

---

Programação Linear Aplicada a  
Estatística

*Alan Henrique de Jesus*

---



UNIVERSIDADE FEDERAL DE SÃO CARLOS  
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA  
PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA UFSCar-USP

*Alan Henrique de Jesus*

## Programação Linear Aplicada a Estatística

Tese apresentada ao Departamento de Estatística – DEs-UFSCar e ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística UFSCar-USP.

Orientador: Prof. Dr. Márcio Alves Diniz

São Carlos  
Dezembro de 2017



*Alan Henrique de Jesus*

# Linear Programming Applied to Statistics

Master dissertation submitted to the Department of Statistics - DEs - UFSCar and to the Institute of Mathematics and Computer Sciences - ICMC - USP, in partial fulfillment of the requirements for the degree of the Master Interagency Program Graduate in Statistics.

Advisor: Prof. Dr. Márcio Alves Diniz

São Carlos  
December 2017





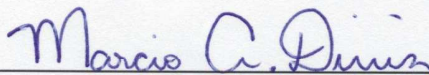
UNIVERSIDADE FEDERAL DE SÃO CARLOS  
Centro de Ciências Exatas e de Tecnologia  
Programa Interinstitucional de Pós-Graduação em Estatística

---

Folha de Aprovação

---

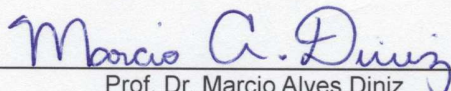
Assinaturas dos membros da comissão examinadora que avaliou e aprovou a defesa de dissertação de mestrado do candidato Alan Henrique de Jesus realizada em 27/11/2017:

  
\_\_\_\_\_  
Prof. Dr. Marcio Alves Diniz  
UFSCar

  
\_\_\_\_\_  
Prof. Dr. Eduardo Fontoura Costa  
ICMC-USP

\_\_\_\_\_  
Profa. Dra. Laura Leticia Ramos Rifo  
UNICAMP

Certifico que a sessão de defesa foi realizada com a participação à distância do membro Profa. Dra. Laura Leticia Ramos Rifo e, depois das arguições e deliberações realizadas, o participante à distância está de acordo com o conteúdo do parecer da comissão examinadora redigido no relatório de defesa do(a) aluno(a) Alan Henrique de Jesus.

  
\_\_\_\_\_  
Prof. Dr. Marcio Alves Diniz  
Presidente da Comissão Examinadora  
UFSCar





# Agradecimentos

Em primeiro lugar, dedico um agradecimento especial a Deus por sempre prover além do necessário durante toda a minha caminhada acadêmica e à minha família, em especial aos meus pais, pelo apoio incondicional, especialmente nos momentos de dificuldades, independentemente da distância.

Além disso, gostaria de agradecer a todos os amigos do Kikos Flat's pelo anos de convivência e diversão. Também a minha turma de mestrado pelos momentos vividos tanto acadêmicos, quanto sociais. Ao pessoal do peladeiros do DC, pelos anos de peladas de sexta à noite. Agradeço também as pessoas que de algum modo, me ajudaram deste o início deste projeto, em particular, ao meu orientador Professor Márcio que me ajudou em todos os momentos e soube, brihamente, me conduzir no decorrer desses trabalho.

Um agradecimento especial a minha namorada Tamyris por todo o suporte e apoio durante o mestrado e por toda a ajuda durante o desenvolvimento deste trabalho, especialmente nas revisões ortográficas.

Para finalizar, agradeço a CAPES pelo apoio dado durante a minha permanência no mestrado.



# Resumo

Determinar probabilidades para eventos os quais temos poucas informações ou intervalos para estas probabilidades, não é tão simples. Para isso desenvolveremos conceitos de programação linear, que nos permite resolver, de certo modo, o problema de determinar uma probabilidade para um evento de interesse, mesmo que nem sempre de maneira única. Apresentaremos alguns exemplos clássicos da estatística, sendo eles: o Problema de *Monty Hall* e o Problema da Probabilidade do Testemunho. Além disso, discutiremos o problema de precificação de uma opção de compra, onde utilizaremos programação linear para determinar o preço que a opção deve ter.

**Palavras chave:** Programação Linear, Estatística, Precificação de Opção.



# Abstract

Determine probabilities for events where we have few information or intervals for probabilities is not so simple. For this we will develop concepts of linear programming, which allows us to solve, in a certain way, the problem of determine a probability for an event of interest, but not always in a unique way. We will present some classic examples of statistics, such as: The Monty Hall Problem and De La Probabilité Des Témoignages. In addition, we will discuss the problem of pricing a call option, where we will use linear programming to solve them.

**Keywords:** Linear programming, Options pringing, Statistics.



# Sumário

<b>Introdução</b>	<b>xi</b>
<b>1 Maximização e Minimização de Esperança Matemática</b>	<b>1</b>
1.1 Quadro geral . . . . .	2
1.2 Limitantes sobre a Esperança Matemática sem Restrição no Conjunto de Soluções . . . . .	3
1.2.1 Teorema Fundamental da Programação Linear . . . . .	4
1.2.2 Teorema de Dualidade . . . . .	7
1.2.3 Aplicação em Problema Estatísticos . . . . .	9
1.3 Limitantes sobre a Esperança Matemática com Restrição no Conjunto de Soluções . . . . .	10
1.3.1 Resultados Gerais . . . . .	11
1.3.2 Restrições de Entropia . . . . .	13
<b>2 Exemplos Estatísticos</b>	<b>17</b>
2.1 Desigualdades . . . . .	17
2.2 Problema de <i>Monty Hall</i> . . . . .	21
2.2.1 Resposta Intuitiva . . . . .	21
2.2.2 Solução para o problema pelo Teorema de Bayes . . . . .	22
2.2.3 Solução por Programação Linear . . . . .	23
2.2.4 Generalização . . . . .	29
2.3 O Problema das Testemunhas . . . . .	33
2.3.1 Caso Com Uma Testemunha . . . . .	34
2.3.2 Caso Com Duas Testemunhas . . . . .	41
2.4 Cadeia de Markov . . . . .	49
<b>3 Precificação de Opções</b>	<b>53</b>
3.1 Conceitos básicos . . . . .	53
3.2 Limitantes para uma opção de compra . . . . .	57
3.2.1 Limitante superior para uma opção de compra . . . . .	58
3.2.2 Limitante inferior para uma opção de compra . . . . .	62

<b>Considerações Finais</b>	<b>67</b>
<b>A Conceitos Básicos</b>	<b>69</b>
A.1 Programação Linear . . . . .	69
A.2 Programação Linear Fracionária . . . . .	70
A.3 Análise Convexa . . . . .	71
A.4 Teoria de Probabilidades . . . . .	72
A.4.1 Probabilidade Condicional . . . . .	75
A.4.2 Esperança . . . . .	76
A.4.3 Distribuição Absolutamente Contínua . . . . .	76
A.5 Algoritmo do Exemplo 1.12 . . . . .	77
<b>Referências Bibliográficas</b>	<b>79</b>



---

# Lista de Figuras

1.1	Gráfico do limitante superior e inferior para a distribuição de probabilidade acumulada. . . . .	10
2.1	Solução para a Desigualdade de Markov. . . . .	19
2.2	Solução para a Desigualdade de Markov. . . . .	19
2.3	Árvore de probabilidade para o problema com três portas. . . . .	23
2.4	Árvore de probabilidade para o problema de Monty Hall com quatro portas. . . . .	30
2.5	Árvore de probabilidade para o caso com uma testemunha. . . . .	36
2.6	Árvore de probabilidades para uma testemunha. . . . .	42
2.7	Árvore de probabilidades para duas testemunhas. . . . .	45
2.8	Cadeia de Markov com três estados. . . . .	50
3.1	Preço da ação para $\omega_1$ e $\omega_2$ . . . . .	57
3.2	Gráfico do preço da opção de compra $C$ para cada valor de $u$ e $d$ . . . . .	58
3.3	$g(S_T)$ para $S_T = S_{T_0}$ . . . . .	60
3.4	$g(S_T)$ para $S_T = 0$ . . . . .	60
3.5	Histograma do histórico de cotação da ação PETR4 . . . . .	65
3.6	Série temporal do histórico de cotação da ação PETR4 . . . . .	65



# Introdução

Uma das questões mais importantes dos mercados financeiros consiste em encontrar o preço de um derivativo, ou seja, precificar um ativo que depende exclusivamente de outro, dadas informações sobre a distribuição do preço do ativo do qual o derivativo depende. Imagine um cenário em que não sabemos qual é esta distribuição. Como devemos proceder para precificar o derivativo? Mais precisamente, como precificamos uma opção de compra europeia, quando não sabemos a distribuição do ativo base, mas conhecemos os dois primeiros momentos dessa distribuição? Neste trabalho apresentamos um limitante superior para o preço de uma opção de compra presente no trabalho de [11]. Motivado por esse limitante superior e sentindo a necessidade de um preço inferior para a opção, desenvolvemos um limitante inferior para o preço de uma opção de compra europeia neste trabalho, utilizando a teoria que será apresentada no Primeiro Capítulo.

Além dessa questão, existem inúmeros outros problemas na estatística em que não temos pleno conhecimento sobre o modelo estatístico que descreve a variável de interesse. Entretanto, é possível que haja informação sobre alguns momentos da distribuição em estudo. Será que essa informação pode nos dizer algo sobre a distribuição de probabilidade a ser construída? Que decisão conseguimos tomar a partir do conhecimento, por exemplo, dos dois primeiros momentos, ou seja, da média e variância, da variável em estudo.

Em Estatística Bayesiana, mesmo em posse de uma amostra com muitas observações, é possível que a informação a priori sobre os parâmetros do modelo seja limitada. Diante disso, o que podemos dizer sobre a distribuição a posteriori dos mesmos parâmetros?

Este trabalho estuda problemas como estes, ou seja, em que temos poucas informações sobre as distribuições envolvidas, ou quando as probabilidades do problema estudado não precisam ser especificadas exatamente, podendo ser dadas como intervalos. O objetivo é, portanto, encontrar um método capaz de fornecer limites máximos e mínimos para probabilidades de eventos de interesse dado um conhecimento restrito sobre o fenômeno ou variável em estudo, por exemplo, quando conhecemos alguns momentos da distribuição.

Para resolver esse tipo de problema, utilizaremos a teoria de programação linear para maximizarmos e minimizarmos, sob certas restrições, esperanças matemáticas (generalizadas) das variáveis aleatórias em estudo. Também faremos uso das propriedades derivadas do cálculo de probabilidades, uma vez que as funções objetivo, no contexto deste trabalho, serão funções lineares que queremos maximizar e minimizar, sujeitas a uma série de condições, por serem integrais tomadas com relação a medidas de probabilidade.

A organização do trabalho é como segue. O primeiro capítulo, que dará todo o suporte teórico para o que será desenvolvido posteriormente, está baseado principalmente nas referências [2] e [21]. Nele desenvolveremos alguns conceitos da teoria de programação linear com o intuito de maximizar e minimizar esperanças matemáticas generalizadas sob um conjunto de restrições também compostas por integrais.

Nos dois capítulos seguintes, apresentamos a aplicação do método exposto no primeiro capítulo a problemas específicos. No segundo capítulo apresentaremos uma outra versão de demonstração para a Desigualdade de Markov, utilizando a programação linear. Além disso, apresentaremos três exemplos clássicos explorados pela teoria probabilística, a saber, o Problema de *Monty Hall*, o Problema das Testemunhas e um Problema de Cadeia de Markov. Em ambos os problemas utilizaremos as técnicas de programação linear para determinar as probabilidades de interesse, dadas certas hipóteses sobre probabilidades de eventos logicamente relacionados ao evento de interesse.

No terceiro capítulo trataremos do problema da precificação de uma opção (européia) de compra quando dispõe-se de informação restrita aos dois primeiros momentos do preço do ativo que baseia a opção. Serão apresentados conceitos fundamentais da teoria de precificação de ativos e, como resultado final, obteremos uma desigualdade que determina um limitante superior e um limitante inferior para o preço da opção de compra. O último capítulo traz algumas considerações finais a respeito do trabalho.

# Capítulo 1

## Maximização e Minimização de Esperança Matemática

Até hoje existem inúmeros problemas estatísticos em que não é possível determinar uma probabilidade exata para um determinado evento. Por exemplo, considere o seguinte caso: temos um dado com a mesma probabilidade de sair cada face, ou seja, a probabilidade de sair cada uma das faces é de  $\frac{1}{6}$ . Imagine agora que tenhamos um outro dado, em que não sabemos o valor exato probabilidade de sair cada face, apenas se sabe que este valor pertence a um intervalo, ou seja, a probabilidade de sair as faces de número  $1, 2, \dots, 6$  são, respectivamente,  $P(1) \in [0.2, 0.5]$ ,  $P(2) \in [0.1, 0.3]$ ,  $P(3) \in [0.1, 0.2]$ ,  $P(4) \in [0.5, 0.6]$ ,  $P(5) \in [0.1, 0.6]$  e  $P(6) \in [0.1, 0.2]$ . Neste caso, a única restrição que temos para determinar qual a probabilidade para cada evento é que a soma das probabilidades de todos os eventos seja igual a um. Diante disso, qual deve ser a probabilidade, por exemplo, de sair a face de número dois ao lançarmos este dado? Essa probabilidade é única? Se sim, sob quais condições?

No mesmo espírito do exemplo anterior, considere que agora temos uma densidade de probabilidade com 100 pontos  $x_i$  equidistantes, com uma distância de 0.02 no intervalo  $[-1, 1]$ . Esses pontos satisfazem as seguintes restrições:

$$\begin{aligned}\mu_1 &= E[X] \in [-0.1, 0.1] \\ \mu_2 &= E[X^2] \in [0.5, 0.6] \\ \mu_3 &= E[3X^3 - 2X] \in [-0.3, 0.4] \\ \mu_4 &= P(X < 0) \in [0.3, 0.4].\end{aligned}\tag{1.1}$$

Juntamente com a restrição de que a soma das probabilidades  $p_i$  seja igual a um, ou seja,  $\sum_{i=1}^{100} p_i = 1$  e todas as probabilidades são maiores ou iguais a zero, ou seja,  $p_i \geq 0$  para  $i = 1, \dots, 100$ . Como podemos utilizar o conjunto de restrição (1.1) para determinar probabilidades para  $P(X \leq x_i)$ ?

Nos problemas acima e em outros similares, a programação linear pode ser muito útil para nos ajudar a determinar tal probabilidade. Neste capítulo, apresentaremos um

método, descrito por Smith(1995), que busca uma solução ótima dentro do conjunto de todas as possíveis soluções para um determinado problema de programação linear, no qual temos informações insuficientes para especificar um único valor. No exemplo 1.12, discutimos mais a respeito de como determinar probabilidades para  $P(X \leq x_i)$ , utilizando a teoria desenvolvida por Smith.

## 1.1 Quadro geral

Seja  $X$  um espaço linear (com elementos  $x$ ) e  $\mathcal{B}$  uma  $\sigma$  - álgebra de subconjuntos mensuráveis de  $X$ .

**Definição 1.1** (Funções de Momentos). *Seja  $f_i : X \rightarrow \mathbb{R}$ , tal que,  $f_i(x)$ ,  $i = 0, 1, \dots, n$ , e definida em  $X$  e  $\mathcal{B}$ -mensuráveis.*

Aqui, vamos assumir que temos  $n + 1$  funções de momentos de valores reais.

**Definição 1.2** (Esperança de Funções de Momentos). *Denotaremos por  $\mu_i$  a esperança das funções de momentos, as quais são dados pela seguinte expressão:*

$$\mu_i = E_P[f_i] \equiv \int_X f_i(x) dP(x) \quad i = 0, 1, \dots, n.$$

As esperanças  $\mu_i$ , serão assumidas conhecidas e finitas, mesmo que possamos não conhecer a distribuição implícita (ou medida não negativa  $P$ ), que gera os ditos momentos.

Por conveniência, tomamos  $f_0(x) \equiv 1$  de modo que  $\mu_0 = E_P[f_0] = 1$  para toda distribuição de probabilidade  $P$  e escrevemos o vetor das funções de momentos como  $\mathbf{f}(x) = (f_0(x), f_1(x), \dots, f_n(x))$ , ou seja,  $\mathbf{f} : X \rightarrow \mathbb{R}^{n+1}$  e o vetor de momentos como  $\boldsymbol{\mu} = (\mu_0, \mu_1, \dots, \mu_n)$ . Assumimos que as funções de momentos são linearmente independentes em  $X$ , ou seja, não existe vetor não nulo  $\boldsymbol{\lambda}$  tal que  $\boldsymbol{\lambda}^T \mathbf{f}(x) = 0$  para todo  $x \in X$ .

**Definição 1.3** (Função Objetivo). *Seja  $X$  um espaço linear e  $\mathcal{B}$ -mensurável, definimos uma função objetivo  $\phi$  como :*

$$\phi : X \rightarrow \mathbb{R},$$

sendo  $\phi$  uma função linear.

**Definição 1.4** (Distribuições Permitidas). *Denotaremos por  $\mathcal{A}$ , o conjunto das distribuições permitidas, dado pela seguinte expressão:*

$$\mathcal{A}(\boldsymbol{\mu}) = \{P : E_P[f_i] = \mu_i, \forall i = 0, \dots, n\}.$$

As distribuições em  $\mathcal{A}$  são definidas em  $(X, \mathcal{B})$  e assumimos que os momentos e as funções-objetivo são integráveis respectivamente, para cada distribuição  $P$  em  $\mathcal{A}$ .

O conjunto  $\mathcal{A}(\mu)$  contém todas as distribuições que podem ser ou não solução do nosso problema. Para efeito de estudo, vamos assumir que o conjunto  $\mathcal{A}$  das distribuições permitidas é não vazio, ou seja,  $\mathcal{A}$  contém ao menos uma distribuição que é solução. Além disso, seja denotado por  $\mathcal{A}(\mu)$  o subconjunto de  $\mathcal{A}$  que contém as distribuições  $P$  em  $\mathcal{A}$  tais que  $\mu = E_P[\mathbf{f}]$ . É conveniente que  $\mathcal{A}$  inclua distribuições com densidades totais não necessariamente iguais a 1, quando ocorrem esses casos aplicamos a restrição de momento  $\mu_o = E_P[f_0] = 1$ . Agora iremos enunciar nosso objetivo de estudo.

Queremos determinar

$$\inf_{P \in \mathcal{A}(\mu)} E_P[\phi] \quad \text{e} \quad \sup_{P \in \mathcal{A}(\mu)} E_P[\phi]. \quad (1.2)$$

Em outras palavras, nosso objetivo é calcular limitantes superior e inferior da esperança de  $\phi$  dado o conjunto de *distribuições permitidas*.

As hipóteses assumidas até aqui são bastante gerais. O conjunto  $X$  pode ser discreto ou contínuo, multidimensional, ou, em alguns casos até com dimensão infinita. As funções momentos e a função-objetivo podem ser quaisquer funções reais, sujeitas apenas aos requisitos de mensurabilidade. Sendo que as funções de momentos serão utilizada para construir os nossos conjunto de restrição e a função objetivo será

## 1.2 Limitantes sobre a Esperança Matemática sem Restrição no Conjunto de Soluções

Primeiro nós consideraremos o problema de calcular limitantes para a esperança, no qual não colocamos nenhuma restrição sobre o conjunto de distribuições que podem resolver nosso problema. Pensando apenas no limitante superior, o problema (1.2) pode ser reescrito como

$$\sup_{P \in \mathcal{D}(\mu)} E_P[\phi], \quad (1.3)$$

onde  $\mathcal{D}(\mu)$  representa o conjunto das *distribuições factíveis*, este conjunto contém apenas as distribuições que são solução para o nosso problema. Com isso, nosso objetivo é determinar qual a distribuição de que maximiza  $\mathcal{D}(\mu)$  nosso problema (1.3). Este problema pode ser visto como um problema de programação linear na forma padrão: a variável de decisão é a massa de probabilidade atribuída a cada ponto  $x$  em  $X$ , e portanto não pode ser negativa; a função-objetivo  $E_P[\phi]$  e as restrições  $\mu = E_P[\mathbf{f}]$  são funções lineares destas variáveis de decisão. Se o espaço  $X$  é finito, (1.3) é um problema de programação linear convencional. Se  $X$  é infinito, (1.3) é um problema de programação linear semi-infinito com um número infinito de variáveis de decisão e um número finito de restrições. Observando a expressão (1.3) como um problema de programação linear podemos encontrar soluções básicas e versões mais gerais do Teorema Fundamental da Programação Linear

e do Teorema de Dualidade da Programação Linear, que serão apresentados a seguir.

### 1.2.1 Teorema Fundamental da Programação Linear

Da mesma maneira que se resolve problemas padrões de Programação Linear, definiremos uma distribuição básica como sendo uma distribuição de probabilidade discreta com pontos suporte  $x_1, x_2, \dots, x_k$  de modo que os vetores  $\mathbf{f}(x_1), \mathbf{f}(x_2), \dots, \mathbf{f}(x_k) \in \mathbb{R}^{n+1}$  sejam linearmente independentes.

Como vimos na seção 1.1, existem  $n + 1$  funções de momentos e, por isso, podemos concluir que estas distribuições básicas podem ter no máximo  $n + 1$  pontos. Portanto, devemos ter  $k \leq n + 1$ . Dessa forma, denotaremos isso por  $\Delta(\boldsymbol{\mu})$ , o conjunto das distribuições básicas factíveis.

Com esta definição de distribuição básica podemos estabelecer uma versão do Teorema Fundamental da Programação Linear que vem a seguir. Mulholland e Rogers (1955) [16] provaram os resultados para um caso específico em que  $X = \mathbb{R}$ . Já Isii (1963) [9] trabalhou na generalização desse caso, mas pensando em distribuições discretas com não mais que  $n + 1$  pontos de suporte e sem considerar a condição de independência linear.

#### **Teorema 1.5.** *Teorema Fundamental*

*Dado o problema*

$$\sup_{P \in \mathcal{D}(\boldsymbol{\mu})} E_P[\phi], \quad (1.4)$$

onde  $\mathcal{D}(\boldsymbol{\mu})$  é o conjunto das distribuições factíveis, as seguintes afirmações são válidas:

- a. *Se existir uma distribuição factível, existe uma distribuição básica factível;*
- b. *Se existir uma distribuição ótima<sup>1</sup>, existe uma distribuição básica ótima;*
- c.

$$\sup_{P \in \mathcal{D}(\boldsymbol{\mu})} E_P[\phi] = \sup_{P \in \Delta(\boldsymbol{\mu})} E_P[\phi].$$

As partes (a) e (b) do teorema parafraseam o Teorema Fundamental convencional da programação linear como encontrado, por exemplo, em Luenberger (1984) (também está enunciado no apêndice). A parte (c) aborda uma preocupação de problemas convencionais de programação linear: mesmo se os limitantes sobre a esperança forem finitos, há a possibilidade de não haver distribuição em  $\mathcal{D}(\boldsymbol{\mu})$  que atinja esse limitante. O principal resultado do teorema—item (c)—é que podemos restringir a busca pela distribuição que atingirá o limitante desejado olhando apenas o conjunto  $\Delta(\boldsymbol{\mu})$  das distribuições básicas factíveis ao invés de buscar no conjunto de todas as distribuições factíveis  $\mathcal{D}(\boldsymbol{\mu})$ .

A prova do Teorema Fundamental se dará provando a seguinte Proposição, que estabelece simultaneamente as três partes do Teorema Fundamental.

<sup>1</sup>A distribuição que otimiza o problema (1.4).



**Proposição 1.6.** Para cada  $P_1 \in \mathcal{D}(\boldsymbol{\mu})$ , existe  $P_2 \in \Delta(\boldsymbol{\mu})$  tal que  $E_{P_2}[\phi] \geq E_{P_1}[\phi]$ .

A prova desta Proposição explora a convexidade do espaço linear dos momentos  $\mathcal{M} \equiv \{(E_P[\mathbf{f}], E_P[\phi]) : P \in \mathcal{D} \text{ e } E_P[1] = 1\}$  e a sua relação com o gráfico de momentos  $\mathcal{F} \equiv \{(\mathbf{f}(x), \phi(x)) : x \in X\}$  e é estabelecida com o auxílio de dois lemas. O primeiro lema diz respeito à representação dos pontos no casco convexo de  $\mathcal{F}$ , denotado como  $\text{conv}(\mathcal{F})$ , e o segundo estabelece a igualdade de  $\mathcal{M}$  e  $\text{conv}(\mathcal{F})$ .

**Lema 1.7.** Para qualquer ponto  $\mathbf{s} \in \text{conv}(\mathcal{F})$ , existem  $q$  ( $q \leq n+2$ ) pontos  $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_q$  em  $\mathcal{F}$  com pesos positivos  $w_1, w_2, \dots, w_q$  tais que  $\mathbf{s} = \sum_{i=1}^q w_i \mathbf{s}_i$ . Além disso, os pontos  $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_q$  podem ser selecionados de modo que sejam linearmente independentes. Desde que  $f_0 \equiv 1$ , os pesos  $w_i$  devem somar 1 e  $\sum_{i=1}^q w_i \mathbf{s}_i$  é uma combinação convexa de pontos de  $\mathcal{F}$ .

*Demonstração.* O fato de  $\mathbf{s}$  poder ser representado como uma combinação convexa de um número finito de pontos em  $\mathcal{F}$  segue a partir de resultados padrão de análise convexa, cujos resultados aqui utilizados encontram-se no apêndice. Para mais detalhes consulte a referência [3]. Agora mostraremos que se os pontos representados  $\mathbf{s}$  não são linearmente independentes,  $\mathbf{s}$  pode ser representado por um pequeno conjunto de pontos que são linearmente independentes. Como  $\mathcal{F} \subseteq \mathbb{R}^{n+2}$ , então não mais que  $n+2$  pontos podem ser linearmente independentes. Suponha que  $\mathbf{s}$  possa ser representado como uma combinação convexa positiva  $\sum_{i=1}^q w_i \mathbf{s}_i$  de pontos  $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_q$  que são linearmente dependentes. Então existem números  $\alpha_1, \alpha_2, \dots, \alpha_{n+2}$  não todos iguais a zero, tais que  $\sum_{i=1}^{n+2} \alpha_i \mathbf{s}_i = \mathbf{0}$ . Para cada  $r$  tal que  $\alpha_r \neq 0$ , podemos escrever  $\mathbf{s}_r$  e  $\mathbf{s}$  como

$$\mathbf{s}_r = \sum_{i \neq r} (\alpha_i / \alpha_r) \mathbf{s}_i \quad \text{e} \quad \mathbf{s} = \sum_{i=1}^q (w_i - (\alpha_i / \alpha_r) w_r) \mathbf{s}_i.$$

Se tomarmos  $r$  como sendo  $(w_r / \alpha_r) = \min\{w_i / \alpha_i : \alpha_i > 0\}$  (se não tivermos  $\alpha_i > 0$ , multiplicamos todos os  $\alpha_i$  por  $-1$ ), todos os  $(w_i - (\alpha_i / \alpha_r) w_r)$  serão não negativos. Assim, temos uma nova combinação convexa com os pesos  $(w_i - (\alpha_i / \alpha_r) w_r)$  e os pontos  $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_k$ . Os pontos  $\mathbf{s}_r$  agora têm peso zero e qualquer outro ponto com peso zero pode ser eliminado para se obter uma combinação convexa positiva com menos pontos. Repetimos este processo de redução até chegarmos a uma combinação convexa positiva com pontos linearmente independentes. Esse processo pode ser realizado, porque quando defini  $\mathbf{s}_r$  e  $\mathbf{s}$ , separei os pontos que tem peso zero.  $\square$

**Lema 1.8.**  $\mathcal{M} = \text{conv}(\mathcal{F})$ .

*Demonstração.* Primeiro mostraremos que  $\text{conv}(\mathcal{F}) \subseteq \mathcal{M}$ . Para cada ponto  $\mathbf{s} \in \mathcal{F}$ , existem distribuições em  $\mathcal{D}$  (degeneradas em  $x$  ou delta de Dirac  $\delta_x$ ) que colocam densidade em um único ponto  $x$  tal que  $(\mathbf{f}, \phi(x)) = \mathbf{s}$ . Assim  $\mathcal{F} \subseteq \mathcal{M}$ . Como  $\mathcal{M}$  é convexo,  $\text{conv}(\mathcal{F}) \subseteq \mathcal{M}$ .

Mostraremos que  $\mathcal{M} \subseteq \text{conv}(\mathcal{F})$  por indução finita sobre  $n$ . Para  $n = 0$ , o resultado  $\mathcal{M} \subseteq \text{conv}(\mathcal{F})$  reduz-se a

$$\{E_P[\phi] : P \in \mathcal{D} \text{ e } E_P[1] = 1\} \subseteq \{\alpha\phi(x_1) + (1 - \alpha)\phi(x_2) : x_1, x_2 \in X, 0 \leq \alpha \leq 1\},$$

o que é verdade para qualquer conjunto  $X$ . Como hipótese de indução assumimos que  $\mathcal{M} \subseteq \text{conv}(\mathcal{F})$  é válido para quaisquer  $n - 1$  funções de momento e para qualquer conjunto  $X$ . Suponha que para algum  $X$ ,  $\mathcal{M} \subseteq \text{conv}(\mathcal{F})$  não vale para o momento  $n$ . Então existe um ponto  $\boldsymbol{\mu}_0 \in \mathcal{M}$  tal que  $\boldsymbol{\mu}_0 \notin \text{conv}(\mathcal{F})$ . Como  $\text{conv}(\mathcal{F})$  é convexo, existe um hiperplano que separa  $\boldsymbol{\mu}_0$  de  $\text{conv}(\mathcal{F})$ , ou seja, existe um vetor não nulo  $\boldsymbol{\lambda} \in \mathbb{R}^{n+2}$  tal que  $\boldsymbol{\lambda}^T \boldsymbol{\mu}_0 \leq 0$  e  $\boldsymbol{\lambda}^T \boldsymbol{\mu} \geq 0$  para todo  $\boldsymbol{\mu} \in \text{conv}(\mathcal{F})$ . Uma vez que  $\boldsymbol{\mu}_0 \in \mathcal{M}$ , existe uma distribuição  $P_0$  tal que  $\boldsymbol{\mu}_0 = E_{P_0}[(\mathbf{f}, \phi)]$ , e isso implica  $E_{P_0}[\boldsymbol{\lambda}^T(\mathbf{f}, \phi)] \leq 0$ . Considerando os pontos  $\boldsymbol{\mu} \in \mathcal{F}$ , temos  $\boldsymbol{\lambda}^T(\mathbf{f}(x), \phi(x)) \geq 0$  para todo  $x \in X$ . Assim, devemos ter  $E_{P_0}[\boldsymbol{\lambda}^T(\mathbf{f}, \phi)] = 0$ , o que implica que  $P_0$  é concentrado no conjunto de pontos  $X' = \{x : \boldsymbol{\lambda}^T(\mathbf{f}(x), \phi(x)) = 0\}$  (ou seja,  $P_0(X') = 1$ ). Note que isto implica que as funções  $(\mathbf{f}, \phi)$  são linearmente dependentes em  $X'$ .

Uma vez que as funções  $(\mathbf{f}, \phi)$  são linearmente dependentes em  $X'$ , podemos tomar uma função momento, digamos  $f_n$ , e aplicar a hipótese de indução com  $X$  substituído por  $X'$ . Seja  $\mathbf{f}' = (f_0, f_1, \dots, f_{n-1})$  denotando o conjunto reduzido de funções momento,  $\mathcal{M}'$  a redução do espaço de momentos  $\{(E_P[\mathbf{f}'], E_P[\phi]) : P \in \mathcal{D} \text{ e } E_P[1] = 1\}$ , e o gráfico de momento reduzido  $\mathcal{F}' = \{(\mathbf{f}'(x), \phi(x)) : x \in X'\}$ . Uma vez que  $\boldsymbol{\mu}'_0 \equiv E_{P_0}[(\mathbf{f}', \phi)] \in \mathcal{M}'$ , a hipótese de indução implica em  $\boldsymbol{\mu}'_0 \in \text{conv}(\mathcal{F}')$ .

Agora mostremos que isto implica em  $\boldsymbol{\mu}_0 \in \text{conv}(\mathcal{F})$ . Uma vez que  $\boldsymbol{\mu}'_0 \in \text{conv}(\mathcal{F}')$ , pelo Lema 1.7 existem  $q$  ( $q \leq n$ ) pontos  $x_1, x_2, \dots, x_q \in X'$ , e pesos positivos  $w_1, w_2, \dots, w_q$  tais que  $\boldsymbol{\mu}'_0 = \sum_{i=1}^q w_i(\mathbf{f}'(x_i), \phi(x_i))$ . Por causa da dependência linear das funções  $(\mathbf{f}, \phi)$  em  $X'$ , existe um vetor  $\mathbf{a} \in \mathbb{R}^{n+1}$  tal que  $f_n(x) = \mathbf{a}^T(\mathbf{f}'(x), \phi(x))$  para todo  $x \in X'$  e  $\boldsymbol{\mu}_{n0} = \mathbf{a}^T \boldsymbol{\mu}'_0$ . Assim

$$\boldsymbol{\mu}_{n0} = \mathbf{a}^T \boldsymbol{\mu}'_0 = \sum_{i=1}^q w_i \mathbf{a}^T(\mathbf{f}'(x_i), \phi(x_i)) = \sum_{i=1}^q w_i f_n(x_i),$$

temos que  $\boldsymbol{\mu}_0 = \sum_{i=1}^q w_i(\mathbf{f}(x_i), \phi(x_i))$  e  $\boldsymbol{\mu}_0 \in \text{conv}(\mathcal{F})$ , o que é uma contradição, pois assumimos que  $\boldsymbol{\mu}_0 \notin \text{conv}(\mathcal{F})$ . Assim  $\mathcal{M} \subseteq \text{conv}(\mathcal{F})$  e, como mostrado anteriormente,  $\text{conv}(\mathcal{F}) \subseteq \mathcal{M}$ , portanto  $\text{conv}(\mathcal{F}) = \mathcal{M}$ .  $\square$

Agora finalmente demonstraremos a Proposição 1.6

*Demonstração da Proposição 1.6.* Considere um ponto qualquer  $(\boldsymbol{\mu}, \gamma) = E_{P_1}[(\mathbf{f}, \phi)] \in \mathcal{M}$  (note que  $\gamma < \infty$  pois assumimos que  $P_1 \in \mathcal{D}$ ). Pelos Lemas 1.7 e 1.8, existem  $q$  ( $q \leq n + 2$ ) pontos  $x_1, x_2, \dots, x_q \in X$  e pesos positivos  $w_1, w_2, \dots, w_q$  tais que os  $q$  vetores  $(\mathbf{f}(x_i), \phi(x_i))$  são linearmente independentes e  $(\boldsymbol{\mu}, \gamma) = \sum_{i=1}^q w_i(\mathbf{f}(x_i), \phi(x_i))$ . Os pontos  $(\mathbf{f}(x_i), \phi(x_i))$  são vértices de um simplex  $\psi$  que contém o ponto  $(\boldsymbol{\mu}, \gamma)$ . A reta que

divide o plano em duas partes  $\zeta \equiv \{(\boldsymbol{\mu}, \tau) : \tau \geq \gamma\}$  deve cruzar um lugar adequado de  $\psi$  de modo que o ponto  $(\boldsymbol{\mu}, \bar{\tau})$  seja o máximo, onde  $\bar{\tau} = \max\{\tau : (\boldsymbol{\mu}, \tau) \in \psi\}$ . Uma vez que este ponto pertence a  $\psi$ , ele pode ser representado como uma combinação convexa positiva de um subconjunto de  $q$  pontos  $(\mathbf{f}(x_i), \phi(x_i))$  contendo não mais que  $n + 1$  pontos. Assim estabelecemos a existência de uma distribuição discreta  $P_2$  com densidade em  $n + 1$  ou menos pontos tal que  $E_{P_2}[\mathbf{f}] = \boldsymbol{\mu}$  e  $E_{P_2}[\phi] = \bar{\tau} \geq E_{P_1}[\phi] = \tau$ .

Ainda temos que mostrar que os vetores  $\mathbf{f}(x_1), \mathbf{f}(x_2), \dots, \mathbf{f}(x_k)$  formados pelos pontos  $x_1, x_2, \dots, x_k$  (envolvidos em  $P_2$ ) são linearmente independentes. Então vamos supor por absurdo, que esses vetores não são linearmente independentes. Assim existe um vetor não nulo  $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_k)$  tal que  $\sum_{i=1}^q \lambda_i \mathbf{f}(x_i) = \mathbf{0}$ . Como os vetores  $(\mathbf{f}(x_i), \phi(x_i))$  são linearmente independentes, devemos ter  $\sum_{i=1}^q \lambda_i \phi(x_i) \neq 0$ . Sem perda de generalidade, podemos assumir  $\sum_{i=1}^q \lambda_i \phi(x_i) > 0$  (caso contrário, troque  $\boldsymbol{\lambda}$  por  $-\boldsymbol{\lambda}$ ). Sejam  $p_1, p_2, \dots, p_k > 0$ , todos denotando a densidade nos pontos  $x_1, x_2, \dots, x_k$  em  $P_2$ . Se  $E_{P_2}[\mathbf{f}] = \boldsymbol{\mu}$ , temos  $\sum_{i=1}^q p_i \mathbf{f}(x_i) = \boldsymbol{\mu}$ . Como as densidades  $p_i$  são não negativas, existe algum  $\epsilon > 0$  tal que  $p_i + \epsilon \lambda_i > 0$  para todo  $i$ . Então temos

$$\sum_{i=1}^q (p_i + \epsilon \lambda_i) \mathbf{f}(x_i) = \boldsymbol{\mu} \quad \text{e} \quad \tau' \equiv \sum_{i=1}^q (p_i + \epsilon \lambda_i) \phi(x_i) > \sum_{i=1}^q p_i \phi(x_i) = \bar{\tau}.$$

Assim  $(\boldsymbol{\mu}, \tau')$  está em  $\psi$  e encontra-se acima da reta  $\zeta$  acima do ponto  $(\boldsymbol{\mu}, \bar{\tau})$ , contradizendo nossa definição de que  $(\boldsymbol{\mu}, \bar{\tau})$  é a máxima intersecção de  $\zeta$  e  $\psi$ . Assim os vetores  $\mathbf{f}(x_1), \mathbf{f}(x_2), \dots, \mathbf{f}(x_k)$  são linearmente independentes e  $P_2 \in \Delta(\boldsymbol{\mu})$ .

□

### 1.2.2 Teorema de Dualidade

Cada problema primário da forma (1.3) tem um problema dual correspondente. Ao invés de procurar uma distribuição factível que maximiza  $E_P[\phi]$ , buscamos um vetor  $\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \dots, \lambda_n)$  que resolva

$$\inf_{\boldsymbol{\lambda}} \{ \boldsymbol{\lambda}^T \boldsymbol{\mu} : \boldsymbol{\lambda}^T \mathbf{f}(x) \geq \phi(x), \quad \forall x \in X \}. \quad (1.5)$$

Podemos notar que a solução para o problema dual da expressão (1.5) proporciona um valor para o limitante superior do problema primário: tomando a esperança em ambos os lados da desigualdade  $\boldsymbol{\lambda}^T \mathbf{f}(x) \geq \phi(x)$  segue que  $\boldsymbol{\lambda}^T \boldsymbol{\mu} \geq E_P[\phi]$  para alguma distribuição  $P$  em  $\mathcal{D}(\boldsymbol{\mu})$ . Esta observação implica o resultado a seguir.

**Lema 1.9.** (*Lema da Dualidade Fraca*)

- a. Se  $P$  e  $\boldsymbol{\lambda}$  são factíveis (soluções) para (1.3) e (1.5), respectivamente, então  $E_P[\phi] \leq \boldsymbol{\lambda}^T \boldsymbol{\mu}$ .
- b. Se  $P$  e  $\boldsymbol{\lambda}$  são factíveis (soluções) para (1.3) e (1.5), respectivamente e  $E_P[\phi] = \boldsymbol{\lambda}^T \boldsymbol{\mu}$ , então  $P$  e  $\boldsymbol{\lambda}$  são ótimos para os respectivos problemas.

O teorema convencional de dualidade de programação linear estabelece a igualdade de solução dos problemas dual e primário: se o problema primário ou dual tem uma solução finita ótima, o mesmo acontece com a outra e os valores correspondentes da função objetivo (nossa função de interesse) são iguais. Em nosso contexto, podemos estabelecer a seguinte versão ligeiramente mais fraca do teorema da dualidade. Este resultado foi estabelecido pela primeira vez em sua forma geral em Isii(1963) e pode ser provado como um caso particular do Teorema da Dualidade que será enunciado a seguir.

**Teorema 1.10** (Teorema da Dualidade). *Se  $\mu$  é um ponto interior, ou seja, um ponto do conjunto  $\{E_P[\mathbf{f}(x)] : P \in \mathcal{D}\}$ , então*

$$\sup_{P \in \mathcal{D}(\mu)} E_P[\phi] = \inf_{\lambda} \{\lambda^T \mu : \lambda^T \mathbf{f}(x) \geq \phi(x), \forall x \in X\}. \quad (1.6)$$

*Além disso, se o problema primário é limitado, o dual tem solução ótima finita.*

Este resultado de dualidade é mais fraco que o resultado convencional porque aqui se exige que  $\mu$  seja um ponto interior do conjunto  $\{E_P[\mathbf{f}(x)] : P \in \mathcal{D}\}$ . A restrição de ponto interior exige que  $\mu$  seja o momento de alguma distribuição, mas exclui, por exemplo, momentos que determinam exclusivamente a distribuição implícita.

**Teorema 1.11** (Condição de Relaxamento Complementar ou *Complementary Slackness Condition*). *Se  $P$  e  $\lambda$  são soluções ótimas dos problemas primário e dual, respectivamente, então  $P$  tem densidade apenas naqueles pontos  $x$  tais que  $\lambda^T \mathbf{f}(x) = \phi(x)$ .*

A condição de relaxamento complementar fornece um valioso método para checar a otimalidade das soluções. Dada uma distribuição de probabilidade factível, podemos contruir um polinômio  $\lambda^T \mathbf{f}$  satisfazendo as condições da Condição de Relaxamento Complementar. Se este polinômio influencia  $\phi$  (ou seja, altera o resultado de  $\phi$ ), então a distribuição é uma solução ótima para (1.3). Reciprocamente, dado um polinômio factível  $\lambda^T \mathbf{f}(x)$ , podemos verificar a otimalidade analisando se é possível a construção de uma distribuição factível com densidade restrita ao pontos de  $x$  onde  $\lambda^T \mathbf{f}(x) = \phi(x)$ .

A perspectiva dual também oferece algumas dicas sobre a função objetivo  $\phi$  que irá criar limitantes menores, ou seja, limitantes mais precisos. O ajustamento dos limitantes depende de quão bem a função  $\phi$  pode ser aproximada por polinômios da forma  $\lambda^T \mathbf{f}$ . Em um extremo, se  $\phi$  é combinação linear de funções momento, então  $E[\phi]$  é precisamente determinada. Se  $\phi$  é fracamente aproximada por polinômios, como é o caso com função indicadora e momentos de potência, os limitantes em  $E[\phi]$  não serão muito justos. Se  $\phi$  é bem aproximado por polinômios, como é o caso com momentos de uma função de probabilidade e função exponencial (quando  $x$  é limitado por baixo), os limitantes em  $E[\phi]$  serão muito mais justos.

### 1.2.3 Aplicação em Problema Estatísticos

Agora para ilustrar a utilização desta teoria em problemas estatísticos, considere o seguinte exemplo, retirado de [3], página 362.

**Exemplo 1.12.** *Temos uma densidade de probabilidade com 100 pontos  $x_i$  equidistantes, com uma distância de 0.02 no intervalo  $[-1, 1]$ . Esses pontos satisfazem as seguintes restrições:*

$$\begin{aligned} \mu_1 &= E[X] \in [-0.1, 0.1] \\ \mu_2 &= E[X^2] \in [0.5, 0.6] \\ \mu_3 &= E[3X^3 - 2X] \in [-0.3, 0.4] \\ \mu_4 &= P(X < 0) \in [0.3, 0.4]. \end{aligned} \tag{1.7}$$

*Juntamente com a restrição de que a soma das probabilidades  $p_i$  seja igual a um, ou seja,  $\sum_{i=1}^{100} p_i = 1$  e todas as probabilidades são maiores ou iguais a zero, ou seja,  $p_i \geq 0$  para  $i = 1, \dots, 100$ . Estas restrições descrevem um poliedro de distribuições de probabilidade.*

*Na Figura 1.1, limitamos as probabilidades, calculando limitantes superior e inferior na distribuição acumulada  $P(X \leq x_i)$ , para  $i = 1, \dots, 100$ . Para cada valor de  $i$  resolvemos dois problemas de programação linear: maximizamos e minimizamos  $P(X \leq x_i)$ , sob as restrições dadas em (1.7). As curvas superior e inferior representam, respectivamente, os limitantes superior e inferior da distribuição acumulada de probabilidade.*

*O Algoritmo 1, representa a lógica utilizada para simulação deste exemplo. A lógica presente nele, é a mesma utilizada para o restante das simulações deste trabalho, diferente normalmente na maneira que se constrói o conjunto restrição. Para maiores detalhes referentes a este desse exemplo, consulte o Anexo A.31.*

---

**Algoritmo 1:** LIMITANTES SUPERIOR E INFERIOR DO EXEMPLO 1.12.

---

**Entrada:**  $x_i, \mu_1, \mu_2, \mu_3, \mu_4$ .

**Saída:**  $\min P(X \leq x_i), \max P(X \leq x_i)$

```

1 início
2    $x_i = -1$ 
3   para cada  $x_i$  faça
4     minimize e maximize  $P(X \leq x_i)$ 
5      $x_i = x_i + 0.02$ 
6   fim
7 fim
    
```

---

Temos que qualquer função de distribuição que esteja entre o limitantes superior e o inferior é uma função de distribuição que satisfaz as hipóteses. Isto sugere a seguinte pergunta: qual distribuição escolher? Que critério utilizar para identificar uma, visto que nem sempre é possível uma única encontrar a solução ótima em um problema? Tratarremos

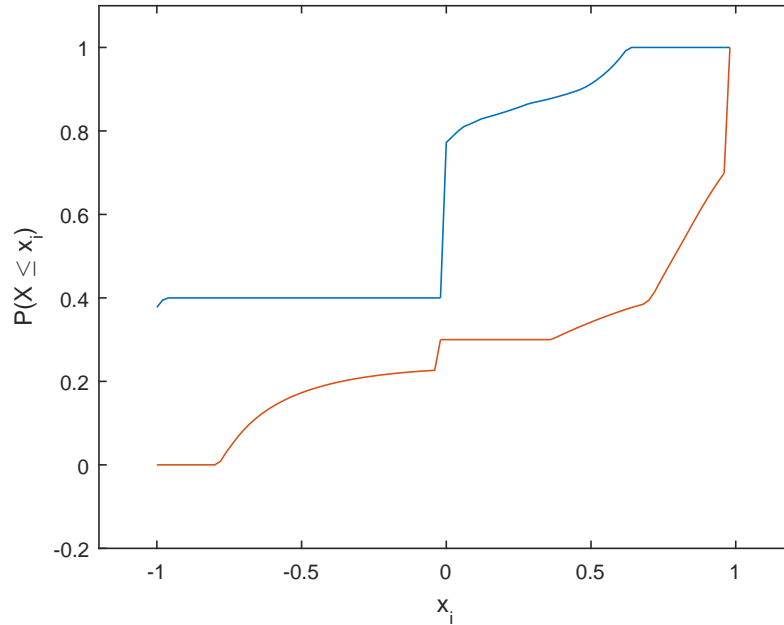


Figura 1.1: Gráfico do limitante superior e inferior para a distribuição de probabilidade acumulada.

destas questões na próxima seção, onde discutiremos maneiras de melhorar os resultados para os nossos limitantes.

### 1.3 Limitantes sobre a Esperança Matemática com Restrição no Conjunto de Soluções

Nas seções anteriores, estudamos problemas de cálculo de limitantes sobre a esperança, onde não colocamos restrições nas distribuições implícitas. Nestes casos, os limitantes foram sempre alcançados (ou aproximados) por uma distribuição discreta e os limitantes resultantes eram muitas vezes bastantes amplos. Para alcançar limitantes melhores (mais precisos) é natural tentar restringir a distribuição implícita para excluir a distribuição discreta. Por exemplo, a estrutura de alguns problemas poderia sugerir que a distribuição implícita é unimodal ou contínua. Nesta seção, estudaremos o problema do cálculo de limitantes onde restringimos a distribuição implícita a um conjunto convexo  $\mathcal{A}\boldsymbol{\mu}$  de distribuições permitidas e focamos no problema de calcular

$$\sup_{P \in \mathcal{A}(\boldsymbol{\mu})} E_P[\phi], \quad (1.8)$$

sendo  $\mathcal{A}(\boldsymbol{\mu}) = \{E_P[f_i] = \mu_i, \forall i = 1, \dots, n\}$ .

Estas restrições do conjunto convexo  $\mathcal{A}$  nos permitem preservar alguma geometria implícita ao problema irrestrito, mesmo (1.8) não sendo um problema linear. Alguns exemplos de conjuntos de restrições convexas que podem ser definidas em  $\mathcal{A}$ :

- a) O conjunto de todas as distribuições  $\mathcal{D}$ ;

- b) O conjunto de todas as distribuições simétricas ao redor de um ponto específico;
- c) O conjunto das distribuições que são unimodais com uma moda específica;
- d) O conjunto de todas as distribuições com entropia maior que um valor específico.

Além disso, como a intersecção de dois conjuntos convexos é novamente um conjunto convexo, estes exemplos podem ser combinados para originar exemplos adicionais de conjuntos de restrições convexas. Agora daremos resultados gerais aplicáveis a todos os conjuntos de restrições convexas e, em seguida, discutiremos os itens (c), (d) e (e) acima para ilustrar estes resultados gerais.

### 1.3.1 Resultados Gerais

Embora (1.8) já não seja um problema linear, ainda possui muitas das mesmas características geométricas: estamos maximizando uma função linear em um conjunto convexo de distribuições factíveis  $\mathcal{A}(\boldsymbol{\mu})$ . No caso da restrição, podemos identificar soluções básicas como os pontos extremos de  $\mathcal{A}(\boldsymbol{\mu})$ , e o Teorema Fundamental nos permite reduzir nossa busca pela solução ótima a um conjunto de pontos extremos. Embora nossa intuição continue a se aplicar neste contexto mais geral, ainda não fizemos nenhuma suposição suficiente sobre  $\mathcal{A}(\boldsymbol{\mu})$  para ter certeza de que possui pontos extremos, muito menos para ter certeza de que as soluções para (1.8) são alcançadas em algum ponto.

Em complemento ao Teorema Fundamental, os resultados de dualidade chegam para completar as definições mais gerais. Para isso vamos definir o problema dual correspondente a (1.8), o qual pode ser escrito da seguinte maneira:

$$\inf_{\boldsymbol{\lambda}} \{ \boldsymbol{\lambda}^T \boldsymbol{\mu} + \sup_{P \in \mathcal{A}} E_P[\phi - \boldsymbol{\lambda}^T \mathbf{f}] \}. \quad (1.9)$$

A justificativa para esta formulação do problema dual, será dado logo a seguir.

Para relacionar este problema com aqueles considerados anteriormente neste trabalho, seja  $\mathcal{A}$  o conjunto que contém distribuições que colocam arbitrariamente grandes massas de probabilidade em qualquer ponto (ou conjunto mensurável), para qualquer  $\boldsymbol{\lambda}$  fixado. Assim

$$\sup_{P \in \mathcal{A}} E_P[\phi - \boldsymbol{\lambda}^T \mathbf{f}] = \begin{cases} 0, & \text{se } \boldsymbol{\lambda}^T \mathbf{f}(x) \geq \phi(x), \quad \forall x \in X \\ \infty, & \text{caso contrário.} \end{cases}$$

Então (1.9) reduz-se a

$$\inf_{\boldsymbol{\lambda}} \{ \boldsymbol{\lambda}^T \boldsymbol{\mu} : \boldsymbol{\lambda}^T \mathbf{f}(x) \geq \phi(x), \quad \forall x \in X \},$$

que é exatamente o problema dual (1.5) apresentado anteriormente neste trabalho.

Como no caso com restrição, podemos notar que as soluções para o problema dual (1.9) fornecem um limitantes superior nas soluções para o problema primário (1.8). Para ver isso, note que para qualquer  $\lambda$  fixado, temos

$$\Phi(\lambda) \equiv \lambda^T \mu + \sup_{P \in \mathcal{A}} E_P[\phi - \lambda^T \mathbf{f}] \geq \lambda^T \mu + \sup_{P \in \mathcal{A}(\mu)} E_P[\phi - \lambda^T \mathbf{f}] = \sup_{P \in \mathcal{A}(\mu)} E_P[\phi].$$

Desse modo, estabeleceremos uma versão generalizada do Lema Fraco da Dualidade.

**Lema 1.13.** *Lema Fraco da Dualidade*

- a) Se  $P$  e  $\lambda$  são factíveis para (1.8) e (1.9), respectivamente, então  $E_P[\phi] \leq \Phi(\lambda)$ .
- b) Se  $P$  e  $\lambda$  são factíveis para (1.8) e (1.9), respectivamente e  $E_P[\phi] = \Phi(\lambda)$ , então  $P$  e  $\lambda$  são ótimos em seus respectivos problemas.

O Teorema de Dualidade e o de Condição de Relaxamento Complementar do caso sem restrição serão generalizados para o caso com restrição, cujas demonstrações para o caso sem restrição são análogas às que serão apresentadas a seguir. Estes resultados foram desenvolvidos pela primeira vez em sua forma geral por [9] mas também podem ser encontrados como uma aplicação do Teorema de Dualidade de Lagrange desenvolvido em [8] e discutido em [12].

**Teorema 1.14** (Teorema de Dualidade). *Se  $\mu$  é um ponto interior de  $\{E_P[\mathbf{f}] : P \in \mathcal{A}\}$ , então*

$$\sup_{P \in \mathcal{A}(\mu)} E_P[\phi] = \inf_{\lambda} \{ \lambda^T \mu + \sup_{P \in \mathcal{A}} E_P[\phi - \lambda^T \mathbf{f}] \}.$$

*Além disso, se o problema primário é limitado, o problema dual tem uma solução finita ótima.*

*Demonstração.* Sejam  $\bar{\gamma}$  e  $\underline{\gamma}$  definidos da seguinte forma:

$$\bar{\gamma} \equiv \sup_{P \in \mathcal{A}(\mu)} E_P[\phi] \quad \text{e} \quad \underline{\gamma} \equiv \inf_{\lambda} \{ \lambda^T \mu + \sup_{P \in \mathcal{A}} E_P[\phi - \lambda^T \mathbf{f}] \}.$$

Queremos mostrar que  $\bar{\gamma} = \underline{\gamma}$ . A prova é dividida em dois casos:  $\bar{\gamma} \geq \underline{\gamma}$  e  $\bar{\gamma} \leq \underline{\gamma}$ . Mostraremos o primeiro caso já que o segundo é análogo.

Quando  $\bar{\gamma} = \infty$  a desigualdade é trivial, e por isso assumimos que  $\bar{\gamma} < \infty$ . O conjunto  $\mathcal{M} \equiv \{(E_P[\mathbf{f}], E_P[\phi]) : P \in \mathcal{A}\}$  é um subconjunto convexo de  $\mathbb{R}^{n+2}$  e o ponto  $(\mu, \bar{\gamma})$  é, pela definição de  $\bar{\gamma}$ , um ponto da fronteira de  $\mathcal{M}$ . Assim, existe um hiperplano que tangência  $\mathcal{M}$  no ponto  $(\mu, \bar{\gamma})$ , ou seja, existe um vetor não nulo  $(w, \alpha) \in \mathbb{R}^{n+2}$  tal que

$$w^T \mu + \alpha \bar{\gamma} \geq w^T E_P[\mathbf{f}] + \alpha E_P[\phi], \quad \forall P \in \mathcal{A}. \quad (1.10)$$



Agora mostraremos que  $\alpha > 0$ . Primeiramente para ver que  $\alpha \geq 0$ , considere o ponto  $(\boldsymbol{\mu}, \bar{\gamma} + \epsilon)$ , para  $\epsilon > 0$ . Por causa da definição de  $\bar{\gamma}$ , este ponto está na parte oposta do plano a  $\mathcal{M}$  e assim,  $\boldsymbol{w}^T \boldsymbol{\mu} + \alpha \bar{\gamma} + \alpha \epsilon \geq \boldsymbol{w}^T E_P[\boldsymbol{f}] + \alpha E_P[\phi]$  para todo  $\epsilon > 0$  implicando que  $\alpha \geq 0$ . Segue também que  $\alpha \neq 0$ , pois, caso contrário,  $\boldsymbol{\mu}$  seria um ponto da fronteira  $\{E_P[\boldsymbol{f}] : P \in \mathcal{A}\}$ , o que contradiz nossa hipótese de que  $\boldsymbol{\mu}$  é um ponto interior e portanto  $\alpha > 0$ .

Definindo  $\boldsymbol{\lambda} = -1/\alpha \boldsymbol{w}$ , podemos escrever a expressão (1.10) como

$$-\boldsymbol{\lambda}^T \boldsymbol{\mu} + \bar{\gamma} \geq -\boldsymbol{\lambda}^T E_P[\boldsymbol{f}] + E_P[\phi]$$

para todo  $P \in \mathcal{A}$  e assim,

$$\bar{\gamma} \geq \boldsymbol{\lambda}^T \boldsymbol{\mu} + \sup_{P \in \mathcal{A}} E_P[\phi - \boldsymbol{\lambda}^T \boldsymbol{f}].$$

Desse modo, temos que  $\bar{\gamma} \geq \underline{\gamma}$  e existe  $\boldsymbol{\lambda}$  que atinge  $\underline{\gamma}$  no caso em que  $\bar{\gamma}$  é finito. □

**Teorema 1.15** (Condição Complementar de Relaxamento). *Se  $P$  e  $\boldsymbol{\lambda}$  são soluções ótimas do problema primário e do problema dual respectivamente, então  $P$  alcança o supremo em (1.9) para um  $\boldsymbol{\lambda}$  específico.*

*Demonstração.* Seja  $P \in \mathcal{A}(\boldsymbol{\mu})$ . Assim

$$E_P[\phi] = \boldsymbol{\lambda}^T \boldsymbol{\mu} + E_P[\phi - \boldsymbol{\lambda}^T \boldsymbol{f}].$$

Pelo Teorema da Dualidade, temos que

$$E_P[\phi] = \boldsymbol{\lambda}^T \boldsymbol{\mu} + \sup_{P \in \mathcal{A}} E_P[\phi - \boldsymbol{\lambda}^T \boldsymbol{f}].$$

Assim

$$E_P[\phi] - \boldsymbol{\lambda}^T \boldsymbol{\mu} = \sup_{P \in \mathcal{A}} E_P[\phi - \boldsymbol{\lambda}^T \boldsymbol{f}].$$

□

### 1.3.2 Restrições de Entropia

Outro conjunto de restrições útil exige que as distribuições tenham entropia maior do que qualquer quantidade específica, ou seja, um valor dado que seja do nosso interesse. Para formalizar essa noção de entropia, temos que dadas distribuições  $P_1$  e  $P_2$ , tal que  $P_1$  é absolutamente contínua em relação a  $P_2$ , a entropia de  $P_1$  relativa a  $P_2$  (para mais detalhes a respeito de distribuição absolutamente contínua, consulte Anexo A.4.3). Assim temos

$$H(P_1, P_2) \equiv - \int_X p(x) \ln p(x) dP_2(x),$$

onde  $p(x)$  é a derivada de Radon-Nikodym de  $P_1$  em relação a  $P_2$  e  $0 \ln 0$  é considerado como sendo igual a zero. A derivada de Radon-Nikodym é igual à derivada tradicional que utilizamos quando a função é absolutamente contínua, que é o nosso caso. Para distribuições  $P_1$  que não são absolutamente contínuas com relação a  $P_2$ ,  $H(P_1, P_2)$  é definida como  $-\infty$ . Para os nossos propósitos, basta interpretar  $H(P_1, P_2)$  como uma medida de quanto  $P_1$  difere da distribuição  $P_2$ : se  $P_1$  e  $P_2$  são idênticas, então  $p(x) = 1$  e  $H(P_1, P_2) = 0$ ; caso contrário  $H(P_1, P_2) < 0$ .

No nosso caso, assumimos que  $P_2$  é dada a priori, que é um limitante inferior  $H_0$  na entropia das distribuições permitidas. O problema (1.8) pode então ser reescrito como:

$$\begin{aligned} & \sup_P E_P[\phi] \\ & \text{sujeito a } E_P[\mathbf{f}] \text{ e } H(P, P_2) \geq H_0. \end{aligned}$$

Uma vez que o conjunto de restrições é convexa e a função objetivo é linear, a solução ótima será um limitante no conjunto de restrição e a restrição de entropia assegurará a equação. Incluindo multiplicadores de Lagrange  $\gamma$  para a restrição de entropia ( $\gamma < 0$ ), podemos escrever o problema dual (1.9) como

$$\inf_{\lambda, \gamma} \sup_P \{E_P[\phi] + \boldsymbol{\lambda}^T (\boldsymbol{\mu} - E_P[\mathbf{f}]) + \gamma(H_0 - H(P, P_2))\}. \quad (1.11)$$

Dado que  $H(P, P_2) = -\infty$  para distribuições  $P$  que não são absolutamente contínuas em relação a  $P_2$ , quando se consideram as soluções ótimas para (1.11), podemos assumir que  $P$  é absolutamente contínua em relação a  $P_2$  e nos concentrar no cálculo  $p^*$ , que é derivada Radon-Nikodym. Derivando (1.11) em relação a  $p$  e tomando o resultado igual a zero obtem-se a forma de  $p^*$  como função de  $\boldsymbol{\lambda}$  e  $\gamma$ ,

$$p^*(x; \boldsymbol{\lambda}, \gamma) = \exp \left( -\frac{1}{\gamma} (\phi(x) - \boldsymbol{\lambda}^T \mathbf{f}(x) + \gamma) \right).$$

Substituindo  $p^*$  em (1.11), reduzimos (1.11) a um problema de otimização equivalente a um problema sem restrição

$$\inf_{\lambda, \gamma} \{-E_{P^*}[\phi] + \boldsymbol{\lambda}^T \boldsymbol{\mu} + \gamma H_0\}, \quad (1.12)$$

onde  $P^*$  denota a distribuição correspondente à derivada  $p^*(x; \boldsymbol{\lambda}, \gamma)$  de Radon-Nikodym. O gradiente de (1.12) é dado por  $[(\boldsymbol{\mu} - E_{P^*}[\mathbf{f}]), (H_0 - H(P^*, P_2))]$ . Assim, as condições de primeira ordem para (1.12) exigem que o gradiente, para ser igual a zero, correspondem à factibilidade do problema original (1.9). A matriz hessiana também é facilmente calculada e pode ser demonstrado que é definida positiva (desde que as funções momentos  $f_i$  sejam linearmente independentes com o suporte de  $P_2$ ). Assim o problema de minimização (1.12)

---

pode ser resolvido utilizando uma variação do método de Newton (mantendo  $\gamma < 0$ ) e aproximações numéricas das integrais envolvidas.



---

## Capítulo 2

# Exemplos Estatísticos

Neste capítulo apresentaremos a clássica Desigualdade de Markov e uma demonstração alternativa para a mesma, uma vez que a demonstração clássica deste resultado é fundamentada na Teoria de Probabilidades. Para tal, utilizaremos a teoria desenvolvida ao longo do primeiro Capítulo. Além disso, apresentaremos resultados que resultam desta desigualdade, mostrando um pouco da abrangência da teoria desenvolvida nesse trabalho.

Posteriormente, apresentaremos três problemas clássicos do cálculo de probabilidades: o problema das testemunhas, o problema de *Monty Hall* e um problema relacionado a cadeia de Markov. Tais problemas consistem, respectivamente, em determinar a probabilidade da testemunha falar a verdade, a probabilidade de o prêmio estar atrás de uma determinada porta dado que o apresentador do programa abriu uma outra porta (vazia) e de determinar as probabilidades de uma matriz de transição de uma cadeia de Markov. Quando analisamos esses problemas em seu modo mais simples, isto é, supondo que as probabilidades a priori assumem valores exatos, podemos solucioná-los utilizando ferramentas de programação linear. Contudo, percebemos que as ferramentas estatísticas (mais precisamente da Teoria de Probabilidades) são mais que suficientes para solucioná-los.

Entretanto, principalmente quando atribuímos intervalos às probabilidades a priori dos dois primeiros problemas (no caso do problema de *Monty Hall*, também generalizamos para um número maior de portas), determinar qual a probabilidade de o prêmio estar atrás de uma porta e a de uma testemunha falar a verdade torna-se complexo. É neste cenário que a programação linear torna-se uma ferramenta muito útil para determinar tais probabilidades.

### 2.1 Desigualdades

Nesta seção, apresentaremos as desigualdades de Markov e Chebyshev e a Lei Fraca dos Grandes Números, ambas já amplamente conhecidas na estatística. Entretanto, nosso objetivo, ao apresentar esses resultados, é o de mostrar a aplicabilidade da teoria desenvol-

vida no primeiro capítulo, demonstrando resultados clássicos importantes da estatística matemática.

**Teorema 2.1.** *Seja  $X$  uma variável aleatória não negativa. Então, dado  $E(X) < \infty$  conhecido, temos que*

$$P(X \geq a) \leq \left\{ \min \frac{E(X)}{a}, 1 \right\}, \quad \forall a > 0.$$

*Demonstração.* Para demonstrar a Desigualdade de Markov, utilizaremos a teoria desenvolvida no primeiro capítulo, principalmente as Seções 1.1 e 1.2.

Dado uma variável aleatória não negativa  $X$ , queremos determinar um limitante superior para  $P(X \geq a)$ ,  $a \geq 0$ , sendo que conhecemos  $E(X) = m$ , ou seja,

$$\begin{aligned} & \sup P(X \geq a) \\ & \text{sujeito a } E(X) = m \text{ e } E(1) = 1. \end{aligned}$$

Então, pelo Teorema da Dualidade (Teorema 1.10), podemos utilizar o problema dual para determinar esse limitante superior. Assim, podemos fazer a seguinte formulação:

$$\sup P(X \geq a) = \inf \{E[f(x)] : f(x) \geq I_{(x \geq a)}(x), f(x) \in \mathcal{L}\},$$

sendo  $\mathcal{L}$  definido da seguinte maneira:

$$\mathcal{L} = \{p(x) : p(x) = a_0 + a_1x\}.$$

Aqui, definimos  $\mathcal{L}$  como um polinômio de primeiro grau porque temos o conhecimento de apenas o primeiro momento de  $X$ . Se tivéssemos o conhecimento de  $n$  momentos, o polinômio seria de grau  $n$ .

Assim, o problema dual é determinar para quais valores de  $a_0$  e  $a_1$  minimizamos a seguinte expressão:

$$\begin{aligned} & \inf_{\{a_0, a_1\}} a_0 + a_1m \\ & \text{sujeito a } a_0 + a_1x \geq I_{(x \geq a)}(x). \end{aligned} \tag{2.1}$$

Pelo Teorema de Relaxamento Complementar (Teorema 1.11), dado um polinômio factível, podemos verificar a otimalidade da solução analisando se é possível a construção de uma distribuição factível com densidade restrita aos pontos de  $x$ , onde  $a_0 + a_1x = \phi(x)$ .

Queremos determinar quais são os valores para os coeficientes  $a_0$  e  $a_1$ , de modo que satisfaça o que foi dito anteriormente. Desse modo,  $a_0 = 0$  e  $a_1 = \frac{1}{a}$  são os valores que minimizam (2.1). A visualização gráfica dessa solução se encontra na Figura 2.1.

Portanto, obtemos a seguinte solução:

$$\sup P(X \geq a) = \frac{m}{a}.$$

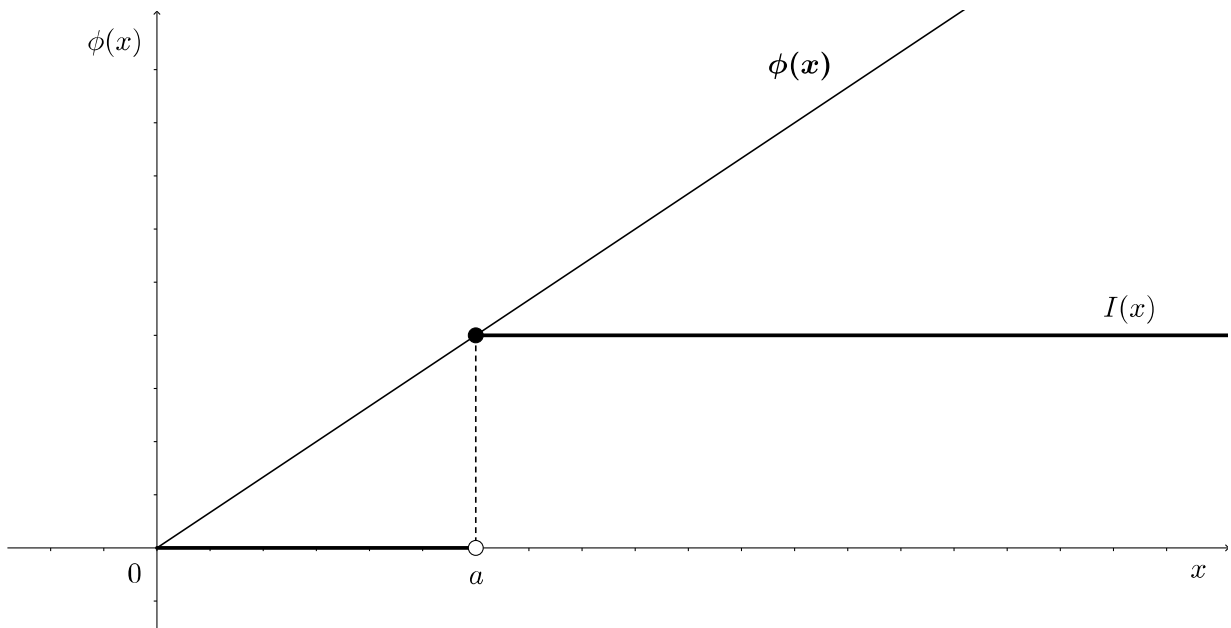


Figura 2.1: Solução para a Desigualdade de Markov.

Além disso, como  $X$  é uma variável aleatória não negativa, podemos ter a seguinte situação:  $P(X \geq a)$ , sendo,  $a > 0$ . Assim, a função indicadora  $I(x)$  terá peso em todos os pontos (Figura 2.2) e, com isso, qualquer função  $\phi(x) \geq 1$  seria uma possível solução para o nosso problema, pois satisfaria o Teorema 1.11. Entretanto, isto contrariaria o axioma da probabilidade, o qual diz que a probabilidade de todo o espaço amostral deve ser um. Para corrigir esse problema, escrevemos a nossa resposta da seguinte maneira:

$$\sup P(X \geq a) = \min \left\{ \frac{m}{a}, 1 \right\}.$$

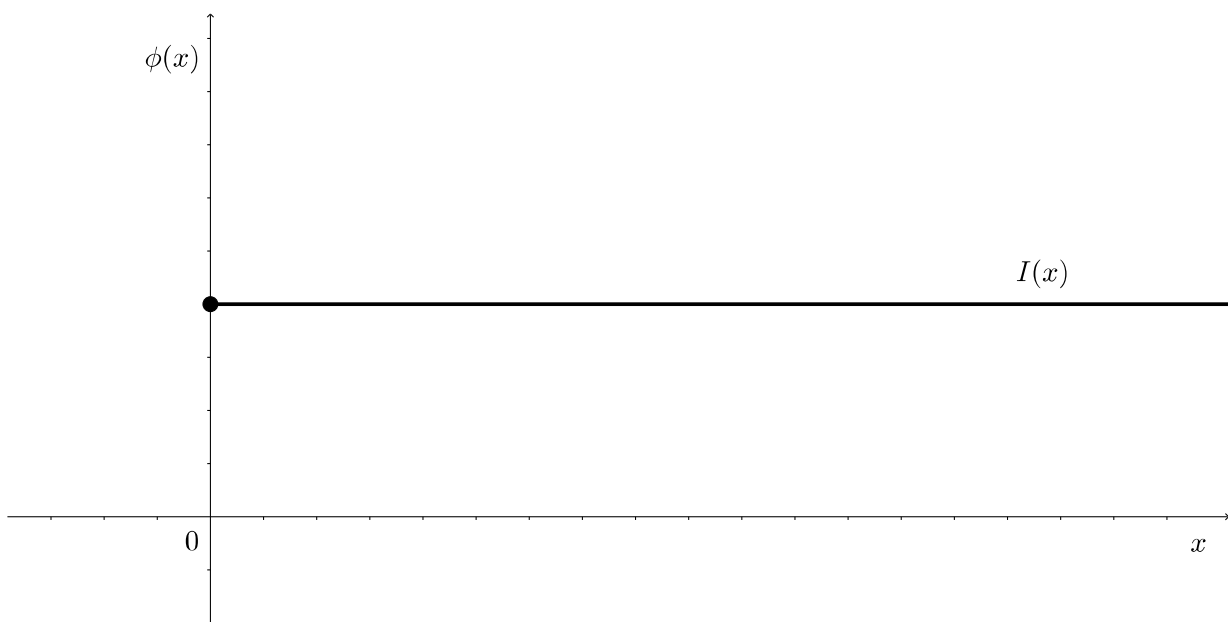


Figura 2.2: Solução para a Desigualdade de Markov.

Portanto, a Desigualdade de Markov é dado por

$$P(X \geq a) = \min \left\{ \frac{E(X)}{a}, 1 \right\}, \quad \forall a \geq 0.$$

Ou a escrita clássica, que é dada por

$$P(X \geq a) \leq \frac{E(X)}{a}, \quad \forall a \geq 0.$$

□

Agora, apresentaremos e demonstraremos a Desigualdade de Chebyshev e a Lei Fraca dos Grandes Números, que são consequências diretas da Desigualdade de Markov.

**Teorema 2.2** (Desigualdade de Chebyshev). *Seja  $X$  uma variável aleatória qualquer e  $a > 0$  uma constante. Então, dados  $E(X) < \infty$  e  $\text{var}(X) < \infty$ , temos*

$$P(|X - E(X)| \geq a) \leq \frac{\text{var}(X)}{a^2}.$$

*Demonstração.* Note que

$$P(|X - E(X)| \geq a) = P(|X - E(X)|^2 \geq a^2).$$

Logo, pela Desigualdade de Markov

$$P(|X - E(X)|^2 \geq a^2) \leq \frac{E[|X - E(X)|^2]}{a^2} = \frac{\text{var}(X)}{a^2}.$$

Portanto,

$$P(|X - E(X)| \geq a) \leq \frac{\text{var}(X)}{a^2}.$$

□

**Teorema 2.3** (Lei Fraca dos Grandes Números). *Seja  $f$  uma função densidade com média  $E(X) < \infty$  e variância  $\text{Var}(X) < \infty$  e seja  $\bar{X}$  a média amostral de uma variável aleatória de tamanho  $n$  com função densidade  $f$ . Sejam  $\epsilon$  e  $\delta$ , com  $\epsilon > 0$  e  $0 < \delta < 1$ . Se  $n$  é qualquer inteiro maior que  $\frac{\text{Var}(X)}{\epsilon^2 \delta}$ , então*

$$P(-\epsilon < \bar{X} - E(X) < \epsilon) \geq 1 - \delta.$$

*Demonstração.* Pela Desigualdade de Markov, temos

$$P(f(X) \geq a) \leq \frac{E(f(X))}{a}, \quad a > 0,$$

para toda variável aleatória  $X$  e toda função não negativa  $f$ . De maneira equivalente, temos



$$P(f(X) < a) \geq 1 - \frac{E(f(X))}{a}, \quad a > 0,$$

Seja  $f(X) = (\bar{X} - E(X))^2$  e  $a = \epsilon^2$ . Então

$$P(-\epsilon < \bar{X} - E(X) < \epsilon) = P(|\bar{X} - E(X)| < \epsilon) = P(|\bar{X} - E(X)|^2 < \epsilon^2).$$

Mas,

$$P(|\bar{X} - E(X)|^2 < \epsilon^2) \geq 1 - \frac{E(\bar{X} - E(X))^2}{\epsilon^2} = 1 - \frac{(1/n)\text{var}(X)}{\epsilon^2} \geq 1 - \delta.$$

Portanto,

$$P(-\epsilon < \bar{X} - E(X) < \epsilon) \geq 1 - \delta.$$

□

## 2.2 Problema de *Monty Hall*

O **problema ou paradoxo de *Monty Hall*** é um problema que surgiu a partir de um concurso televisivo dos Estados Unidos da América chamado *Let's Make a Deal*, exibido na década de 1970.

O jogo consistia no seguinte: Monty Hall (o apresentador do programa) apresentava três portas ao competidor, sabendo que atrás de uma delas estava um bom prêmio (por exemplo, um carro) e que nas demais havia um prêmio de baixo valor. O jogo então tinha as seguintes regras:

1. Na 1ª etapa o concorrente escolhia uma porta (que ainda não é aberta).
2. Em seguida Monty abria uma das outras duas portas que o competidor não havia escolhido, sendo aberta uma porta que não continha o prêmio valioso.
3. Com uma porta aberta, o concorrente tinha que escolher se mudava ou permanecia na porta por ele escolhida inicialmente.

Qual é a estratégia ótima para este problema: ficar com a porta escolhida inicialmente ou mudar de porta? Com qual das duas portas ainda fechadas o concorrente tem mais probabilidade de ganhar? Por quê?

### 2.2.1 Resposta Intuitiva

A resposta intuitiva ao problema aqui apresentado é a de que, quando o apresentador revelou uma porta não premiada, o concorrente teria à frente um novo dilema com apenas

duas portas e um prêmio, e portanto as chances de que o prêmio esteja em qualquer uma das duas portas seriam de 50%. O apresentador teria ajudado o competidor, já que a probabilidade de ganhar o prêmio subiria de  $1/3$  para  $1/2$ . Nesse caso, não faria diferença para o concorrente trocar ou não a porta uma vez que ambas teriam a mesma probabilidade de esconder o prêmio. No entanto, sob certas condições iniciais, esta resposta está errada, pois a porta que o apresentador abre depende da porta que o concorrente escolhe inicialmente. O apresentador sabe desde o começo onde está o prêmio (ele nunca abrirá uma porta premiada). Ao abrir uma porta, ele não está criando um jogo todo novo, mas está dando informações valiosas sobre o jogo original. É por esse motivo que a resposta parece intuitiva, porque parece que o apresentador abriu uma porta aleatoriamente, mas isso não é o que acontece, porque se o concorrente escolher uma porta inicialmente que não contém o prêmio, então o apresentador não tem liberdade de escolha e só pode abrir uma porta.

### 2.2.2 Solução para o problema pelo Teorema de Bayes

Considere os seguintes eventos:

- A: o prêmio está atrás da porta 1.
- B: o prêmio está atrás da porta 2.
- C: o prêmio está atrás da porta 3.
- RA: o apresentador revela o conteúdo vazio da porta 1.
- RB: o apresentador revela o conteúdo vazio da porta 2.
- RC: o apresentador revela o conteúdo vazio da porta 3.

Sem perda de generalidade, vamos supor que o concorrente escolheu inicialmente a porta 1 e a porta vazia aberta pelo apresentador foi a 2. Desse modo, estamos interessados na probabilidade do evento  $A$  dado o evento  $RB$ , ou seja, em  $P(A|RB)$ . Vamos supor, inicialmente as seguintes probabilidade para cada evento:

- $P(A) = P(B) = P(C) = \frac{1}{3}$ ,
- $P(RB|B) = 0$ ,
- $P(RB|C) = 1$ .

Portanto, pelo Teorema de Bayes temos:

$$\begin{aligned}
 P(A|RB) &= \frac{P(RB|A)P(A)}{P(RB|A)P(A) + P(RB|B)P(B) + P(RB|C)P(C)} \\
 &= \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2} \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}} = \frac{1}{3}.
 \end{aligned}$$

Como temos a restrição de que  $P(A|RB) + P(B|RB) + P(C|RB) = 1$  e  $P(B|RB) = 0$ , segue que

$$\begin{aligned}
 P(C|RB) &= 1 - P(A|RB) \\
 &= 1 - 1/3 \\
 &= 2/3.
 \end{aligned}$$

Portanto, a probabilidade do competidor escolher a porta com o prêmio é maior se ele trocar de porta, de acordo com as probabilidades iniciais descritas acima, o que contradiz o senso comum.

### 2.2.3 Solução por Programação Linear

Considere a seguir uma árvore de probabilidade na qual  $X = i$  representa o evento do prêmio estar atrás da porta  $i$ ,  $Y = i$  o evento do concorrente escolher a porta  $i$  e  $Z = i$  a probabilidade do apresentador abrir a porta  $i$ , com  $i = 1, 2, 3$ . De acordo com o que foi dito acima, continuamos supondo que a probabilidade inicial do prêmio estar atrás de cada uma das portas é igual ( $1/3$ ), assim como a probabilidade do concorrente escolher determinada porta.

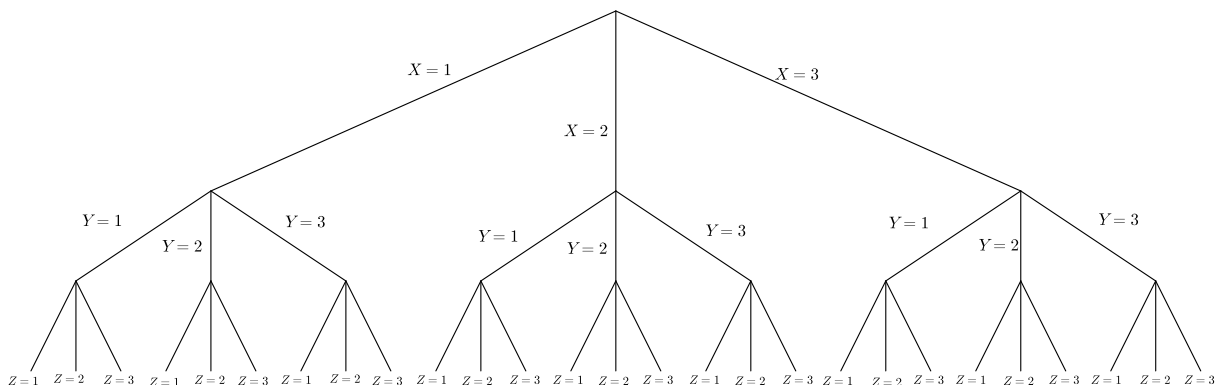


Figura 2.3: Árvore de probabilidade para o problema com três portas.

Com o intuito de simplificar a notação utilizada, considere

- $P(X = 1, Y = 1, Z = 2) = p_1$
- $P(X = 1, Y = 1, Z = 3) = p_2$
- $P(X = 1, Y = 2, Z = 3) = p_3$
- $P(X = 1, Y = 3, Z = 2) = p_4$
- $P(X = 2, Y = 1, Z = 3) = p_5$
- $P(X = 2, Y = 2, Z = 1) = p_6$
- $P(X = 2, Y = 2, Z = 3) = p_7$
- $P(X = 2, Y = 3, Z = 1) = p_8$
- $P(X = 3, Y = 1, Z = 2) = p_9$
- $P(X = 3, Y = 2, Z = 1) = p_{10}$
- $P(X = 3, Y = 3, Z = 1) = p_{11}$
- $P(X = 3, Y = 3, Z = 2) = p_{12}$

**Observação 2.4.** *Os eventos com probabilidade nula, por exemplo,  $P(X = 1, Y = 1, Z = 1) = 0$ , foram desconsiderados para a formulação do problema.*

A partir da árvore de probabilidade da Figura 2.3, é possível extrair um conjunto de restrições. Neste caso necessitamos de doze restrições, pois nosso problema contém doze variáveis,  $p_1, p_2, \dots, p_{12}$ , e sabemos que a solução existe e é única.

Podemos nos perguntar quais restrições estamos procurando? A resposta é: procuramos equações que descrevam cada probabilidade do nosso problema, ou seja, equações que representam a probabilidade do prêmio estar atrás de uma porta, a probabilidade do concorrente escolher uma porta e a probabilidade do apresentador abrir uma das portas dado que o concorrente escolhe outra. Porém, como estamos em uma árvore de probabilidade, esses eventos a partir do segundo nó da árvore, necessariamente serão probabilidades condicionais. A partir disso, obtemos o seguinte conjunto de restrições:

$$\left\{ \begin{array}{l}
 P(X = 1) = p_1 + p_2 + p_3 + p_4 = \frac{1}{3}, \\
 P(X = 2) = p_5 + p_6 + p_7 + p_8 = \frac{1}{3}, \\
 P(Y = 1|X = 1) = \frac{p_1 + p_2}{p_1 + p_2 + p_3 + p_4} = \frac{1}{3} \\
 P(Y = 2|X = 1) = \frac{p_3}{p_1 + p_2 + p_3 + p_4} = \frac{1}{3} \\
 P(Y = 1|X = 2) = \frac{p_5}{p_5 + p_6 + p_7 + p_8} = \frac{1}{3} \\
 P(Y = 2|X = 2) = \frac{p_6 + p_7}{p_5 + p_6 + p_7 + p_8} = \frac{1}{3} \\
 P(Y = 1|X = 3) = \frac{p_9}{p_9 + p_{10} + p_{11} + p_{12}} = \frac{1}{3} \\
 P(Y = 2|X = 3) = \frac{p_{10}}{p_9 + p_{10} + p_{11} + p_{12}} = \frac{1}{3} \\
 P(Z = 2|X = 1, Y = 1) = \frac{p_1}{p_1 + p_2} = \frac{1}{2}, \\
 P(Z = 1|X = 2, Y = 2) = \frac{p_6}{p_6 + p_7} = \frac{1}{2}, \\
 P(Z = 1|X = 3, Y = 3) = \frac{p_{11}}{p_{11} + p_{12}} = \frac{1}{2}, \\
 \sum_{i=1}^{12} p_i = 1, \\
 p_i \geq 0, \quad i = 1, 2, \dots, 12.
 \end{array} \right. \quad (2.2)$$

O conjunto de equações (2.2), contém doze equações além das restrições  $p_i \geq 0$ . A primeira e a segunda linhas representam a probabilidade do prêmio estar atrás das portas 1 ou 2, respectivamente, o que neste caso é  $\frac{1}{3}$ . Entre a terceira e a oitava linhas, representam a probabilidade do concorrente escolher uma das portas, dado que o prêmio esteja atrás de uma das portas. Entre a nona e a décima primeira linhas, representam a probabilidade do apresentador abrir uma das portas, quando o concorrente escolhe a porta em que está o prêmio. Finalmente, a última equação é a restrição para que a soma de todas as probabilidade seja igual a um (esta equação sempre estará presente quando tratamos de um problema de probabilidade).

Realizando algumas manipulações algébricas no conjunto de restrição (2.2), obtemos o seguinte conjunto:



$$\mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \\ p_6 \\ p_7 \\ p_8 \\ p_9 \\ p_{10} \\ p_{11} \\ p_{12} \end{bmatrix}, \mathbf{b} = \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Com a construção do problema realizada, considerando o conjunto de restrição acima, queremos saber qual é a probabilidade de o prêmio estar atrás da porta de número  $X = 3$ , dado que o concorrente escolheu a porta  $Y = 1$  e o apresentador abriu a porta  $Z = 2$ . Resolvendo o sistema matricial, obtemos as seguintes probabilidades para cada evento  $p_i$ .

$$\begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \\ p_6 \\ p_7 \\ p_8 \\ p_9 \\ p_{10} \\ p_{11} \\ p_{12} \end{bmatrix} = \begin{bmatrix} \frac{1}{18} \\ \frac{1}{18} \\ \frac{1}{9} \\ \frac{1}{9} \\ \frac{1}{9} \\ \frac{1}{18} \\ \frac{1}{18} \\ \frac{1}{9} \\ \frac{1}{9} \\ \frac{1}{9} \\ \frac{1}{18} \\ \frac{1}{18} \end{bmatrix}.$$

Portanto,  $P(X = 3|Y = 1, Z = 2) = \frac{p_9}{p_1 + p_9} = \frac{2}{3}$ , que é exatamente o mesmo resultado que encontramos utilizando o Teorema de Bayes.

Mas, por que realizar esta construção de um problema que já tem uma resolução bem definida na Teoria de Probabilidades? Esta construção é importante porque, através dessa abordagem podemos extrapolar o problema, ou seja, conseguimos maximizar e minimizar a probabilidade de se ganhar o prêmio dado que o concorrente já tenha escolhido uma porta e o apresentador aberto uma outra porta. Além disso, podemos também impor intervalos para cada probabilidade do evento, entre outras coisas. Diante isso, podemos agora trabalhar com o seguinte problema:

$$\text{minimize/maximize } P(X = 3|Y = 1, Z = 2) = \frac{p_9}{p_1 + p_9}$$

sujeito, por exemplo, ao seguinte conjunto de restrições.

$$\left\{ \begin{array}{l} p_1 + p_2 + p_3 + p_4 \in [0.1, 0.4], \\ p_5 + p_6 + p_7 + p_8 \in [0.1, 0.4], \\ \frac{2p_1}{3} + \frac{2p_2}{3} - \frac{p_3}{3} - \frac{p_4}{3} \in [0.1, 0.3], \\ -\frac{p_1}{3} - \frac{p_2}{3} + \frac{2p_3}{3} - \frac{p_4}{3} \in [0.1, 0.3], \\ \frac{2p_5}{3} - \frac{p_6}{3} - \frac{p_7}{3} - \frac{p_8}{3} \in [0.1, 0.3], \\ -\frac{p_5}{3} + \frac{2p_6}{3} + \frac{2p_7}{3} - \frac{p_8}{3} \in [0.1, 0.3], \\ \frac{2p_9}{3} - \frac{p_{10}}{3} - \frac{p_{11}}{3} - \frac{p_{12}}{3} \in [0.1, 0.3], \\ -\frac{p_9}{3} + \frac{2p_{10}}{3} + \frac{2p_{11}}{3} - \frac{p_{12}}{3} \in [0.1, 0.3], \\ \frac{p_1}{2} - \frac{p_2}{2} \in [0.3, 0.7], \\ \frac{p_6}{2} - \frac{p_7}{2} \in [0.3, 0.7], \\ \frac{p_{11}}{2} - \frac{p_{12}}{2} \in [0.3, 0.7], \\ \sum_{i=1}^{12} p_i = 1, \\ p_i \geq 0, \quad i = 1, 2, \dots, 12. \end{array} \right. \quad (2.4)$$

Note que, neste problema, nossa função objetivo  $P(X = 3|Y = 1, Z = 2) = \frac{p_9}{p_1 + p_9}$  não é função linear em  $p$ , uma vez que é uma fração. Dessa maneira estamos lidando com um Problema de Programação Linear Fracionária. Entretanto, como nossa função objetivo é o quociente de duas função lineares, é possível transformar o Problema de Programação Linear Fracionária em um Problema de Programação Linear usual (para saber como é realizada está transformação consulte o Anexo A.2).

Primeiramente, resolvendo o problema de minimização, obtemos o seguinte resultado:

$$\min P(X = 3|Y = 1, Z = 2) = 0.1648$$

Agora, resolvendo o problema de maximização, obtemos o seguinte resultado:

$$\max P(X = 3|Y = 1, Z = 2) = 0.8145.$$

**Observação 2.5.** *Este método de construção do problema, que consiste em extrair o conjunto de restrições a partir da árvore de probabilidades será recorrente neste capítulo. Outro fato importante a destacar é de que a quantidade de restrições que necessitamos,*



é exatamente igual ao número de variáveis no problema em questão, se quisermos obter uma solução única.

**Observação 2.6.** Sempre que nossa função objetivo for uma probabilidade condicional, utilizaremos a transformação apresentada no Anexo A.2 para trabalharmos com um Problema de Programação Linear. Isso será fortemente utilizado até o final deste capítulo.

## 2.2.4 Generalização

Suponha que o problema de Monty Hall foi modificado da seguinte forma: ao invés de três, temos quatro portas, por exemplo, e o apresentador não abre todas portas vazias mas somente uma. Diante desse problema, obtemos agora uma árvore de probabilidade bem maior que as anteriores, que pode ser vista na Figura 2.4.

Listaremos a seguir os eventos, sendo o primeiro nó referente à probabilidade do prêmio estar atrás da porta  $i$ , o qual denotaremos por  $X = i$ , o próximo nó  $Y = i$ , que representa probabilidade do concorrente escolher a porta  $i$  e por fim,  $Z = i$  a probabilidade do apresentador abrir a porta  $i$ , sempre com  $i = 1, 2, 3, 4$ .

Seguem os seguintes eventos, com a simplificação de notação que utilizaremos:

$$\begin{array}{l|l}
 P(X = 1, Y = 1, Z = 2) = p_1 & P(X = 3, Y = 1, Z = 2) = p_{19} \\
 P(X = 1, Y = 1, Z = 3) = p_2 & P(X = 3, Y = 1, Z = 4) = p_{20} \\
 P(X = 1, Y = 1, Z = 4) = p_3 & P(X = 3, Y = 2, Z = 1) = p_{21} \\
 P(X = 1, Y = 2, Z = 3) = p_4 & P(X = 3, Y = 2, Z = 4) = p_{22} \\
 P(X = 1, Y = 2, Z = 4) = p_5 & P(X = 3, Y = 3, Z = 1) = p_{23} \\
 P(X = 1, Y = 3, Z = 2) = p_6 & P(X = 3, Y = 3, Z = 2) = p_{24} \\
 P(X = 1, Y = 3, Z = 4) = p_7 & P(X = 3, Y = 3, Z = 4) = p_{25} \\
 P(X = 1, Y = 4, Z = 2) = p_8 & P(X = 3, Y = 4, Z = 1) = p_{26} \\
 P(X = 1, Y = 4, Z = 3) = p_9 & P(X = 3, Y = 4, Z = 2) = p_{27} \\
 P(X = 2, Y = 1, Z = 3) = p_{10} & P(X = 4, Y = 1, Z = 2) = p_{28} \\
 P(X = 2, Y = 1, Z = 4) = p_{11} & P(X = 4, Y = 1, Z = 3) = p_{29} \\
 P(X = 2, Y = 2, Z = 1) = p_{12} & P(X = 4, Y = 2, Z = 1) = p_{30} \\
 P(X = 2, Y = 2, Z = 3) = p_{13} & P(X = 4, Y = 2, Z = 3) = p_{31} \\
 P(X = 2, Y = 2, Z = 4) = p_{14} & P(X = 4, Y = 3, Z = 1) = p_{32} \\
 P(X = 2, Y = 3, Z = 1) = p_{15} & P(X = 4, Y = 3, Z = 2) = p_{33} \\
 P(X = 2, Y = 3, Z = 4) = p_{16} & P(X = 4, Y = 4, Z = 1) = p_{34} \\
 P(X = 2, Y = 4, Z = 1) = p_{17} & P(X = 4, Y = 4, Z = 2) = p_{35} \\
 P(X = 2, Y = 4, Z = 3) = p_{18} & P(X = 4, Y = 4, Z = 3) = p_{36}.
 \end{array}$$

Aqui, vamos supor que as probabilidades em todos os eventos são equiprováveis, ou seja, a probabilidade do prêmio estar atrás de uma das portas é  $\frac{1}{4}$ . Então, é possível obter um conjunto de restrições, assim como foi feito no caso clássico do Problema de *Monty*

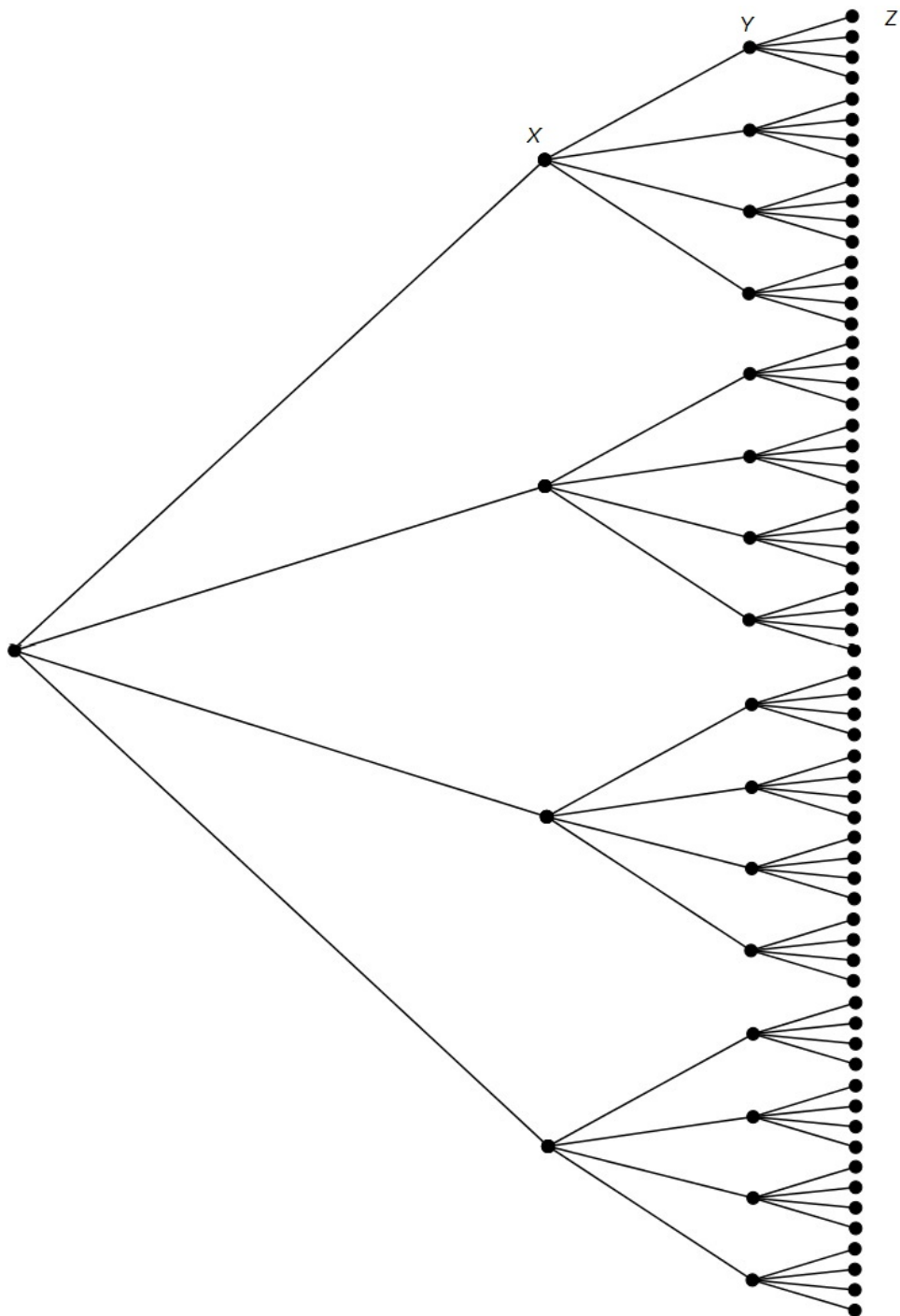


Figura 2.4: Árvore de probabilidade para o problema de Monty Hall com quatro portas.

*Hall.* Logo temos as seguintes restrições:

$$\left\{ \begin{array}{l} p_i \geq 0, \quad i = 1, 2, \dots, 12, \\ \sum_{i=1}^{12} p_i = 1, \\ P(X = 1) = \sum_{i=1}^9 p_i = \frac{1}{4}, \\ P(X = 2) = \sum_{i=10}^{18} p_i = \frac{1}{4}, \\ P(X = 3) = \sum_{i=19}^{27} p_i = \frac{1}{4}. \end{array} \right. \quad (2.5)$$

$$\left\{ \begin{array}{l} P(Y = 1|X = 1) = \frac{p_1 + p_2 + p_3}{\sum_{i=1}^9 p_i} = \frac{1}{4}, \\ P(Y = 2|X = 1) = \frac{p_4 + p_5}{\sum_{i=1}^9 p_i} = \frac{1}{4}, \\ P(Y = 3|X = 1) = \frac{p_6 + p_7}{\sum_{i=1}^9 p_i} = \frac{1}{4}, \\ P(Y = 1|X = 2) = \frac{p_{10} + p_{11}}{\sum_{i=10}^{18} p_i} = \frac{1}{4}, \\ P(Y = 2|X = 2) = \frac{p_{12} + p_{13} + p_{14}}{\sum_{i=10}^{18} p_i} = \frac{1}{4}, \\ P(Y = 3|X = 2) = \frac{p_{15} + p_{16}}{\sum_{i=10}^{18} p_i} = \frac{1}{4}, \\ P(Y = 1|X = 3) = \frac{p_{19} + p_{20}}{\sum_{i=19}^{27} p_i} = \frac{1}{4}, \\ P(Y = 2|X = 3) = \frac{p_{21} + p_{22}}{\sum_{i=19}^{27} p_i} = \frac{1}{4}, \\ P(Y = 3|X = 3) = \frac{p_{23} + p_{24} + p_{25}}{\sum_{i=19}^{27} p_i} = \frac{1}{4}, \\ P(Y = 1|X = 4) = \frac{p_{28} + p_{29}}{\sum_{i=28}^{36} p_i} = \frac{1}{4}, \\ P(Y = 2|X = 4) = \frac{p_{30} + p_{31}}{\sum_{i=28}^{36} p_i} = \frac{1}{4}, \\ P(Y = 3|X = 4) = \frac{p_{32} + p_{33}}{\sum_{i=28}^{36} p_i} = \frac{1}{4}. \end{array} \right. \quad (2.6)$$

$$\left\{ \begin{array}{l}
P(Z = 2|X = 1, Y = 1) = \frac{p_1}{p_1 + p_2 + p_3} = \frac{1}{3}, \\
P(Z = 3|X = 1, Y = 1) = \frac{p_2}{p_1 + p_2 + p_3} = \frac{1}{3}, \\
P(Z = 1|X = 2, Y = 2) = \frac{p_{12}}{p_{12} + p_{13} + p_{14}} = \frac{1}{3}, \\
P(Z = 3|X = 2, Y = 2) = \frac{p_{13}}{p_{12} + p_{13} + p_{14}} = \frac{1}{3}, \\
P(Z = 1|X = 3, Y = 3) = \frac{p_{23}}{p_{23} + p_{24} + p_{25}} = \frac{1}{3}, \\
P(Z = 2|X = 3, Y = 3) = \frac{p_{24}}{p_{23} + p_{24} + p_{25}} = \frac{1}{3}, \\
P(Z = 1|X = 4, Y = 4) = \frac{p_{34}}{p_{34} + p_{35} + p_{36}} = \frac{1}{3}, \\
P(Z = 2|X = 4, Y = 4) = \frac{p_{35}}{p_{34} + p_{35} + p_{36}} = \frac{1}{3}, \\
P(Z = 3|X = 1, Y = 2) = \frac{p_4}{p_4 + p_5} = \frac{1}{2}, \\
P(Z = 2|X = 1, Y = 3) = \frac{p_6}{p_6 + p_7} = \frac{1}{2}, \\
P(Z = 2|X = 1, Y = 4) = \frac{p_8}{p_8 + p_9} = \frac{1}{3}, \\
P(Z = 3|X = 2, Y = 1) = \frac{p_{10}}{p_{10} + p_{11}} = \frac{1}{2}, \\
P(Z = 1|X = 2, Y = 3) = \frac{p_{15}}{p_{15} + p_{16}} = \frac{1}{2}, \\
P(Z = 1|X = 2, Y = 4) = \frac{p_{17}}{p_{17} + p_{18}} = \frac{1}{2}, \\
P(Z = 2|X = 3, Y = 1) = \frac{p_{19}}{p_{19} + p_{20}} = \frac{1}{2}, \\
P(Z = 1|X = 3, Y = 2) = \frac{p_{21}}{p_{21} + p_{22}} = \frac{1}{2}, \\
P(Z = 1|X = 3, Y = 4) = \frac{p_{26}}{p_{26} + p_{27}} = \frac{1}{2}, \\
P(Z = 2|X = 4, Y = 1) = \frac{p_{28}}{p_{28} + p_{29}} = \frac{1}{2}, \\
P(Z = 1|X = 4, Y = 2) = \frac{p_{30}}{p_{30} + p_{31}} = \frac{1}{2}, \\
P(Z = 1|X = 4, Y = 3) = \frac{p_{32}}{p_{32} + p_{33}} = \frac{1}{2}.
\end{array} \right. \quad (2.7)$$

As equações de (2.6) representam a probabilidade do concorrente escolher alguma das quatro portas, sendo essas probabilidades equiprováveis e iguais a  $\frac{1}{4}$ . Ainda, as equações de (2.7) representam a probabilidade do apresentador abrir umas das portas após a escolha do concorrente, sendo estas escolhas equiprováveis também. Entretanto, o valor da probabilidade depende da quantidade de portas que o apresentador pode abrir, dados

os eventos anteriores. Por exemplo, se o prêmio está atrás da porta de número dois e o concorrente escolhe essa porta, o apresentador pode abrir qualquer uma das três portas restantes, e assim, a probabilidade é de  $\frac{1}{3}$ . Por outro lado, se o prêmio está atrás da porta de número dois e o concorrente escolhe a porta de número um, o apresentador poderá abrir apenas as portas de número três ou quatro e, neste caso, a probabilidade é de  $\frac{1}{2}$ .

Feita a construção e a explicação do conjunto de restrições, voltamos nosso foco para a seguinte pergunta: suponha que o concorrente ao prêmio tenha escolhido a porta de número 1 e o apresentador tenha aberto a porta de número 2. O que o concorrente deve fazer: manter-se na porta escolhida originalmente ou mudar de porta? Se ele tiver que mudar de porta, para qual?

Para responder a essa pergunta é preciso construir dois Problemas de Programação Linear Fracionária, os quais resolveremos como Problemas de Programação Linear (PPL) utilizando o Anexo A.2. Desse modo,

$$\text{minimize/maximize } P(X = 3|Y = 1, Z = 2) = \frac{p_{19}}{p_1 + p_{19} + p_{28}}$$

sujeito as restrições (2.5) – (2.7).

Resolvendo o problema acima, concluímos que a probabilidade do prêmio estar atrás da porta de número três é:

$$\text{minimize/maximize } P(X = 3|Y = 1, Z = 2) = 0.375.$$

O segundo PPL é dado pela seguinte expressão:

$$\text{minimize/maximize } P(X = 4|Y = 1, Z = 2) = \frac{p_{28}}{p_1 + p_{19} + p_{28}}$$

sujeito as restrições (2.5) – (2.7).

Resolvendo o problema chegamos na probabilidade do prêmio estar atrás da porta de número quatro é:

$$\text{minimize/maximize } P(X = 4|Y = 1, Z = 2) = 0.375$$

Portanto, o concorrente deve trocar de porta, pois a probabilidade de o prêmio estar atrás da porta três ou quatro é (probabilidade de 0.375) maior que estar na porta um (probabilidade de 0.25). Neste caso, é vantajoso mudar para a porta três ou para a porta quatro.

## 2.3 O Problema das Testemunhas

Pierre Simon Laplace (1749-1827) foi matemático, astrônomo e físico francês. Além dessas grandes áreas da ciência, Laplace também teve bastante interesse na Teoria de

Probabilidades na qual escreveu um famoso texto a respeito de um problema jurídico. Neste texto, seu interesse consistia em determinar a probabilidade de o testemunho de uma pessoa ser verdadeiro. A seguir apresentaremos uma releitura desse texto, trazendo o mesmo para uma linguagem mais moderna e simplificada, do que a utilizada por Laplace.

### 2.3.1 Caso Com Uma Testemunha

Desde os primórdios os julgamentos de qualquer natureza são baseados em testemunho (ou evidência), e por isso é muito importante apresentarmos provas de que este testemunho é verídico. Muitas vezes é difícil calcular a probabilidade do testemunho ser verdadeiro, por conta da dificuldade em estimar a veracidade das testemunhas e também pelo grande número de circunstâncias que acompanham os fatos atestados pela mesma. Apesar disso, em muitos casos, pode-se realizar uma aproximação para a veracidade do testemunho, cujo processo será cuidadosamente deduzido a seguir.

Inicialmente iremos considerar o caso mais simples, em que há apenas uma testemunha no julgamento. No modelo proposto neste trabalho, seguimos as ideias de Laplace [10], que sugere que a probabilidade do testemunho de uma determinada testemunha ser verdadeiro, é composto pela probabilidade de veracidade do testemunho e a possibilidade de erro da testemunha. Para simplificar o problema, imagine que uma bola é retirada de uma urna que contém  $n$  bolas numeradas de 1 a  $n$  e que uma testemunha anuncia que foi sorteada a bola de número  $i$ , tal que,  $1 \leq i \leq n$ . Seja  $p$  a probabilidade do testemunho ser verdadeiro, ou seja, a probabilidade de que a testemunha não está mentindo, e  $r$  a probabilidade de que ela não está enganada a respeito do que ela viu. As retiradas de todas as bolas são consideradas equiprováveis.

Resumindo, temos as seguintes variáveis:

$p$  : Probabilidade de veracidade do testemunho.

$r$  : Probabilidade da testemunha não estar enganada.

$n$  : Número de bolas na urna.

A partir disso, temos as seguintes possibilidades sobre o testemunho:

- 1º) A testemunha não engana e não está enganada, ou seja, a testemunha não mente e as informações que ela tem sobre o fato são verdadeiras.
- 2º) A testemunha não engana e está enganada, ou seja, a testemunha não mente mas as informações que ela tem sobre o fato são falsas.
- 3º) A testemunha engana e não está enganada, ou seja, a testemunha mente e sabe o real número da bola sorteada.

- 4º) A testemunha engana e está enganada, ou seja, a testemunha mente, entretanto está mentindo sobre algo que não é verdade.

Considerando as possibilidades acima citadas para os possíveis testemunhos, podemos construir uma árvore de probabilidade, na qual temos todas essas possibilidades presentes. Para isso, considere os eventos abaixo, em que modelamos o problema da probabilidade do testemunho, como um problema de urnas e bolas descrito anteriormente. Assim, temos os seguintes eventos:

$E_1$  : Número da bola sorteada.

$E_2$  : Número da bola avistada pela testemunha.

$E_3$  : O testemunho é verdadeiro, ou seja, a testemunha não mente.

$E_4$  : A testemunha declara que a bola  $i$  foi sorteada.

A árvore de probabilidade resultante destes quatro eventos é apresentada na Figura 2.5. Nesta figura, cada nó da árvore é a realização de cada um dos eventos  $E_1, E_2, E_3$  e  $E_4$ , com todas as possíveis possibilidades para cada evento. Agora, explicaremos como são compostas as probabilidades  $p_i, i = 1, \dots, 13$  da Figura 2.5.

A probabilidade do testemunho para cada evento é composto da seguinte maneira:

$p_1$  : No primeiro nó temos a probabilidade da bola  $i$  ser sorteada assumida como  $\frac{1}{n}$ , e no segundo temos a probabilidade  $r$  da testemunha ter avistado que a bola  $i$  foi sorteada. No terceiro nó temos a probabilidade  $p$  da testemunha não enganar. Assim, se a testemunha viu a bola  $i$  e seu testemunho é verídico, temos no ultimo nó, com probabilidade 1, que a testemunha anunciará que a bola  $i$  foi sorteada. Portanto, temos a seguinte probabilidade resultante:  $p_1 = \frac{pr}{n}$ . Para  $p_2$  e  $p_3$  a dedução é similar.

$p_4$  e  $p_5$  : As probabilidades  $p_4$  e  $p_5$  são compostas pelas probabilidades da bola  $i$  ser sorteada, entretanto, agora a testemunha está enganada com probabilidade  $1 - r$ , ou seja, ela não viu que a bola  $i$  foi sorteada, ou seja, viu uma bola  $j$ , sendo que  $j \neq i$ . Nesse evento a testemunha pretende enganar e, desse modo, como ela viu uma bola  $j \neq i$ , podem acontecer duas situações no seu testemunho: a primeira situação consiste em ela anunciar involuntariamente que a bola  $i$  foi sorteada, ou seja, como ela estava enganada e mente, precisa escolher uma bola diferente da que ela viu para anunciar e, acidentalmente, pode anunciar que saiu a bola  $i$ , (caso  $p_4$ ). Já a segunda situação difere apenas pelo fato de a testemunha anunciar uma bola diferente de  $i$  e  $j$ , ou seja, ela anuncia uma bola  $h$ , (caso  $p_5$ ). Portanto,  $p_4 = \frac{(1-r)(1-p)}{n(n-1)}$  e

$$p_5 = \frac{(1-r)(1-p)(n-2)}{n(n-1)}.$$

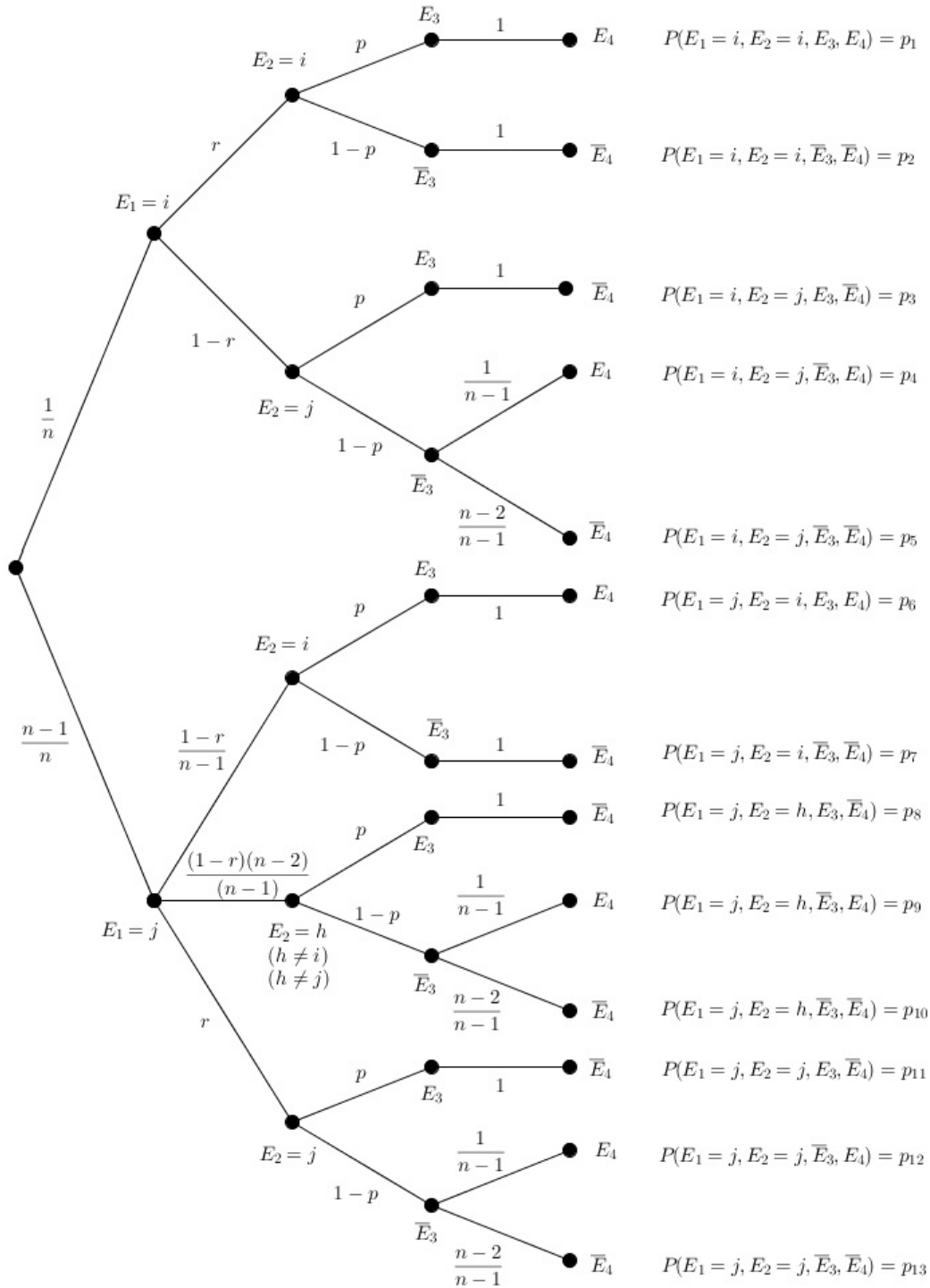


Figura 2.5: Árvore de probabilidade para o caso com uma testemunha.

$p_6, \dots, p_{13}$  : Na parte inferior da árvore estão os casos em que a bola  $i$  não foi sorteada, ou seja, foi sorteada alguma bola  $j$ , sendo  $j \neq i$ . No nó subsequente temos a informação sobre a testemunha estar enganada ou não em relação ao sorteio da bola  $j$ . Se ela está enganada, temos duas opções para o que ela pode ter visto: visto a bola  $i$  ou alguma outra bola  $h$ , sendo  $h \neq j$  e  $h \neq i$ .

Para os nós subsequentes na parte inferior da árvore, o pensamento é análogo aos



mencionados na parte superior da árvore.

Dada a construção dos eventos e as explicações para os mesmos, podemos perguntar qual é a nosso evento de interesse. Queremos saber qual é a probabilidade da bola  $i$  ter sido sorteada, dado que a testemunha declarou que foi sorteada a bola  $i$ . Escrevendo esse evento em notação matemática, temos a seguinte probabilidade

$$P(E_1 = i|E_4) = \frac{P((E_1 = i) \cap E_4)}{P(E_4)} = \frac{p_1 + p_4}{p_1 + p_4 + p_6 + p_9 + p_{12}}.$$

Podemos descobrir a expressão para a probabilidade do testemunho substituindo os valores para cada  $p_i$  presente na fórmula. Entretanto, não conseguimos delimitar intervalos para as probabilidades  $p$  e  $r$  e para o tamanho de  $n$ . Para fazer isso, utilizaremos a teoria de programação linear, uma vez ela permite que encontremos limitantes superior e inferior para  $P(E_1 = i|E_4)$  de acordo com  $p$ ,  $r$  e  $n$ .

Para utilizarmos programação linear é necessário que tenhamos uma função objetivo, no caso,  $P(E_1 = i|E_4)$  e um conjunto de restrições que possibilitem que variemos os valores de  $p$ ,  $r$  e  $n$ , de modo a maximizar e/ou minimizar  $P(E_1 = i|E_4)$ .

É necessário<sup>1</sup> um conjunto de restrições que contenha dezesseis equações, pelo fato que temos dezesseis variáveis (probabilidades) presentes em nossa árvore. O conjunto de restrições será apresentado abaixo e logo em seguida as explicações de cada equação.

### Conjunto de Restrições

- (1)  $P(E_2 = i|E_1 = i) = \frac{p_1 + p_2}{p_1 + p_2 + p_3 + p_4 + p_5} = r,$
- (2)  $P(E_2 = j|E_1 = j) = \frac{p_{11} + p_{12} + p_{13}}{p_6 + p_7 + p_8 + p_9 + p_{10} + p_{11} + p_{12} + p_{13}} = r,$
- (3)  $P(E_2 = i|E_1 = j) = \frac{p_6 + p_7 + p_{13}}{p_6 + p_7 + p_8 + p_9 + p_{10} + p_{11} + p_{12} + p_{13}} = \frac{1 - r}{n - 1},$
- (4)  $P(E_4|E_1 = i, E_2 = j, \bar{E}_3) = \frac{p_4}{p_4 + p_5} = \frac{1}{n - 1},$
- (5)  $P(E_4|E_1 = j, E_2 = h, \bar{E}_3) = \frac{p_9}{p_9 + p_{10}} = \frac{1}{n - 1},$
- (6)  $P(E_4|E_1 = j, E_2 = j, \bar{E}_3) = \frac{p_{12}}{p_{12} + p_{13}} = \frac{1}{n - 1},$
- (7)  $P(E_1) = p_1 + p_2 + p_3 + p_4 + p_5 = \frac{1}{n},$
- (8)  $P(E_3|E_1 = i, E_2 = i) = \frac{p_1}{p_1 + p_2} = p,$
- (9)  $P(E_3|E_1 = i, E_2 = j) = \frac{p_3}{p_3 + p_4 + p_5} = p,$

<sup>1</sup>Porque existe solução e ela é única

$$(10) P(E_3|E_1 = j, E_2 = i) = \frac{p_6}{p_6 + p_7} = p,$$

$$(11) P(E_3|E_1 = j, E_2 = h) = \frac{p_8}{p_8 + p_8 + p_{10}} = p,$$

$$(12) P(E_3|E_1 = j, E_2 = j) = \frac{p_1}{p_1 + p_2} = p,$$

$$(13) \sum_{i=1}^{13} p_i = 1.$$

A interpretação do conjunto de restrições acima será apresentado a seguir.

(1), (2) → Essas probabilidades são iguais a  $r$ .

(3) → Equação que descreve a probabilidade da testemunha avistar a bola  $i$ , dado que foi sorteada a bola  $j$ .

(4), (5), (6) → Equações que descrevem a probabilidade da testemunha anunciar a bola  $i$ , dado que ela pretende enganar.

(7) → Probabilidade de sortear a bola  $i$ .

(8), ..., (12) → Essas probabilidades são iguais a  $p$ .

(13) → Restrição para que a soma de todas as probabilidades seja 1.

### Problema de Programação Linear (PPL)

Dado o conjunto de restrições apresentado anteriormente, temos o seguinte problema de otimização:

$$\begin{aligned} &\text{minimize/maximize } P(E_1 = i|E_4) = \frac{p_1 + p_4}{p_1 + p_4 + p_6 + p_9 + p_{12}} \\ &\text{sujeito a (1) – (13).} \end{aligned}$$

Podemos reescrever nosso problema de otimização na forma matricial, uma vez que desta maneira, é possível atribuir intervalo para cada variável. Assim, temos

$$P(E_1 = i|E_4) = \frac{p_1 + p_4}{p_1 + p_4 + p_6 + p_9 + p_{12}}$$

sujeito a  $\mathbf{Ap} = \mathbf{b}$ , sendo

$$\mathbf{A} = \begin{bmatrix} 1-r & 1-r & -r & -r & -r & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -r & -r & -r & -r & -r & 1-r & 1-r & 1-r & 1-r \\ 0 & 0 & 0 & 0 & 0 & \frac{n-2+r}{n-1} & \frac{n-2+r}{n-1} & \frac{r-1}{n-1} & \frac{r-1}{n-1} & \frac{r-1}{n-1} & \frac{r-1}{n-1} & \frac{r-1}{n-1} & \frac{r-1}{n-1} & \frac{r-1}{n-1} \\ 0 & 0 & 0 & \frac{n-2}{n-1} & \frac{-1}{n-1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{n-2}{n-1} & \frac{-1}{n-1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{n-2}{n-1} & \frac{-1}{n-1} & \frac{-1}{n-1} \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1-p & -p & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1-p & -p & -p & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1-p & -p & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-p & -p & -p & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-p & -p & -p & -p \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix},$$

e

$$\mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \\ p_6 \\ p_7 \\ p_8 \\ p_9 \\ p_{10} \\ p_{11} \\ p_{12} \\ p_{13} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \frac{1}{n} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Perceba que nossa função objetivo não é linear em  $p$  porque é uma fração composta pelos  $p_i$ 's. Desse modo, temos um Problema de Programação Linear Fracionária. Entretanto, como nossa função objetivo é o quociente de duas funções lineares em  $p$  é possível transformar o Problema de Programação Fracionária em um Problema de Programação Linear. Para detalhes de como é realizada esta transformação, consulte o Anexo A.2.

**Observação 2.7.** *Como é possível transformar nosso problema fracionário em um problema linear, chamaremos sempre nosso problema como Problema de Programação Linear.*

Resolvendo o problema de programação linear analiticamente, obtemos

$$P(E_1 = i|E_4) = \frac{p_1 + p_4}{p_1 + p_4 + p_6 + p_9 + p_{12}} = pr + \frac{(1-p)(1-r)}{n-1}$$

Se  $r$  é igual a 1, ou seja, a testemunha não está enganando, a probabilidade de sair a bola de número  $i$  será de  $p$ , que é a probabilidade de veracidade da testemunha.

Se  $n \rightarrow \infty$ , esta probabilidade convergirá para  $pr$ . Isto quer dizer que quando tivermos muitas bolas em nossa urna, a probabilidade resultante dependerá fortemente da veracidade da testemunha e de ela não estar enganada.

### Probabilidade Máxima e Mínima

O interessante da utilização da programação linear, como dito anteriormente, é de que podemos colocar intervalos para as equações, justamente o que faremos agora. Queremos que nossa variável  $p$  variem da seguinte forma, com  $n$  e  $r$  fixos:

$$- p \in [0.3, 0.7],$$

$$- r = 0.3,$$

$$- n = 11$$

Aplicando esses intervalos para o nosso conjunto de restrições, obtemos o seguinte novo conjunto de restrições:

$$(1) P(E_2 = i|E_1 = i) = \frac{p_1 + p_2}{p_1 + p_2 + p_3 + p_4 + p_5} = 0.3,$$

$$(2) P(E_2 = j|E_1 = j) = \frac{p_{11} + p_{12} + p_{13}}{p_6 + p_7 + p_8 + p_9 + p_{10} + p_{11} + p_{12} + p_{13}} = 0.3,$$

$$(3) P(E_2 = i|E_1 = j) = \frac{p_6 + p_7 + p_{13}}{p_6 + p_7 + p_8 + p_9 + p_{10} + p_{11} + p_{12} + p_{13}} = 0.07,$$

$$(4) P(E_4|E_1 = i, E_2 = j, \bar{E}_3) = \frac{p_4}{p_4 + p_5} = 0.1,$$

$$(5) P(E_4|E_1 = j, E_2 = h, \bar{E}_3) = \frac{p_9}{p_9 + p_{10}} = 0.1,$$

$$(6) P(E_4|E_1 = j, E_2 = j, \bar{E}_3) = \frac{p_{12}}{p_{12} + p_{13}} = 0.1,$$

$$(7) P(E_1) = p_1 + p_2 + p_3 + p_4 + p_5 = 0.0909$$

$$(8) P(E_3|E_1 = i, E_2 = i) = \frac{p_1}{p_1 + p_2} \in [0.3, 0.7],$$

$$(9) P(E_3|E_1 = i, E_2 = j) = \frac{p_3}{p_3 + p_4 + p_5} \in [0.3, 0.7],$$

$$(10) P(E_3|E_1 = j, E_2 = i) = \frac{p_6}{p_6 + p_7} \in [0.3, 0.7],$$

$$(11) \quad P(E_3|E_1 = j, E_2 = h) = \frac{p_8}{p_8 + p_8 + p_{10}} \in [0.3, 0.7],$$

$$(12) \quad P(E_3|E_1 = j, E_2 = j) = \frac{p_1}{p_1 + p_2} \in [0.3, 0.7],$$

$$(13) \quad \sum_{i=1}^{13} p_i = 1.$$

Assim, temos o seguinte PPL:

$$\begin{aligned} &\text{minimize/maximize } P(E_1 = i|E_4) = \frac{p_1 + p_4}{p_1 + p_4 + p_6 + p_9 + p_{12}} \\ &\text{sujeito a (1) – (13).} \end{aligned}$$

Resolvendo nosso PPL, obtemos os seguintes resultados:

$$\rightarrow \text{Mínimo } P(E_1 = i|E_4) = 0.2635.$$

$$\rightarrow \text{Máximo } P(E_1 = i|E_4) = 0.6608.$$

### 2.3.2 Caso Com Duas Testemunhas

O objetivo desta seção é o de generalizar o problema de determinar a probabilidade do testemunho para o caso com duas testemunhas. Para tal, vamos simplificar um pouco nossas hipóteses, fazendo uma analogia com urnas que contém bolas, agora coloridas. Apresentaremos uma nova abordagem para o caso com uma testemunha, para logo depois ampliarmos para o caso com duas testemunhas.

#### Abordagem com uma testemunha

Considere que tenhamos um urna contendo  $n - 1$  bolas pretas e um bola branca. Considere ainda que tenha sido extraída a única bola branca da urna, e que uma testemunha anuncia que ela foi sorteada. Queremos determinar a probabilidade desta extração, ou seja, queremos determinar a probabilidade da bola branca ter sido sorteada, dada que a testemunha anunciou que ela foi sorteada. Dessa maneira, podemos listar os seguintes eventos:

$E_1$  : Uma bola branca foi retirada.

$E_2$  : A testemunha não se engana.

$E_3$  : A testemunha é verdadeira.

$E_4$  : A testemunha anuncia que uma bola branca foi retirada.

**Observação 2.8.** *As notações  $\overline{E}_1$  e  $\overline{E}_2$ , representam respectivamente que uma bola não branca foi sorteada e que a testemunha se engana.*

A partir da definição desses quatro eventos, é possível construir uma árvore de probabilidade, representada na Figura 2.6.

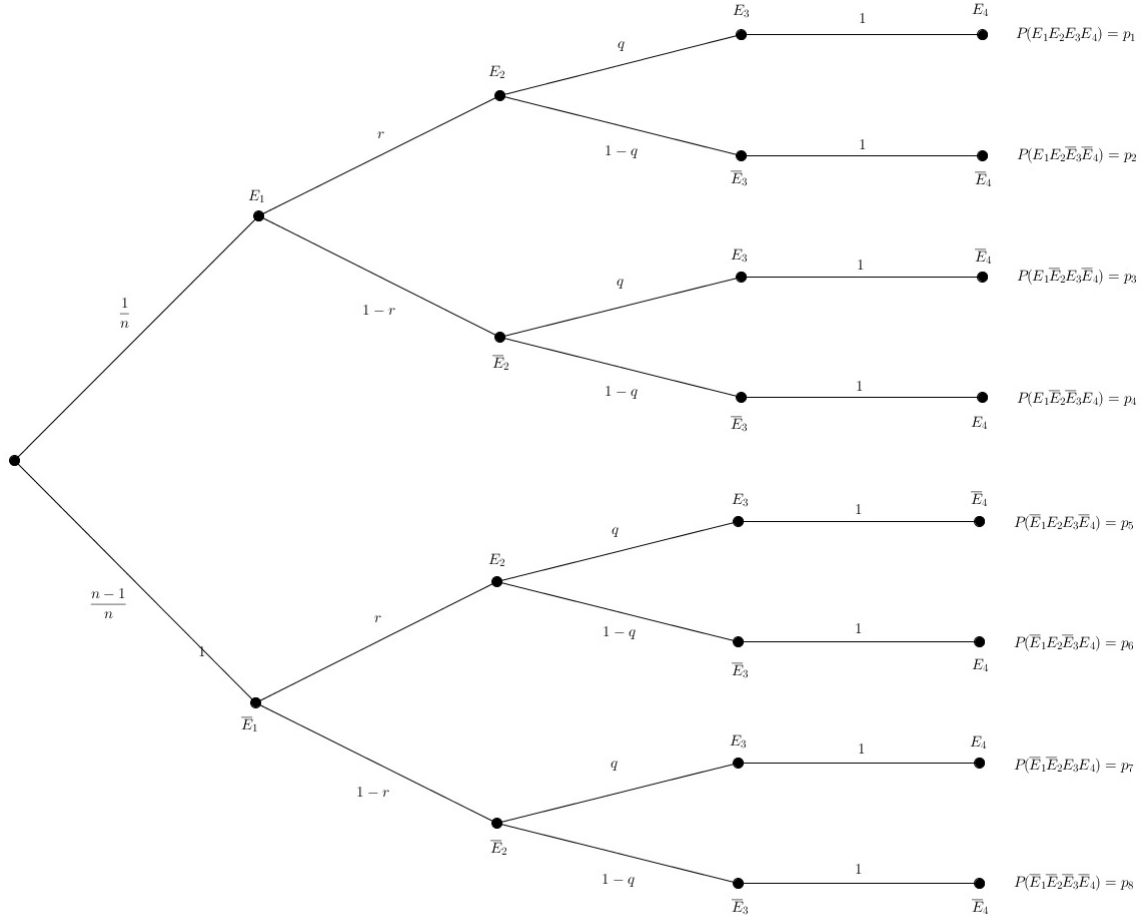


Figura 2.6: Árvore de probabilidades para uma testemunha.

Assim como fizemos na seção anterior, através da árvore de probabilidade, é possível deduzirmos um conjunto de restrições de modo a obtermos um problema de programação fracionária, que, utilizando a transformação adequada, conseguimos tratar como um problema de programação linear. Desse modo, temos o seguinte conjunto de restrições:

$$(1) \sum_{i=1}^8 p_i = 1,$$

$$(2) P(E_1) = p_1 + p_2 + p_3 + p_4 = \frac{1}{n},$$

$$(3) P(E_2|E_1) = \frac{p_1 + p_2}{p_1 + p_2 + p_3 + p_4} = r,$$

$$(4) P(E_2|\bar{E}_1) = \frac{p_5 + p_6}{p_5 + p_6 + p_7 + p_8} = r,$$

$$(5) P(E_3|E_1E_2) = \frac{p_1}{p_1 + p_2} = p,$$

$$(6) P(E_3|E_1\bar{E}_2) = \frac{p_3}{p_3 + p_4} = p,$$

$$(7) P(E_3|\bar{E}_1E_2) = \frac{p_5}{p_5 + p_6} = p,$$

$$(8) P(E_3|\bar{E}_1\bar{E}_2) = \frac{p_7}{p_7 + p_8} = p.$$

$$\begin{aligned} \text{minimize/maximize } P(E_1|E_4) &= \frac{p_1 + p_4}{p_1 + p_4 + p_6 + p_7} \\ \text{sujeito a } \mathbf{A}\mathbf{p} &= \mathbf{b}. \end{aligned}$$

onde,

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1-r & 1-r & -r & -r & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1-r & 1-r & -r & -r \\ 1-p & -p & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1-p & -p & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1-p & -p & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1-p & -p \end{bmatrix}, \mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \\ p_6 \\ p_7 \\ p_8 \end{bmatrix} \text{ e } \mathbf{b} = \begin{bmatrix} 1 \\ \frac{1}{n} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Resolvendo o problema analiticamente obtemos uma expressão algébrica que nos permite calcular a probabilidade do testemunho. Essa expressão tem a seguinte forma:

$$P(E_1|E_4) = \frac{p_1 + p_4}{p_1 + p_4 + p_6 + p_7} = \frac{pr + (1-r)(1-p)}{pr + (1-r)(1-p) + (n-1)(1-p)r + (n-1)(1-r)p},$$

sendo

$$q = pr + (1-r)(1-p) \text{ e } 1-q = p(1-r) + (1-p)r,$$

obtemos,

$$P(E_1|E_4) = \frac{q}{q + (n-1)(1-q)}.$$

### Probabilidade Máxima e Mínima

Como realizamos para o caso problema anterior, podemos colocar intervalos para as equações, e será justamente isso o que faremos agora. Queremos que nossas variáveis  $p$ ,  $r$  e  $n$  variem da seguinte forma:

$$- p \in [0.4, 0.8],$$

$$- r \in [0.3, 0.7],$$

$$- n \in [10, 15].$$

Aplicando esses intervalos para o nosso conjunto de restrições, obtemos o seguinte novo conjunto de restrições:

$$(1) \sum_{i=1}^8 p_i = 1,$$

$$(2) P(E_1) = p_1 + p_2 + p_3 + p_4 \in [10, 15],$$

$$(3) P(E_2|E_1) = \frac{p_1 + p_2}{p_1 + p_2 + p_3 + p_4} \in [0.3, 0.7],$$

$$(4) P(E_2|\bar{E}_1) = \frac{p_5 + p_6}{p_5 + p_6 + p_7 + p_8} \in [0.3, 0.7],$$

$$(5) P(E_3|E_1E_2) = \frac{p_1}{p_1 + p_2} \in [0.4, 0.8],$$

$$(6) P(E_3|E_1\bar{E}_2) = \frac{p_3}{p_3 + p_4} \in [0.4, 0.8],$$

$$(7) P(E_3|\bar{E}_1E_2) = \frac{p_5}{p_5 + p_6} \in [0.4, 0.8],$$

$$(8) P(E_3|\bar{E}_1\bar{E}_2) = \frac{p_7}{p_7 + p_8} \in [0.4, 0.8].$$

Assim, temos o seguinte PPL:

$$\begin{aligned} &\text{minimize/maximize } P(E_1|E_4) = \frac{p_1 + p_4}{p_1 + p_4 + p_6 + p_7} \\ &\text{sujeito a (1) - (8).} \end{aligned}$$

Resolvendo nosso PPL, obtemos os seguintes resultados:

$$\rightarrow \text{Mínimo } P(E_1|E_4) = 0.1659.$$

$$\rightarrow \text{Máximo } P(E_1|E_4) = 0.7610.$$

### Abordagem com duas testemunhas

Agora, utilizando a mesma ideia de bolas coloridas em urnas, queremos generalizar nosso problema para o caso em que temos duas testemunhas. Para fazermos isso, simplificaremos um pouco nosso problema, ao invés de termos duas probabilidades ( $r$  e  $p$ ), teremos apenas a probabilidade da testemunha anunciar que a bola branca foi retirada, que será representado por  $q$  para a primeira testemunha e  $q'$  para a segunda.

Para tal, considere que temos duas urnas,  $A$  e  $B$ , em que a urna  $A$  contém  $n$  bolas brancas e a urna  $B$  contém  $n$  bolas pretas. Em seguida, retiramos uma bola de uma das urnas e substituímos por outra bola da outra urna. Feito isso, obtemos os seguintes eventos:



$E_1$  : Uma bola branca é sorteada na primeira retirada da urna  $A$ .

$T_1$  : A primeira testemunha afirma que a primeira bola retirada foi branca.

$E_2$  : Uma bola branca é sorteada na segunda retirada da urna  $B$ .

$T_2$  : A primeira testemunha afirma que a segunda bola retirada foi branca.

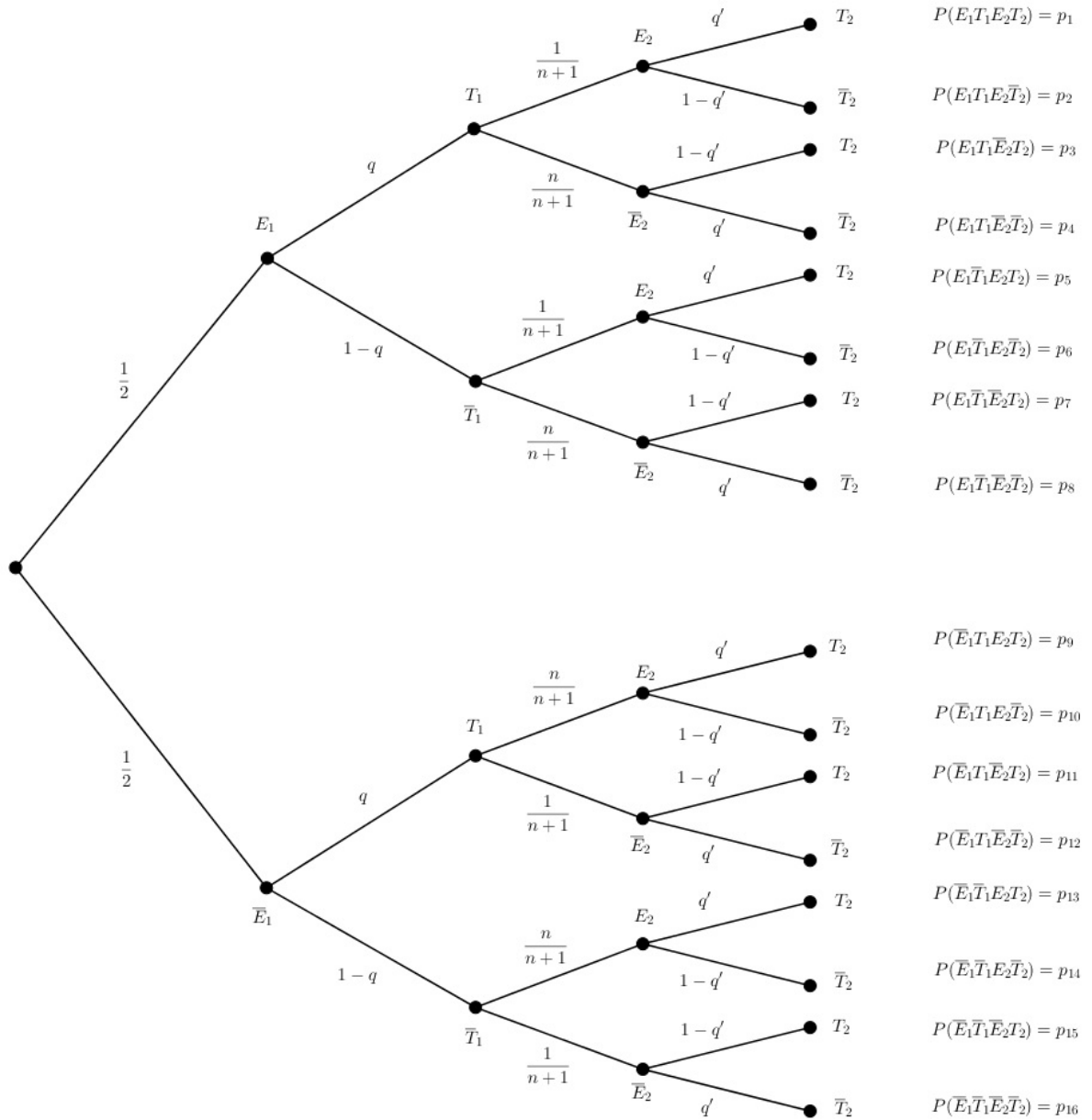


Figura 2.7: Árvore de probabilidades para duas testemunhas.

A seguir apresentaremos o conjunto de restrições que extraímos através da árvore de probabilidades, usando a mesma lógica dos problemas anterior, a única diferença aqui se dá com o intuito de simplificar a notação, para a escrita de maneira matricial denotaremos  $1 - q = r$  e  $1 - q' = r'$ . Sendo assim, temos o seguinte conjunto de restrições:

$$(1) \sum_{i=1}^{16} p_i = 1,$$

$$(2) P(E_1) = \sum_{i=1}^8 p_i = \frac{1}{2},$$

$$(3) P(T_1|E_1) = \frac{p_1 + p_2 + p_3 + p_4}{p_1 + p_2 + p_3 + p_4 + p_5 + p_6 + p_7 + p_8} = q,$$

$$(4) P(T_1|\bar{E}_1) = \frac{p_9 + p_{10} + p_{11} + p_{12}}{p_9 + p_{10} + p_{11} + p_{12} + p_{13} + p_{14} + p_{15} + p_{16}} = 1 - q = r,$$

$$(5) P(E_2|E_1T_1) = \frac{p_1 + p_2}{p_1 + p_2 + p_3 + p_4} = \frac{1}{n+1},$$

$$(6) P(E_2|E_1\bar{T}_1) = \frac{p_5 + p_6}{p_5 + p_6 + p_7 + p_8} = \frac{1}{n+1},$$

$$(7) P(E_2|\bar{E}_1T_1) = \frac{p_9 + p_{10}}{p_9 + p_{10} + p_{11} + p_{12}} = \frac{n}{n+1},$$

$$(8) P(E_2|\bar{E}_1\bar{T}_1) = \frac{p_{13} + p_{14}}{p_{13} + p_{14} + p_{15} + p_{16}} = \frac{n}{n+1},$$

$$(9) P(T_2|E_1T_1E_2) = \frac{p_1}{p_1 + p_2} = q',$$

$$(10) P(T_2|E_1T_1\bar{E}_2) = \frac{p_3}{p_3 + p_4} = 1 - q' = r',$$

$$(11) P(T_2|E_1\bar{T}_1E_2) = \frac{p_5}{p_5 + p_6} = q',$$

$$(12) P(T_2|E_1\bar{T}_1\bar{E}_2) = \frac{p_7}{p_7 + p_8} = 1 - q' = r',$$

$$(13) P(T_2|\bar{E}_1T_1E_2) = \frac{p_9}{p_9 + p_{10}} = q',$$

$$(14) P(T_2|\bar{E}_1T_1\bar{E}_2) = \frac{p_{11}}{p_{11} + p_{12}} = 1 - q' = r',$$

$$(15) P(T_2|\bar{E}_1\bar{T}_1E_2) = \frac{p_{13}}{p_{13} + p_{14}} = q',$$

$$(16) P(T_2|\bar{E}_1\bar{T}_1\bar{E}_2) = \frac{p_{15}}{p_{15} + p_{16}} = 1 - q' = r'.$$

Considerando esse conjunto de restrições, podemos formular o seguinte PPL:

$$\text{minimize/maximize } P(E_1E_2|T_1T_2) = \frac{p_1}{p_1 + p_3 + p_9 + p_{11}}$$

sujeito a  $\mathbf{Ap} = \mathbf{b}$ .

onde,

$$\mathbf{A} = \begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1-q & 1-q & 1-q & 1-q & -q & -q & -q & -q & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-r & 1-r & 1-r & 1-r & -r & -r & -r & -r \\
\frac{n}{n+1} & \frac{n}{n+1} & \frac{-1}{n+1} & \frac{-1}{n+1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \frac{n}{n+1} & \frac{n}{n+1} & \frac{-1}{n+1} & \frac{-1}{n+1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{n+1} & \frac{1}{n+1} & \frac{-n}{n+1} & \frac{-n}{n+1} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{n+1} & \frac{1}{n+1} & \frac{-n}{n+1} & \frac{-n}{n+1} \\
1-q' & -q' & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1-r' & -r' & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1-q' & -q' & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1-r' & -r' & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-q' & -q' & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-r' & -r' & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-q' & -q' & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-r' & -r'
\end{bmatrix}$$

$$\mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \\ p_6 \\ p_7 \\ p_8 \\ p_9 \\ p_{10} \\ p_{11} \\ p_{12} \\ p_{13} \\ p_{14} \\ p_{15} \\ p_{16} \end{bmatrix} \quad \text{e} \quad \mathbf{b} = \begin{bmatrix} 1 \\ \frac{1}{2} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Resolvendo o problema analiticamente, obtemos uma expressão algébrica que nos permite calcular a probabilidade do testemunho. Essa expressão tem a seguinte forma:

$$P(E_1 E_2 | T_1 T_2) = \frac{qq'}{qq' + (1-q)(1-q') + [q(1-q') + q'(1-q)]n}.$$

### Probabilidade Máxima e Mínima

Como realizamos para o caso problema anterior, podemos colocar intervalos para as equações, e será justamente isso o que faremos agora. Queremos que nossas variáveis  $q, r'$ , com  $n$  fixo, variem da seguinte forma:

- $q \in [0.4, 0.8]$ ,
- $q' \in [0.2, 0.6]$ ,
- $n = 20$ .

Aplicando esses intervalos para o nosso conjunto de restrições, obtemos o seguinte novo conjunto de restrições:

$$(1) \sum_{i=1}^{16} p_i = 1,$$

$$(2) P(E_1) = \sum_{i=1}^8 p_i = \frac{1}{2},$$

$$(3) P(T_1|E_1) = \frac{p_1 + p_2 + p_3 + p_4}{p_1 + p_2 + p_3 + p_4 + p_5 + p_6 + p_7 + p_8} \in [0.4, 0.8],$$

$$(4) P(T_1|\bar{E}_1) = \frac{p_9 + p_{10} + p_{11} + p_{12}}{p_9 + p_{10} + p_{11} + p_{12} + p_{13} + p_{14} + p_{15} + p_{16}} \in [0.2, 0.6],$$

$$(5) P(E_2|E_1T_1) = \frac{p_1 + p_2}{p_1 + p_2 + p_3 + p_4} = \frac{1}{21},$$

$$(6) P(E_2|E_1\bar{T}_1) = \frac{p_5 + p_6}{p_5 + p_6 + p_7 + p_8} = \frac{1}{21},$$

$$(7) P(E_2|\bar{E}_1T_1) = \frac{p_9 + p_{10}}{p_9 + p_{10} + p_{11} + p_{12}} = \frac{20}{21},$$

$$(8) P(E_2|\bar{E}_1\bar{T}_1) = \frac{p_{13} + p_{14}}{p_{13} + p_{14} + p_{15} + p_{16}} = \frac{20}{21},$$

$$(9) P(T_2|E_1T_1E_2) = \frac{p_1}{p_1 + p_2} \in [0.2, 0.7],$$

$$(10) P(B_2|E_1B_1\bar{E}_2) = \frac{p_3}{p_3 + p_4} \in [0.3, 0.8],$$

$$(11) P(T_2|E_1\bar{T}_1E_2) = \frac{p_5}{p_5 + p_6} \in [0.2, 0.7],$$

$$(12) P(T_2|E_1\bar{T}_1\bar{E}_2) = \frac{p_7}{p_7 + p_8} \in [0.3, 0.8],$$

$$(13) P(T_2|\bar{E}_1T_1E_2) = \frac{p_9}{p_9 + p_{10}} \in [0.2, 0.7],$$

$$(14) P(T_2|\bar{E}_1\bar{T}_1\bar{E}_2) = \frac{p_{11}}{p_{11} + p_{12}} \in [0.3, 0.8],$$

$$(15) P(T_2|\bar{E}_1\bar{T}_1E_2) = \frac{p_{13}}{p_{13} + p_{14}} \in [0.2, 0.7],$$

$$(16) P(T_2|\bar{E}_1\bar{T}_1\bar{E}_2) = \frac{p_{15}}{p_{15} + p_{16}} \in [0.3, 0.8].$$

Assim, temos o seguinte PPL:

$$\begin{aligned} & \text{minimize/maximize } P(E_1E_2|T_1T_2) = \frac{p_1}{p_1 + p_3 + p_9 + p_{11}} \\ & \text{sujeito a (1) - (16).} \end{aligned}$$

**Observação 2.9.** Note que os intervalos para  $r$  e  $r'$  foram colocados apenas para visualizar o intervalo em que essas variáveis variam, entretanto no momento de simular essas restrições, ela depende da variação de  $q$  e  $q'$ , uma vez que  $r = 1 - q$  e  $r' = 1 - q'$ .

Resolvendo nosso PPL, obtemos os seguintes resultados:

→ Mínimo  $P(E_1E_2|T_1T_2) = 0.2167$ .

→ Máximo  $PP(E_1E_2|T_1T_2) = 0.6696$ .

## 2.4 Cadeia de Markov

Na Estatística, Cadeia de Markov é um tópico bastante importante, seja em uma área mais teórica, como Processos Estocásticos, ou então em áreas mais computacionais, como no famoso método *Markov Chain Monte Carlo* (MCMC), muito utilizado, por exemplo, para simular valores de variáveis aleatórias com certas distribuições.

Podemos aplicar a teoria apresentada no primeiro capítulo para determinar probabilidades para eventos de interesse, para os casos em que temos restrições nas probabilidades de transição de uma cadeia. Sendo assim, considere algumas definições que utilizaremos na composição desse exemplo.

Descrevemos uma **cadeia de Markov** da seguinte maneira: seja  $S = \{s_1, s_2, \dots, s_r\}$  um conjunto de estados. O processo começa em um desses estados e move-se sucessivamente de um estado para outro. Cada movimento é chamado de passo. Se a cadeia está atualmente no estado  $s_i$ , então ela se move para o estado  $s_j$  no próximo passo com uma probabilidade denotada por  $s_{ij}$ , e essa probabilidade não depende dos estados ocorridos nos passos anteriores, apenas do estado atual.

A probabilidade  $p_{ij}$  é chamada **probabilidade de transição** do estado  $i$  para o estado  $j$ . O processo pode permanecer no estado que se encontra e isso ocorre com probabilidade  $p_{ii}$ .

**Definição 2.10.** *Considere uma cadeia de Markov com estados  $1, 2, \dots, N$ . Seja  $p_{ij}$  a probabilidade de transição do estado  $i$  para o estado  $j$ . Então a matriz  $P_{N \times N}$  com entradas  $p_{ij}$  denomina-se matriz de transição da cadeia de Markov.*

Agora, considere a cadeia de Markov com três estados, representada graficamente na Figura 2.8, que também pode ser representada pela seguinte matriz de transição:

$$\begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix}.$$

Aqui, não sabemos quais são os valores das probabilidades  $p_{ij}$ , com  $i = 1, 2, 3$  e  $j = 1, 2, 3$ . As informações que temos a respeito dessas probabilidades são algumas restrições, que são dadas a seguir:

$$(1) \quad p_{11} + p_{12} + p_{13} = 1,$$

$$(2) \quad p_{21} + p_{22} + p_{23} = 1,$$

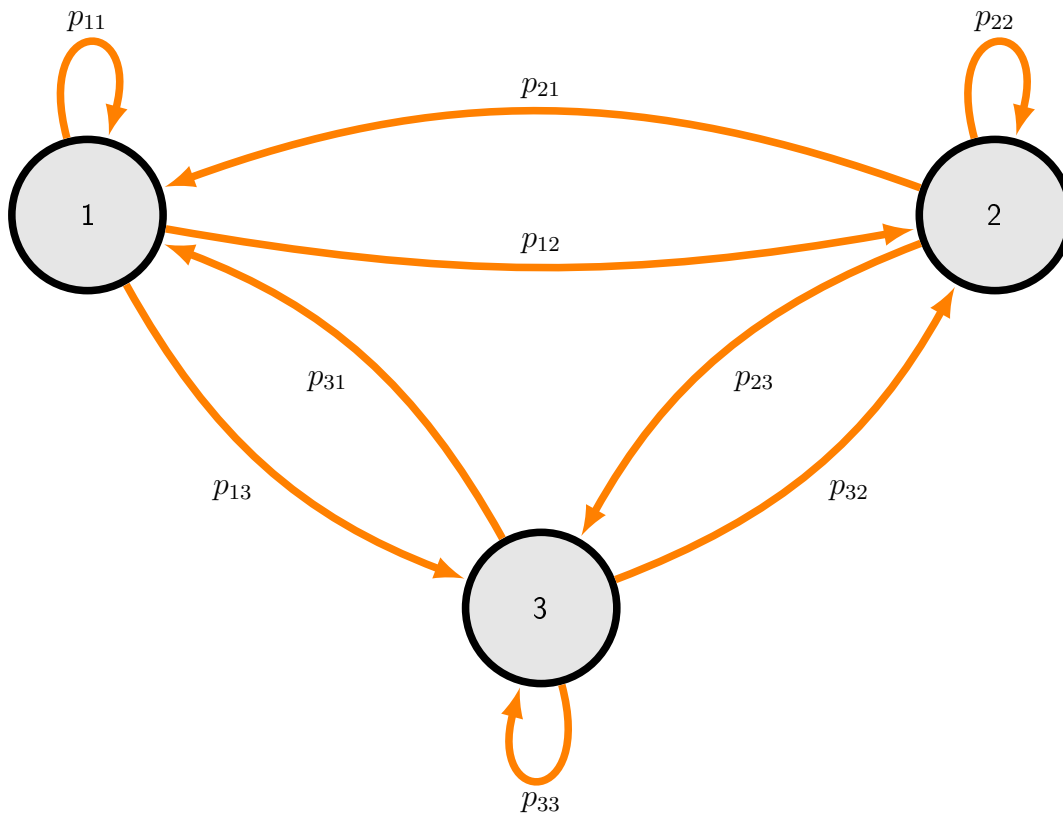


Figura 2.8: Cadeia de Markov com três estados.

$$(3) \quad p_{31} + p_{32} + p_{33} = 1,$$

$$(4) \quad p_{12} + p_{13} = 0.5,$$

$$(5) \quad p_{22} + p_{23} = 0.5,$$

$$(6) \quad p_{31} + p_{32} = 0.7,$$

$$(7) \quad p_{11} + p_{12} = 0.6,$$

$$(8) \quad p_{33} \in [0.2, 0.5],$$

$$(9) \quad p_{21} \in [0.3, 0.4].$$

Dadas as nove restrições acima, sendo as três primeiras restrições necessárias e as restantes arbitrárias conseguimos obter um Problema de Programação Linear Fracionário, da seguinte maneira:

$$\begin{aligned} & \text{minimize/maximize} \quad \frac{p_{11}}{p_{11} + p_{12} + p_{21}} \\ & \text{sujeito a (1) - (9).} \end{aligned}$$

Resolvendo nosso PPL, obtemos os seguintes resultados:

$$\rightarrow \text{Mínimo } \frac{p_{11}}{p_{11} + p_{12} + p_{21}} = 0.5.$$

$$\rightarrow \text{Máximo } \frac{p_{11}}{p_{11} + p_{12} + p_{21}} = 0.5556.$$





## Capítulo 3

# Precificação de Opções

Quando nos deparamos com situações em que queremos determinar a probabilidade para um evento de nosso interesse, duas perguntas importantes surgem: (i) qual a sua distribuição de probabilidade do evento? (ii) quais os valores dos parâmetros desta distribuição? Quando conseguimos responder tais perguntas, geralmente, conseguimos determinar a probabilidade para o nosso evento de interesse. Entretanto, quando não temos as respostas para as perguntas (i) e (ii), como devemos proceder para encontrar a probabilidade para um evento de interesse ?

No mesmo espírito da demonstração apresentada no capítulo anterior para a Desigualdade de Markov, utilizaremos a teoria apresentada no Capítulo 1 para encontrar uma desigualdade que determina um limitante superior para o preço de uma opção de compra europeia desenvolvida em [1], quando temos conhecimento da média e da variância do preço do ativo em que a opção de compra se baseia. Inspirado por este resultado, desenvolveremos um limitante inferior, sob as mesmas hipóteses do limitante superior. Para isso, definiremos alguns conceitos básicos da teoria de precificação de ativos e discutiremos em que condições garantimos a não existência de oportunidades de arbitragem e como determinar uma carteira ótima. Por último, compararemos os limitantes superior e inferior para o preço da opção de compra, obtidos através das desigualdades, com o preço da opção de compra obtido pela equação de Black e Scholes (referência [5]).

### 3.1 Conceitos básicos

Nesta seção, falaremos um pouco sobre como podemos aplicar a teoria desenvolvida no Capítulo 1 à precificação de opções europeias. Para tal precisamos primeiramente definir alguns conceitos. O espaço amostral que utilizaremos é composto de todos os possíveis estados da natureza que podem acontecer depois de um determinado período de tempo, que é denominado **espaço dos estados da natureza** ou **espaço de estados**.

**Definição 3.1 (Definição de ativo).** *Ativo é um termo básico utilizado para expressar os bens, valores, créditos, direitos e assemelhados que, num determinado momento, formam*

o patrimônio de um indivíduo, que são avaliados pelos respectivos custos.

**Definição 3.2.** O preço de um ativo na data  $t$ , é uma variável aleatória  $X(t)$  definida no espaço de estados da natureza  $\omega$ .

Para o nosso contexto, o número real associado a cada estado da natureza  $\omega$  através da variável aleatória  $X(t)$  é que o preço do ativo  $X$  em um período de tempo  $t$  na ocorrência de cada estado da natureza  $\omega$ .

**Definição 3.3.** O preço de um ativo livre de risco  $X(t)$  é uma variável aleatória degenerada, ou seja, é uma função que associa cada elemento  $\omega$  do espaço de estados da natureza  $\Omega$  a um mesmo número real, para qualquer  $t$ .

**Definição 3.4.** Se um ativo não é livre de risco, dizemos que ele é um **ativo de risco**.

A definição acima nos diz que um ativo é livre de risco se seu preço em um período de tempo  $t$  é o mesmo do período de tempo inicial, ou seja, o preço é o mesmo independentemente de qual estado da natureza ocorra. Portanto, não há incerteza a respeito de seu preço. Uma observação deve ser feita quando se considera uma taxa de juros no período de tempo  $[t - 1, t)$ : a menos da taxa de juros, o valor do ativo livre de risco no tempo  $t$  é o mesmo que no tempo  $t - 1$ , ou seja, os preços em diferentes tempos só diferem por conta da taxa de juros.

Considere um mercado que opera em um único período, onde  $n$  ativos diferentes são negociados, com  $\Omega$  discreto e finito. Dependendo dos eventos durante esse único período, existem  $m$  possíveis estados da natureza no final do período. Se investirmos uma unidade monetária (u.m.) em algum ativo  $i$  e o estado da natureza vir a ser  $\omega$ , receberemos um *payoff* de  $r_{\omega i}$ . Assim, para cada ativo  $i$  é descrito um vetor de *payoff*  $(r_{1i}, \dots, r_{mi})$ . Segue a matriz *payoffs*  $m \times n$  que dá os *payoffs* de cada um dos  $n$  ativos para cada um dos  $m$  estados da natureza é:

$$\mathbf{R} = \begin{bmatrix} r_{11} & \dots & r_{1n} \\ \vdots & \ddots & \vdots \\ r_{m1} & \dots & r_{mn} \end{bmatrix}.$$

Seja  $x_i$  a quantidade do ativo  $i$  em posse do agente ou investidor. Um portfólio ou carteira de ativos é então um vetor  $\mathbf{x} = (x_1, \dots, x_n)$ . Os componentes de um portfólio  $\mathbf{x}$  podem ser positivos ou negativos. Um valor positivo de  $x_i$  indica que foram compradas  $x_i$  unidades do ativo  $i$  e, portanto, o direito de receber  $r_{\omega i}x_i$  se o estado  $\omega$  ocorrer. Um valor negativo de  $x_i$  indica venda do ativo  $i$ .

O ganho no estado  $\omega$  que resulta de um portfólio  $\mathbf{x}$  é dado por

$$z_{\omega} = \sum_{i=1}^n r_{\omega i}x_i.$$

Então temos o vetor  $\mathbf{z} = (z_1, \dots, z_m)$ , que pode ser escrito como

$$\mathbf{z} = \mathbf{R}\mathbf{x}.$$

Seja  $p_i$  o preço do ativo  $i$  no início do período e seja  $\mathbf{p} = (p_1, \dots, p_n)$  o vetor de preços dos ativos. Então, o custo para adquirir um portfólio  $\mathbf{x}$  é dado por  $\mathbf{p}^T \mathbf{x}$ .

O problema principal da precificação de ativos é determinar qual o preço de um certo ativo. A fim de resolver este problema, iremos introduzir a condição de ausência de arbitragem, implícita em toda teoria financeira: o preço dos ativos deve ser tal que um investidor não possa fazer investimento negativo ou nulo e, pelo menos para algum estado da natureza, ele tenha chance de lucrar. Matematicamente falando, a condição de ausência de arbitragem pode ser expressa da seguinte maneira no contexto deste capítulo:

$$\text{se } \mathbf{R}\mathbf{x} \geq \mathbf{0} \text{ então, devemos ter } \mathbf{p}^T \mathbf{x} \geq 0.$$

Dado um conjunto de ativos, descrito pela matriz de *payoff*  $\mathbf{R}$ , apenas determinados  $\mathbf{p}$ 's são consistentes com a ausência de arbitragem. O que caracteriza esses preços? Quais restrições devemos colocar sobre os preços dos ativos para garantir que não haja arbitragem? Para responder estas perguntas, utilizaremos o Lema de Farkas apresentado no Apêndice (Teorema A.2).

**Teorema 3.5.** *Não há condição de arbitragem, se e somente, existir um vetor não negativo  $\mathbf{q} = (q_1, \dots, q_m)$  (um vetor cujo valores são maiores ou iguais a zero), tal que o preço de cada ativo  $i$  é dado por*

$$p_i = \sum_{s=1}^m q_s r_{si}.$$

*Demonstração.* A condição de ausência de arbitragem implica que não existe vetor  $\mathbf{x}$  tal que  $\mathbf{x}^T \mathbf{R}^T \geq \mathbf{0}^T$  e  $\mathbf{x}^T \mathbf{p} < 0$ . Esta é da mesma forma que a condição (b) na demonstração do Lema de Farkas (Teorema A.2). Aqui, temos que  $\mathbf{p}$  está no lugar de  $\mathbf{b}$  e  $\mathbf{R}^T$  no lugar de  $\mathbf{A}^T$ . Portanto, pelo Lema de Farkas, a condição de ausência de arbitragem é assegurada se, e somente se, existir algum vetor não negativo  $\mathbf{q}$  tal que  $\mathbf{R}^T \mathbf{q} = \mathbf{p}$ , que é exatamente a mesma condição do Teorema A.2.  $\square$

**Definição 3.6.** *Uma opção de compra europeia<sup>1</sup> é um contrato estipulado em  $t = 0$  que dá o direito de comprar (sem perda de generalidade) em  $t = 1$  um ativo qualquer disponível num mercado (chamado de ativo base) por um preço  $K$  (chamado strike price ou preço de exercício) fixado em  $t = 0$ .*

**Exemplo 3.7.** *Em  $t = 1$  supomos que o estado da natureza  $\omega_i$  ocorra. Desse modo, o ativo  $x$  valerá  $x_i$  e assim, aquele que contratou o direito de comprar  $x$  por  $K$  só o fará se*

<sup>1</sup>Sempre que estiver escrito opção de compra, estaremos nos referindo a opção de compra européia.

$x_i > K$  pois, caso contrário, convém abandonar a opção. Desse modo, o ganho daquela que comprou tal opção é:

$$x_c = \begin{cases} x - K & \text{se } x > K \\ 0 & \text{se } x \leq K \end{cases}$$

onde  $x$  é o preço do ativo em  $t = 1$ .

**Definição 3.8.** Uma **opção de venda europeia** consiste em um contrato estipulado em  $t = 0$  que dá o direito, sem implicar o dever, de vender um ativo (sem perda de generalidade) em  $t = 1$  por um valor  $K$  fixado em  $t = 0$ .

**Exemplo 3.9.** Dado que ocorra o estado  $\omega_i$ , se  $x_i \leq K$  convém exercer a opção e vender o ativo  $x$ , caso contrário, convém abandoná-la.

Desse modo, o ganho da opção é a variável aleatória:

$$x_v = \begin{cases} 0 & \text{se } x_i > K \\ K - x_i & \text{se } x_i \leq K \end{cases}$$

onde  $x_i$  é o preço do ativo  $x$  em  $t = 1$ .

Uma pergunta importante é: qual deve ser o preço dessa opção em  $t = 0$ ? Agora vamos usar o teorema 3.5, para determinar tal preço. Para isso considere o exemplo a seguir.

**Exemplo 3.10.** Considere um mercado que opera para um único período e que envolva três ativos: uma ação, um ativo livre de risco e uma opção de compra.

Seja  $S$  o preço da ação no início do período, em  $t = 0$ . O preço  $\bar{S}$  no final do período é aleatório e assumimos que pode ser igual a  $Su$ , com probabilidade  $\beta$  ou  $Sd$  com probabilidade  $1 - \beta$ , (representado na figura 3.1), em que  $u$  e  $d$  são escalares tais que  $0 < d < 1 < u$ . Finalmente a opção nos dá o direito de comprar, no final do período, uma ação pelo preço  $K$  (strike price) pré determinado.

Se o preço  $\bar{S}$  for maior que  $K$ , exercemos a opção e então imediatamente vendemos a ação no mercado de ações, obtendo um ganho de  $\bar{S} - K$ . Se  $\bar{S} < K$ , não exercemos a opção de compra. Assim, o valor da opção no final do período é igual ao  $\max\{0, \bar{S} - K\}$ . Uma vez que a opção de compra também é um ativo, ela também deve ter um preço em  $t = 0$ . Sob a condição de ausência de arbitragem, o preço da opção de compra  $C$  é dado pela seguinte expressão:

$$C = \gamma \max\{0, Su - K\} + \delta \max\{0, Sd - K\},$$

onde  $\gamma$  e  $\delta$  são soluções do seguinte sistema de equações lineares, pois respeita o Teorema 3.5.

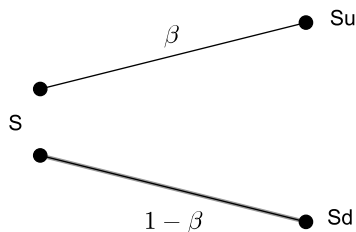


Figura 3.1: Preço da ação para  $\omega_1$  e  $\omega_2$ .

$$\begin{cases} u\gamma + d\delta = 1 \\ \gamma + \delta = \frac{1}{r} \end{cases}$$

Para ilustrar este problema, iremos considerar que a ação tem o preço atual, ou seja, em  $t = 0$ , de  $S = 10$  com um strike price  $K = 12$ , taxa de juros  $r = 1.05$  e limitamos os valores de  $u$  e  $d$  em , respectivamente,  $-8 < d < 1$  e  $1 < u < 10$ , com intervalo de 0.01 entre cada valor de  $u$  e  $d$ . A seguir, temos o gráfico 3.2 que mostra os valores da opção  $C$  para cada valor de  $u$  e  $d$ . O maior valor encontrado para  $C$  nos intervalos dados foi:  $C = 504.9589$  para  $u = 10$  e  $d = 0.99$  e o menor valor encontrado para  $C$  foi  $C \simeq 0$  para  $u = 4.2$  e  $d = -4.41$ .

## 3.2 Limitantes para uma opção de compra

Nesta seção enunciaremos e provaremos dois resultados que nos auxiliarão na busca de um limitante superior e um inferior para o valor da opção de compra quando conhecemos a média e a variância do preço do ativo base, caso que se enquadra nas hipóteses da Desigualdade de Chebyshev. Enunciada e demonstrada no início do segundo capítulo, ela fornecia um limitante para uma probabilidade dada uma distribuição de probabilidade  $X$  qualquer, onde tínhamos conhecimento da média e da variância desta distribuição. O limitante superior foi desenvolvido por [11] e está presente no trabalho de [1] e o limitante inferior foi desenvolvido neste trabalho, pela necessidade de se obter um preço inferior para opção de compra. O desenvolvimento do limitante inferior se deu utilizando a técnica utilizada para demonstrar a Desigualdade de Markov, apresentada no Segundo Capítulo.

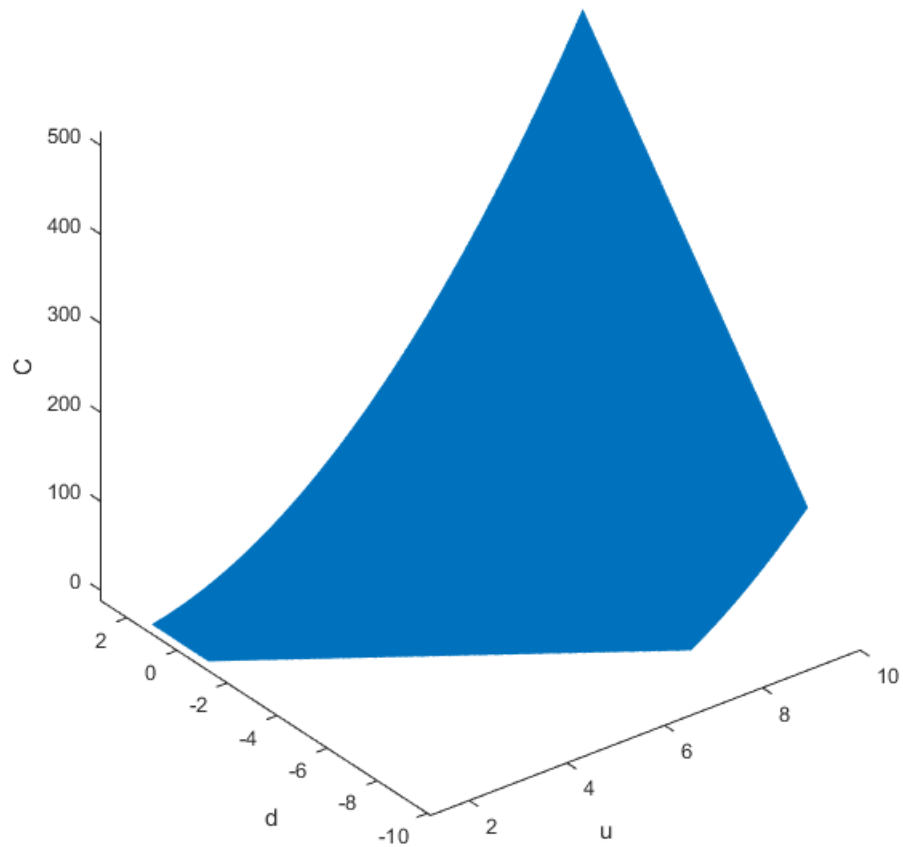


Figura 3.2: Gráfico do preço da opção de compra  $C$  para cada valor de  $u$  e  $d$ .

### 3.2.1 Limitante superior para uma opção de compra

**Teorema 3.11.** *O limitante superior do preço de uma opção de compra no tempo  $t$  com strike price  $K$  sobre uma ação com preço no vencimento com média  $\mu$  e variância  $\sigma^2$  conhecidas, é dado por*

$$\max_{S_T \sim (\mu, \sigma^2)^+} e^{-rt} E[\max(0, S_T - K)] = \begin{cases} \frac{e^{-rt}}{2} \left[ (\mu - K) + \sqrt{\sigma^2 + (\mu - K)^2} \right], & K \geq \frac{\mu^2 + \sigma^2}{2\mu} \\ e^{-rt} \left[ \mu - K + K \frac{\sigma^2}{\mu^2 + \sigma^2} \right], & K < \frac{\mu^2 + \sigma^2}{2\mu}. \end{cases}$$

*Demonstração.* Para demonstrar esse teorema, iremos utilizar a teoria apresentada no pri-

meiro capítulo, principalmente o Teorema da Dualidade, que diz que para cada problema de otimização linear, existe um problema dual correspondente a ele. No caso deste teorema, estamos maximizando o valor esperado, então partiremos do respectivo problema dual, que será o de minimização.

O limitante superior ótimo de uma opção de compra europeia com *strike price*  $K$  é dado pela solução do seguinte problema de encontrar  $\{a_i\}$  tais que

$$\begin{aligned} & \text{minimize} && \sum_{i=0}^n a_i S_T, \\ & \text{sujeito a} && \sum_{r=0}^n a_r S_T^r \geq \max(0, S_T - k), \forall S_T \geq 0. \end{aligned}$$

Para o nosso caso de interesse, reformularemos tal problema com as seguintes variáveis duais:  $a_0, a_1$  e  $a_2$ , pois são dados apenas  $E(S_T)$  e  $\text{var}(S_T)$ . Então, pelo Teorema da Dualidade apresentado no Capítulo 1, podemos fazer a seguinte formulação dual para o problema mostrado acima:

$$\begin{aligned} & \text{minimize} && (\mu^2 + \sigma^2)a_2 + \mu a_1 + a_0, \\ & \text{sujeito a} && g(S_T) = a_2 S_T^2 + a_1 S_T + y_0 \geq \max(0, S_T - k), \forall S_T \geq 0. \end{aligned} \tag{3.1}$$

A função dual factível  $g(\cdot)$  é uma função quadrática qualquer, que está no quadrante positivo (é não negativa e encontra-se acima da linha  $(S_T - K)_+$ ). Em uma solução ótima, a função quadrática deve ser tangente à linha  $(S_T - K)_+$ . As possibilidades de tangência da função quadrática na linha  $(S_T - K)_+$  são ilustradas nas figuras 3.3 e 3.4.

Como a linha  $(S_T - K)_+$  é tangente à função quadrática, utilizando a definição de reta tangente, podemos escrever:

$$g(S_T) - (S_T - K) = a(S_T - b)^2, \text{ para algum } a \geq 0.$$

A restrição de não negatividade sobre  $g(\cdot)$  pode ser expressa como  $a(S_T - b)^2 + S_T - K \geq 0$ , para todo  $S_T \geq 0$ . Como  $g(S_T)$  é uma função quadrática, com a concavidade virada para cima, o seu ponto de mínimo é atingido no seu vértice. Assim, temos que  $S_{T_0} = b - \frac{1}{2a}$  é o ponto de mínimo desta função quadrática.

Aqui, dependemos do valor de  $S_{T_0}$ , o qual pode ser ou não um valor não negativo. Como estamos tratando de preço de opção de compra, temos que  $S_{T_0}$  é não negativo. Logo, teremos as seguintes duas inequações possíveis: ou  $S_T = S_{T_0}$  ou  $S_T = 0$  e uma das duas precisa ser válida para termos uma solução ótima, temos essas duas hipóteses porque a função que cobre a outra, ou corta o eixo  $S_T$  no  $S_T = 0$  ou corta em algum outro ponto depois desse  $S_T = S_{T_0}$ . Assim, como  $S_{T_0}$  deve ser não negativo, temos os dois casos seguintes:

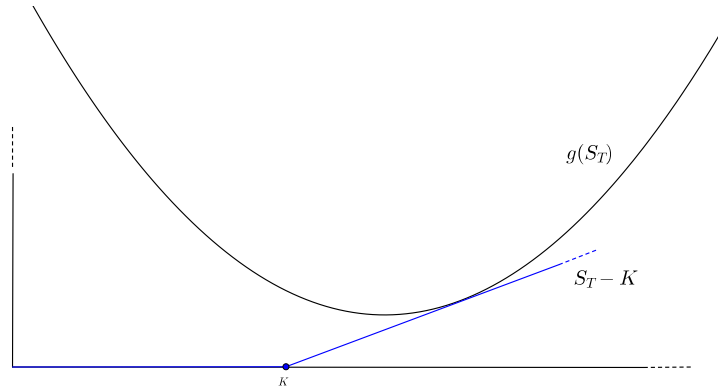


Figura 3.3:  $g(S_T)$  para  $S_T = S_{T_0}$

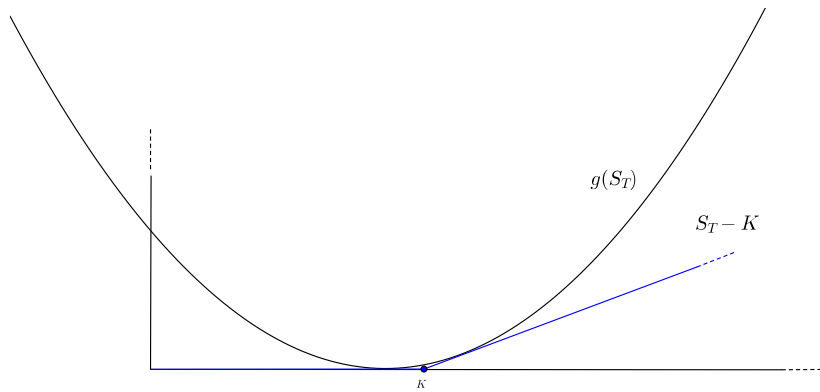


Figura 3.4:  $g(S_T)$  para  $S_T = 0$

a) Se  $b \geq \frac{1}{2a}$ , então  $-\frac{1}{4a} + b - K = 0$ , restrição vinculada a  $x = x_0$ .

Isolando  $a$ , obtemos a expressão  $a = \frac{1}{4(b - K)}$ .

Considere agora a seguinte função objetivo, obtida através da manipulação da reta tangente.

$$g(S_T) = aS_T^2 + (1 - 2ab)S_T + ab^2 - K.$$

Na função objetivo é possível identificar  $a_0$ ,  $a_1$  e  $a_2$ , que são respectivamente  $ab^2 - K$ ,  $(1 - 2ab)$  e  $a$ . Agora, como sabemos quem são essas variáveis, podemos substituir os seus respectivos valores e o valor de  $a$  na equação 3.1. Logo, temos



$$\begin{aligned}
\max_{S_T \sim (\mu, \sigma^2)^+} e^{-rt} E[\max(0, S_T - K)] &= \min_b \frac{(\mu^2 + \sigma^2)}{4(b - K)} + \mu \left(1 - \frac{2b}{4(b - K)}\right) + \frac{b^2}{4(b - K)} - K \\
&= \min_b \frac{(\mu^2 + \sigma^2) + \mu[4(b - K)] - 2b\mu + b^2 - 4K(b - K)}{4(b - K)} \\
&= \min_b \frac{(\mu - K)^2 + 2(\mu - K)(b - K) + (b - K)^2 + \sigma^2}{4(b - K)} \\
&= \min_b \frac{[(\mu - K) + (b - K)]^2 + \sigma^2}{4(b - K)}.
\end{aligned}$$

Também, precisamos determinar qual é o mínimo e, para isso, precisamos derivar a função acima e igualar a zero para podermos encontrar os possíveis candidatos a ponto de mínimo. Assim,

$$\begin{aligned}
f'(b) &= \left[ \frac{[(\mu - K) + (b - K)]^2 + \sigma^2}{4(b - K)} \right]' \\
&= \frac{(b - K)^2 - (\mu - K)^2 - \sigma^2}{4(b - K)^2}.
\end{aligned}$$

Como queremos  $f'(b) = 0$ , basta então termos  $(b - K)^2 - (\mu - K)^2 - \sigma^2 = 0$ . Temos que  $a \geq 0$  e  $b \geq \frac{1}{2a}$ , logo  $b_0 = \sqrt{(\mu - K)^2 + \sigma^2} + K$ . Portanto,

$$\begin{aligned}
\max_{S_T \sim (\mu, \sigma^2)^+} e^{-rt} E[\max(0, S_T - K)] &= \min_b \frac{[(\mu - K) + (b - K)]^2 + \sigma^2}{4(b - K)} \\
&= \frac{[(\mu - K) + (\sqrt{(\mu - K)^2 + \sigma^2} + K - K)]^2 + \sigma^2}{4(\sqrt{(\mu - K)^2 + \sigma^2} + K - K)} \\
&= \frac{1}{2} \left[ (\mu - K) + \sqrt{\sigma^2 + (\mu - K)^2} \right].
\end{aligned}$$

Seja  $a_0 = \frac{1}{4(b_0 - K)}$  e considere  $b_0 \geq \frac{1}{2a_0}$ . Substituindo  $a_0$  em  $b_0 \geq \frac{1}{2a_0}$ , obtemos a seguinte expressão  $b_0 \geq 2(b_0 - K)$ . Agora isolando  $K$ , chegamos que o limitante superior que deduzimos é válido sempre que  $\frac{\mu^2 + \sigma^2}{2\mu} \leq K$ .

- b) Se  $b < \frac{1}{2a}$ , temos que  $S_{T_0} = b - \frac{1}{2a} = 0$ , pois  $S_T$  é não negativa. Desse modo,  $ab^2 - K = 0$  (vinculada à restrição  $S_T = 0$ ).

Realizando o mesmo processo do caso (a), substituímos  $a = \frac{K}{b^2}$  na função objetivo, o que nos dá

$$\begin{aligned} \max_{S_T \sim (\mu, \sigma^2)^+} e^{-rt} E[\max(0, S_T - K)] &= \min_b (\mu^2 + \sigma^2) \frac{K}{b^2} + \mu \left( 1 - 2 \frac{K}{b^2} b \right) + \frac{K}{b^2} b^2 - K \\ &= \min_b (\mu^2 + \sigma^2) \frac{K}{b^2} - 2 \frac{K}{b} \mu + \mu. \end{aligned}$$

Novamente, realizando o mesmo processo de minimização do caso (a), concluímos que o mínimo é atingido em  $b_0 = \frac{\mu^2 + \sigma^2}{\mu}$  e, assim obtemos

$$\begin{aligned} \max_{S_T \sim (\mu, \sigma^2)^+} e^{-rt} E[\max(0, S_T - K)] &= (\mu^2 + \sigma^2) \frac{K}{\left( \frac{\mu^2 + \sigma^2}{\mu} \right)^2} - 2 \frac{K\mu}{\left( \frac{\mu^2 + \sigma^2}{\mu} \right)} + \mu \\ &= \mu - K \frac{\mu^2}{\mu^2 + \sigma^2}. \end{aligned}$$

Seja  $a_0 = \frac{K}{b_0^2}$  e considere  $b_0 < \frac{1}{2a_0}$ . Substituindo  $a_0$  em  $b_0 < \frac{1}{2a_0}$ , obtemos a seguinte expressão  $b_0 < \frac{b_0^2}{2K}$ . Agora isolando  $K$ , chegamos que o limitante superior que deduzimos é válido sempre que  $\frac{\mu^2 + \sigma^2}{2\mu} > K$ .

Para finalizar a demonstração, basta multiplicarmos ambos os membros por  $e^{-rt}$ , para trazer o preço a valor presente. Assim concluímos nossa prova.  $\square$

### 3.2.2 Limitante inferior para uma opção de compra

**Teorema 3.12.** *O limitante inferior do preço de uma opção de compra com strike price  $K$  sobre uma ação com preço no vencimento com média  $\mu$  e variância  $\sigma^2$  conhecidas, é dado por*

$$\min_{S_T \sim (\mu, \sigma^2)^+} e^{-rt} E[(S_T - K)_+] = \begin{cases} e^{-rt}(\mu - K) & \text{se } \mu \geq K \\ 0, & \text{se } \mu < K. \end{cases}$$

*Demonstração.* A estratégia que utilizaremos para demonstrar o limitante inferior se assemelha fortemente com a estratégia utilizada para demonstrar a Desigualdade de Markov. Desse modo, utilizaremos o Teorema de Dualidade 1.10 e o Teorema de Condição Complementar de Relaxamento 1.11 em nossa demonstração.

Assim, queremos determinar

$$\begin{aligned} \min \quad & E[(S_T - K)_+] \\ \text{sujeito a} \quad & E(S_T) = \mu, \\ & E(S_T^2) = \mu^2 + \sigma^2. \end{aligned}$$

Então, pelo Teorema da Dualidade, podemos reescrever nosso problema da seguinte maneira

$$\min[(S_T - K)_+] = \max[E[f(S_T)] : f(S_T) \leq (S_T - K)_+, f(S_T) \in \mathcal{L}],$$

sendo  $\mathcal{L}$  definido como segue:

$$\mathcal{L} = \{p(t) : p(t) = a_0 + a_1 t + a_2 t^2\}$$

Note que aqui, diferente da Desigualdade de Markov,  $p(t)$  pode ser uma função de até segundo grau, pois o primeiro e o segundo momentos de  $S_T$  são conhecidos.

Agora, queremos determinar  $a_0, a_1$  e  $a_2$ , de tal modo que

$$\begin{aligned} \max_{\{a_0, a_1, a_2\}} \quad & a_0 + a_1 \mu + a_2 (\mu^2 + \sigma^2) \\ \text{sujeito a} \quad & a_0 + a_1 S_T + a_2 S_T^2 \leq (S_T - K)_+. \end{aligned}$$

Temos que determinar  $a_0, a_1$  e  $a_2$ , de tal forma que satisfaça a condição do Teorema 1.11 para que tenhamos otimalidade. Como estamos tratando do preço de uma opção de compra, nossa variável aleatória não pode assumir valores negativos. Além disso, ela assume o valor zero até o *strike price*  $K$  e após  $K$  ela assume o valor de  $S_T - K$  (definição de opção de compra).

Agora, queremos determinar  $a_0, a_1$  e  $a_2$  de modo que satisfaça o Teorema 1.11 e a Definição de Opção de Compra. Diferente do limitante superior, o qual determinamos pela aproximação por uma função quadrática, no caso do limitante inferior, isso não é possível, pois não conseguimos construir uma função quadrática com concavidade para baixo com dois pontos de tangência. Sendo assim, utilizaremos uma função afim.

Como a função tem que estar definida em todos os pontos de  $\phi$  (Teorema 1.11), vamos determinar uma função que seja igual a zero até  $K$  e que, após  $K$ , seja  $S_T - K$  porque a função que melhor aproxima  $S_T - K$  é ela própria. Assim, tomando  $a_0 = -K, a_1 = 1$  e  $a_2 = 0$  e multiplicando por  $e^{-rt}$ , obtemos

$$\min e^{-rt} E[(S_T - K)_+] = \begin{cases} e^{-rt}(\mu - K) & \text{se } \mu \geq K \\ 0, & \text{se } \mu < K. \end{cases}$$

□

### Exemplo de aplicação dos limitantes para uma opção de compra

A fim de ilustrar a utilização deste último resultado, vamos apresentar o modelo de precificação de opções europeias de Black & Scholes, que é um dos modelos de precificação de opções mais conhecido no mercado financeiro.

**Definição 3.13** (Modelo de Precificação de Opções de Black & Scholes (Cox 1979).). *Seja  $S_t$  o preço do ativo base no tempo  $t$ ,  $K$  o strike price,  $r$  a taxa de juros livre de risco (por exemplo, uma taxa anual),  $N(x)$  a distribuição acumulada de uma normal padrão avaliada em  $x$  e  $\sigma$  a volatilidade medida geralmente como o desvio-padrão (anualizado) dos retornos da ação. Então, temos que o preço da opção de compra no tempo  $t$  é dado por:*

$$C_t = S_t N(d_1) - K e^{-r(T-t)} N(d_2)$$

onde

$$d_1 = \frac{\ln(S_t/K) + (r + \frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}}$$

e

$$d_2 = d_1 - \sigma\sqrt{T-t}.$$

### Exemplo de Comparação entre Modelo de Precificação de Black & Scholes e Limitantes Superior e Inferior para uma Opção de Compra

Para comparar os limitantes superior e inferior desenvolvido nesse trabalho com o Modelo de Precificação de Black & Scholes, vamos utilizar a ação *PETRA4*, que é a ação da Petrobras na bolsa de valores.

Os dados utilizados nessa simulação foram retirados dos sites [4] e [7]. Em [4] encontramos presente os valores praticados no mercado no dia 03/10/2017 e em [7] é possível consultar o histórico de preço das ações e opções de compra.

Foram utilizados dados do período de 02/04/2017 até 02/10/2017. Como o mercado financeiro funciona apenas em dias úteis, temos 128 cotações para o preço da *PETRA4*. Para ilustrar o comportamento da ação *PETRA4* nesse período, observamos o histograma das cotações no período e a série temporal das cotações nas Figuras 3.5 e 3.6 respectivamente.

Para utilizarmos os limitantes desenvolvidos, necessitamos da média ( $\mu$ ), variância ( $\sigma^2$ ) e do *strike price* ( $K$ ). A média e a variância obtidas a partir da análise dos dados

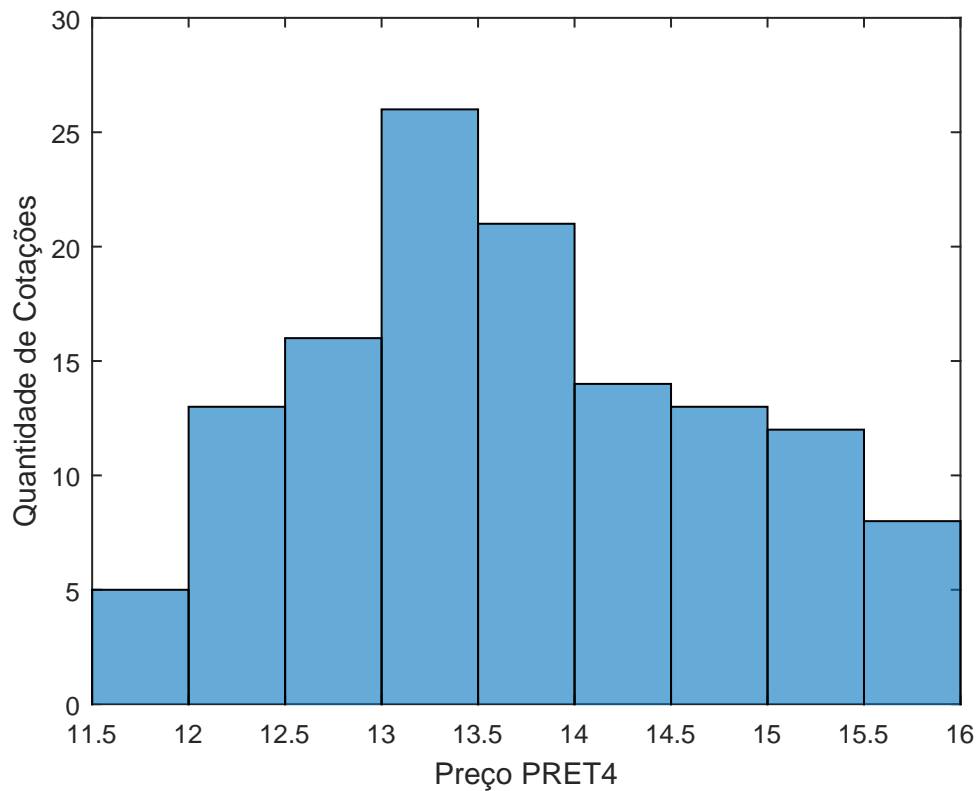


Figura 3.5: Histograma do histórico de cotação da ação PETR4

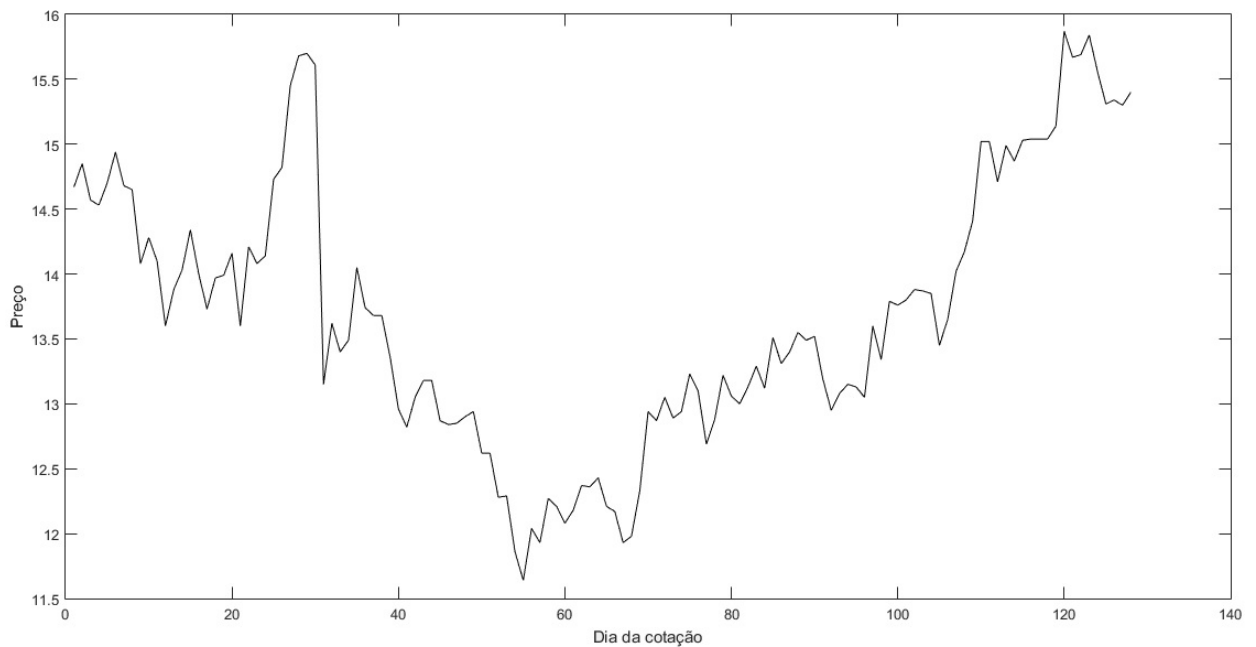


Figura 3.6: Série temporal do histórico de cotação da ação PETR4

foram, respectivamente,  $\hat{\mu} = 13.6965$  e  $\hat{\sigma}^2 = 1.0979$ , que consideramos válidos na data de vencimento. Para determinarmos o *strike price*, precisamos escolher uma opção de compra disponível no mercado. Escolhemos a opção PETRJ14 PN, cujo *strike price* é

$K = 14$ . Sendo assim, temos todas as hipóteses para calcularmos os limitantes.

Agora, resta determinar as hipóteses para o cálculo do Black & Scholes. Os dados a seguir foram retirados do site da IBOVESPA, referência [4]. Os dados foram coletados no dia 03/10/2017 e a opção escolhida PETRJ14-PN tinha seu vencimento programado para o dia 16/10/2017. Assim, tínhamos nove dias úteis como nosso período. O preço da ação PETR4 no dia da cotação era de 15,40, desvio padrão de 0.0238, taxa de juros de 8.5% a.a. e volatilidade anualizada de 37,81%.

Primeiro vamos calcular os limitantes superior e inferior da opção. Para isso, considere os seguintes limitantes:

$$\max_{S_T \sim (\mu, \sigma^2)_+} e^{-rt} E[\max(0, S_T - K)] = \begin{cases} \frac{e^{-rt}}{2} \left[ (\mu - K) + \sqrt{\sigma^2 + (\mu - K)^2} \right], & K \geq \frac{\mu^2 + \sigma^2}{2\mu} \quad (I) \\ e^{-rt} \left[ \mu - K + K \frac{\sigma^2}{\mu^2 + \sigma^2} \right], & K < \frac{\mu^2 + \sigma^2}{2\mu}. \quad (II) \end{cases}$$

Nosso limitante superior será dado pela equação (I), porque se substituirmos  $\hat{\mu}$  e  $\hat{\sigma}^2$  na expressão  $K \geq \frac{\mu^2 + \sigma^2}{2\mu}$ , chegamos no seguinte resultado  $K \geq 6.889$ , como  $K = 14$ , logo utilizaremos a equação (I). Desse modo,

$$\max_{S_T \sim (\mu, \sigma^2)_+} e^{-rt} E[\max(0, S_T - K)] = 0.3936.$$

E o limitante inferior é dado por

$$\min_{S_T \sim (\mu, \sigma^2)_+} e^{-rt} E[(S_T - K)_+] = \begin{cases} e^{-rt}(\mu - K) & \text{se } \mu \geq K \quad (I) \\ 0, & \text{se } \mu < K. \quad (II) \end{cases}$$

E o nosso limitante superior será dado pela equação (II), pois  $\mu < K \rightarrow 13.6965 < 14$ . Sendo assim,

$$\min_{S_T \sim (\mu, \sigma^2)_+} e^{-rt} E[(S_T - K)_+] = 0.$$

Portanto, obtemos o intervalo  $C_t \in [0, 0.3936]$  de preços para nossa opção de compra PETRJ14-PN.

Por último, resta calcularmos o preço da opção de compra dado pelo Modelo de Black & Scholes. Para isso, foi utilizado a função *blsprice* do Matlab, com os dados de preço da ação, volatilidade, *strike price*, período e taxa de juros dados acima. O resultado obtido foi de  $C_t = 0.3157$ .

Podemos concluir então, que quando temos conhecimento da média e da variância do ativo base da opção de compra, os limitantes superior e inferior apresentam uma boa aproximação do valor dado pelo Modelo de Black & Scholes.

## Considerações Finais

Neste trabalho mostramos como é possível encontrar limitantes superiores e inferiores para integrais tomadas com relação a uma determinada medida de probabilidade (esperanças generalizadas) quando dispomos de informações limitadas a respeito dessa medida. Mais especificamente, obtemos esses limitantes quando essa informação também é disponibilizada na forma de outras integrais, que podem assumir um valor específico ou um determinado valor em um intervalo. Para resolver esse tipo de problema utilizamos a teoria de programação linear, exposta no primeiro capítulo desta dissertação.

Na sequência, ilustramos como a programação linear pode ser aplicada para resolver problemas de inferência estatística. No Capítulo 2, tratamos de dois problemas clássicos da teoria do cálculo de probabilidades, o Problema de *Monty Hall* e o Problema das Testemunhas. No capítulo três, aplicamos os resultados desenvolvidos no primeiro capítulo para obter os limitantes superior e inferior para o preço de uma opção de compra quando se conhece apenas a média e a variância do preço do ativo base da opção na data de vencimento  $T$ .

Após aplicarmos a teoria de programação linear nos problemas apresentados no Capítulos 2 e 3, podemos perceber que tal teoria, mesmo sendo uma metodologia pouco usual para resolver problemas estatísticos, se mostra bastante útil quando nos deparamos com problemas em que temos pouca informação. Isso ocorre nas situações em que não temos conhecimento de qual é a distribuição de probabilidade de nosso evento de interesse, mas, sabemos apenas quais são os primeiro e segundo momentos, ou seja, a média e a variância do nosso evento de interesse. Como temos apenas o conhecimento da média e da variância, não conseguimos calcular uma probabilidade precisa para nosso problema. Entretanto, utilizando programação linear, é possível obter limitantes superiores e inferiores para probabilidade de nosso interesse. Isto é muito bom, pois, mesmo não estando em posse da informação completa do problema, conseguimos encontrar uma boa resposta aproximada.

Além disso, é possível, através da programação linear, generalizar problemas com soluções já conhecidas, para casos com mais variáveis e/ou quando delimitados intervalos para as probabilidades. Delimitar esses intervalos só se tornou possível porque conseguimos extrair equações das árvores de probabilidades para obter os conjuntos de restrições necessárias para os problemas propostos no Capítulo 2. Entretanto, isso foi feito analisando particularmente cada caso, e uma das sugestões para a continuação deste trabalho

seria o desenvolvimento de um algoritmo que permitesse ter a árvore de decisão como entrada e o conjunto de restrições como saída. Desse modo, poderíamos generalizar, por exemplo, o Problema de *Monty Hall* para  $n$  portas e  $m$  prêmios. Visto que para quatro portas tivemos um conjunto de restrições com trinta e seis equações, fazer isso manualmente para  $n$  seria algo extremamente complicado. Isto também é válido para a generalização da probabilidade do testemunho quando a quantidade de testemunhas for igual a  $n > 2$ .

Em suma, a teoria de programação linear é bastante utilizada para resolver problemas determinísticos e, neste trabalho, apresentamos aplicações desta teoria em problemas estatísticos e financeiros.



# Apêndice A

## Conceitos Básicos

Neste apêndice apresentamos alguns conceitos de Programação Linear, Teoria de Probabilidades e Análise Convexa utilizados no decorrer do trabalho.

### A.1 Programação Linear

**Teorema A.1** (Teorema da Fundamental da Programação Linear). *Dado um modelo de programação linear na forma padrão, ou seja, no qual  $\mathbf{A}$  é uma matriz ( $m \times n$ ),*

$$\begin{aligned} & \text{minimize } \boldsymbol{\beta}^T \mathbf{x} \\ & \text{sujeito a } \mathbf{Ax} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned}$$

*Assim,*

*A) Se existe uma solução factível, então existe uma solução básica factível.*

*B) Se existe uma solução factível ótima, existe uma solução básica factível ótima.*

*Demonstração.* A demonstração deste teorema, pode ser encontra em Luenberger (1984). □

**Teorema A.2** (Lema de Farkas). *Seja  $\mathbf{A}$  uma matriz  $m \times n$  e seja  $\mathbf{b}$  um vetor em  $\mathbb{R}^m$ . Então uma das duas alternativas a seguir é válida:*

*a) Existe algum  $\mathbf{x} \geq \boldsymbol{\beta}$  tal que  $\mathbf{Ax} = \mathbf{b}$ .*

*b) Existe algum vetor  $\boldsymbol{\lambda}$  tal que  $\boldsymbol{\lambda}^T \mathbf{A} \geq \boldsymbol{\beta}^T$  e  $\boldsymbol{\lambda}^T \mathbf{b} < 0$ .*

*Demonstração.* A demonstração se dará mostrando que, se acontece uma das alternativas, a outra não pode ser satisfeita.

Se existem algum  $\mathbf{x} \geq \mathbf{0}$  satisfazendo  $\mathbf{Ax} = \mathbf{b}$ , e se  $\boldsymbol{\lambda}^T \mathbf{A} \geq \boldsymbol{\beta}^T$ , então  $\boldsymbol{\lambda}^T \mathbf{b} = \boldsymbol{\lambda}^T \mathbf{Ax} \geq 0$ , o que mostra que a segunda alternativa não é satisfeita.

Agora assumimos que não existe um vetor  $\mathbf{x} \geq \mathbf{0}$  satisfazendo  $\mathbf{Ax} = \mathbf{b}$ . Considere o seguinte par de problemas

$$\begin{aligned} & \text{maximize } \boldsymbol{\beta}^T \mathbf{x} \\ & \text{sujeito a } \mathbf{Ax} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}, \end{aligned}$$

e

$$\begin{aligned} & \text{minimize } \boldsymbol{\lambda}^T \mathbf{b} \\ & \text{sujeito a } \boldsymbol{\lambda}^T \mathbf{A} \geq \mathbf{0}^T, \end{aligned}$$

note que o primeiro problema é o problema dual do segundo. O problema de maximização é infactível, o que implica que o problema de minimização ou é ilimitado ou também infactível. Uma vez que  $\boldsymbol{\lambda} = \mathbf{0}$  é uma solução factível do problema de minimização, então este problema só pode ser ilimitado, ou seja,  $\boldsymbol{\lambda}^T \mathbf{b} \rightarrow -\infty$ . Portanto, existe algum  $\boldsymbol{\lambda}$  que é factível, ou seja,  $\boldsymbol{\lambda}^T \mathbf{A} \geq \mathbf{0}^T$  e cujo valor de  $\boldsymbol{\lambda}$  é negativo, ou seja,  $\boldsymbol{\lambda}^T \mathbf{b} < 0$ .

□

## A.2 Programação Linear Fracionária

O problema de programação especificado é não linear por causa da função objetivo, que é um quociente de duas funções lineares e o conjunto de restrições (equações e/ou inequações) são lineares. Problemas deste tipo são chamados de *Problema de Programação Linear Fracionária*. Felizmente, este problema pode ser transformado em um problema de programação linear, como descrito por [24]. Usando uma notação genérica, a transformação ocorre da seguinte maneira:

Suponha que  $\mathbf{n}$ ,  $\mathbf{d}$  e  $\mathbf{x}$  são vetores coluna de mesma dimensão, e  $\mathbf{A}$  e  $\mathbf{b}$  são matrizes, tais que,  $\mathbf{Ax} = \mathbf{b}$ . O problema de programação fracionária será o de encontrar o vetor  $\mathbf{x}$ , tal que,

$$\text{maximize } f(\mathbf{x}) = \frac{\mathbf{n}^T \mathbf{x}}{\mathbf{d}^T \mathbf{x}}$$

sujeito a

$$\begin{cases} \mathbf{Ax} = \mathbf{b} \\ \mathbf{x} \geq \mathbf{0} \text{ e } \mathbf{d}^T \mathbf{x} > 0. \end{cases}$$

Agora considere a transformação  $y(\mathbf{x}) = \frac{1}{\mathbf{d}^T \mathbf{x}}$  e  $\mathbf{z}(\mathbf{x}) = \mathbf{x}y(\mathbf{x})$ . Essas definições implicam uma restrição linear sobre o vetor  $\mathbf{z}$ , ou seja,  $\mathbf{d}^T \mathbf{z}(\mathbf{x}) = 1$ , pois  $\mathbf{d}^T \mathbf{z}(\mathbf{x}) = \mathbf{d}^T \mathbf{x}y(\mathbf{x}) = \mathbf{d}^T \mathbf{x} / \mathbf{d}^T \mathbf{x} = 1$ . Agora o problema de programação fracionária pode ser escrito como um problema de programação linear usual em termos de  $\mathbf{z}$  e  $y$  da seguinte maneira:

Encontre o vetor  $(\mathbf{z}^T y)^T$ , tal que,

$$\text{maximize } f(\mathbf{z}, y) = (\mathbf{n}^T \ 0)(\mathbf{z}^T \ y)^T = \mathbf{n}^T \mathbf{z} = \mathbf{n}^T \mathbf{x}y(\mathbf{x}) = \frac{\mathbf{n}^T \mathbf{x}}{\mathbf{d}^T \mathbf{x}}$$

sujeito a

$$\begin{cases} \mathbf{A}\mathbf{z} = \mathbf{b}y \text{ e } \mathbf{d}^T \mathbf{z} = 1 \\ \mathbf{z} \geq \mathbf{0} \text{ e } y > 0. \end{cases}$$

Também podemos escrever o conjunto de restrições na forma matricial

$$\begin{pmatrix} \mathbf{A} & -\mathbf{b} \\ \mathbf{d}^T & 0 \end{pmatrix} \begin{pmatrix} \mathbf{z} \\ y \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix}.$$

Este problema de programação linear é idêntico ao problema fracionário em termos de  $\mathbf{A}$ ,  $\mathbf{b}$ ,  $\mathbf{n}$ , e  $\mathbf{d}$ , desde que a função objetivo e o conjunto de restrições sejam idênticos, como visto por substituição direta. Se  $(\mathbf{z}^T y)^T$  é solução do problema transformado, então  $\mathbf{x} = \frac{\mathbf{z}}{y}$  é solução para o problema original.

**Observação A.3.** *Esta transformação é válida também para problemas de minimização, e quando o conjunto de restrições contém igualdades e/ou desigualdades.*

## A.3 Análise Convexa

Nesta seção definiremos alguns conceitos de Análise Convexa, utilizados principalmente no primeiro capítulo. O espaço que estaremos trabalhando, será um espaço vetorial real.

**Definição A.4** (Combinação Linear). *Dizemos que um vetor  $x \in \mathbb{R}^n$  é uma combinação linear de vetores  $x_1, \dots, x_k$  se existem escalares  $\lambda_1, \dots, \lambda_k \in \mathbb{R}$  adequados, tais que  $x = \lambda_1 x_1 + \dots + \lambda_k x_k$ . Além disso,*

- Se  $\sum_{i=1}^k \lambda_i = 1$ , então dizemos que  $x$  é uma combinação afim dos vetores  $x_k$ .
- Se  $\lambda_i \geq 0$  para todo  $1 \leq i \leq k$ , então dizemos que  $x$  é uma combinação positiva dos vetores  $x_k$ .

**Definição A.5** (Conjunto Afim). *Um conjunto  $A \subseteq \mathbb{R}^n$  é afim se a linha que une quaisquer dois pontos distintos em  $A$  estiver em  $A$ , ou seja, se para quaisquer  $x_1, x_2 \in A$  e  $\lambda \in \mathbb{R}$ , temos que  $\lambda x_1 + (1-\lambda)x_2 \in A$ . Em outras palavras,  $A$  contém as combinações lineares de quaisquer dois pontos em  $A$ , contanto que a soma dos coeficientes na combinação linear seja igual a um.*

**Definição A.6** (Conjunto Convexo). *Um conjunto  $D \subset \mathbb{R}^n$  é dito convexo quando dados dois pontos  $x_1, x_2 \in D$ , então o ponto genérico  $x = \lambda x_1 + (1 - \lambda)x_2 \in D$  para qualquer  $\lambda \in [0, 1]$ .*

Em outras palavras, podemos dizer que um conjunto é convexo se contém o segmento de reta formado por quaisquer dois pontos ao longo de uma trajetória retilínea desobstruída entre eles, onde o meio desobstruído está no conjunto. Todo conjunto afim também é convexo, pois ele contém as retas inteiras entre quaisquer dois pontos distintos nele, e, portanto, o segmento de reta entre os pontos.

**Definição A.7** (Combinação Convexa). *Dados  $x_i \in \mathbb{R}^n, \alpha_i \in [0, 1], i = 1, \dots, p$ , tais que  $\sum_{i=1}^p \alpha_i = 1$ . Então  $\sum_{i=1}^p \alpha_i x_i$  denomina-se a combinação convexa dos pontos  $x_i \in \mathbb{R}^n$  com parâmetros  $\alpha_i, i = 1, \dots, p$ . Em outras palavras, uma combinação é convexa quando ela é concomitantemente combinação afim e positiva.*

**Definição A.8** (Casco Convexo). *Dado um conjunto  $A$  qualquer, o menor conjunto convexo que contém  $A$  é denominado casco convexo, ou seja, o casco convexo de  $A$  é a interseção de todos os conjuntos convexos contendo  $A$ , ou a combinação convexa de todos pontos em  $A$ .*

## A.4 Teoria de Probabilidades

Considere um experimento aleatório, ou seja, um experimento cujo resultado é desconhecido antes de sua realização.

**Definição A.9.** *Dado um experimento aleatório, o conjunto de todos os resultados possíveis desse experimento, o qual denotaremos por  $\Omega$ , é denominado **espaço amostral**.*

Neste trabalho, nos referiremos ao espaço amostral como sendo o **espaço de estados da natureza** ou de **espaço de estados**.

**Exemplo A.10.** *Considere o lançamento de uma moeda em que estamos interessados em saber se o resultado é cara ou coroa. Temos que tal experimento é aleatório uma vez que não sabemos, antes de realizá-lo, qual será o resultado. Contudo, podemos afirmar que os resultados possíveis são cara ou coroa. Assim,  $\Omega = \{Cara, Coroa\}$ .*

**Definição A.11.** *Seja  $\Omega$  um conjunto não vazio. Uma  $\sigma$ -álgebra de  $\Omega$  é uma coleção  $\mathcal{F}$  de subconjuntos de  $\Omega$  tal que*

- i)  $\emptyset \in \mathcal{F}$ ;
- ii) Se  $A \in \mathcal{F}$ , então  $A^c = \Omega - A \in \mathcal{F}$ ;
- iii) Se  $A_1, A_2, \dots$  é uma coleção enumerável de elementos de  $\mathcal{F}$ , então  $\cup_{n=1}^{\infty} A_n \in \mathcal{F}$ .

O par  $(\Omega, \mathcal{F})$  é chamado *espaço mensurável*.

**Definição A.12.** Na Teoria de Probabilidades, os elementos de uma  $\sigma$ -álgebra são chamados de *eventos aleatórios* ou simplesmente *eventos observáveis*.

Se um experimento aleatório tem um conjunto de possíveis resultados  $\Omega$ , então dizemos que o evento  $A \subseteq \Omega$  ocorreu se o resultado  $\omega \in \Omega$  de tal experimento pertence a  $A$ . Informalmente, os eventos que pertencem a uma  $\sigma$ -álgebra são aqueles que podemos decidir se ocorreram ou não dada a informação disponível sobre experimentos já realizados. Desse modo, podemos dizer que  $\sigma$ -álgebras modelam informação.

**Exemplo A.13.** Se  $\Omega$  é um conjunto não vazio de resultados possíveis de um experimento, então,  $\{\emptyset, \Omega\}$  e o conjunto das partes<sup>1</sup> de  $\Omega$ , o qual denotaremos por  $\mathcal{P}(\Omega)$  são, respectivamente, a menor e a maior  $\sigma$ -álgebras de  $\Omega$ . A primeira corresponde a não ter informação nenhuma, ou seja, sabe-se apenas que um dos resultados possíveis pode ser observado enquanto que a última contém toda informação referente ao resultado do experimento, ou seja, tudo é observável.

**Definição A.14.** Considere o conjunto  $\mathbb{R}$  dos números reais. A menor  $\sigma$ -álgebra que contém todos os intervalos abertos de  $\mathbb{R}$  é chamada  $\sigma$ -álgebra de Borel e é denotada por  $\mathcal{B}$ .

A  $\sigma$ -álgebra de Borel é a  $\sigma$ -álgebra natural com a qual trabalhar quando se trata de  $\mathbb{R}$ . Pense em termos de um experimento aleatório, como escolher um número da reta real. Tal  $\sigma$ -álgebra é capaz de responder as questões mais básicas que podem ser feitas a respeito de tais experimentos. Isso ficará claro no próximo exemplo.

**Exemplo A.15.** Considere um experimento aleatório que corresponde a selecionar um número  $X$  pertencente a  $\mathbb{R}$ . Imagine que queremos saber se  $X$  está entre números reais quaisquer  $a$  e  $b$ . Como o conjunto aberto  $(a, b)$  está em  $\mathcal{B}$ , podemos decidir se  $(a, b)$  ocorreu ou não. Se ele ocorreu então  $X$  está entre  $a$  e  $b$ , caso contrário não.

E se quisermos saber se  $X \geq a$ ? A visualização do evento  $[a, \infty)$  como um conjunto de Borel é como segue: para cada  $n \in \mathbb{N}$ , note que  $B_n = (a - \frac{1}{n}, \infty)$  é uma união enumerável dos intervalos abertos  $B_n = \bigcup_k (a - \frac{1}{n}, k)$ . Portanto, cada  $B_n \in \mathcal{B}$ . Agora,  $[a, \infty) = \bigcap_n B_n$  é uma interseção enumerável de elementos de  $\mathcal{B}$  e portanto também pertence a  $\mathcal{B}$ . Assim, também conseguimos responder a essa pergunta.

As demais perguntas que podem ser feitas com relação ao valor de  $X$  em  $\mathbb{R}$  também podem ser respondidas através da  $\sigma$ -álgebra de Borel, o que a faz ser a mais natural, visto que qualquer outra  $\sigma$ -álgebra que contenha todos os intervalos possíveis de  $\mathbb{R}$  e, portanto, possa responder todos esses tipos de perguntas, deve conter todos os elementos de  $\mathcal{B}$ .

<sup>1</sup>Dado um conjunto  $\Omega$ , o conjunto formado por todos os subconjuntos de  $\Omega$  é chamado conjunto das partes de  $\Omega$ .

**Observação A.16.** *Sejam  $\mathcal{F}$  e  $\mathcal{G}$  duas  $\sigma$ -álgebras definidas num mesmo espaço amostral. Se  $\mathcal{F} \subseteq \mathcal{G}$  dizemos que  $\mathcal{G}$  é **mais fina** do que  $\mathcal{F}$ . Segue que  $\mathcal{G}$  contém mais informação do que  $\mathcal{F}$ , visto que possui mais eventos. Isso significa, informalmente, que  $\mathcal{G}$  pode “responder” a mais perguntas do que  $\mathcal{F}$ .*

**Definição A.17.** *Suponha que  $(\Omega, \mathcal{F})$  é um espaço mensurável. Uma função  $Q : \mathcal{F} \rightarrow [0, 1]$  é uma **medida de probabilidade** em  $(\Omega, \mathcal{F})$  se:*

$$i) Q(\Omega) = 1;$$

ii) *Se  $A_1, A_2, \dots, A_n, \dots$  é uma família enumerável de elementos de  $\mathcal{F}$  os quais são dois a dois disjuntos (isto é,  $A_n \cap A_m = \emptyset$  se  $n \neq m$ ), então*

$$Q\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} Q(A_n).$$

*Neste caso dizemos que  $Q$  é  $\sigma$ -aditiva.*

*A tripla  $(\Omega, \mathcal{F}, Q)$  é chamada **espaço de probabilidade**.*

**Definição A.18.** *Dizemos que evento  $A \in \mathcal{F}$  tem medida nula se  $Q(A) = 0$ .*

**Definição A.19.** *Suponha que  $(S, \mathcal{A})$  e  $(T, \mathcal{B})$  são espaços mensuráveis. Uma função  $f : S \rightarrow T$  é dita  $\mathcal{A}/\mathcal{B}$ -**mensurável** (ou simplesmente mensurável) se, e somente se,*

$$\forall B \in \mathcal{B} \Rightarrow f^{-1}(B) = \{s \in S \mid f(s) \in B\} \in \mathcal{A}.$$

**Definição A.20.** *Seja  $(\Omega, \mathcal{F}, Q)$  um espaço de probabilidade. Uma **variável aleatória** é uma função*

$$X : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}),$$

*que é  $\mathcal{F}/\mathcal{B}$ -mensurável.*

**Teorema A.21.** *Somas e produtos de funções mensuráveis são mensuráveis.*

*Demonstração.* A demonstração deste resultado pode ser encontrada na página 30 da referência [23]. □

**Corolário A.22.** *Somas e produtos de variáveis aleatórias são variáveis aleatórias.*

As **funções indicadoras** são uma classe especial de variáveis aleatórias. Se  $A \subseteq \Omega$ , definimos a função indicadora  $I_A$  por

$$I_A(\omega) = \begin{cases} 1 & \text{se } \omega \in A \\ 0 & \text{se caso contrário.} \end{cases}$$

Note que  $I_A : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B})$  é uma variável aleatória se, e somente se,  $A \in \mathcal{F}$

**Definição A.23.** *Seja  $X : \Omega \rightarrow \mathbb{R}$  uma função real. Definimos a  $\sigma$ -álgebra gerada por  $X$ , denotada por  $\sigma(X)$  pela família*

$$\sigma(X) = \{X^{-1}(B) \mid B \in \mathcal{B}\}.$$

*Similarmente, dada uma família  $\mathcal{X}$  de funções de  $\Omega$  em  $\mathbb{R}$ , definimos a  $\sigma$ -álgebra gerada por todos os eventos  $X^{-1}(B)$ , onde  $X \in \mathcal{X}$  e  $B \in \mathcal{B}$ , a qual denotamos por  $\sigma(\mathcal{X})$ .*

### A.4.1 Probabilidade Condicional

Suponha que lancemos dois dados e que cada um dos 36 pares possíveis sejam equiprováveis, ou seja, tenham a mesma probabilidade de ocorrer. Assim, para cada  $\omega \in \Omega$ , em que  $\Omega$  é formado pelos 36 resultados possíveis,  $Q(\omega) = \frac{1}{36}$ . Além disso, digamos que o primeiro resultado seja 3. Então, dado que conhecemos essa informação, qual é a probabilidade de que a soma dos dois dados seja igual a 8?

Para calcular essa probabilidade, devemos pensar do seguinte modo: sabendo que saiu o número 3 no primeiro dado, existirão no máximo seis resultados possíveis para o nosso experimento, os quais são (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6). Como cada um desses resultados tinha originalmente a mesma probabilidade de ocorrência, os resultados deveriam continuar a ter probabilidades iguais. Isto é, se o primeiro lançamento resultou em 3, a probabilidade (condicionada a esse fato) de cada um dos seis resultados possíveis é  $\frac{1}{6}$ , enquanto que a probabilidade (também condicionada) dos outros 30 outros eventos do espaço amostral é 0. Com isso, a probabilidade desejada será igual a  $\frac{1}{6}$ .

Se  $A$  e  $B$  representam, respectivamente, o evento em que o primeiro dado é 3 e o evento em que a soma dos dados é 8, então a probabilidade que acabamos de obter é chamada de **probabilidade condicional** de que  $B$  ocorra dado que  $A$  ocorreu e é representada por  $P(B \mid A)$ .

No caso geral, se um evento qualquer  $A$  ocorrer, então, para que um outro evento  $B$  ocorra é necessário que a ocorrência real seja um ponto tanto em  $A$  quanto em  $B$ , isto é, ele deve estar em  $A \cap B$ . Assim, como sabemos que  $A$  ocorreu, tem-se que  $A$  se torna nosso novo, e agora reduzido, espaço amostral. Com isso, a probabilidade de que o evento  $A \cap B$  ocorra será igual à probabilidade de  $A \cap B$  relativa à probabilidade de  $A$ . Formalmente, temos a seguinte definição.

**Definição A.24.** *Se  $A$  e  $B$  são eventos e  $Q(A) \neq 0$ , definimos a **probabilidade condicional de  $B$  dado  $A$**  por*

$$Q(B \mid A) = \frac{Q(A \cap B)}{Q(A)}.$$

*Se  $(\Omega, \mathcal{F}, Q)$  é um espaço de probabilidade, e se  $A \in \mathcal{F}$  com  $Q(A) > 0$ , então  $(\Omega, \mathcal{F}, Q(\cdot \mid A))$  é também um espaço de probabilidade.*

### A.4.2 Esperança

**Definição A.25.** Uma variável aleatória  $X$  é dita **discreta** se o número de valores possíveis para  $X$  for enumerável.

**Observação A.26.** Em palavras, uma variável aleatória discreta só pode assumir, no máximo, um número enumerável de valores possíveis  $x_1, x_2, \dots$ .

**Definição A.27.** Se  $X$  é uma variável aleatória discreta com medida de probabilidade  $Q$ , então a **esperança** ou **valor esperado** de  $X$ , denotada por  $E(X)$ , é definida por

$$E(X) = \sum_{j=1}^{\infty} x_j Q(A_j),$$

em que cada  $A_j \in \mathcal{F}$ .

Em palavras, a esperança de  $X$  é uma média ponderada dos possíveis valores que  $X$  pode receber, com cada valor sendo ponderado pela probabilidade de que  $X$  seja igual a esse valor.

### A.4.3 Distribuição Absolutamente Contínua

**Definição A.28.** Seja  $X$  uma variável aleatória. Suponha que o contradomínio ( $\mathbb{R}_x$ ) de  $X$  seja um intervalo ou uma coleção de intervalos. Então diremos que  $X$  é uma variável aleatória contínua.

**Definição A.29.** Dizemos que  $X$  é uma variável aleatória absolutamente contínua se existe uma função  $f_X : \mathbb{R} \rightarrow [0, +\infty)$  denominada função de densidade de probabilidade que satisfaz às seguintes propriedades:

- $f(x) \geq 0$ , para todo  $x \in \mathbb{R}_x$ ,
- $\int_{-\infty}^{\infty} f(x) dx = 1$ .

Além disso, definimos para qualquer  $c, d \in \mathbb{R}_x$ , com  $c < d$  que

$$P(c < X < d) = \int_c^d f(x) dx.$$

Note que, da forma como a probabilidade foi definida, a probabilidade de um ponto isolado é sempre zero. Dessa forma, podemos concluir que, quando  $X$  é uma variável aleatória contínua, a probabilidade de ocorrer um valor específico é zero.

**Observação A.30.** Se  $X$  é uma variável aleatória absolutamente contínua, então

$$\frac{\partial F_X(x)}{\partial x} = f_X(x).$$



## A.5 Algoritmo do Exemplo 1.12

A seguir, está o algoritmo mais detalhado de como foi realizado a simulação do exemplo 1.12 do primeiro capítulo, o software utilizado foi o MATLAB.

### Algoritmo A.31.

Entrada

n : números de pontos

bineq : matriz dos coeficientes para os casos com intervalos

beq : matriz dos coeficientes para igualdade

f : função objetivo

Saída

z : vetor com os valor da função objetivo

x : vetor com os valores para cada probabilidade que determina  
o valor da função objetivo

Passos

```
%% loop para criar o conjunto de restrições
```

```
for 1:n
```

```
%% Condição para  $P(X < 0)$ 
```

```
if  $P(X < 0)$ 
```

```
    Aineq: preencher a matriz de restrição, extraídas das esperanças  
           e colocar peso um para a probabilidade  $P(X < 0)$ 
```

```
    Aeq: preencher a matriz de restrição para que as probabilidades somem um
```

```
    lb: limitante inferior para que as probabilidades sejam maiores ou igual a zero
```

```
else
```

```
    Aineq: preencher a matriz de restrição, extraídas das esperanças  
           e colocar peso zero para a probabilidade  $P(X < 0)$ 
```

```
    Aeq: preencher a matriz de restrição para que as probabilidades somem um
```

```
    lb: limitante inferior para que as probabilidades sejam maiores ou igual a zero
```

```
end
```

```
end
```

```
%% loop para minimização e maximização
```

```
for 1:n
```

f: colocar peso 1 em cada loop

```
[x z] linprog(f, Aineq, bineq, Aeq, beq,lb) %% função de otimização do MATLAB
```

```
end
```

```
%% plotar gráfico
```

---

## Referências Bibliográficas

- [1] BERTSIMAS, D.; POPESCU, I. *On The Relation Between Option And Stock Prices: A Convex Optimization Approach*, Operations Research, Vol. 50, N° 2, March-April 2002, pp. 358 - 374.
- [2] BERTSIMAS, D.; TSITSIKLIS, J. N. *Introduction to LINEAR OPTIMIZATION*. Athena Scientific, Belmont, Massachusetts, 1997.
- [3] BOYD, S.; VANDENERGHE, L. *Convex Optimization*, Cambridge University Press, 2009.
- [4] BM&F BOVESPA- <[www.bmfbovespa.com.br](http://www.bmfbovespa.com.br)> Acessado em 03/10/2017.
- [5] COX, J. C.; ROSS, S.A.; RUBINSTEIN, M; *Option Princig: A Simplified Approach*, Journal Of Financial Economics 7, 1979.
- [6] FLORÊNCIO, P. H. B.; NETO, A. S. S.; DANTAS, M. J. P. *Análise do problema de Mont Hall: um enfoque bayesiano* - SAEPRO, 2014.
- [7] FINANCE YAHOO, <<https://finance.yahoo.com/quote/PETR4.SA>>. Acessado em 03/10/2017.
- [8] HURWICZ, L. *Programming in Linear Spaces. In Studies in Linear and Non-Linear Programming*. Stanford University Press, Stanford, Calif. 1958.
- [9] ISII, K. 1963. *On Sharpness of Tchebycheff-Type Inequalities*. Ann. Inst. Stat. Math. 14, 185-197.
- [10] LAPLACE, P. S. *De La Probabilité Des Témoignages.*, pp. 455-470.
- [11] LO, A. *Semiparametric upper bounds for option prices and expected payoffs*. J. Financial Econom. 19 373–388. 1987.
- [12] LUENBERGER, D. *Optimization by Vector Space Methods*. John Wiley, New York. 1969.
- [13] LUENBERGER, D. G. *Introduction to Linear and Nonlinear Programming*, Addison Wesley, 1984.

- 
- [14] MAO, J. C. T. *Quantitative analysis of financial decisions*. Toronto, Ontario, Collier-MacMillan Canada, 1969.
- [15] MARKOWITZ, H. *Portfolio selection: efficient diversification of investments*. New York, John Wiley and Sons, 1959.
- [16] MULHOLLAND, H.P. and ROGERS, C. A. *Representation Theorems for Distribution Functions*. Proc. London Math. Soc. (3) 8, 177 - 223. 1955.
- [17] OUWEHAND, P. *Foundations of Stochastic Finance* - Department of Mathematical Sciences Stellenbosch University. 2008.
- [18] PLISKA, S. R. *Introduction to Mathematical Finance - Discrete Time Models*, 1997.
- [19] ROSS, S. *Probabilidade : um curso moderno com aplicações*. 8. ed. São Paulo: Bookman, 2010. 608p.
- [20] SHAMBLIN, J. E. & STEVENS JR, G.T. *Pesquisa Operacional – Uma Abordagem Básica*. Editora Atlas, São Paulo/SP; 1979.
- [21] SMITH, J. E. *Generalized Chebychev Inequalities: Theory and Applications in Decision Analysis* - Operations Research, 1995.
- [22] VASARHELYI, M. A. *A utilização de modelos em administração financeira*. R. Adm. Emp., Rio de Janeiro, 1976.
- [23] WILLIAMS, D. *Probability with Martingales*. Statistical Laboratory, DPMMS. Cambridge University.
- [24] WHITTLE, P. *Non-linear programming*, in Handbook of Applicable Mathematics, volume IV, ch. 15, W. Ledermann and S. Vajda (eds.), New York: Wiley, 1982.