012, 75 p.; 20,5cm
- (Notas em Matemática Aplicada; v. 33)

e-ISBN 978-85-86883-95-8

1. Simulação Estocástica. 2. Matemática Atuarial. 3. Finanças.

I. Vieira, Hélio. II. Frery, Alejandro C. III. Vereda, Luciano.

IV. Título. III. Série

CDD - 51

Esta é uma republicação em formato de e-book do livro original do mesmo título publicado em 2008 nesta mesma série pela SBMAC.

Conteúdo

1	Introdução	13
1.1	Matemática atuarial	13
1.2	Motivação	14
1.3	Previdência privada	14
1.4	Objetivos	16
1.5	Estrutura do texto	16
2	Variáveis Aleatórias e a Plataforma R	19
2.1	Transformações de variáveis aleatórias	20
2.2	Geradores de números pseudo-aleatórios	27
2.3	A distribuição normal multivariada	28
2.4	Exemplos de uso do R	29
3	Cenários de vida de um indivíduo	33
3.1	Conceitos básicos de matemática atuarial	33
3.1.1	Modelo probabilístico para sobrevivência de um indivíduo	33
3.1.2	Tábuas de mortalidade	35
3.2	Apresentação do problema	37
3.3	Modelagem	37
3.4	Simulação usando o R	38
3.5	Exercício	39
4	Cenários para variáveis financeiras	43
4.1	Análise de séries temporais	43
4.2	Introdução ao modelo VAR	44
4.3	As variáveis financeiras	45
4.4	Apresentação do problema	46
4.5	Modelagem	46
4.5.1	Estimativa do modelo VAR das variáveis financeiras	46

4.6	Simulação usando o R	50
4.7	Exercício	51
5	Benefícios de um plano de aposentadoria	53
5.1	Conceitos básicos da matemática financeira	53
5.1.1	Taxa de juros efetiva	54
5.1.2	Taxa de juros nominal	55
5.1.3	Pagamentos antecipados	56
5.1.4	Perpetuidades	57
5.1.5	Anuidades	58
5.1.6	Anuidades de vida inteira	59
5.2	Apresentação do problema	60
5.3	Modelagem	61
5.4	Exercício	62
6	Soma de Sinistros	65
6.1	Processos de Poisson	65
6.2	Apresentação do problema	66
6.3	Modelagem	66
6.4	Simulação usando o R	67
6.5	Exercício	69
7	Conclusão	71

Lista de Figuras

2.1	Densidade e função de distribuição acumulada da lei exponencial padrão	22
2.2	Densidade e função de distribuição acumulada da lei normal padrão	23
2.3	Densidades long-normal e beta	24
2.4	Histograma de cem observações uniformes	29
2.5	Histograma de cem observações exponenciais	30
2.6	Gráfico de pares das ocorrências de normais multivariadas	31
2.7	Gráfico de pares e histogramas das ocorrências de gaussianas multivariadas	32
6.1	Taxa de sinistros em função do tempo	67
6.2	Função de distribuição acumulada dos tempos de ocorrência do processo de Poisson não homogêneo	68

Lista de Tabelas

3.1	Tábua de mortalidade americana masculina de 2004: idades entre 0 e 39 anos	40
3.2	Tábua de mortalidade americana masculina de 2004: idades entre 40 e 79 anos	41
3.3	Tábua de mortalidade americana masculina de 2004: idades entre 79 e 115 anos	42
4.1	Estatísticas das séries de dados mensais de janeiro/2000 a dezembro/2005	46
4.2	Séries financeiras nos anos de 2000 e 2001	47
4.3	Séries financeiras nos anos de 2002 e 2003	48
4.4	Séries financeiras nos anos de 2004 e 2005	49
4.5	Estimativas do modelo VAR.	50

Códigos

2.1	Histograma de uniformes	29
2.2	Histograma de exponenciais	30
2.3	Geração de gaussianas multivariadas	31
3.1	Carregando a tábua de mortalidade através da leitura de um arquivo	38
3.2	Geração de cenários de vida	39
4.1	Iniciação das variáveis do modelo	51
4.2	Iniciação das variáveis do modelo	52
4.3	Iniciação das variáveis do modelo	52
6.1	Geração de um processo de Poisson não homogêneo	67
6.2	Geração da soma de sinistros	68

Capítulo 1

Introdução

1.1 Matemática atuarial

A *Matemática Atuarial* é uma área aplicada da matemática que utiliza conhecimentos de probabilidade, estatística, otimização e economia na análise de problemas financeiros em seguros, fundos de pensão e investimentos.

Oficialmente, ela surgiu na Inglaterra no final da primeira metade do século XIX, embora existam casos de estudos atuariais com datas anteriores a essa. Por exemplo, no século XVII, na Inglaterra e na Holanda, a coroa vendia títulos públicos que asseguravam ao seu comprador uma renda vitalícia. E para a coroa não ter prejuízo com esse negócio, foi necessário determinar com a maior precisão a importância em dinheiro que deveria ser cobrada (o que chamamos de prêmios) para esses títulos. Esse trabalho ficou como tarefa para os principais matemáticos desses dois países.

Hoje, no Brasil, estamos passando por uma fase onde essa área de matemática aplicada está em alta, devido aos diversos desafios existentes no mercado de seguros e previdência. Principalmente após o enorme crescimento do mercado de previdência privada nos últimos anos devido ao fim da aposentadoria integral dos servidores públicos e, agora em 2008, a quebra do monopólio do mercado de resseguros no nosso país (até então tarefa feita somente pelo IRB - Instituto de Resseguros do Brasil).

Um profissional que queira trabalhar em matemática atuarial deve possuir uma formação muito multidisciplinar (que inclui matemática, computação, economia e direito). Sua principal atividade é a criação de modelos matemáticos para efetuar cálculos de probabilidades de eventos, avaliar riscos, fixar valores de prêmios de seguro ou de indenizações, entre outras atribuições. E para isso, a sua principal ferramenta de avaliação é a simu-

lação estocástica, que é o tema central desse trabalho.

1.2 Motivação

A área matemática atuarial é dividida em duas grande subáreas: a matemática atuarial ramo vida, que estuda os problemas ligados a seguros de vida e previdência, e a matemática atuarial ramo não-vida, que estuda os problemas ligados aos outros ramos de seguro, tais como os de residência, automóveis, etc.

Nesse texto trataremos basicamente de duas aplicações de simulação estocástica em problemas atuariais, um no ramo vida e outro no ramo não-vida. O primeiro deles é o da geração de cenários de fluxo de caixa de um plano de previdência privada de um indivíduo. O segundo é a geração de cenários para o volume de dinheiro a ser pago numa carteira de seguros de automóveis. Apresentam-se a seguir os principais aspectos sobre previdência privada que motivam o presente trabalho e logo depois a motivação da geração de cenários para avaliação do risco relativo ao montante de sinistros a serem pagos.

1.3 Previdência privada

A *previdência social*, obrigatória e oferecida pelo Estado, tem como objetivo garantir a subsistência do indivíduo, ou seja, o mínimo de preservação de qualidade de vida de modo condizente com a justiça social. Já a *previdência privada*, complementar e facultativa, é utilizada para preservar um determinado padrão de vida.

Os participantes dos fundos de *previdência privada* têm seus direitos garantidos pela legislação do setor. As empresas são reguladas pela Superintendência de Seguros Privados (SUSEP), órgão responsável pelo controle e fiscalização dos mercados de seguro, previdência privada aberta, capitalização e resseguro.

A previdência privada, que poderia ser classificada como um seguro de renda, pode ser aberta ou fechada.

A *previdência privada fechada*, também conhecida como fundos de pensão, opera apenas dentro de uma empresa ou grupo de empresas do mesmo empregador visando à prestação de benefícios complementares e semelhantes aos da previdência social. Os planos são, em sua maioria, custeados pelas empresas e funcionários e de contribuição variável de acordo com os cálculos atuariais e a política da empresa, podendo ser feito também individualmente, por pessoa física.

Em todas as empresas que têm fundos de pensão, um dos principais temas de discussão interna é a qualidade da gestão das carteiras, pois, se ela não for eficiente, não será possível assegurar o valor das aposentadorias.

A *previdência privada aberta* é constituída pelas instituições abertas e participação pública, visando a prestação de benefícios opcionais, de caráter mais individual. As entidades abertas administram diversos fundos ao mesmo tempo e qualquer pessoa pode participar de seus planos.

Os planos de previdência “tradicionais” garantem remuneração igual ao da caderneta de poupança, no mínimo. As contribuições são mensais por um determinado número de anos, de acordo com a renda futura esperada e da idade previamente estabelecida. Os benefícios podem ser recebidos de diversas formas: vitalício, por um determinado período de tempo (por exemplo, 20 anos) ou em forma de capital (todo o valor de uma só vez).

No caso da previdência complementar aberta, os principais planos oferecidos pelas seguradoras são: o PGBL (Plano Gerador de Benefício Livre) e o VGBL (Vida Gerador de Benefício Livre). Ambos permitem o acúmulo de recursos por um prazo determinado. Existem diversas modalidades de fundos e planos e os rendimentos das aplicações financeiras são repassados integralmente para os clientes. O *período de diferimento* e o *período de benefício* compõem as duas fases destes planos. O primeiro ocorre quando as contribuições estão sendo aplicadas. Já o período de benefício é definido a partir do momento que o indivíduo escolhe para começar a receber o dinheiro acumulado.

A principal diferença entre estes planos está na tributação. O plano VGBL é mais indicado a autônomos e profissionais liberais. Não é dedutível do imposto de renda, a tributação ocorre apenas sobre o ganho de capital no momento do resgate. O VGBL é ideal para quem não pode se beneficiar das vantagens fiscais oferecidas pelos planos de previdência, seja porque declara o Imposto de Renda pelo formulário simplificado ou porque já atingiu o limite máximo de contribuição permitido para dedução fiscal, i.e., 12% da renda bruta anual tributável. Já o plano PGBL é mais vantajoso para quem faz a declaração do imposto de renda pelo formulário completo. Até 12% da renda bruta pode ser deduzida do cálculo do imposto. A tributação deste valor, investido em PGBL, será paga no momento do resgate.

Um aspecto fundamental para as seguradoras é a gestão dos ativos (contribuições recebidas de indivíduos) e dos passivos (benefícios pagos aos segurados) de um plano de previdência é o de modelar o risco associado ao negócio. Isto é, predizer as taxas de juros dos rendimentos de sua carteira de investimentos e a vida futura dos indivíduos segurados por ela. E, na verdade, essas incertezas são parâmetros de entrada para qualquer sistema de gestão de ativos e passivos de uma seguradora. Elas podem

ser representadas pela geração de um grande número de cenários (obtidos por simulação estocástica), os quais refletem os comportamentos possíveis desses parâmetros. Utilizando esses cenários é que decisões no modelo de gestão são tomadas.

1.4 Objetivos

O principal objetivo desse livro é o de introduzir conceitos de matemática atuarial e simulação estocástica motivados por problemas atuariais e financeiros. Para isso serão apresentados quatro problemas, onde para cada um deles será discutida uma estratégia a ser usada na modelagem probabilística; a solução computacional será feita utilizando os recursos da plataforma R.

1.5 Estrutura do texto

Esse texto está organizado da seguinte forma:

- O capítulo 2 introduz os conceitos básicos de simulação estocástica e da plataforma R que serão utilizados, tais como: geração de números pseudo-aleatórios, geração de variáveis aleatórias discretas e contínuas e geração de vetores aleatórios normais multivariados.
- O capítulo 3 descreve os princípios de matemática atuarial e estuda a geração de cenários de vida de indivíduo que contribui para um plano de previdência privada do tipo PGBL.
- O capítulo 4 introduz um modelo de séries temporais para variáveis financeiras a ser utilizado no capítulo seguinte.
- O capítulo 5 introduz um simulador para geração de cenários do valor de benefício de um plano de previdência PGBL para um indivíduo que contribui mensalmente.
- O capítulo 6 propõe um simulador estocástico para a geração de cenários que estima a quantia de sinistros ocorridos numa carteira de apólices de automóveis.
- Finalmente, o capítulo 7 conclui o trabalho e apresenta diretrizes para estudos futuros.

Os capítulos 3, 4 e 5 propõem problemas e soluções baseadas na dissertação de mestrado da aluna Carla Jardim Dias [4], orientada pelos professores Hélio Côrtes Vieira e Luciano Vereda, defendida no programa de pós-graduação em matemática na PUC-Rio.

Nos endereços da Web <http://www.sbmac.org.br/notas.php>, <http://www.mat.puc-rio.br/~lopes> e <http://acfrery.googlepages.com/> se encontram todas as implementações em R apresentadas nesse livro, assim como texto suplementares e exercícios.

Capítulo 2

Variáveis Aleatórias e a Plataforma R

R é uma linguagem de programação de alto nível que oferece recursos para a produção de gráficos de qualidade bem como para a realização de cálculos estatísticos com boa precisão numérica.

Dentre as virtudes de R não podemos deixar de mencionar que ela obedece os padrões de software livre e que está disponível para uma boa variedade de plataformas de hardware e sistemas operacionais. Há uma grande comunidade de desenvolvedores contribuindo permanentemente para o desenvolvimento de novas funcionalidades, bem como para a correção dos eventuais problemas que R apresenta. A comunidade R_STAT, disponível em http://br.groups.yahoo.com/group/R_STAT, é um exemplo disso para usuários de R no Brasil. O número de bibliotecas de qualidade para as mais diversas aplicações cresce constantemente aumentando, com isso, a aplicabilidade deste sistema em diversas áreas.

R pode ser obtido tanto compilado quanto para compilar pelo sítio Web <http://www.r-project.org>.

Na seção seguinte apresentaremos os elementos de probabilidade necessários para construir técnicas de simulação estocástica. Embora R forneça funções para a geração de ocorrências de variáveis aleatórias com diversas distribuições de interesse, consideramos apropriado discutir as técnicas subjacentes a essas rotinas.

2.1 Transformações de variáveis aleatórias

A distribuição uniforme no intervalo unitário ocupa um lugar de destaque no universo da simulação estocástica. Ela é definida pela densidade

$$f(x) = \mathbb{1}_{(0,1)}(x), \quad (2.1.1)$$

onde $\mathbb{1}_A$ é a função indicadora do conjunto A , isto é, $\mathbb{1}_A(x) = 1$ se $x \in A$, 0 caso contrário. A probabilidade de qualquer evento dentro do intervalo unitário depende apenas do comprimento do intervalo, isto é, $\Pr(a \leq X \leq b) = 1/(b - a)$ para todos $0 < a < b < 1$.

A maioria dos sistemas computacionais oferecem rotinas para a geração de eventos desta distribuição como, por exemplo, a função `rand` da linguagem de programação C. Esse é o assunto da seção 2.2. Nesta seção veremos resultados que nos permitirão obter ocorrências de variáveis aleatórias de qualquer distribuição discreta, contínua ou mista a partir de ocorrências de variáveis aleatórias uniformemente distribuídas no intervalo unitário. Outras técnicas, dentre as quais se destacam os métodos de rejeição e de composição, podem ser vistas nos textos [3, 16].

Neste ponto é conveniente lembrar que conhecer a distribuição de uma variável aleatória consiste em ser capaz de calcular a probabilidade de qualquer evento dentro da classe de eventos possíveis, suas uniões e interseções. A função de distribuição acumulada, definida por $F(t) = \Pr(X \leq t)$ para todo $t \in \mathbb{R}$, é capaz de caracterizar a distribuição de qualquer variável aleatória X .

As propriedades das funções de distribuição acumulada são as seguintes:

1. São contínuas à direita, isto é, se $x_n \downarrow x$ então $F(x_n) \rightarrow F(x)$.
2. Se $x \rightarrow -\infty$ então $F(x) \rightarrow 0$, e se $x \rightarrow \infty$ então $F(x) \rightarrow 1$.
3. Se $x < y$ então $F(x) \leq F(y)$.

Reciprocamente, toda função satisfazendo as três propriedades acima é função de distribuição acumulada, isto é, ela caracteriza a distribuição de alguma variável aleatória real.

Caso exista uma função $f: \mathbb{R} \rightarrow \mathbb{R}_+$ para a qual $F(t) = \int_{-\infty}^t f(x) dx$, então dizemos que F caracteriza a distribuição de uma variável aleatória contínua. Nesse caso, f recebe o nome de densidade de probabilidade, ou simplesmente de densidade, e também caracteriza a distribuição da variável aleatória.

Toda densidade de probabilidade possui duas propriedades:

1. Ela é não negativa, e o conjunto $\{x \in \mathbb{R}: f(x) > 0\}$ recebe o nome de suporte da densidade ou da distribuição, e

2. a sua integral é unitária, isto é, $\int_{\mathbb{R}} f = 1$.

A função de distribuição acumulada de variáveis aleatórias contínuas é estritamente crescente no suporte e, portanto, inversível nesse conjunto.

Uma informação muito importante que decorre do conhecimento da distribuição de uma variável aleatória é a dos seus *momentos*. O momento de ordem k da variável aleatória X com distribuição caracterizada pela densidade f é dado pela integral

$$\mathbb{E}(X^k) = \int_{\mathbb{R}} x^k f(x) dx,$$

se ela existir. Caso se trate de uma variável aleatória discreta, a integral é substituída por uma soma. O primeiro momento recebe o nome de *esperança* ou *média*, e a diferença entre o segundo momento e o quadrado da esperança é chamado *variância*, isto é

$$\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2.$$

Caso o espaço amostral da variável aleatória seja finito ou enumerável, dizemos se tratar de uma variável aleatória discreta. Nesse caso, basta conhecer a probabilidade de cada evento atômico para caracterizar a sua distribuição, isto é, a função $(x_i, p_i = \Pr(X = x_i))$ para todo x_i evento elementar. A função (x_i, p_i) recebe o nome de função de probabilidade, e é possível ver que a função de distribuição acumulada, neste caso, é constante por partes: $F(t) = \sum_{x_i \leq t} p_i$.

Veremos a seguir dois casos importantes de cada tipo de variável aleatória.

A distribuição da variável aleatória X chama-se exponencial padrão se a sua densidade é $f(t) = \exp\{-t\} \mathbb{1}_{\mathbb{R}_+}(t)$. Com isso, a sua função de distribuição acumulada é $F(t) = (1 - \exp\{-t\}) \mathbb{1}_{\mathbb{R}_+}(t)$. A distribuição da variável aleatória $Y = \lambda X$, com X exponencial padrão e $\lambda > 0$ chama-se exponencial de média λ , e a sua densidade é

$$f_Y(t) = \frac{1}{\lambda} \exp\{-t/\lambda\} \mathbb{1}_{\mathbb{R}_+}(t). \quad (2.1.2)$$

A figura 2.1 mostra a densidade e a função de distribuição acumulada desta distribuição, para $\lambda = 1$, restringindo o suporte ao intervalo $[0, 3]$ para fins de exibição.

A distribuição da variável aleatória X é chamada normal padrão se a sua densidade é

$$f(t) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}t^2\right\},$$

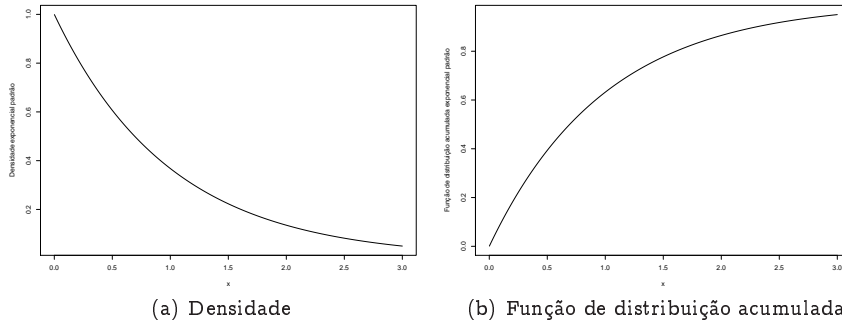


Figura 2.1: Densidade e função de distribuição acumulada da lei exponencial padrão

para todo $t \in \mathbb{R}$. Não dispomos de uma forma explícita para a função de distribuição acumulada da distribuição normal padrão, mas há uma relação entre ela e uma importante função especial chamada “função de erro”:

$$\Phi(t) = \frac{2}{\sqrt{\pi}} \int_0^t \exp\{-x^2\} dx.$$

É frequente encontrar a notação $\Phi(t) = \text{erf}(t)$, e vale que $\Phi(t) = 2F(\sqrt{2}t) - 1$, com F a função de distribuição acumulada da lei normal padrão. A figura 2.2 mostra a densidade e a função de distribuição acumulada desta distribuição, restringindo o suporte ao intervalo $[-3, 3]$ para fins de exibição.

Se X segue uma distribuição normal padrão, então $Y = \sigma X + \mu$, com $\mu \in \mathbb{R}$ e $\sigma > 0$ segue uma distribuição normal de média μ e desvio padrão σ (ou, equivalentemente, variância σ^2). Neste caso, a densidade de Y é

$$f_Y(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(t - \mu)^2\right\}.$$

Uma vez definida a distribuição normal de média μ e variância σ^2 , estamos em condições de definir uma outra lei de interesse para a modelagem de dados positivos: a distribuição log-normal. Se X segue uma lei de média μ e variância σ^2 , então $Y = e^X$ segue uma distribuição log-normal de parâmetros μ e σ^2 . A densidade que caracteriza esta distribuição é

$$f_Y(y) = \frac{1}{y\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{\log(y) - \mu}{\sigma}\right)^2\right\}. \quad (2.1.3)$$

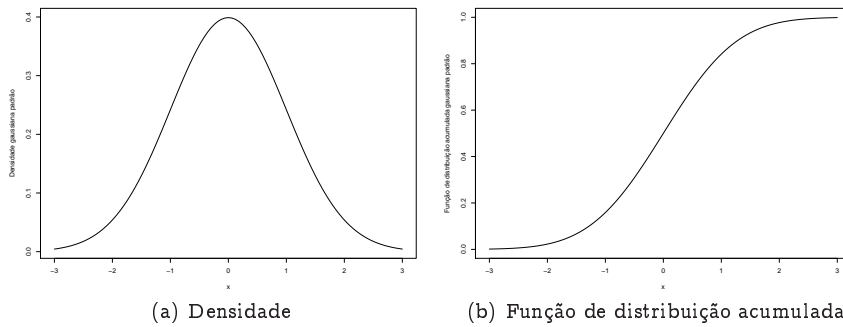


Figura 2.2: Densidade e função de distribuição acumulada da lei normal padrão

Note que μ e σ^2 são apenas os parâmetros que indexam esta lei; não são a média e a variância da distribuição. A média desta distribuição é

$$\mathbb{E}(Y) = \exp\left\{\mu + \frac{\sigma^2}{2}\right\},$$

enquanto a variância é dada por

$$\text{Var}(Y) = \exp\{2\mu + \sigma^2\} \left(\exp\{\sigma^2\} - 1\right).$$

A figura 2.3(a) mostra a densidade da equação (2.1.3) para $\mu = 0$ e $\sigma^2 = 1$.

As distribuições vistas até agora (exponencial, normal e log-normal) possuem suporte infinito, sendo a segunda toda a reta enquanto as outras duas possuem suporte sobre os reais positivos. Veremos a seguir um importante caso de distribuição com suporte limitado, a lei beta.

A variável aleatória $X: \Omega \rightarrow (0, 1)$ possui distribuição beta com parâmetros $p, q > 0$ se a sua densidade é

$$f_X(x) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} x^{p-1} (1-x)^{q-1} \mathbb{1}_{(0,1)}(x). \quad (2.1.4)$$

A distribuição beta é muito flexível e conveniente para descrever fenômenos definidos sobre um suporte limitado, tal como veremos na seção 6.3. Dentre os casos particulares descritos por esta lei podemos mencionar a distribuição uniforme, caracterizada pela densidade dada na equação (2.1.1), obtida fazendo $p = q = 1$. A figura 2.3(b) mostra quatro densidades

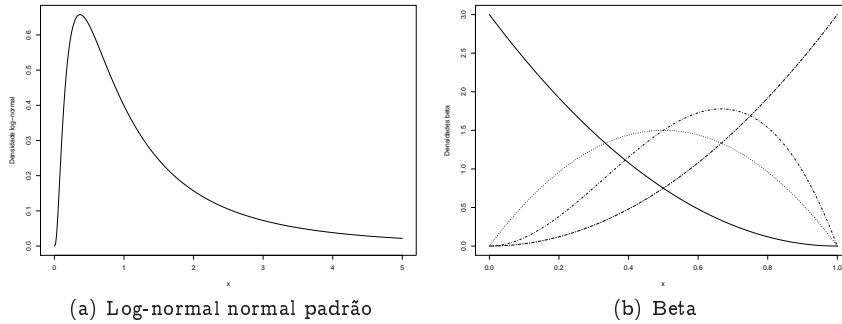


Figura 2.3: Densidades long-normal e beta

beta: $(p, q) = (1, 3)$ em traço contínuo, $(p, q) = (2, 2)$ em pontilhado, $(p, q) = (3, 2)$ em ponto-e-traço e $(p, q) = (3, 1)$ em traços curtos e longos.

Uma das distribuições discretas mais simples é a lei de Bernoulli. Uma variável aleatória discreta segue a lei de Bernoulli com parâmetro $p \in [0, 1]$ se ela adota o valor 1 com probabilidade p ou o valor 0 com probabilidade $1 - p$. Desta forma, a função de probabilidade que caracteriza esta lei é $((0, 1 - p), (1, p))$. A soma de n variáveis aleatórias independentes e identicamente distribuídas, cada uma obedecendo uma distribuição de Bernoulli com parâmetro p segue uma lei Binomial com parâmetro (n, p) . A função de probabilidade que caracteriza esta distribuição é $(k, \binom{n}{k} p^k (1 - p)^{n-k})$ para todo $0 \leq k \leq n$.

A distribuição de Poisson caracteriza a contagem do número de eventos que, isoladamente, são raros mas para os quais são feitas muitas observações independentes. Imaginemos, por exemplo, a situação de observar em um pequeno intervalo de tempo a chegada de um freguês em um banco, ou a ocorrência de um sinistro para uma empresa de seguros devido a um roubo de carro; a probabilidade disso ocorrer é pequena, mas quando a observação é realizada ao longo de um intervalo maior ela se torna razoável. Nessas condições, e com algumas restrições adicionais que podem ser vistas em [6], podemos modelar o número de eventos com a variável aleatória X cuja lei é dada por

$$\Pr(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad (2.1.5)$$

para todo $k \in \mathbb{N}_0$, onde $\lambda > 0$ é a média e o parâmetro que caracteriza a distribuição. Tal como definido acima, trata-se do processo de Poisson homogêneo; na seção 6.1 veremos uma forma mais geral deste importante

modelo.

Temos, assim, sete exemplos de variáveis aleatórias que não obedecem a distribuição uniforme: duas contínuas (exponencial, normal, log-normal e beta) e três discretas (Bernoulli, binomial e Poisson). Veremos a seguir como transformar variáveis aleatórias uniformes em variáveis aleatórias que obedecem outra distribuição. O resultado que será apresentado a seguir é um dos pilares da simulação estocástica.

Teorema 2.1 (Inversão – Caso Geral). *Sejam $F: \mathbb{R} \rightarrow [0, 1]$ a função de distribuição acumulada de uma variável aleatória e F^- a sua inversa generalizada, dada por $F^-(t) = \inf\{x \in \mathbb{R} : t \leq F(x)\}$. Se U é uma variável aleatória que segue uma lei uniforme no intervalo $(0, 1)$ então F é a função de distribuição acumulada da variável aleatória resultante da transformação $V = F^-(U)$.*

Lema 2.1 (Inversão – Caso Contínuo). *Quando a variável aleatória de interesse é contínua vale que $F^-(t) = F^{-1}(t)$ para todo $t \in \mathbb{R}$.*

A aplicabilidade deste resultado geral restringe-se somente à disponibilidade de boas implementações de F^{-1} ou de F^- . Casos importantes para os quais não há formas explícitas simples são as distribuições normal, log-normal e beta.

No caso da distribuição exponencial com parâmetro λ , a função de distribuição acumulada é dada por $F(t) = (1 - \exp\{-t/\lambda\}) \mathbb{1}_{\mathbb{R}_+}(t)$, logo a inversa é dada por $F^{-1}(t) = -\lambda \log(1 - t)$. Com isso, a variável aleatória $Y = -\lambda \log(1 - X)$ segue uma distribuição exponencial com média λ quando X segue uma lei uniforme em $(0, 1)$.

Já no caso gaussiano, como não dispomos de uma forma explícita nem para a função de distribuição acumulada nem para sua inversa, devemos lançar mão de outra técnica ou de rotinas numéricas.

Para gerar ocorrências da distribuição de Bernoulli podemos usar o Teorema 2.1 e fazer $Y = \mathbb{1}_{(0,p)}(U)$. Com isso, Y irá obedecer a lei desejada com parâmetro p . As ocorrências de variáveis aleatórias binomiais podem ser construídas como a soma de eventos de variáveis aleatórias Bernoulli independentes.

A geração de ocorrências de variáveis aleatórias Poisson pode parecer inviável com este método pois, em princípio, requereria o cálculo de infinitos valores de probabilidade. Esse problema pode ser contornado fazendo uso de um algoritmo iterativo que, dada a ocorrência da variável aleatória uniforme $X(\omega) = x$, calcule o valor inicial $p_0 = \Pr(Y = 0) = e^{-\lambda}$ e faça a comparação com apenas o primeiro “degrau” da inversa generalizada da função de distribuição acumulada. Se $p_0 \geq x$ então retorne 0, caso contrário calcule $p_1 = \Pr(Y = 1) = \lambda e^{-\lambda}$ e faça a comparação com $p_0 + p_1$, e assim

por diante. Esse algoritmo se torna eficiente se os valores p_0, p_1, \dots são armazenados entre chamadas à rotina de simulação.

Uma conexão interessante entre as distribuições exponencial e de Poisson decorre do fato de considerar o tempo entre os eventos de um processo de Poisson. A equação (2.1.5) descreve apenas o número de eventos que ocorre em um intervalo de tempo padronizado. Resta modelar em quais instantes esses eventos ocorrem e, para isso, podemos lançar mão de uma propriedade adicional do processo de Poisson: a falta de memória.

Tal como foi formulado, esse processo possui a seguinte propriedade para T , o tempo entre ocorrências sucessivas:

$$\Pr(T > t + s \mid T > s) = \Pr(T > t), \quad (2.1.6)$$

para todo $s, t > 0$. Em palavras, a equação (2.1.6) mostra que o fato de já haver esperado um tempo superior a $s > 0$ em nada influi quanto tempo mais deveremos esperar até observar mais um evento. A distribuição que satisfaz essa propriedade é a exponencial, caracterizada pela densidade dada na equação (2.1.2).

A propriedade de falta de memória é útil como forma alternativa para geração de ocorrências de variáveis aleatórias com distribuição de Poisson, bem como para a simulação de modelos mais gerais.

Outra propriedade importante do processo de Poisson é a falta de informação sobre a localização exata dos eventos. Se sabemos que ocorreram exatamente n chegadas no intervalo (t_1, t_2) , com $t_1 < t_2$, então as suas localizações seguem distribuições independentes e identicamente distribuídas segundo leis uniformes nesse intervalo.

De volta à distribuição gaussiana, podemos, em princípio, utilizar o seguinte resultado discutido em [6]: sejam U_1, U_2 variáveis aleatórias independentes e identicamente distribuídas segundo uma lei uniforme no intervalo unitário. Definamos $R = \sqrt{-2 \log U_1}$ e $\Theta = \pi(2U_2 - 1)$. As variáveis aleatórias X, Y definidas por $X = R \cos \Theta$ e $Y = R \sin \Theta$ são independentes com distribuição comum gaussiana padrão.

A plataforma R oferece diversas rotinas para a geração de ocorrências de variáveis aleatórias com estas e muitas outras distribuições. Dentre elas podemos mencionar `runif`, para ocorrências de uniformes, `rbinom` para binomiais (e, como caso particular, de Bernoulli), `rnorm` para normais e `rpois` para Poisson.

Tal como vimos nesta seção, a geração de ocorrências de variáveis aleatórias uniformes é essencial para a obtenção de eventos de variáveis aleatórias com outras distribuições. A seção seguinte discute esse problema.

2.2 Geradores de números pseudo-aleatórios

A geração de números aleatórios é o alicerce de qualquer sistema de simulação estocástica. Porém, nos computadores digitais as conhecidas funções que geram números aleatórios não são efetivamente aleatórias. Números realmente aleatórios são gerados por um processo físico. Para isso, são construídos dispositivos físicos que analisam fenômenos microscópicos ou quânticos e através de um conversor digital conseguem gerar um número aleatório.

Na prática o que se usa em simulação estocástica são os geradores de números pseudo-aleatórios. Esses geradores produzem uma seqüência determinística de números inteiros ou em ponto flutuante na precisão do computador, que imita uma seqüência de variáveis aleatórias independentes e uniformemente distribuídas entre 0 e 1.

A essência de uma seqüência de números pseudo-aleatórios é a sua imprevisibilidade, no sentido de que ninguém é capaz de, ao vê-la, dizer qual é a regra determinística que a produz e conseguir prever qual é o próximo número da seqüência.

O método mais conhecido que gera uma seqüência de números pseudo-aleatórios é o LCG (*Linear Congruential Generator*).

O LCG gera uma seqüência $\{u_i\}$ de números reais entre 0 e 1. Para isso, ele usa a seguinte equação de recorrência para gerar números inteiros entre 0 e m :

$$v_{i+1} = (a \times v_i + b) \bmod m, \quad i \geq 0,$$

onde a , b e m são inteiros e \bmod é a operação módulo, isto é, o resto da divisão inteira. O termo v_0 dessa recorrência é chamado de semente do gerador e é um outro parâmetro de entrada para o método. A seqüência $\{u_i\}_{i \geq 0}$ é obtida fazendo $u_i = v_i/m$.

Uma desvantagem do método LCG é que, por essa fórmula de construção, ele gera uma seqüência periódica. O ideal é fazer uma boa escolha dos parâmetros de entrada a fim de que o período seja o maior possível. O período do LCG é no máximo m , e na maioria dos casos é menor do que isso. Para se obter um período máximo deve-se fazer as escolhas para a , b , m e v_0 de acordo com os seguintes critérios:

1. $a > 0$, $b > 0$;
2. $m > \max(a, b, v_0)$;
3. b e m devem ser primos entre si;
4. $a - 1$ deve ser divisível por todos os fatores primos de m ;

5. $a - 1$ deve ser um múltiplo de 4 se m for um múltiplo de 4.

Uma boa escolha, segundo Press et al. em [15], é $a = 1664525$, $b = 1013904223$, e $m = 2^{32}$.

O método LCG é capaz de gerar boas seqüências pseudo-aleatórias, mas é conhecido que suas propriedades estão longe do ideal. Se for necessário um gerador melhor, sugere-se utilizar o *Mersenne Twister* (MT) proposto por Matsumoto e Nishimura em 1977 [11]. Hoje, existem algoritmos rápidos baseados no MT que tem período $2^{19937} - 1$. No endereço <http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/emt.html> encontra-se uma excelente implementação para o gerador MT. Para maiores detalhes em algoritmos de geração de números pseudo-aleatórios veja [15].

Na plataforma R existem vários outros geradores que são muito eficientes, não é o objetivo desse texto detalhá-los.

2.3 A distribuição normal multivariada

Um dos modelos mais importantes na hora de descrever o comportamento de duas ou mais variáveis relacionadas entre si é o normal multivariado.

Diremos que um vetor de variáveis aleatórias segue uma distribuição normal multivariada se a sua densidade é dada por

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{m/2} |\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})\right\},$$

onde $\mathbf{x} = (x_1, \dots, x_m)'$ é a observação em \mathbb{R}^m , $\boldsymbol{\mu} = (\mu_1, \dots, \mu_m)' \in \mathbb{R}^m$ é o vetor de médias e Σ é a matriz de covariância (que é de tamanho $m \times m$ e positiva definida, logo Σ^{-1} existe). O vetor de médias é o ponto de máximo da densidade, isto é, a sua moda.

A matriz de covariância é da forma $\Sigma = (\sigma_{ij})$ onde

$$\begin{aligned} \sigma_{ij} &= \sigma_{ji} \text{ é a covariância entre as componentes } i \text{ e } j, \\ \sigma_{ii} &= \sigma_i^2 > 0 \text{ é a variância da componente } i. \end{aligned}$$

A covariância entre as componentes i e j pode também ser escrita em termos da correlação $-1 < \rho_{ij} < 1$ como $\sigma_{ij} = \rho_{ij} \sigma_i \sigma_j$. Dentre outras referências, os textos [10, 13, 17] fornecem detalhes desta importante distribuição.

Um resultado importante desta distribuição que serve para gerar ocorrências de variáveis aleatórias que seguem esta lei é o seguinte. Se $\mathbf{X} = (X_1, \dots, X_n)$ é um vetor formado por n variáveis aleatórias independentes e identicamente distribuídas com distribuição normal padrão, $\boldsymbol{\mu} \in \mathbb{R}^n$ é um vetor n -dimensional e A é uma matriz de dimensão $n \times n$, então o vetor

de variáveis aleatórias $\mathbf{Y} = \mathbf{AX} + \boldsymbol{\mu}$ segue uma lei normal multivariada de média $\boldsymbol{\mu}$ e matriz de covariância $\boldsymbol{\Sigma} = \mathbf{AA}'$.

A função `mvnorm` do R permite simular ocorrências desta lei multivariada.

2.4 Exemplos de uso do R

Nesta seção veremos alguns exemplos de uso da plataforma R para a geração de ocorrências de variáveis aleatórias com as distribuições definidas nas seções 2.1 e 2.3.

Código 2.1: Histograma de uniformes

```

1 > x = runif(n=100, min=0, max=1)
2 > hist(x, breaks=30, main='Histograma de Observacoes
   Uniformes', ylab="Frequencia")

```

Na primeira linha do código 2.1 atribuímos ao vetor `x` cem ($n=100$) observações de variáveis aleatórias independentes identicamente distribuídas segundo uma lei uniforme no intervalo $(0, 1)$. O intervalo pode ser modificado alterando os parâmetros `min` e `max` na chamada à função `runif`. Na segunda linha deste código geramos um histograma com `breaks=30` intervalos, e o exibimos com a legenda que é visível na Figura 2.4. Deixamos para o leitor curioso descobrir a enorme variedade de opções que podem ser dadas para construir gráficos em R.

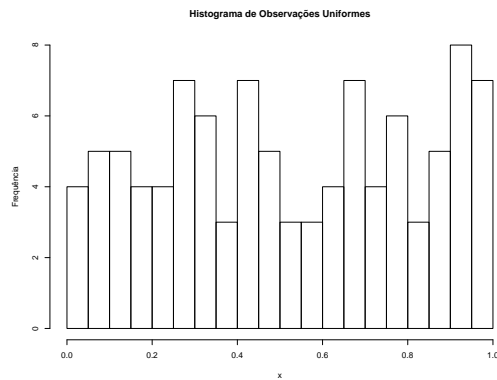


Figura 2.4: Histograma de cem observações uniformes

O código 2.2 mostra como gerar cem ocorrências de variáveis aleatórias independentes e identicamente distribuídas segundo uma lei exponencial de taxa $\lambda = 2$. A visualização do histograma desses dados, gerada pela segunda linha deste código, mostra-se na Figura 2.5.

Código 2.2: Histograma de exponenciais

```

1 > x = rexp(n=100, rate=2)
2 > hist(x, breaks=30, main='Histograma de Observacoes
   Exponenciais', ylab='Frequencia')

```

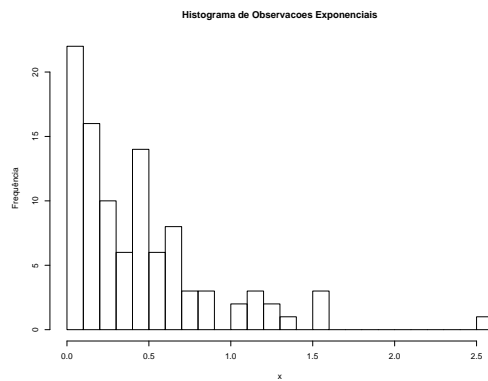


Figura 2.5: Histograma de cem observações exponenciais

Para ilustrar a simulação de ocorrências de variáveis aleatórias normais multivariadas lançaremos mão de um dos exemplos de análise de imagens mostrado em [18]. Nesse texto são analisados dados de diversas regiões de uma pintura, e a estimação dos parâmetros μ e Σ , supondo a distribuição gaussiana multivariada, para uma delas, a que corresponde a um cacto, é

$$\hat{\mu}_{\text{cacto}} = \begin{pmatrix} 0.281 \\ 0.420 \\ 0.252 \end{pmatrix}, \hat{\Sigma}_{\text{cacto}} = 10^{-3} \begin{pmatrix} 2.001 & 1.938 & 1.783 \\ 1.938 & 3.001 & 1.247 \\ 1.783 & 1.247 & 2.610 \end{pmatrix}.$$

O código 2.3 mostra uma sequência de instruções para gerar e visualizar cem ocorrências de variáveis aleatórias distribuídas segundo a lei gaussiana multivariada com os parâmetros $\hat{\mu}_{\text{cacto}}$ e $\hat{\Sigma}_{\text{cacto}}$.

A linha 1 carrega o pacote MASS, que fornece a função para gerar este tipo de ocorrências. A linha 2 atribui à variável μ o valor da média, e à

Código 2.3: Geração de gaussianas multivariadas

```

1 > library(MASS)
2 > mu = c(0.281, 0.420, 0.252)
3 > Sigma = matrix(c(2.001, 1.938, 1.783, 1.938, 3.001,
4   1.247, 1.783, 1.247, 2.610), 3, 3)
5 > y = mvrnorm(n=100, mu, Sigma)
6 > pairs(y)

```

variável Sigma a matriz de covariância de interesse. Note como é estipulada a estrutura de uma matriz de tamanho 3×3 aos dados. A linha 4 produz a geração, enquanto a linha 5 mostra como obter o gráfico exibido na figura 2.6.

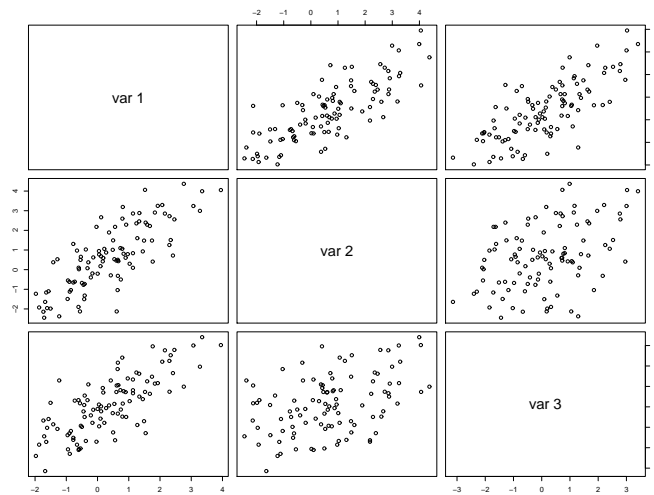


Figura 2.6: Gráfico de pares das ocorrências de normais multivariadas

Empregando recursos um pouco mais avançados do sistema R é possível colocar os histogramas de cada componente nas diagonais do diagrama de pares. A figura 2.7 mostra o resultado dessa modificação. O código para produzir este gráfico, com pequenas alterações, é o que consta na página de ajuda da função `pairs`.

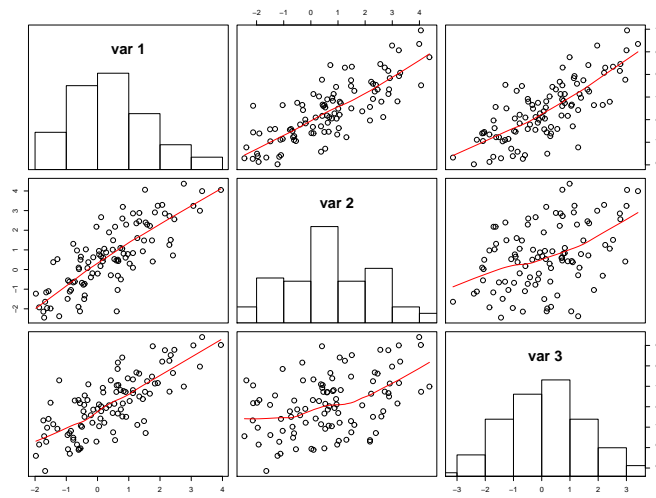


Figura 2.7: Gráfico de pares e histogramas das ocorrências de gaussianas multivariadas

Capítulo 3

Cenários de vida de um indivíduo

Esse capítulo tem como objetivo introduzir os conceitos básicos de atuária para elaborar um simulador estocástico que gere cenários de vida de um indivíduo. Para isso, o capítulo começa apresentando alguns conceitos básicos e depois apresenta o problema que será discutido, sua modelagem e a solução utilizando a plataforma R.

3.1 Conceitos básicos de matemática atuarial

Um dos elementos básicos dos seguros que lidam com a vida humana é a tábua de mortalidade. Essa seção apresenta inicialmente um modelo probabilístico para a sobrevivência de um indivíduo. Depois, introduz-se as tábuas de mortalidade.

3.1.1 Modelo probabilístico para sobrevivência de um indivíduo

Considere um indivíduo com x anos e T seu tempo futuro de vida, denotado por $T(x)$. Portanto, a idade em que o indivíduo morre é $x + T(x)$.

O tempo futuro de vida T de um indivíduo é uma variável aleatória com uma função de distribuição

$$G(t) = \Pr(T \leq t), t \geq 0.$$

A função $G(t)$ representa a probabilidade que o indivíduo morrerá dentro de t anos, para qualquer valor de t . Assume-se que a função de distribuição de probabilidade de T , denotada por G , é conhecida e é contínua, e que G possui uma função de densidade g ($g(t) = G'(t)$).

Dentro desse contexto, a probabilidade que um indivíduo na idade x morra dentro de t anos, denotada por ${}_tq_x$, pode ser escrita da seguinte maneira:

$${}_tq_x = G(t).$$

Similarmente, a probabilidade de um indivíduo sobreviver por mais t anos, denotada por ${}_tp_x$ é:

$${}_tp_x = 1 - {}_tq_x.$$

A probabilidade que um indivíduo com x anos sobreviva por mais s anos e depois morra dentro de t anos é:

$${}_{s|t}q_x = \Pr(s < T < s + t) = G(s + t) - G(s) = {}_{s+t}q_x - {}_sq_x.$$

A probabilidade condicional de que um indivíduo com x anos sobreviva mais t anos depois de ter chegado a idade $x + s$ é:

$${}_tp_{x+s} = \Pr(T > s + t \mid T > s) = \frac{1 - G(s + t)}{1 - G(s)}.$$

Da mesma forma, pode-se dizer que a probabilidade de um indivíduo morrer dentro de t anos dado que ele atingiu a idade $x + s$ é:

$${}_tq_{x+s} = \Pr(T \leq s + t \mid T > s) = \frac{G(s + t) - G(s)}{1 - G(s)}.$$

O valor esperado do tempo restante de um indivíduo com idade x é $E[T]$, será denotado por $\overset{\circ}{e}_x$, e sua definição é:

$$\overset{\circ}{e}_x = E[T] = \int_0^{\infty} t g(t) dt.$$

De agora em diante, se $t = 1$ o índice t é omitido nas notações apresentadas.

Define-se agora a variável aleatória K que representa o número de anos completos futuros vividos por um indivíduo com uma idade x . Assim, pode-se dizer que

$$\Pr(K = k) = \Pr(k \leq T < k + 1) = {}_kp_x \ q_{x+k}, \quad k = 0, 1, \dots$$

O valor esperado de K é chamado de valor esperado de truncamento do tempo de vida futura de um indivíduo com idade x , e é denotado por e_x . Portanto,

$$e_x = E[K] = \sum_{k=1}^{\infty} k \Pr(K = k) = \sum_{k=1}^{\infty} k {}_k p_x q_{x+k}.$$

Ou então,

$$e_x = \sum_{k=1}^{\infty} \Pr(K \geq k) = \sum_{k=1}^{\infty} {}_k p_x.$$

Seja S uma fração do ano durante o qual o indivíduo com idade inicial x morre, isto é,

$$T = K + S.$$

A variável aleatória S tem distribuição contínua entre 0 e 1. Como usualmente é feito, nesse trabalho considera-se que a variável S é distribuída uniformemente entre 0 e 1. Será considerado também nesse trabalho, como também é de costume na literatura, que as variáveis K e S são independentes.

Assim, pode-se mostrar que:

$$\overset{\circ}{e}_x = e_x + \frac{1}{2},$$

e que

$$\text{Var}[T] = \text{Var}[K] + \frac{1}{12}.$$

Para inteiros positivos m , pode-se definir a variável aleatória

$$S^{(m)} = \frac{1}{m}[mS + 1],$$

que é obtida como um arredondamento para o menor múltiplo de $1/m$ maior que S . A função de probabilidade de $S^{(m)}$ está definida nos pontos $\frac{1}{m}, \frac{2}{m}, \dots, 1$. Observe que a independência entre K e S implica na independência entre K e $S^{(m)}$. Mais ainda, como foi assumido que S é uniforme entre 0 e 1, então pode-se provar que $S^{(m)}$ tem uma distribuição discreta uniforme.

3.1.2 Tábuas de mortalidade

Partindo-se de um número fechado de participantes, denominado “raiz”, em que o gênero pode ser levado em consideração, a *tábua de mortalidade* revela a quantidade de pessoas vivas anualmente em cada idade, ou seja,

trata-se de uma tábua determinada pelas taxas anuais de mortalidade ou de sobrevivência.

Existem vários tipos de tábuas atuariais, como a AT-49, AT-83 e a AT-2000, onde AT quer dizer *annuity table* e o número refere-se ao ano em que as estatísticas passadas começaram a valer. Tábuas mais modernas, como a AT-2000, são adotadas em várias empresas no Brasil, principalmente para planos mais flexíveis como o PGBL. Ela possui o tempo de vida médio mais elevado que as outras, reduzindo assim, o valor do benefício a ser pago pelas empresas.

Por exemplo, na AT-49 a expectativa de sobrevida para alguém com 60 anos é de 18,5 anos, na AT-83 é de 22,6 e na AT-2000 é de 24,6. As tábuas mais modernas não devem ser vistas como injustas ou benéficas apenas às seguradoras. Refletem as mudanças que a sociedade vem sofrendo, como o aumento da expectativa de vida, melhores condições sanitárias e avanços na medicina. Nos Estados Unidos, por exemplo, a quantidade de pessoas centenárias passou de 4000 em 1940, para mais de 61000, em 1997. Em resumo, se uma empresa trabalhar com tábuas desatualizadas, ela corre o risco de não ter como pagar no futuro a renda mensal vitalícia aos beneficiários.

A construção dessas tábuas pode ser oriunda da experiência das seguradoras ou se valer dos dados dos censos demográficos. Não existe ainda uma tábua de mortalidade genuinamente brasileira. O principal motivo é que ainda não deu tempo para construí-la, pois seria preciso observar a vida de um grupo de pessoas por um longo período de tempo.

As tábuas, sob condição universal, são compostas de seis colunas:

x : coluna de idades, em anos;

l_x : número de pessoas vivas na idade x do grupo em estudo, de um total inicial hipotético de 100.000 novos nascimentos;

d_x : número de pessoas do grupo que morrem entre as idades x e $x + 1$;

q_x : probabilidade anual de morte, ou seja, a razão entre o número de pessoas do grupo que morrem numa idade w (d_w) pelo número de pessoas da idade de w (l_w). Matematicamente, tem-se: $q_x = \frac{d_x}{l_x}$ ou $q_x = \frac{l_x - l_{x+1}}{l_x}$;

p_x : probabilidade anual de sobrevivência, obtida pela relação $p_x = 1 - q_x$;

e_x : expectativa de vida para um indivíduo com idade x .

Observe que a distribuição de K e outras variáveis podem ser calculadas pela tabela. Por exemplo,

$${}_k p_x = p_x p_{x+1} p_{x+2} \dots p_{x+k-1}, \quad k = 1, 2, 3, \dots$$

Para se obter a distribuição de T através da tabela deve-se usar uma interpolação. Um tipo de interpolação supõe que ${}_u q_x$ é uma função linear em u , para u entre 0 e 1, e x é um número inteiro. Em outras palavras, isso significa que é assumida uma distribuição uniforme de sobrevivência ao longo do ano.

Apresenta-se nas tabelas 3.1, 3.2 e 3.3 a tabela de mortalidade americana masculina de 2004, disponível no endereço http://www.cdc.gov/nchs/products/pubs/pubd/nvsr/51/51_03.htm.

3.2 Apresentação do problema

Uma das grandes incertezas que um gerente de ativos e passivos de uma seguradora no ramo vida deseja modelar é a sobrevivência de um indivíduo que possui uma apólice com ela. Assim, nesse trabalho, o primeiro problema que será considerado nesse livro é o de gerar cenários de uma variável aleatória M que representa o número de meses que um indivíduo sobrevive a partir do momento em que ele fez um seguro de previdência. Para simplificar o problema, assumamos que esse indivíduo aderiu ao plano no dia de seu n -ésimo aniversário e que a probabilidade de sobrevivência desse indivíduo obedece a lei de distribuição definida pela tabela de mortalidade americana masculina de 2004.

3.3 Modelagem

Para modelar a incerteza em relação à sobrevivência do indivíduo utilizando a tabela de mortalidade americana masculina de 2004, o simulador gera uma variável aleatória distribuída por uma Bernoulli(q_x), onde q_x é a probabilidade do indivíduo na idade x morrer entre o ano x e $x + 1$. Para isto é gerado um número aleatório uniformemente distribuído entre $[0, 1]$. Se o número gerado for maior do que q_x , então o indivíduo sobrevive e o processo é repetido para o ano seguinte. Caso contrário, se o número gerado for menor ou igual a q_x , o indivíduo morre. Resta saber em qual mês ocorre a morte. Para isto, o simulador gera um número aleatório discreto distribuído uniformemente entre 1 e 12 e o mês em questão corresponderá ao mês da morte do indivíduo.

3.4 Simulação usando o R

O primeiro passo para implementar o simulador é o de carregar a tábua de mortalidade no R. A maneira mais fácil é carregá-la no sistema através de um arquivo, que pode ser criado copiando a tábua linha por linha, separando cada valor por um único espaço em branco. Colocamos também na primeira linha do arquivo os seguintes caracteres: “x l d p q”. Esse cabeçalho indica que a primeira coluna se refere à idade, a segunda ao número de pessoas vivas, a terceira ao número de pessoas que morreram no decorrer daquele ano, a quarta à probabilidade de sobreviver por mais um ano dada a idade “x” (ou seja, o valor de p_x), e, finalmente, a quinta indicando a probabilidade de não sobreviver (q_x).

Para carregar a tabela no R existe uma função que lê de um arquivo texto (formato ASCII) os seus dados. Essa função chama-se `read.table`, e ela retorna um tipo de dado, que no R é chamado de “data frame”. Os argumentos básicos dessa função são o nome do arquivo, um flag informando se existe uma linha de cabeçalho (*header*), o caracter que representa o separador dos números da tabela, e o caracter que indica o ponto decimal (“.” ou “,”). No código 3.1, a primeira linha lê a tabela do arquivo `AT2004.txt` armazenando-a na variável `AT2004M`. A segunda linha desse código mostra o comando para ver o conteúdo da tabela. Na terceira linha, o conteúdo da coluna “p” da tabela é carregado na variável `p_x`.

Código 3.1: Carregando a tábua de mortalidade através da leitura de um arquivo

```

1 > AT2004M ← read.table("./AT2004.txt", header=T, sep=" ",
   dec=".")
2 > AT2004M
3 > p_x ← AT2004M$p

```

Agora que as probabilidades p_x estão carregadas na memória do R, podemos desenvolver o simulador como uma função que retorna o número de meses que o indivíduo ainda viverá, dada a sua idade no momento de sua adesão ao plano. Esse simulador será implementado como uma função cujo nome é `CenarioDeSobrevida` e que possui os seguintes parâmetros:

- `IdadeInicial`: idade inicial que o indivíduo entra no plano;
- `p_x`: um vetor cuja entrada de índice i corresponde a probabilidade p_{i-1} de uma tábua de mortalidade.

E a função retorna um cenário para o número de meses que o indivíduo sobrevive. Seu código está na listagem 3.2.

Código 3.2: Geração de cenários de vida

```
1 > CenarioDeSobrevida ← function(IdadeInicial ,p_x){
2 +   N ← 0
3 +   Idade ← IdadeInicial
4 +   vivo ← TRUE
5 +   while(vivo) {
6 +     u ← runif(1)
7 +     if (u < p_x[Idade-1]) {
8 +       N ← N + 12
9 +       Idade ← Idade + 1
10 +    }
11 +    else {
12 +      u ← runif(1)
13 +      N ← N + as.integer(12*u) + 1
14 +      vivo ← FALSE
15 +    }
16 +  }
17 +  N
18 +}
```

Para gerar um cenário para um indivíduo que inicia o plano com 40 anos execute o seguinte comando: `CenarioDeSobrevida(40,p_x)`.

3.5 Exercício

(1) Gere 1000 cenários para os números de meses de sobrevida de um indivíduo que inicia o plano com 40 anos, calcule a média desse número de meses, compare o resultado dessa média com o valor de e_{40} que se encontra na última coluna da tábua de mortalidade, e, finalmente, visualize o histograma desses cenários.

x	l_x	d_x	p_x	q_x	e_x
0	100000	747	0.992526	0.007474	74.83
1	99253	51	0.999487	0.000513	74.40
2	99202	33	0.999672	0.000328	73.43
3	99169	24	0.999753	0.000247	72.46
4	99145	21	0.999795	0.000205	71.47
5	99124	18	0.999811	0.000189	70.49
6	99106	18	0.999818	0.000182	69.50
7	99088	17	0.999828	0.000172	68.52
8	99071	16	0.999847	0.000153	67.53
9	99055	12	0.999874	0.000126	66.54
10	99043	10	0.999898	0.000102	65.55
11	99033	11	0.999896	0.000104	64.55
12	99022	15	0.999844	0.000156	63.56
13	99007	27	0.999727	0.000273	62.57
14	98980	43	0.999565	0.000435	61.59
15	98937	61	0.999387	0.000613	60.61
16	98876	77	0.999218	0.000782	59.65
17	98799	92	0.999065	0.000935	58.70
18	98707	105	0.998939	0.001061	57.75
19	98602	115	0.998838	0.001162	56.81
20	98487	124	0.998736	0.001264	55.88
21	98363	134	0.998639	0.001361	54.95
22	98229	140	0.998578	0.001422	54.02
23	98089	141	0.998562	0.001438	53.10
24	97948	139	0.998580	0.001420	52.17
25	97809	136	0.998608	0.001392	51.25
26	97673	134	0.998632	0.001368	50.32
27	97539	131	0.998651	0.001349	49.38
28	97408	131	0.998659	0.001341	48.45
29	97277	131	0.998656	0.001344	47.52
30	97146	131	0.998648	0.001352	46.58
31	97015	133	0.998633	0.001367	45.64
32	96882	136	0.998596	0.001404	44.70
33	96746	142	0.998533	0.001467	43.76
34	96604	150	0.998445	0.001555	42.83
35	96454	160	0.998338	0.001662	41.89
36	96294	172	0.998218	0.001782	40.96
37	96122	184	0.998082	0.001918	40.04
38	95938	198	0.997932	0.002068	39.11
39	95740	214	0.997765	0.002235	38.19

Tabela 3.1: Tábua de mortalidade americana masculina de 2004: idades entre 0 e 39 anos

x	l_x	d_x	p_x	q_x	e_x
40	95526	232	0.997580	0.002420	37.28
41	95294	250	0.997371	0.002629	36.36
42	95044	272	0.997137	0.002863	35.46
43	94772	297	0.996873	0.003127	34.56
44	94475	322	0.996582	0.003418	33.67
45	94153	352	0.996268	0.003732	32.78
46	93801	381	0.995933	0.004067	31.90
47	93420	414	0.995576	0.004424	31.03
48	93006	446	0.995195	0.004805	30.17
49	92560	483	0.994792	0.005208	29.31
50	92077	520	0.994343	0.005657	28.46
51	91557	562	0.993866	0.006134	27.62
52	90995	600	0.993405	0.006595	26.79
53	90395	635	0.992973	0.007027	25.96
54	89760	670	0.992543	0.007457	25.14
55	89090	705	0.992079	0.007921	24.33
56	88385	749	0.991533	0.008467	23.52
57	87636	799	0.990879	0.009121	22.71
58	86837	861	0.990088	0.009912	21.92
59	85976	931	0.989173	0.010827	21.13
60	85045	1008	0.988142	0.011858	20.36
61	84037	1090	0.987034	0.012966	19.60
62	82947	1171	0.985877	0.014123	18.85
63	81776	1252	0.984688	0.015312	18.11
64	80524	1334	0.983433	0.016567	17.38
65	79190	1424	0.982024	0.017976	16.67
66	77766	1521	0.980436	0.019564	15.96
67	76245	1624	0.978709	0.021291	15.27
68	74621	1728	0.976838	0.023162	14.59
69	72893	1838	0.974783	0.025217	13.93
70	71055	1957	0.972467	0.027533	13.27
71	69098	2082	0.969869	0.030131	12.64
72	67016	2210	0.967022	0.032978	12.01
73	64806	2338	0.963914	0.036086	11.41
74	62468	2468	0.960494	0.039506	10.81
75	60000	2605	0.956585	0.043415	10.24
76	57395	2743	0.952211	0.047789	9.68
77	54652	2867	0.947536	0.052464	9.14
78	51785	2973	0.942587	0.057413	8.62
79	48812	3065	0.937211	0.062789	8.11

Tabela 3.2: Tábua de mortalidade americana masculina de 2004: idades entre 40 e 79 anos

x	l_x	d_x	p_x	q_x	e_x
80	45747	3149	0.931164	0.068836	7.62
81	42598	3226	0.924276	0.075724	7.15
82	39372	3286	0.916534	0.083466	6.70
83	36086	3325	0.907856	0.092144	6.26
84	32761	3335	0.898197	0.101803	5.84
85	29426	3310	0.887532	0.112468	5.45
86	26116	3242	0.875836	0.124164	5.08
87	22874	3132	0.863083	0.136917	4.73
88	19742	2976	0.849246	0.150754	4.40
89	16766	2778	0.834296	0.165704	4.09
90	13988	2543	0.818211	0.181789	3.80
91	11445	2278	0.800981	0.199019	3.54
92	9167	1993	0.782604	0.217396	3.29
93	7174	1699	0.763094	0.236906	3.06
94	5475	1410	0.742475	0.257525	2.86
95	4065	1130	0.721969	0.278031	2.68
96	2935	875	0.701889	0.298111	2.52
97	2060	654	0.682568	0.317432	2.38
98	1406	472	0.664345	0.335655	2.25
99	934	329	0.647562	0.352438	2.13
100	605	224	0.629940	0.370060	2.02
101	381	148	0.611437	0.388563	1.91
102	233	95	0.592009	0.407991	1.81
103	138	59	0.571610	0.428390	1.71
104	79	36	0.550190	0.449810	1.61
105	43	20	0.527700	0.472300	1.52
106	23	11	0.504085	0.495915	1.43
107	12	6	0.479289	0.520711	1.35
108	6	3	0.453253	0.546747	1.26
109	3	2	0.425916	0.574084	1.19
110	1	1	0.397212	0.602788	1.11
111	0	0	0.367072	0.632928	1.04
112	0	0	0.335426	0.664574	0.97
113	0	0	0.302197	0.697803	0.91
114	0	0	0.267307	0.732693	0.84
115	0	0	0.230673	0.769327	0.78
116	0	0	0.192206	0.807794	0.72
117	0	0	0.151817	0.848183	0.67
118	0	0	0.109408	0.890592	0.62
119	0	0	0.064878	0.935122	0.57
120	0	0	0.000000	1.000000	0.00

Tabela 3.3: Tábua de mortalidade americana masculina de 2004: idades entre 79 e 115 anos

Capítulo 4

Geração de cenários para variáveis financeiras

O objetivo desse capítulo é o de apresentar um modelo econométrico para as variáveis financeiras que servem de entrada para o modelo estocástico de fluxo de caixa que será apresentado no capítulo 5. Conhecer os valores futuros desses parâmetros é fundamental e é a base dos modelos de gestão de ativos e passivos.

Nesse modelo econométrico são utilizadas variáveis econômicas que se mostraram adequadas para definir a composição da carteira de ativos que formam um fundo de previdência.

Primeiramente, é feita uma pequena introdução à séries temporais. Em particular, é introduzido o modelo multivariado auto-regressivo. Depois, as variáveis a serem estudadas são apresentadas e analisadas sob o ponto de vista estatístico. Em seguida, apresenta-se o problema de gerar cenários das trajetórias futuras do modelo VAR utilizando a plataforma R.

4.1 Análise de séries temporais

A *análise de séries temporais* é uma área da Estatística dedicada ao estudo de observações que apresentam dependência no tempo. É muito comum encontrar estudos sobre esse assunto que envolvem a utilização de dados acerca da realidade econômica e financeira.

O enfoque de Box-Jenkins [1] para análise de séries temporais tem como objetivo principal a realização da previsão. Essa metodologia permite que valores futuros de uma série sejam previstos tomando por base apenas seus

valores presentes e passados. Usualmente, isso é feito explorando a correlação temporal que existe entre os valores observados da série.

A relação temporal considerada pelo enfoque de Box-Jenkins é representada formalmente por um conjunto de processos estocásticos genericamente denominados modelos ARIMA. Por envolverem apenas uma série de tempo, eles são classificados como modelos univariados.

Os modelos ARIMA resultam da combinação de três componentes: Auto-Regressivo (AR), de Integração (I) e de médias móveis (MA, por ser a abreviação de *Moving Average*).

Na metodologia de Box-Jenkins, uma série de tempo pode conter as três componentes ou apenas um subconjunto, resultando daí vários métodos para análise.

4.2 Introdução ao modelo VAR

É de interesse desse trabalho estudar as trajetórias futuras de várias variáveis ao mesmo tempo. A metodologia de Box-Jenkins é extensível também para esse caso.

O modelo de série temporal multivariado para as variáveis financeiras a ser utilizado nesse capítulo é um *modelo de vetores auto-regressivos* (VAR-*Vector Auto-Regressive models*), ele é baseado somente na componente auto-regressiva.

Num modelo VAR, o vetor $\mathbf{y}_t \in \mathbb{R}^n$ no instante t é descrito apenas por seus valores passados e pelo vetor de ruído branco $\epsilon_t \in \mathbb{R}^n$. O modelo VAR mais geral é o de ordem p , nele o vetor \mathbf{y}_t depende de $\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \dots, \mathbf{y}_{t-p}$ e do vetor de resíduos ϵ_t , que estão correlacionados entre eles no instante t mas não em momentos anteriores a t . A sua representação algébrica é a seguinte:

$$\mathbf{y}_t = \mathbf{c} + \sum_{k=1}^p \mathbf{A}_k \mathbf{y}_{t-k} + \epsilon_t,$$

onde $\mathbf{c} \in \mathbb{R}^n$ é o vetor de intercepto, $\mathbf{A}_k \in \mathbb{R}^{n \times n}$, $k = 0, 1, 2, \dots, p$, são as matrizes dos coeficientes do modelo e $\epsilon_t \in \mathbb{R}^{n \times n}$ é o vetor aleatório dos resíduos, que obedece a lei da normal multivariada e que satisfaz $E[\epsilon_t] = \mathbf{0} \in \mathbb{R}^n$, $E[\epsilon_t \epsilon_t^T] = \Sigma \in \mathbb{R}^{n \times n}$.

Os estimadores de Mínimos Quadrados Ordinários são os melhores estimadores dado que os resíduos das diferentes equações não são correlacionados e que são independentes e igualmente distribuídos ao longo do tempo. Na presença de correlação entre os resíduos das equações, conhecida como correlação contemporânea, o melhor estimador é o método de regressão aparentemente não-relacionada (SUR - *Seemly Unrelated Regression*). Para

maiores detalhes dos métodos de estimação veja o livro de Judge et al. de 1985 [7].

4.3 As variáveis financeiras

No modelo de gestão de passivos e passivos, os ativos devem ser investidos todos os meses tanto no período de diferimento quanto no período de benefícios. Para isso, foi considerada uma carteira onde o capital é aplicado em renda fixa, renda variável e em títulos públicos. Foram utilizadas as seguintes variáveis financeiras como *proxies* (aproximações) para a rentabilidade desses investimentos:

CDI: Certificado de Depósito Interfinanceiro: acumulada no mês anualizada com periodicidade mensal para representar renda fixa;

IBOVESPA: Índice de Lucratividade da Bolsa de Valores de São Paulo: com variação e periodicidades mensais para representar renda variável;

IGP-M: Índice Geral de Preços / Mercado: com variação e periodicidades mensais para representar inflação;

SWAP 180: Derivativo financeiro que tem por finalidade promover a troca (simultaneamente) de ativos financeiros entre os agentes econômicos envolvidos, por exemplo: Uma empresa possui um ativo financeiro indexado a variação do dólar comercial e deseja trocar a variação deste ativo financeiro (dólar comercial) por uma determinada taxa pré-fixada sem se desfazer do ativo financeiro, neste caso ela poderá através de um swap de taxas realizar tal operação: taxa anual, periodicidade mensal, representa a rentabilidade de títulos com maturidade de seis meses.

O período de análise está compreendido desde Janeiro de 2000 até Dezembro de 2005. Portanto foram utilizados 72 dados mensais de cada série. As taxas CDI e Swap 180 foram convertidas em dados mensais tanto para os testes como para a simulação.

Estatísticas das variáveis financeiras escolhidas. Nas tabelas 4.2, 4.3 e 4.4 são exibidas as séries das variáveis financeiras utilizadas no trabalho. E na tabela 4.1, estão apresentadas as principais estatísticas das quatro variáveis financeiras escolhidas.

Ibovespa e IGP-M registram os maiores desvios padrões. Estes dois rendimentos também possuem os maiores valores máximos e os menores

mínimos, sendo os únicos que registram valores negativos em todo período estudado. Swap e CDI se caracterizam por ter, aproximadamente, os mesmos valores das estatísticas, sendo Swap superior em todas, exceto no valor máximo, que é igual ao CDI. Em geral, a rentabilidade e o risco de títulos longos é superior a rentabilidade e o risco de títulos curtos; este fato estilizado é confirmado na tabela abaixo.

Estatística	CDI(%)	Ibovespa(%)	IGP-M(%)	Swap(%)
Média	1.435755	1.285417	0.885972	1.509065
Mediana	1.394218	1.625000	0.715000	1.467933
Valor máximo	2.075377	17.92000	5.190000	2.075377
Valor mínimo	1.009570	-17.17000	-1.000000	1.033354
Desvio padrão	0.215506	8.387782	1.022399	0.250056

Tabela 4.1: Estatísticas das séries de dados mensais de janeiro/2000 a dezembro/2005

4.4 Apresentação do problema

Quando um indivíduo adere a um plano de previdência, a seguradora investe o que ele contribui a cada mês em renda variável ou em renda fixa. O objetivo agora é gerar cenários para as trajetórias futuras dessas variáveis financeiras citadas utilizando um modelo VAR.

4.5 Modelagem

O primeiro passo para se modelar essas trajetórias futuras é o de estimar o modelo VAR. Depois, na próxima seção, será discutida a forma de como gerar os cenários.

4.5.1 Estimativa do modelo VAR das variáveis financeiras

Para se estimar o modelo, o primeiro passo será definir quantas defasagens para trás serão consideradas, ou seja, escolher o valor de p para o modelo VAR(p). Os critérios de informação de Akaike e o de Schwarz [7] têm como objetivo escolher, segundo uma estatística, qual é a defasagem p mais adequada. No caso em estudo, para os dados durante esse período de observação, ambos indicaram que o modelo VAR(1) é o melhor.

	CDI(%)	Ibovespa(%)	IGP-M(%)	Swap(%)
Jan/2000	1.44	-4.11	1.24	1.54
Fev/2000	1.44	7.81	0.35	1.49
Mar/2000	1.44	0.91	0.15	1.44
Abr/2000	1.28	-12.81	0.23	1.33
Mai/2000	1.42	-3.74	0.31	1.56
Jun/2000	1.39	11.84	0.85	1.44
Jul/2000	1.30	-1.63	1.57	1.36
Ago/2000	1.39	5.42	2.39	1.42
Set/2000	1.22	-8.17	1.16	1.24
Out/2000	1.28	-6.66	0.38	1.32
Nov/2000	1.22	-10.63	0.29	1.30
Dez/2000	1.19	14.84	0.63	1.24
Jan/2001	1.26	15.81	0.62	1.27
Fev/2001	1.01	-10.08	0.23	1.03
Mar/2001	1.25	-9.14	0.56	1.41
Abr/2001	1.18	3.32	1	1.48
Mai/2001	1.33	-1.8	0.86	1.69
Jun/2001	1.27	-0.61	0.98	1.52
Jul/2001	1.50	-5.53	1.48	1.96
Ago/2001	1.60	-6.65	1.38	1.96
Set/2001	1.32	-17.17	0.31	1.62
Out/2001	1.53	6.85	1.18	1.86
Nov/2001	1.39	13.79	1.1	1.50
Dez/2001	1.39	4.47	0.22	1.47
Fonte	BCB-DEPEC	Bovespa	FGV	BM&F

Tabela 4.2: Séries financeiras nos anos de 2000 e 2001

Considere o modelo VAR(1) abaixo indexado pelo mês t :

$$\begin{bmatrix} \text{CDI} \\ \text{IBO} \\ \text{IGPM} \\ \text{SWAP} \end{bmatrix}_t = \mathbf{c} + \mathbf{A}_1 \begin{bmatrix} \text{CDI} \\ \text{IBO} \\ \text{IGPM} \\ \text{SWAP} \end{bmatrix}_{t-1} + \begin{bmatrix} \epsilon_{\text{CDI}} \\ \epsilon_{\text{IBO}} \\ \epsilon_{\text{IGPM}} \\ \epsilon_{\text{SWAP}} \end{bmatrix}_t$$

Foi utilizado o software Eviews[®] para o cálculo das estimativas, e o modelo final é:

$$\begin{bmatrix} \text{CDI} \\ \text{IBO} \\ \text{IGPM} \\ \text{SWAP} \end{bmatrix}_t = \begin{bmatrix} 0.003067 \\ -0.070048 \\ 0.005284 \\ 0.004704 \end{bmatrix} +$$

	CDI(%)	Ibovespa(%)	IGP-M(%)	Swap(%)
Jan/2002	1.53	-5.83	0.36	1.56
Fev/2002	1.25	10.31	0.06	1.25
Mar/2002	1.37	-5.55	0.09	1.34
Abr/2002	1.48	-1.28	0.56	1.50
Mai/2002	1.40	-1.71	0.83	1.45
Jun/2002	1.31	-13.39	1.54	1.59
Jul/2002	1.53	-12.36	1.95	1.96
Ago/2002	1.44	6.35	2.32	1.87
Set/2002	1.38	-16.95	2.4	1.74
Out/2002	1.64	17.92	3.87	2.08
Nov/2002	1.53	3.35	5.19	1.90
Dez/2002	1.73	7.23	3.75	2.07
Jan/2003	1.97	-2.9	2.33	2.05
Fev/2003	1.83	-6.04	2.28	1.99
Mar/2003	1.77	9.66	1.53	1.83
Abr/2003	1.87	11.38	0.92	1.80
Mai/2003	1.96	6.89	-0.26	1.83
Jun/2003	1.85	-3.35	-1	1.69
Jul/2003	2.08	4.62	-0.42	1.83
Ago/2003	1.76	11.81	0.38	1.60
Set/2003	1.67	5.51	1.18	1.50
Out/2003	1.63	12.32	0.38	1.50
Nov/2003	1.34	12.24	0.49	1.25
Dez/2003	1.37	10.17	0.61	1.28
Fonte	BCB-DEPEC	Bovespa	FGV	BM&F

Tabela 4.3: Séries financeiras nos anos de 2002 e 2003

$$\begin{bmatrix} 0.681981 & -0.001377 & 0.009502 & 0.093059 \\ 14.35822 & -0.033621 & 0.342870 & -8.486075 \\ -1.340442 & -0.002985 & 0.718172 & 1.078787 \\ -0.216747 & -0.004738 & 0.016571 & 0.883706 \end{bmatrix} \begin{bmatrix} \text{CDI} \\ \text{IBO} \\ \text{IGPM} \\ \text{SWAP} \end{bmatrix}_{t-1} + \begin{bmatrix} \epsilon_{\text{CDI}} \\ \epsilon_{\text{IBO}} \\ \epsilon_{\text{IGPM}} \\ \epsilon_{\text{SWAP}} \end{bmatrix}_t \quad (4.5.1)$$

A matriz de variância e covariância dos resíduos do modelo estimado é:

$$\Sigma = \begin{bmatrix} 1.99 \times 10^{-06} & 1.61 \times 10^{-05} & 2.01 \times 10^{-07} & 1.89 \times 10^{-06} \\ 1.61 \times 10^{-05} & 0.007178 & 6.07 \times 10^{-05} & -1.55 \times 10^{-06} \\ 2.01 \times 10^{-07} & 6.07 \times 10^{-05} & 3.27 \times 10^{-05} & 1.69 \times 10^{-06} \\ 1.89 \times 10^{-06} & -1.55 \times 10^{-06} & 1.69 \times 10^{-06} & 2.25 \times 10^{-06} \end{bmatrix}$$

Na tabela 4.5.1 são apresentados os resultados da estimação do VAR(1).

	CDI(%)	Ibovespa(%)	IGP-M(%)	Swap(%)
Jan/2004	1.26	-1.73	0.88	1.21
Fev/2004	1.08	-0.44	0.69	1.05
Mar/2004	1.37	1.78	1.13	1.33
Abr/2004	1.17	-11.45	1.21	1.14
Mai/2004	1.22	-0.32	1.31	1.30
Jun/2004	1.22	8.21	1.38	1.31
Jul/2004	1.28	5.62	1.31	1.34
Ago/2004	1.29	2.09	1.22	1.38
Set/2004	1.24	1.94	0.69	1.33
Out/2004	1.21	-0.83	0.39	1.28
Nov/2004	1.25	9.01	0.82	1.32
Dez/2004	1.48	4.25	0.74	1.52
Jan/2005	1.38	-7.05	0.39	1.44
Fev/2005	1.22	15.56	0.3	1.26
Mar/2005	1.52	-5.43	0.85	1.55
Abr/2005	1.41	-6.64	0.86	1.43
Mai/2005	1.50	1.47	-0.22	1.50
Jun/2005	1.58	-0.62	-0.44	1.56
Jul/2005	1.51	3.96	-0.34	1.46
Ago/2005	1.65	7.69	-0.65	1.58
Set/2005	1.50	12.62	-0.53	1.43
Out/2005	1.40	-4.4	0.6	1.34
Nov/2005	1.38	5.71	0.4	1.29
Dez/2005	1.47	4.82	-0.01	1.38
Fonte	BCB-DEPEC	Bovespa	FGV	BM&F

Tabela 4.4: Séries financeiras nos anos de 2004 e 2005

Os valores entre parênteses são os erros padrões das estimações dos coeficientes e os valores entre colchetes são as estatísticas “t”.

A influência na primeira defasagem do CDI no valor corrente dessa série já era esperado, pois é conhecida a inércia de seqüências envolvendo taxa de juros. A mesma propriedade também pode ser observada na série de SWAP, que depende negativamente do IBOVESPA. Isso pode ser explicado observando que em períodos de mercado calmo, o IBOVESPA tende a subir e as taxas longas tendem a cair ou no máximo permanecem constantes (ou seja, há uma suave relação negativa entre essas duas séries). Como também há inércia na inflação, era de se esperar a influência da primeira defasagem do IGPM na sua observação corrente; adicionalmente, a relação negativa entre a primeira defasagem do CDI e o IGPM reflete a influência do juro

sobre a atividade econômica, e desta sobre a inflação (uma alta no juro gera uma queda na atividade econômica, que tende a inibir a alta nos preços).

Variáveis Independentes	Variáveis Dependentes			
	CDI	Ibovespa	IGP-M	Swap
CDI(-1)	0.681981 (0.15762) [4.32666]	14.35822 (9.46424) [1.51710]	-1.340442 (0.63842) [-2.09963]	-0.216747 (0.16741) [-1.29472]
Ibovespa(-1)	-0.001377 (0.00215) [-0.64008]	-0.033621 (0.12920) [-0.26022]	-0.002985 (0.00872) [-0.34253]	-0.004738 (0.00229) [-2.07317]
IGP-M(-1)	0.009502 (0.02291) [0.41468]	0.342870 (1.37587) [0.24920]	0.718172 (0.09281) [7.73805]	0.016571 (0.02434) [0.68091]
Swap(-1)	0.093059 (0.15123) [0.61536]	-8.486075 (9.08025) [-0.93456]	1.078787 (0.61251) [1.76124]	0.883706 (0.16062) [5.50198]

Tabela 4.5: Estimativas do modelo VAR.

4.6 Simulação usando o R

Para implementar o gerador de cenários, primeiramente, é necessário iniciar as variáveis que definem o modelo VAR estabelecido na última seção, que são: o vetor de intercepto C a matriz A , a matriz de variância e covariância dos resíduos Σ e o vetor Y_0 que corresponde ao vetor da condição inicial. Essa iniciação das variáveis se encontram no código 4.1.

Na listagem 4.2 encontra-se a função `nextVAR1` que gera um cenário para o próximo vetor das variáveis financeiras cujos parâmetros são a matriz A , o vetor de intercepto C , a matriz Σ e a condição inicial Y .

Observe que para multiplicar a matriz A pelo vetor Y utiliza-se o comando `%*%`. Ela usa a função `mvnorm` para gerar o vetor de resíduo que possui distribuição normal multivariada com média zero e matriz de variância e covariância Σ .

A função `rVAR` gera cenários para N meses do modelo VAR(1) com vetor

Código 4.1: Iniciação das variáveis do modelo

```

1 > C ← c(0.003067, -0.070048, 0.005284, 0.004704)
2 > dim(C) ← c(4,1)
3 >
4 > A ← c(0.681981, 14.358220, -1.340442, -0.216747,
5         -0.001377, -0.033621, -0.002985, -0.004738,
6         0.009502, 0.342870, 0.718172, 0.016571,
7         0.093059, -8.486075, 1.078787, 0.883706)
8 > dim(A) ← c(4,4)
9 >
10 > Sigma ← c(1.99*10(-6), 1.61*10(-5), 2.01*10(-7), 1.89*
11            10(-6),
12            1.61*10(-5), 0.0071780000, 6.07*10(-5), -1.55*
13            10(-6),
14            2.01*10(-7), 6.07*10(-5), 3.27*10(-5), 1.69*
15            10(-6),
16            1.89*10(-6), -1.55*10(-6), 1.69*10(-6), 2.25*
17            10(-6))
> dim(Sigma) ← c(4,4)
>
> Y ← c(1.47, 4.82, -0.01, 1.38)
> dim(Y) ← c(4,1)

```

de intercepto C , com a matriz A , matriz de variância e covariância Sigma , e condição inicial Y . Essa função encontra-se na listagem 4.3. Ela utiliza a função `nextVAR1` para gerar cada passo a partir do vetor do passo anterior e armazena todos os cenários numa matriz Z de dimensão $N \times 4$.

Para executar a função `rVAR` a fim de gerar cenários para 12 meses seguidos a partir da condição inicial Y_0 use o comando: `rVAR(12,A,C,Sigma,Y_0)`.

4.7 Exercício

(1) Gere 1000 cenários para as variáveis financeiras para 30 anos, para cada uma delas: tire a média desses cenários e visualize num histograma o resultado dessa simulação.

Código 4.2: Iniciação das variáveis do modelo

```
1 >library(MASS)
2 >
3 >nextVAR1 ← function (A,C,Sigma,Y) {
4 +   epsilon ← mvrnorm(1,c(0,0,0,0),Sigma)
5 +   Z ← A%%Y + C + epsilon
6 +   Z
7 +}
```

Código 4.3: Iniciação das variáveis do modelo

```
1 >rVAR ← function(N,A,C,Sigma,Y) {
2 +   Z ← matrix(nrow = 4,ncol = N,byrow=FALSE)
3 +   Y0 ← Y
4 +   for (j in 1:N) {
5 +     Y0 ← nextVAR1(A,C,Sigma,Y0)
6 +     for (i in 1:4){
7 +       Z[i,j] ← Y0[i]
8 +     }
9 +   }
10 +   Z
11 +}
```

Capítulo 5

Benefícios de um plano de aposentadoria

As principais incertezas na modelagem de um plano de previdência são a idade em que o segurado vai morrer e as incertezas referentes a economia. Nesse capítulo será feito o uso das funções desenvolvidas nos dois últimos capítulos para gerar cenários para o valor do benefício que o indivíduo irá receber ao se aposentar. Inicialmente, será feita uma introdução aos conceitos básicos de matemática financeira. Depois será apresentado o problema, sua modelagem e a implementação de sua solução usando o R.

5.1 Conceitos básicos da matemática financeira

Juro é a remuneração do capital empregado. Quando aplicamos um capital durante um período de tempo determinado, esperamos obter um certo *rendimento*. Após esse período, o *capital principal* investido se transformará em um valor capitalizado, chamado de *montante*, que será o capital principal acrescido do rendimento obtido durante o período de aplicação. A *taxa de juros* é a razão entre o rendimento e o capital principal.

Uma taxa de juros é sempre associada a uma *unidade básica de tempo*, por exemplo, 6% ao ano. Deve também ser conhecido o *período de aplicação*, que pode ser por exemplo diário, mensal, anual, bi-anual etc. Uma taxa de juros é chamada *Taxa efetiva efetiva* se o período de aplicação é idêntico à unidade básica de tempo, nesse caso o rendimento é creditado no fim da unidade básica de tempo.

No *regime de capitalização simples*, os rendimentos de cada período

são calculados sobre o capital principal. Portanto eles não incidem sobre os rendimentos acumulados. Esse regime só faz algum sentido em um contexto não-inflacionário, que não é o caso de interesse desse trabalho.

No *regime de capitalização composto*, o rendimento gerado pela aplicação será incorporado, passando a participar do cálculo de rendimento dos períodos de aplicação que vêm a seguir.

5.1.1 Taxa de juros efetiva

Seja j uma taxa de juros efetiva anual. Para simplificar assumamos que essa taxa é constante para todos os anos. Considere um fundo onde o capital principal C_0 é investido e que um capital adicional A_k é aplicado no mesmo fundo no fim de cada ano k , para $k = 1, \dots, n$. O balanço do fundo no fim do ano k , denotado por C_k , incluindo o pagamento adicional A_k é dado por:

$$C_k = C_{k-1} + jC_{k-1} + A_k, \quad k = 1, \dots, n, \quad (5.1.1)$$

que pode ser reescrito na forma:

$$C_k - (1+j)C_{k-1} = A_k. \quad (5.1.2)$$

Ao multiplicar essa equação por $(1+j)^{n-k}$ em ambos os lados e somar sobre todos os valores de k , obtêm-se que:

$$C_n = (1+j)^n C_0 + \sum_{k=1}^n (1+j)^{n-k} A_k. \quad (5.1.3)$$

As potências de $(1+j)$ são chamadas de *fatores de acumulação*. Portanto, o montante acumulado ou o *valor futuro* de um investimento com capital principal C após n anos é $(1+j)^n C$.

Desconto pode ser entendido como a diferença entre o valor do resgate do investimento e seu valor inicial. O *fator de desconto* é definido como

$$v = \frac{1}{1+j}.$$

A equação 5.1.3 pode ser reescrita na forma:

$$v^n C_n = C_0 + \sum_{k=1}^n v^k A_k. \quad (5.1.4)$$

Assim, o *valor presente* de um capital C devido a um investimento de n anos é $v^n C$.

Segue uma outra forma de escrever a equação (5.1.1):

$$C_k - C_{k-1} = jC_{k-1} + A_k.$$

Se somarmos a equação acima sobre todos os k 's, tem-se que:

$$C_n - C_0 = \sum_{k=1}^n jC_{k-1} + \sum_{k=1}^n A_k$$

Em conclusão, o rendimento total do fundo é a soma total dos juros creditados com o total dos depósitos feitos durante o período.

5.1.2 Taxa de juros nominal

Quando o período de aplicação não é idêntico à unidade básica de tempo, a taxa de juros é chamada de *nominal*. Por exemplo, considere uma taxa anual de 6% e um período de aplicação de 3 meses. Nesse caso, o rendimento creditado no fim de cada trimestre correspondente a uma taxa de juros de $6\%/4 = 1,5\%$. Dessa forma, um capital inicial de 1 unidade monetária renderia no fim de 4 trimestres o correspondente a $(1.015)^4 = 1.06136$. Observe que a taxa de juros anual nominal de 6% quando convertida trimestralmente é equivalente a uma taxa anual efetiva de 6.136%.

Considere uma taxa de juros anual j , define-se $j^{(m)}$ como a taxa de juros nominal convertida para um período de m vezes ao ano equivalente a j . Para determinar $j^{(m)}$ utiliza-se a seguinte equação:

$$\left(1 + \frac{j^{(m)}}{m}\right)^m = 1 + j.$$

Explicitando $j^{(m)}$, tem-se que:

$$j^{(m)} = m[(1 + j)^{1/m} - 1]. \quad (5.1.5)$$

Pode-se interpretar $j^{(m)}$ quando $m \rightarrow \infty$ como uma taxa de juros continuamente composta. Seja

$$\delta = \lim_{m \rightarrow \infty} j^{(m)},$$

que é denominada força dos juros equivalente a j . A equação (5.1.5) pode ser reescrita na forma:

$$j^{(m)} = \frac{(1 + j)^{1/m} - (1 + j)^0}{1/m}.$$

Conseqüentemente, pode-se dizer que δ é a derivada de $(1 + j)^x$ no ponto $x = 0$. Assim,

$$\delta = \ln(1 + j)$$

ou

$$e^\delta = 1 + j.$$

Portanto, o fator de acumulação para um período de h anos, $h \in \mathbb{R}$, é $(1 + j)^h = e^{\delta h}$ e o fator de desconto para o mesmo período é $v^h = e^{-\delta h}$.

5.1.3 Pagamentos antecipados

Até agora foi assumido que os rendimentos são creditados no fim do período de aplicação. Mas algumas vezes é útil assumir que os rendimentos são creditados no início de cada período de aplicação. Rendimentos creditados dessa forma são também conhecidos como *descontos*, e a taxa de juros correspondente é chamada de *taxa de desconto* ou *taxa de juros antecipada*.

Seja d uma taxa anual de desconto efetiva. Uma pessoa que investir uma quantidade C de capital principal vai ter creditado imediatamente a quantia dC , e o capital investido C vai ser retornado no fim do período. Investindo o rendimento dC nas mesmas condições, o investidor vai receber um rendimento adicional de $d(dC) = d^2C$. Novamente, ao aplicar a quantia d^2C , o investidor receberá d^3C como retorno de seu investimento. Repetindo esse processo infinitamente, o investidor vai receber no fim de um ano a seguinte quantia

$$C + dC + d^2C + d^3C + \dots = \frac{1}{1 - d}C$$

como retorno de seu investimento inicial C .

A taxa de juros equivalente j para esse tipo de investimento é obtida pela equação

$$\frac{1}{1 - d} = 1 + j,$$

que é equivalente dizer que:

$$d = \frac{j}{1 + j}$$

ou

$$j = \frac{d}{1 - d}.$$

Seja $d^{(m)}$ a taxa de desconto nominal equivalente a d só que creditada m vezes ao ano. O investidor, então, recebe no início de cada período de conversão $\frac{d^{(m)}}{m}C$ e o seu capital principal C é retornado no fim do período.

Usando a igualdade

$$\frac{1}{1 - \frac{d^{(m)}}{m}} = 1 + \frac{j^{(m)}}{m} = (1 + j)^{1/m},$$

pode-se dizer que:

$$d^{(m)} = m[1 - (1 + j)^{-1/m}].$$

Por um pequeno algebrismo, pode-se dizer também que:

$$d^{(m)} = \frac{j^{(m)}}{1 + \frac{j^{(m)}}{m}}.$$

Daí, pode-se tirar uma relação muito simples entre $j^{(m)}$ e $d^{(m)}$:

$$\frac{1}{d^{(m)}} = \frac{1}{m} + \frac{1}{j^{(m)}}.$$

Com essa relação, conclui-se que a diferença entre a taxa de desconto e a taxa de juros some quando elas são consideradas contínuas, pois:

$$\lim_{m \rightarrow \infty} d^{(m)} = \lim_{m \rightarrow \infty} j^{(m)} = \delta.$$

5.1.4 Perpetuidades

Essa seção introduz certos tipos de seqüências de pagamentos perpétuos (*perpetuidades*) e calcula seus valores presentes.

Inicialmente, considere as perpetuidades de pagamentos anuais de uma unidade monetária. Se o primeiro pagamento ocorre no tempo 0, a perpetuidade é chamada de *perpetuidade antecipada* e seu valor presente é dado por

$$a_{\infty|} = 1 + v + v^2 + \dots = \frac{1}{1 - v} = \frac{1}{d}.$$

Se o primeiro pagamento ocorre no fim do primeiro ano, a perpetuidade é chamada de *perpetuidade postecipada* e seu valor presente é dado por

$$a_{\infty|} = v + v^2 + v^3 + \dots = \frac{v}{1 - v} = \frac{1}{j}.$$

Considere agora que pagamentos de $\frac{1}{m}$ unidades monetárias são feitas m vezes ao ano. Se os pagamentos são feitos antecipadamente (primeiro pagamento no tempo 0), então o valor presente desse tipo de perpetuidades é:

$$a_{\infty|}^{(m)} = \frac{1}{m} + \frac{1}{m}v^{1/m} + \frac{1}{m}v^{2/m} + \dots = \frac{1}{m} \frac{1}{1 - v^{1/m}} = \frac{1}{d^{(m)}}.$$

Por sua vez, se o primeiro pagamento é feito no fim do período $1/m$, então o valor presente da perpetuidade é:

$$a_{\infty|}^{(m)} = \frac{1}{m}v^{1/m} + \frac{1}{m}v^{2/m} + \frac{1}{m}v^{3/m} + \dots = \frac{1}{m} \frac{v^{1/m}}{1 - v^{1/m}} = \frac{1}{j^{(m)}}.$$

5.1.5 Anuidades

Uma *anuidade* é definida como uma seqüência de pagamentos durante um tempo limitado, que é denotado por n . Na prática, as anuidades são mais usadas que as perpetuidades.

O valor presente de uma anuidade com n pagamentos anuais de uma unidade monetária começando no instante 0 é:

$$a_{n|} = 1 + v + v^2 + \dots + v^{n-1}.$$

Pode-se representar a anuidade como a diferença entre duas perpetuidades, uma iniciada no instante 0 e outra no instante n . Assim, tem-se que:

$$a_{n|} = a_{\infty|} - v^n a_{\infty|} = \frac{1}{d} - v^n \frac{1}{d} = \frac{1 - v^n}{d}.$$

De forma similar, pode-se obter a fórmula para o valor presente da anuidade com o primeiro pagamento feito no final do primeiro ano da seguinte maneira:

$$a_{n|} = \frac{1 - v^n}{j}.$$

A fórmula para o valor presente de uma anuidade com m pagamentos parciais de $1/m$ unidade monetária iniciada no tempo 0 é:

$$a_{n|}^{(m)} = \frac{1 - v^n}{d^{(m)}}.$$

E, por fim, a fórmula para o valor presente de uma anuidade com m pagamentos parciais de $1/m$ unidade monetária iniciada no tempo 1 é:

$$a_{n|}^{(m)} = \frac{1 - v^n}{j^{(m)}}.$$

Observe que o denominador varia dependendo da forma de pagamento e da freqüência que eles são feitos.

O valor acumulado das anuidades é também de interesse. As fórmulas para cada um dos tipos acima são respectivamente:

$$s_{n|} = \frac{(1 + j)^n - 1}{d},$$

$$s_{n|} = \frac{(1+j)^n - 1}{j},$$

$$s_{n|}^{(m)} = \frac{(1+j)^n - 1}{d^{(m)}},$$

e

$$s_{n|}^{(m)} = \frac{(1+j)^n - 1}{j^{(m)}}.$$

No caso de previdência e de seguros do ramo vida, sob o ponto de vista atuarial, as contas das anuidades devem levar em consideração a probabilidade do indivíduo estar vivo para fazer o pagamento. A seguir serão introduzidos conceitos básicos atuariais para tal fim.

5.1.6 Anuidades de vida inteira

Uma *anuidade de vida inteira* consiste numa série de pagamentos que são feitos por um indivíduo com idade inicial x enquanto ele estiver vivo. Portanto, ela possui uma dependência do tempo de sobrevivência do indivíduo. Assim, o valor presente da anuidade deve ser considerado como uma variável aleatória. Essa variável será denotada por Y . Considere novamente que os pagamentos são de uma unidade monetária, e que são feitos nos tempos $0, 1, \dots, K$. O valor presente desses pagamentos é, portanto:

$$Y = 1 + v + v^2 + \dots + v^K = a_{k+1|}.$$

A função de probabilidade de Y é dada por:

$$\Pr(Y = a_{(k+1)|}) = \Pr(K = k) = {}_k p_x q_{x+k}.$$

O valor esperado de Y é denotado por a_x . Pode-se escrever que:

$$a_x = E[Y] = \sum_{k=0}^{\infty} a_{k+1|} {}_k p_x q_{x+k}.$$

Por outro lado, o valor Y pode também ser reescrito na forma:

$$Y = \sum_{k=0}^{\infty} v^k I_{\{K \geq k\}},$$

onde $I_{\{K \geq k\}}$ é uma variável indicadora para o evento $K \geq k$. Conseqüentemente,

$$a_x = E[Y] = \sum_{k=0}^{\infty} v^k {}_k p_x,$$

onde $v = \frac{1}{1+i}$ é o fator de desconto para uma taxa i .

O valor presente atuarial de uma anuidade de vida de uma unidade monetária por ano, pagável em prestações de $1/m$ no início de cada m -ésima parte de um ano enquanto um indivíduo com x anos sobrevive, é denotado por $a_x^{(m)}$:

$$a_x^{(m)} = \frac{1}{m} \sum_{h=0}^{\infty} v^{h/m} {}_{h/m}p_x.$$

A tradicional aproximação de $a_x^{(m)}$ pode ser escrita na forma:

$$a_x^{(m)} = a_x - \frac{m-1}{2m}.$$

Para $m = 12$, a anuidade de vida mensal é dada por :

$$a_x^{(12)} = a_x - \frac{11}{24}. \quad (5.1.6)$$

Todos os demais tipos de anuidades de vida inteira são deduzidos dessa mesma forma.

Para maiores detalhes sobre a teoria atuarial do ramo vida veja o livro do Gerber [5].

5.2 Apresentação do problema

O objetivo desse capítulo é o de gerar cenários para a variável aleatória S que corresponde ao valor do saldo final da conta de um plano PGBL de um indivíduo, no mês de sua aposentadoria para poder calcular o valor de seu benefício. O modelo a ser trabalhado deve considerar como hipóteses os seguintes fatos: O indivíduo possui um plano de contribuição definida, que não estabelece o benefício no ato da contratação, mas sim a contribuição que é constante e igual a $R\$C$, ou seja não é alterada durante o período de diferimento (que corresponde ao intervalo de meses do início do plano até o momento de sua aposentadoria). O benefício só será conhecido ao final da vida laborativa do participante e é vitalício. Os resgates e as portabilidades ficam bloqueadas e os encargos não são atrelados ao plano. Durante o período de diferimento, os pagamentos são mensais e são realizados no fim de cada mês. É assumido que o participante inicia o plano no dia do seu aniversário, e que ele não morre no período de diferimento. A tábua de mortalidade utilizada é a americana masculina de 2004, apresentada no capítulo 3 e o plano adota uma taxa de 5% ao ano para o desconto dado ao benefício no momento da aposentadoria.

5.3 Modelagem

Modelos para o plano de previdência de um indivíduo com as hipóteses feitas acima possuem dois períodos bem definidos, o de diferimento e o de benefícios, que correspondem ao período do mês seguinte a sua aposentadoria até a sua morte.

Primeiramente será apresentado o simulador para o período de diferimento, com ele será possível gerar cenários de quanto o segurado terá de saldo em sua conta. Tendo esse valor na mão pode-se aplicar a fórmula que as seguradoras usam para calcular o valor de benefício.

Suponha que o participante deposite na conta do seu plano uma quantia C no fim de cada mês e que o montante existente na conta no início de cada mês é investido de acordo com a seguinte carteira: 40% em renda fixa (RF), 30% em renda variável (RV) e 30% em títulos (TIT). Isso representa uma restrição muito forte no modelo, mas que em trabalhos futuros será substituída por um otimizador estocástico com múltiplos estágios para definir a melhor carteira a longo prazo. Para maiores detalhes desse problema de otimização veja [2, 9, 14, 19].

Portanto se existiam M_t unidades monetárias no início do mês t na conta do participante, no início do mês $t + 1$ a conta terá o saldo de:

$$M_{t+1} = 0.4 \times M_t \times (1 + RF_t) + 0.3 \times M_t \times (1 + RV_t) + 0.3 \times M_t \times (1 + TIT_t) + C, \quad (5.3.7)$$

onde RF_t , RV_t e TIT_t correspondem, respectivamente, aos juros do mês t dos ativos de renda fixa, renda variável e dos títulos. Essa fórmula recursiva se mostra bem adequada porque esses juros variam com o tempo. Eles são estimados pelo modelo VAR das variáveis financeiras apresentadas no capítulo 4.

A taxa de juros RF_t referente a renda fixa é representada pela variável de certificados de depósitos interfinanceiro CDI_t . Para a taxa da renda variável RV_t foi utilizada a variável índice do IBOVESPA, IBO_t . Finalmente, para a taxa dos títulos TIT_t foi utilizado a taxa de SWAP de 180 dias, $SWAP_t$. Como todas essas taxas são nominais e o que importa para o segurado é o poder aquisitivo de seus benefícios em termos reais, todas as rentabilidades devem ser corrigidas pela inflação, ou seja descontadas pelo IGPM. A conversão da taxa nominal em taxa efetiva de cada uma delas é feita através da seguinte fórmula:

$$\text{Taxa_Efetiva}_t = \frac{1 + \text{Taxa_Nominal}_t}{1 + \text{IGPM}_t} - 1.$$

Os dados de entrada da primeira parte do simulador são:

- A idade inicial, denotada por $Idade_Inicial$, do participante.
- A idade em que o participante irá se aposentar, $Idade_Aposentadoria$.
- O valor da contribuição mensal C .

O simulador do período de diferimento seguirá o seguinte algoritmo:

Passo 1: $M_0 = 0$; $t = 0$; $Idade = Idade_Inicial$.

Passo 2: Estime o vetor $[\text{CDI} \quad \text{IBO} \quad \text{IGPM} \quad \text{SWAP}]_t^T$ para o mês t usando o modelo VAR descrito na equação (4.5.1).

Passo 3: Calcule M_{t+1} usando a equação 5.3.7.

Passo 4: $t = t + 1$.

Passo 5: Se $t \bmod 12 = 0$, então $Idade = Idade + 1$.

Passo 6: Se $Idade < Idade_Aposentadoria$ então volte ao Passo 2, senão $M_{final} = M_t$; e termine.

Assim, essa primeira parte do simulador calcula um cenário para o montante final de um indivíduo, cujo valor é armazenado na variável M_{final} .

Os *benefícios* são pagamentos constantes que os participantes recebem em função da sua aposentadoria e de sua expectativa de vida futura. O cálculo do benefício do primeiro ano, denotado por B_1 , é dado por

$$B_1 = \frac{M_{final}}{12a_x^{(12)}}, \quad (5.3.8)$$

onde $a_x^{(12)}$ é a anuidade de vida mensal de um indivíduo com uma idade x considerando a taxa de desconto de 5%, e sua fórmula é dada pela equação (5.1.6). Para questão de esclarecimentos, os benefícios, no tempo futuro depois da aposentadoria, são ajustados ao final de cada ano pela inflação acumulada no ano, sendo essa representada pela série $IGPM_t$. Assim, o benefício no ano a é dado por:

$$B_a = B_{a-1} \prod_{s=1}^{12} (1 + IGPM_{12(a-1)+s}) \quad (5.3.9)$$

5.4 Exercício

Como esse problema é uma aplicação dos capítulos anteriores, propomos o seguinte exercício:

(1) Faça um programa em R que gere um cenário para o valor do benefício de um indivíduo que entre no plano com 30 anos e contribui mensalmente com o valor de R\$ 400,00 por mês seguindo as hipóteses citadas na descrição do problema.

Capítulo 6

Soma de Sinistros

O objetivo desse capítulo é o de modelar a soma total dos sinistros, a serem pagos por uma seguradora, que ocorreram num período fixo de tempo para um determinado conjunto de contratos de seguro. Para esse problema devem ser modeladas probabilisticamente dois tipos de variáveis aleatórias: uma que define a quantidade de sinistro que ocorreram num determinado período de tempo e outra que define o valor de cada um deles. Sendo assim, iniciamos o capítulo com uma introdução a processos de Poisson.

6.1 Processos de Poisson

Tal como vimos anteriormente, a equação (2.1.5) modela o número de chegadas em um intervalo de tempo unitário quando o fenômeno de interesse pode ser descrito por um processo de Poisson.

Em algumas situações é de interesse generalizar esse processo, fazendo com que a taxa de chegadas ou de ocorrências varie ao longo do tempo. Um caso evidente é o do número de chegadas de fregueses em um restaurante: será maior, tipicamente, nos horários próximos ao almoço e ao jantar. No que segue definiremos o processo de Poisson não homogêneo com taxa de ocorrência $\lambda: \mathbb{R}_+ \rightarrow \mathbb{R}_+$ determinística.

Um processo de Poisson não homogêneo com taxa de ocorrência $\lambda(t)$ é caracterizado pelas seguintes propriedades:

1. A probabilidade de mais de uma ocorrência no intervalo de tempo $(t, t + dt)$ é infinitesimal, igual a $o(dt)$.
2. O número esperado de eventos no intervalo $(t, t + dt)$ é $E[N(t, t + dt)] = \lambda(t) dt + o(dt)$ e, com isso, o número esperado de eventos no intervalo

$(0, t)$ é dado por $E[N(0, t)] = \int_0^t \lambda(u) du$.

3. Dado que houve $n \geq 1$ ocorrências no intervalo $(0, t)$, os instantes em que aconteceram estão distribuídos de forma independente e idêntica segundo a lei caracterizada pela densidade $\lambda(t) / \int_0^t \lambda(u) du$.

6.2 Apresentação do problema

Um dos grande problemas das seguradoras é estimar o quanto ela vai gastar no futuro para pagar os sinistros que ocorreram durante a vigência dos contratos que ela possui em sua carteira. Esse valor pode ser representado pelo o que é chamada de soma total de sinistros, denotada por S , que possui N parcelas de valores individuais X_1, X_2, \dots, X_N . Assim

$$S = X_1 + X_2 + \dots + X_N, \quad N = 0, 1, 2, \dots,$$

onde $S = 0$ quando $N = 0$.

Em atuária, o *modelo de risco coletivo* possui essa representação com X_i , $i \geq 1$, sendo considerados como variáveis aleatórias independentes e identicamente distriuídas, a não ser em casos especiais. É geralmente assumido também que as variáveis X_i s são independentes de N . A variável N é chamada de variável aleatória de número de sinistros, e sua distribuição é geralmente de distribuição de frequência de sinistros. Os valores de X_i s são chamados de sinistros individuais.

O problema desse capítulo é gerar um cenário para a variável $S(T)$ que representa o valor da soma total de sinistros num período de $T = 24$ horas a serem pagos por uma seguradora de automóveis.

6.3 Modelagem

Suponha que essa seguradora descobriu que a taxa de acidentes de um certo grupo de motoristas assegurados num período de 24 horas aumenta de meia noite ao meio dia e que então diminui até a próxima meia noite. Ela conseguiu modelar que os eventos dos sinistros obedecem a um processo de Poisson não-homogêneo onde a intensidade do processo em função de t é dado por

$$\lambda(t) = 1/6 - (12 - t)^2/1152, \quad (6.3.1)$$

onde t é o número de horas depois de meia noite. A figura 6.1 ilustra esta taxa em função do tempo.

Considere que os valores de cada sinistro são independentes e identicamente distribuídos obedecendo uma distribuição exponencial com taxa 500.

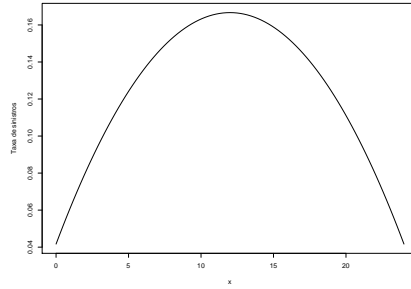


Figura 6.1: Taxa de sinistros em função do tempo

E além disso, suponha que os valores desses sinistros são independentes do horário que eles ocorrem.

6.4 Simulação usando o R

Veremos a seguir como simular eventos do processo de Poisson não homogêneo com a taxa caracterizada pela equação (6.3.1). Para tanto, será útil conferir que o número esperado de sinistros no período de um dia é dado por

$$\int_0^{24} \lambda(t) dt = 3. \quad (6.4.2)$$

De posse dessa informação, obtenhamos a ocorrência de uma variável aleatória com distribuição Poisson e taxa 3.

Código 6.1: Geração de um processo de Poisson não homogêneo

```
1 > rpois(1, 3)
2 [1] 5
```

Usando a equação 6.4.2, sabemos que

$$f_{\Lambda}(t) = 1/18 - (12 - t)^2/3456, \quad (6.4.3)$$

é uma densidade de probabilidade e, portanto

$$F_{\Lambda}(t) = \begin{cases} 0 & \text{se } t < 0, \\ \int_0^t f_{\Lambda}(x) dx = \frac{t}{72} + \frac{t^2}{288} - \frac{t^3}{10368} & \text{se } 0 \leq t \leq 24, \\ 1 & \text{se } t > 24. \end{cases} \quad (6.4.4)$$

A figura 6.2 mostra esta função no intervalo $(0, 24)$. Como ela é suave, a sua inversão (numérica ou analítica) é imediata e, com isso, a simulação das ocorrências dos instantes do processo de Poisson não homogêneo com taxa de ocorrência $\lambda(t)$.

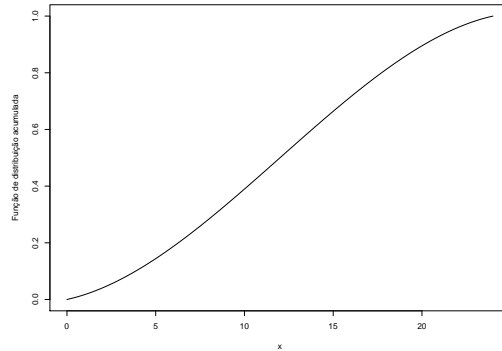


Figura 6.2: Função de distribuição acumulada dos tempos de ocorrência do processo de Poisson não homogêneo

No entanto, e voltando a código 6.1, o nosso interesse reside em simular o total de sinistros ocorridos no dia. Já sabemos que houve cinco ocorrências (linhas 1 e 2 do código acima). Simulemos agora outras tantas ocorrências da distribuição exponencial com média 500 e calculemos a soma.

Código 6.2: Geração da soma de sinistros

```

1 > sinistros ← rexp(n=5, rate=1/500)
2 > sinistros
3 [1] 1017.46970 131.37854 349.67077 1113.96588
   76.54605
4 > sum(sinistros)
5 [1] 2689.031

```

A linha 1 gera o vetor `sinistros` de tamanho cinco, sendo cada entrada a ocorrência de uma variável aleatória com distribuição exponencial e média 500. Esses valores são exibidos com o comando da linha 2. O total é calculado e exibido pelo comando da linha 4.

6.5 Exercício

- (1) Modifique o código acima utilizando ao invés da exponencial com média 500, uma log-normal com os parâmetros $\mu = \ln 500$ e $\sigma = \ln 100$.

Capítulo 7

Conclusão

Nesse texto foram estudados modelos probabilísticos simples com aplicações em atuária e finanças. Eles foram apresentados na forma de problemas e suas implementações foram feitas utilizando o software R.

O R é uma plataforma de desenvolvimento para aplicações e análise estatísticas que é gratuita. Ela oferece muitos recursos e por isso ela deve ser explorada. Só uma pequena amostra desse potencial foi exemplificada nesse texto.

Sugestões de estudos futuros. Uma pessoa que deseja aprofundar os seus estudos em matemática financeira e atuarial deve começar obtendo uma boa base em probabilidade e estatística, e para isso existem muitos bons livros. Após essa base, é sugerido o livro do Gerber [5] para matemática atuarial ramo vida, que é adotado em vários cursos no mundo. Para se aprofundar mais em modelagem de riscos é sugerido o livro do Klugman et al. [8]; nele são descritos vários modelos probabilísticos para a modelagem de seguros e resseguros. Por fim, recomendamos o livro do McNeil et al. [12] para modelos de análise de riscos mais sofisticados.

Agradecimentos. Os autores Hélio Côrtes Vieira e Luciano Vereda gostariam de agradecer a FAPERJ (Projeto Pensa-Rio e Prioridade-Rio) pelo apoio a realização desse trabalho. Hélio Côrtes Vieira gostaria de agradecer ao CNPq pela bolsa de produtividade de pesquisa concedida, que auxilia também a execução desse trabalho. O autor Alejandro C. Frery agradece ao CNPq pela bolsas de pós-doutorado sênior, durante cuja vigência foi gestado este projeto, e pela bolsa de produtividade em pesquisa, que também colaborou. Um agradecimento muito especial é para o Prof. Francisco

Cribari-Neto, do Departamento de Estatística da Universidade Federal de Pernambuco, co-autor do texto que serviu de ponto de partida para vários trechos deste livro e que é um dos pioneiros no uso do sistema R.

Bibliografia

- [1] G. E. Box and G. M. Jenkins. *Time series analysis: forecasting and control*. San Francisco: Holden Day, 1976.
- [2] G. Consigli and M. A. H. Dempster. Dynamic stochastic programming for asset-liability management. *Annals of Operations Research*, 81(0):131–161, 1998.
- [3] L. Devroye. *Non-Uniform Random Variate Generation*. Springer-Verlag, New York, 1986.
- [4] C. J. Dias. Um modelo estocástico para o fluxo de caixa de um plano de previdência de um indivíduo. Dissertação de Mestrado em Matemática 1, Pontifícia Universidade Católica do Rio de Janeiro, 2006.
- [5] H. U. Gerber. *Life Insurance Mathematics*. Springer, 3rd. edition, 1997.
- [6] B. James. *Probabilidade: um Curso em Nível Intermediário*. Projeto Euclides. Instituto de Matemática Pura e Aplicada, Rio de Janeiro, 1981.
- [7] G. G. Judge, R. C. Hill, W. E. Griffith, H. Lutkepohl, and T. Lee. *The Theory and Practice of Econometrics*. John Willey and Sons, 1985.
- [8] S. Klugman, H. Panjer, and G. Willmot. *Loss Models: From Data to Decisions*. Willey, New York, 2004.
- [9] R. Kouwenberg and S. A. Zenios. Stochastic programming models for asset liability managements. Working Paper 1, School of Economics and Management at the University of Cyprus, 2001.

- [10] W. J. Krzanowski. *Principles of Multivariate Analysis: a User's Perspective*. Oxford Statistical Science Series. Clarendon Press, Oxford, 1988.
- [11] M. Matsumoto and T. Nishimura. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Trans. Model. Comput. Simul.*, 8(1):3–30, 1998.
- [12] A. J. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management*. Princeton Press, New York, 2005.
- [13] R. J. Muirhead. *Aspects of Multivariate Statistical Theory*. Wiley Series in Probability and Mathematical Statistics. Wiley, New York, 1982.
- [14] J. M. Mulvey and B. Shetty. Financial planning via multi-stage stochastic optimization. *Comput. Oper. Res.*, 31(1):1–20, 2004.
- [15] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C: The art of scientific computing*. Cambridge, 2nd. edition, 1992.
- [16] C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer Texts in Statistics. Springer, New York, 2000.
- [17] Y. L. Tong. *The Multivariate Normal Distribution*. Springer Series in Statistics. Springer-Verlag, New York, 1990.
- [18] L. Velho, J. Miranda, and A. C. Frery. *Image Processing for Computer Graphics and Vision*. Springer, 2008.
- [19] W. Ziemba and J. Mulvey. *Worldwide asset and liability modeling*. Cambridge, 1998.

Índice

- Anuidades de vida inteira, 59
- Benefícios, Cálculo dos, 62
- Capital principal, 53
- Capitalização composta, 54
- Capitalização simples, 53
- Desconto, 54
- Distribuições de Probabilidade
 - Distribuição Bernoulli, 24
 - Distribuição Beta, 23
 - Distribuição Binomial, 24
 - Distribuição Exponencial, 21
 - Distribuição Log-Normal, 22
 - Distribuição Normal, 21
 - Distribuição Normal
 - Multivariada, 28
 - Distribuição Poisson, 24
 - Distribuição Uniforme, 20
- Distribuição Acumulada, 20
- Função densidade de probabilidade, 20
- Gerador LCG, 27
- Gerador Mersenne Twister, 28
- Gráfico de pares, 31, 32
- Juro, 53
- Matemática atuarial, 13, 33
- Momentos, 21
- Montante, 53
- Números pseudo-aleatórios, 27
- Pagamentos, 56
- Período de aplicação, 53
- Perpetuidades, 57
- Plano PGBL, 15
- Plano VGBL, 15
- Previdência privada, 14
- Previdência social, 14
- Processo de Poisson, 26
- Processo de Poisson não homogêneo, 65
- R, Plataforma, 19
- Rendimento, 53
- Soma de Sinistros, 66
- Tábua de Mortalidade, 35
- Taxa de juros contínua, 55
- Taxa de juros nominal, 55
- Teorema de Inversão, 25
- Unidade básica de tempo, 53