Inferência para o modelo Bernoulli na presença de adversários

Victor Junji Takara

Dissertação apresentada
ao
Instituto de Matemática e Estatística
da
Universidade de São Paulo
para
obtenção do título
de
Mestre em Ciências

Programa: Mestrado em Estatística

Orientador: Prof. Dr. Luís Gustavo Esteves

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro da CNPq

São Paulo, Março de 2021

Inferência para o modelo Bernoulli na presença de adversários

Esta é a versão original da dissertação/tese elaborada pelo candidato (Victor Junji Takara)

Agradecimentos

Vejo este trabalho como resultado de um esforço coletivo, cujo mérito pode ser atribuído, além de mim, àqueles que se dedicaram em minha formação direta ou indiretamente (que incluem desde meus pais que me deram condições para tal, professores que me ensinaram e até mesmo os pagadores de impostos que permitiram a manutenção da instituição em que estudei e da bolsa paga através da CNPq) e aos que se empenharam em desenvolver diversas ferramentas que utilizei.

Confesso que, por várias vezes desenvolvendo este trabalho, senti um verdadeiro alívio por encontrar resultados matemáticos que respondiam exatamente o que estava buscando e, portanto, começo agradecendo aos cientistas que tornaram essas passagens possíveis.

Optarei por não escrever nomes, pois da última vez que o fiz, pude notar, só depois, a ausência de pessoas muito importantes para a fase em que passei.

Agradeço à minha família e aos meus amigos, que me dão energia para seguir com todas as minhas atividades com mais entusiasmo, além de me ajudarem de inúmeras formas a transpor momentos difíceis, como o que todos estamos passando de 2020 para cá.

Agradeço aos meus professores, tanto àqueles que me ajudaram a compor esse trabalho com críticas, sugestões e discussões, quanto àqueles que me ensinaram a base de tudo o que foi utilizado. Em especial, agradeço ao meu orientador por todo o seu empenho, paciência e sagacidade, com quem pude discutir e amadurecer diversos tópicos, além de ter me apresentado de forma clara tantos temas da estatística que sempre me foram obscuros. Agradeço também como amigo, pelo apoio e pelas inúmeras conversas motivantes que tivemos.

Resumo

TAKARA, V. J. **Inferência para o modelo Bernoulli na presença de adversários**. 2021. 57 f. Dissertação (Mestrado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2021. A teoria da decisão com adversários se originou na tentativa de solucionar problemas na área de aprendizado de máquina. Nessa teoria, supõe-se a existência de adversários que têm como intuito a perturbação dos dados (ou do mecanismo gerador dos mesmos). Uma vez que ela é baseada em inferência bayesiana, a todas as incertezas são atreladas medidas de probabilidade, inclusive às possíveis ações realizadas por adversários. No entanto, pela natureza aplicada da teoria, ela foi criada e estudada com enfoque na teoria da decisão, sem muita preocupação com formalismos na área de estatística. Assim, o objetivo desse trabalho foi estudar elementos inferenciais importantes, como a estimação pontual e o teste de hipóteses para o modelo Bernoulli na presença de adversários. Ilustramos como essas alterações impactam a estimativa pontual e o teste de hipótese bayesiano, além da própria distribuição dos dados observáveis e de componentes importantes, como o comportamento do risco bayesiano e regiões críticas.
Palavras-chave: teoria da decisão, inferência bayesiana, aprendizado de máquina

Abstract

TAKARA, V.J. **Inference for Bernoulli model in presence of adversaries**. 2021. 57 f. Dissertation (Master) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2021. The adversarial decision theory originated in the attempt to solve issues in the area of machine learning. In this theory, it is assumed that there are adversaries whose intention is to disturb the data (or the mechanism which generates them). Since it is based on Bayesian inference, probabilities measures are attached to all uncertain quantities and to possible actions taken by opponents. However, due to the applied nature of this theory, it was created and studied focused on decision theory and its applications, without much concern with statistical formalisms. Thus, the objective of this work was to study important inferential concepts, such as point estimation and hypothesis testing for the Bernoulli model in presence of adversaries. We illustrate how these changes impact the point estimate and the Bayesian hypothesis test, besides the distribution of observable data and important statistical elements such as the behavior of Bayesian risk and critical regions.
Keywords: decision theory, bayesian inference, machine learning.

Sumário

Lista de Abreviaturas

CIID	Condicionalmente independentes e identicamente distribuídas (*Conditionally independent and identically distributed*)
IID	Idenpendentes e identicamente distribuídas (*independent and identically distributed*)
VA	Variável(is) aleatória(s) (*random variable(s)*)
TD	Teoria da decisão (*Decision theory*)
TDA	Teoria da decisão com adversários (*Adversarial decision theory*)

Lista de Símbolos

Ω	Espaço Paramétrico
$\mathcal{X}$	Espaço amostral
k	$\sum_{i=0}^{n} y_i$
$\boldsymbol{y}$	Vetor amostral $(y_1, y_2, ..., y_n)$
$\boldsymbol{x}$	Vetor amostral $(x_1, x_2, ..., x_n)$
$\overset{q.c.}{\to}$	Converge quase certamente
$\mathbb{1}_A(x)$	Função indicadora na variável x para a região A
$\beta(a, b)$	Função beta aplicada em (a, b)
$B(z, a, b)$	Função beta incompleta aplicada em (z, a, b)
$\psi(x)$	Função digama aplicada ao ponto x
H_n	n-ésimo número harmônico clássico
$F_{p,q}$	Função hipergeométrica generalizada de p parâmetros do numerador e q parâmetros do denominador

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

1.1 Considerações Preliminares

Imagine um cenário no qual existem duas entidades envolvidas com interesses antagônicos, sendo que uma assume a posição de defesa e a outra a posição de ataque. O agente defensor tem interesse em tomar alguma decisão com base nos dados que tem disponíveis, enquanto o atacante realizará perturbações (que podem ser tanto nos dados diretamente quanto no mecanismo que os gera) com a intenção de fazer com que o defensor tome decisões equivocadas.

Todavia, o agente decisor (defensor) pode ter a perspicácia de que seus dados não estão íntegros e utilizar um modelo que considera a possível ação do atacante. O que descrevemos brevemente até então é o contexto do estudo da teoria da decisão (TD) com adversários (Insua *et al.* (2018)), que se originou do estudo da resolução de problemas na área de aprendizado de máquina (Biggio e Roli (2018)).

A teoria da decisão com adversários (TDA) é, em linhas gerais, bem ampla e engloba em princípio todos os modelos nos quais a incerteza acerca da ação do atacante pode ser expressa com probabilidades.

Esse tema foi apresentado em artigos científicos apenas recentemente (a partir de 2004, segundo Biggio e Roli (2018)) e é utilizada em problemas importantes como análise de riscos (Ríos e Insua (2012)), segurança cibernética (Insua *et al.* (2019)), segurança biométrica (Biggio *et al.* (2015)) e também na área médica (Singpurwalla *et al.* (2016)).

Muitos desses trabalhos fazem uso de modelos com adversários sob abordagem da área de aprendizado de máquina, que tende a ser diferente da abordagem estatística, uma vez que o aprendizado de máquina é um método composto pela combinação de técnicas estatísticas, ciência da computação e otimização, empenhando todos esses elementos na construção de um algoritmo de aprendizagem (Mohri *et al.* (2012)).

Problemas básicos e importantes em estatística, como a estimação e teste de hipóteses, foram apresentados apenas recentemente (González-Ortega *et al.* (2019); Insua *et al.* (2018)), havendo portanto ainda muito a ser explorado.

1.2 Objetivos

O objetivo deste trabalho é o estudo dos problemas de estimação e testes para o modelo Bernoulli sob a perspectiva da TDA. Mais precisamente, consideram-se contextos de ataques baseados em

uma situação apresentada por González-Ortega *et al.* (2019).

1.3 Organização do Trabalho

O trabalho inicia, no capítulo 2, com a introdução de conceitos importantes para o desenvolvimento das soluções propostas nos últimos capítulos. Serão apresentadas algumas ideias de como a probabilidade, como uma função matemática que mede incerteza, pode ser construída a partir da crença, além de uma breve discussão sobre a noção de permutabilidade em modelos estatísticos.

Ainda no capítulo 2, relembramos alguns conceitos básicos em Inferência Bayesiana (para a realização da operação bayesiana) e TD (para estudo dos problemas de estimação e teste de hipóteses). Mais precisamente, no capítulo 3, apresentamos os modelos com adversários e o TDA. Nesse capítulo, após introduzida a base conceitual, são apresentadas soluções para os problemas de estimação e testes de hipóteses para o modelo Bernoulli segundo a TD para algumas situações com adversários.

No capítulo 4, apresentamos um exemplo de aplicação do modelo estudado e, no capítulo 5, concluímos. Além disso, demonstrações e resultados matemáticos necessários se encontram no apêndice.

Capítulo 2

Conceitos básicos

2.1 Probabilidade, crença e incerteza subjetivas

Talvez um dos exemplos mais corriqueiros para se introduzir a noção de incerteza seja o do resultado do lançamento de uma moeda. Os resultados possíveis de lançamento residem na percepção do indivíduo que contempla o cenário. Um indivíduo poderia contemplar, por exemplo, que a moeda poderia cair com face cara ou coroa ou mesmo nenhuma das duas, caindo apoiada em sua borda. Pode ser razoável pensar que nem todas as pessoas terão as mesmas impressões acerca do experimento e que, talvez, nem todas concebam sequer a possibilidade da moeda não cair com uma das faces voltada para cima. Dessa forma, torna-se evidente que a incerteza dependerá do indivíduo com o cenário que contempla.

A reflexão acerca dessa subjetividade da incerteza pode ser mais bem aprofundada em Lindley (2014). Ela é também um dos alicerces em Inferência Bayesiana. Sob essa óptica, o grau de incerteza sobre os eventos são mensurados a partir de uma função matemática que chamamos de probabilidade[1]. Assim, a probabilidade nunca poderá existir sem haver atrelada a si um evento aleatório e um indivíduo que o contemple.

Do ponto de vista matemático, a probabilidade será uma função que ligará um conjunto de eventos considerados por um indivíduo a um conjunto dos números reais no intervalo $[0, 1]$, de acordo com suas crenças, sendo $\mathbb{P}(.)$ a notação que utilizaremos para denotá-la. Mais detalhes sobre a Teoria de Probabilidade podem ser encontrados em Schervish (1995).

Ao considerarmos um dado evento A, caso um indivíduo creia que esse evento certamente ocorrerá, ele deveria atribuir $\mathbb{P}(A) = 1$. Por outro lado, caso ele atribuísse $\mathbb{P}(A) = 0$, significaria que a percepção do indivíduo é de que A não poderia ocorrer de forma alguma.

Em DeGroot (2004), podemos ver como é possível formalizar essa relação entre a crença e a incerteza, esta traduzida por uma medida de probabilidade. Essa construção, na qual o autor apresenta algumas suposições necessárias para se estabelecer a conexão entre a medida de probabilidade e a relação de crença pessoal e nos permitir o uso da função $\mathbb{P}$ para a medição da incerteza. Essas ideias são revisadas brevemente no que segue.

Assim, primeiro, introduziremos algumas notações. Sejam A_1 e A_2 dois eventos, subconjuntos de um conjunto $\Omega \neq \emptyset$, Definimos as relações sobre os eventos da seguinte maneira:

- $A_1 \prec A_2$: o indivíduo crê mais em A_2 do que em A_1 (isto é,$A_2 \succ A_1$ equivale a $A_1 \prec A_2$)

[1]Note que, se assumirmos que a incerteza é subjetiva, a probabilidade também será

- $A_1 \succ A_2$: o indivíduo crê mais em A_1 do que em A_2
- $A_1 \sim A_2$: o indivíduo crê em A_1 tanto quanto em A_2

A partir dessas notações, definimos ainda:

- $A_1 \precsim A_2$ se e somente se $A_1 \sim A_2$ ou $A_1 \prec A_2$
- $A_1 \succsim A_2$ se e somente se $A_1 \sim A_2$ ou $A_1 \succ A_2$

Considerando as relações apresentadas, algumas condições são impostas, como segue:
Para A_1, A_2 eventos,

1. apenas uma única das relações pode valer: $A_1 \prec A_2$, ou $A_1 \succ A_2$ ou $A_1 \sim A_2$
2. Se $A_1 \cap A_2 = B_1 \cap B_2 = \emptyset$, $A_1 \precsim B_1$ e $A_2 \precsim B_2$, então $A_1 \cup A_2 \precsim B_1 \cup B_2$. Além disso, se $A_1 \prec B_1$ ou $A_2 \prec B_2$, então $A_1 \cup A_2 \prec B_1 \cup B_2$
3. $\emptyset \precsim A$ e $\emptyset \prec \Omega$, para todo evento A
4. Se $A_1 \supset A_2 \supset A_3 \supset ...$ é uma sequência decrescentes de eventos e B é um evento fixado tal que $A_i \succsim B$, $\forall i \in \mathbb{N}^*$, então $\bigcap_{i=1}^{\infty} A_i \succsim B$

A partir dessas suposições, DeGroot (2004) faz uma construção na qual demonstra a existência de uma única medida de probabilidade associada à crença $\precsim$ sobre os eventos contemplados pelo indivíduo no sentido de que $A \precsim B \iff \mathbb{P}(A) \leq \mathbb{P}(B)$ e uma função de probabilidade. Podem ser provados diversos outros resultados a partir dessas suposições na mesma referência.

Considerando então essa construção da relação entre crença e incerteza, podemos aplicá-la a um exemplo encontrado em Lindley (2014), no qual o autor relaciona e compara a incerteza (probabilidade) de um dado evento com as probabilidades de extrações de bolas de uma urna. Iremos ilustrar essa construção utilizando o exemplo da moeda apresentado.

Exemplo 2.1.1 *Suponha que existam: uma urna com n bolas, sendo r bolas vermelhas e $n-r$ brancas, $n \in \mathbb{N}^*$, $r \in \mathbb{N}$, $n \geq r$, um determinado evento V, correspondente a retirar uma bola vermelha da urna ao acaso e um evento E, correspondente ao resultado "cara" no lançamento de uma moeda, cuja ocorrência é incerta para um indivíduo. Se esse indivíduo considerar que todas as bolas são indistinguíveis entre si, então, a probabilidade de se retirar uma bola vermelha da urna ao acaso deverá ser a proporção de bolas vermelhas na urna, ou seja, $\mathbb{P}(V) = \frac{r}{n}$, pois a retirada de qualquer uma será equiprovável.*

Sendo assim, caso ele creia mais no resultado "cara" no lançamento da moeda em questão do que na retirada de uma bola vermelha da urna ($E \succ V$), pelas suposições, temos que $\mathbb{P}(E) > \mathbb{P}(V) = \frac{r}{n}$, ou seja, o resultado "cara" no lançamento da moeda é menos incerto do que a retirada de uma bola vermelha ao acaso nessa urna hipotética.

Caso ele creia que ambos os eventos sejam igualmente incertos ($E \sim V$), consideraria $\mathbb{P}(E) = \mathbb{P}(V) = \frac{r}{n}$ e, caso ele creia que o evento E é mais incerto do que a retirada de uma bola vermelha, $\mathbb{P}(E) < \frac{r}{n}$.

No parágrafo anterior, podemos ver que as possíveis probabilidades de retiradas de bolas vermelhas da urna são sempre números racionais no intervalo [0,1], enquanto a incerteza poderia assumir qualquer valor real nesse intervalo. Porém, a característica de que o conjunto dos números racionais é denso no conjunto dos reais (Lima (2019)) permite que consigamos uma aproximação de um número real a partir de um racional tão boa quanto gostaríamos e, portanto, podemos descrever a incerteza utilizando urnas com uma aproximação tão boa quanto for desejável (tomando-se urnas com grande número de bolas).

Por exemplo, supondo que a incerteza acerca do evento E é $\frac{\pi}{4} = 0,785398...$, uma urna com 8 bolas vermelhas e 2 bolas brancas nos daria uma probabilidade de $0,8$ de retirada de uma bola vermelha, dando uma margem de erro de $0,014602...$ em relação à incerteza real do indivíduo sobre E. Porém, poderíamos considerar a urna com 7854 bolas vermelhas e 2146 bolas brancas, resultando que a probabilidade de retirar uma bola vermelha ao acaso seria de $0,7854$, uma margem de erro muito menor ($< 0,0001$) do que a anterior em relação à probabilidade de E. Poderíamos, assim, obter uma aproximação tão precisa quanto fosse de interesse para a probabilidade de qualquer evento E simplesmente mudando a composição e o número de bolas na urna.

2.2 Modelos probabilísticos permutáveis

Nesta seção, faremos uma breve discussão sobre o conceito de permutabilidade e modelos permutáveis, uma vez que todos os modelos que desenvolveremos neste trabalho serão dessa classe.

Definição 2.2.1 *Um conjunto finito de VA $X_1, X_2..., X_n$ é dito permutável se qualquer permutação de $(X_1, X_2..., X_n)$ possuir a mesma distribuição conjunta de $(X_1, X_2..., X_n)$. Uma coleção infinita é permutável se toda subcoleção finita é permutável.*

Isso significa que, dado um conjunto de VA permutáveis, a ordem em que ocorreram as realizações é uma informação irrelevante para se fazer inferência.

Vamos apresentar a versão do Teorema da Representação de De Finetti para VA de Bernoulli, descrita em Schervish (1995). Sua demonstração, assim como sua versão mais geral para demais VA pode ser encontrada na mesma referência.

Teorema 2.2.1 *Uma sequência infinita de VA $(X_n)_{n\geq 1}$ de Bernoulli é dita permutável se, e somente se, existe uma VA Θ com suporte em $[0,1]$, tal que, condicionada em $\Theta = \theta$, $(X_n)_{n\geq 1}$ são VA IID Bernoulli(θ). Além disso, se a sequência é permutável, então a distribuição de Θ é única e $\sum_{i=1}^{n}(\frac{X_i}{n})_{n\geq 1}$ converge quase certamente para Θ*

De forma geral, teoremas da representação tipo de De Finetti nos dão uma motivação para o uso de modelos permutáveis: mesmo que diferentes estatísticos com diferentes crenças sobre $(X_n)_{n\geq 1}$ calculem probabilidades distintas para a mesma sequência $(X_n)_{n\geq 1}$ de VA de Bernoulli, caso eles creiam que essa sequência é permutável, então todos eles creem que existe uma VA Θ tal que, condicionado a $\Theta = \theta$, as VA são IID Bernoulli(θ) (Schervish (1995)).

Além disso, como consequência de 2.2.1, eles acreditam que $\lim_{n\to\infty} \sum_{i=1}^{n} \frac{X_i}{n} \overset{q.c.}{\to} \Theta$. Ou seja, independente das crenças sobre θ estabelecidas à priori para eles, $\sum_{i=1}^{n} \frac{X_i}{n}$ convergirá.

Lindley e Novick (1981) apresentam a ideia de permutabilidade referida na definição 2.2.1 como um conceito relacionado à ideia de subpopulação apresentada por Fisher (1956). Essa relação será explicada através do exemplo a seguir, no qual foram usados dados reais de um estudo na área médica apresentado em Julious e Mulee (1994).

Exemplo 2.2.1 *Pesquisadores estavam tentando comparar dois tratamentos diferentes, cirurgia aberta e NLPC*[2]*, para a eliminação de cálculos renais de dois diâmetros diferentes (*$< 2cm$ *e* $\geq 2cm$*). Desta forma, foi considerado sucesso quando o tratamento levou à eliminação das pedras e fracasso, caso contrário. Os resultados são apresentados na* ***Tabela 2.1****.*

Diâmetro	Tratamento	
	Cirurgia aberta	NLPC
$<$ 2cm	93%(81/87)	83%(234/270)
$\geq$ 2cm	73%(192/263)	69%(55/80)

Tabela 2.1: ***Taxa de sucessos por tratamento***

Ao analisarmos os números, aparentemente, a cirurgia aberta apresenta maiores taxas de sucesso para ambos os tamanhos de cálculos renais, $< 2cm$ *(*93% *contra* 83%*) e* $\geq 2cm$ *(*73% *contra* 69%*). Porém, se agruparmos os sucessos por tratamento, teremos que a cirurgia aberta apresenta uma taxa de sucesso de* 78% *(*273/350*) e a NLPC apresenta uma taxa de sucesso de* 83% *(289/350), o que aparenta ser contraditório. A esse fenômeno, damos o nome de* ***Paradoxo de Simpson****.*

Lindley e Novick (1981) tentam explicar que o efeito do Paradoxo de Simpson decorreria da ausência de permutabilidade. Porém, Pearl (2009) relata que isso se deve, na verdade, ao conceito de causalidade[3], ideia que ainda não era formalmente tratada pelas ferramentas matemáticas da época.

Lindley e Novick (1981) defendem que as suposições relativas à causalidade devam ser feitas primeiro, para então ser considerada a suposição de permutabilidade. Assim, no exemplo 2.2.1, vemos que existe uma relação causal comum entre a escolha entre cirurgia aberta ou NLPC e os tamanhos dos cálculos renais (médicos tenderam a realizar mais cirurgias abertas em casos em que o cálculo renal era maior e mais NLPC nos caos em que ele era menor), de forma que a hipótese de permutabilidade condicional ao tamanho do cálculo renal ou ao tratamento individualmente não parecem razoáveis.

Pearl (2009) descreve, portanto, que a suposição de permutabilidade em situações práticas depende da estrutura de causa e efeito entre variáveis (mecanismo causal) envolvida na geração dos dados. Dessa forma, se realizássemos um experimento com n indivíduos, ao realizarmos a observação de um novo (n+1-ésimo) indivíduo, a manutenção da suposição de permutabilidade dependeria das condições do experimento, no qual, apenas caso fossem idênticas para todos os $n + 1$ indivíduos, poderia ser assumida.

Uma forma bastante prática de analisar se a suposição de permutabilidade se sustenta é verificar se existe o que Greenland e Robins (1986) chamam de equivalência de resposta. Para entender

[2]NLPC (Nefrolitotomia percutânea) é uma cirurgia que permite a remoção de cálculos renais a partir de uma pequena incisão, sendo considerada menos invasiva do que a cirurgia aberta

[3]De forma simplista, tomando dois eventos A e B quaisquer, dizemos que existe a relação de causalidade se A causa B ou se B causa A

essa ideia, vamos considerar um experimento em que temos grupos balanceados em que cada indivíduo recebeu ou não um dado tratamento e com isso é medida uma variável resposta. Caso pudéssemos inverter os indivíduos que receberam ou não o tratamento de modo que todos que receberam deixariam de receber e todos que não receberam passassem a recebê-lo, sob a suposição de permutabilidade entre indivíduos, deveríamos ter a mesma distribuição dos dados.

2.3 Inferência bayesiana

No paradigma bayesiano, todas as quantidades incertas devem ser tratadas como VA e, portanto, devemos atribuir medidas de probabilidade para elas (Schervish (1995)). Sob essa abordagem, uma das ideias centrais é o uso da operação bayesiana, baseado no teorema de Bayes, apresentado na forma discreta em (DeGroot e Schervish):

Teorema 2.3.1 *Considere os eventos $B_1, ..., B_k$ que compõem a partição de um espaço S de eventos, tal que $\mathbb{P}(B_j) > 0, \forall j \in \{1, ..., k\}$ e seja A um evento tal que $\mathbb{P}(A) > 0$. Então, para $i \in \{1, ..., k\}$,*

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(B_i)\mathbb{P}(A|B_i)}{\sum_{j=1}^{k} \mathbb{P}(B_j)\mathbb{P}(A|B_j)}$$

Sem perda de generalidade, podemos utilizar uma versão mais geral desse teorema (que é demonstrado em Schervish (1995)), também para medidas absolutamente contínuas ou mistas. Esse teorema estabelece uma relação entre eventos que são partições do espaço ($B_i, i \in \{1, ..., k\}$) e um dado evento qualquer (A) no espaço.

Se tomarmos o evento B_i como sendo $\Theta = \theta$ e A a realização da amostra $\boldsymbol{X} = \boldsymbol{x}$, temos uma operação que atualiza uma crença inicial sobre Θ com as informações dadas pela amostra $\boldsymbol{X} = \boldsymbol{x}$ através do conjunto de conceitos apresentados em 2.1 para a construção da medida de incerteza P_θ.

Denotaremos o nome **distribuição a priori** (ou, simplesmente, priori) àquela que representa essa crença inicial sobre Θ, medida P_θ, com densidade $p_\Theta(\theta)$. A partir da priori e da distribuição condicional de $X|\theta$, podemos realizar a operação bayesiana:

$$p_{\Theta|\boldsymbol{X}}(\theta|\boldsymbol{x}) = \frac{p_{\boldsymbol{X}|\Theta=\theta}(\boldsymbol{x}|\theta)p_\Theta(\theta)}{\int_\Theta p_{\boldsymbol{X}|\Theta=\theta}(\boldsymbol{x}|\boldsymbol{\theta}))\, dP_\theta} \tag{2.1}$$

Chamamos a distribuição correspondente da densidade calculada em 2.1 de **distribuição a posteriori**.

No teorema 2.2.1, verificamos que, caso consideremos a amostra $X|\Theta = \theta$ como resultado de um processo de Bernoulli(θ), então é assegurada a existência de uma única medida P_θ. Como mencionado anteriormente, o teorema apresentado possui generalizações para demais tipos de VA em Schervish (1995).

Uma possível interpretação da inferência bayesiana é uma estrutura matemática de como medimos e atualizamos a crença sobre o parâmetro Θ com base nos dados $\boldsymbol{x}$. No entanto, trabalhar com o conjunto $\boldsymbol{x}$ de dados pode ser matematicamente mais difícil. Contornando esse problema, existem muitos casos em que pode ser mais fácil trabalhar com funções de $\boldsymbol{x}$, o que nos remete à ideia de

utilizarmos estatísticas. Matematicamente, estatísticas são funções $T : \mathcal{X} \to \mathcal{T}$, com $\mathcal{T}$ mensurável (ou seja, $\mathcal{T}$ também é VA).

No entanto, ao trabalharmos com estatísticas, é possível que haja alguma perda de informação relevante para fazermos inferência, o que nos leva ao próximo conceito a ser apresentado: a suficiência.

2.3.1 Suficiência

A ideia da suficiência estatística está atrelada à capacidade de síntese dos dados. O exemplo abaixo encontrado em DeGroot e Schervish será apresentado de forma heurística e dá uma boa ideia do que é uma estatística suficiente.

Exemplo 2.3.1 *Suponha que, para um dado problema, dois estatísticos, A e B, desejam obter informações a respeito de um certo parâmetro θ. Além disso, temos que o conjunto de observações $(\boldsymbol{X} = X_1, X_2, ..., X_n)$ é uma amostra aleatória simples observável para A, mas não para B. Contudo, vamos supor que B tem acesso a uma certa estatística $T(\boldsymbol{X})$. A estatística $T(.)$ usada por B será dita suficiente para o modelo, se e somente se, a informação obtida através dela por B for a mesma de A, que possui a amostra inteira e, portanto, acesso a qualquer estatística de $\boldsymbol{X}$.*

Formalmente, podemos utilizar a definição encontrada em Schervish (1995):

Definição 2.3.1 ***Estatística suficiente bayesiana***
Considere $\mathcal{P}_0$ uma família de distribuições paramétricas em ($\mathcal{X}$,$\mathcal{B}$), (Ω,τ) o espaço paramétrico e $\Theta : \mathcal{P}_0 \to \Omega$. Uma estatística $T : \mathcal{X} \to \mathcal{T}$ é dita suficiente bayesiana se, para qualquer priori P_Θ, existem versões das posterioris $P_{\Theta|\boldsymbol{X}}$ e $P_{\Theta|T}$ tais que, para todo $B \in \tau$, $P_{\Theta|\boldsymbol{X}}(B|\boldsymbol{x}) = P_{\Theta|T}(B|T(\boldsymbol{x}))$, $P_{\boldsymbol{X}}$-quase-certamente.

Em outras palavras, num modelo bayesiano, uma estatística suficiente T é tal que, para qualquer que seja a escolha da distribuição à priori, as distribuições à posteriori de θ condicionadas à amostra inteira $\boldsymbol{x}$ e à $T(\boldsymbol{x})$ são as mesmas a menos para elementos de medida nula dentre aquelas que compõem a família de distribuições paramétricas adotada.

2.4 Teoria da decisão bayesiana

A teoria da decisão (TD) é uma vasta área de estudos que envolve o entendimento do julgamento humano e de modelos normativos para a tomada de decisão, além de estudos experimentais de comportamento individual e coletivo (Mendoza e Gutiérrez-Peña (2010)). Em estatística, o termo se refere a uma classe de problemas estatísticos nos quais os estatísticos devem ganhar informação sobre alguns parâmetros de interesse para que estejam aptos a tomar as decisões mais efetivas em situações cujas consequências das decisões dependerão dos valores desses parâmetros (DeGroot (2004)).

Nessa abordagem, pressupõe-se a existência de um ou mais agentes racionais, que terão que tomar decisões baseando-se na maximização de uma função denominada utilidade esperada, que seria uma forma de mensurar o ganho ou perda média derivados de vantagens e desvantagens por conta da ocorrência hipótetica de cada evento (Myerson (1991)).

É reconhecido que a hipótese de racionalidade dos tomadores de decisão não pode ser sempre satisfeita, porém, o estudo sob essa hipótese nos ajuda a ter subsídios para fabricar modelos simplificados (mas ainda informativos) de situações muito mais complexas (Myerson (1991), Neumann e Oskar (1955)).

Mesmo sendo uma simplificação do que observamos no mundo, a TD ainda pode abranger diferentes formas de modelagem. Particularmente, estamos interessados numa classe de modelos também compreendida como a teoria das decisões estatísticas ótimas, sobretudo, a teoria da decisão estatística subjetiva ou Bayesiana, como é denominada em DeGroot (2004). Através dela, podemos aplicar os conceitos já apresentados de probabilidade e incerteza subjetivas como elementos da função de utilidade esperada, a ser apresentada.

2.4.1 Utilidade e perda

A utilidade é um conceito amplamente discutido na filosofia e na economia e passou por diversas transformações ao longo da história. Ele deu origem a diversas áreas de estudo, como a teoria da decisão (Moscati (2020)) e novas linhas de estudo na área da economia aplicada (Stigler (1950)), como por exemplo, a economia comportamental (Moscati (2020)).

Bentham (1780) apresentou a utilidade como um princípio relativo à "propriedade que todo o indivíduo tem de tender a produzir, em algum sentido, benefício, vantagem, prazer ou felicidade ou, de forma análoga no sentido oposto, de prevenir prejuízo, desvantagem, dor ou infelicidade".

Através desse princípio, Bentham sugeriu que haveria como realizar a medição de dores e prazeres com o propósito de construir um sistema civil mais racional(Stigler (1950)). Apesar desse movimento ter falhado na época, considera-se que ele tenha dado origem a várias outras linhas de reflexão por parte de outros filósofos e pesquisadores, o que resultou, cerca de 100 anos depois, nas primeiras representações matemáticas da utilidade por Jevons, Menger e Walras (Stigler (1950), Moscati (2020)).

Reconhecidamente, houve ainda muitos problemas na forma matemática apresentada inicialmente e, portanto, foram realizadas diversas tentativas de desenvolver ferramentas para mensurar a utilidade de maneira mais formal (Stigler (1950)). Uma ferramenta mais consolidada para a sua mensuração só viria a ser desenvolvida em 1944 (Moscati (2020)), quando a teoria da utilidade esperada foi apresentada formalmente por John von Neumann e Oskar Morgenstern que pode ser encontrada em Neumann e Oskar (1955).

Dentre as diversas construções matemáticas de utilidade, relembraremos brevemente a construção encontrada em DeGroot (2004). Primeiro, vamos introduzir algumas notações, seguindo então com as definições. Considere um agente decisor com um conjunto R de recompensas. Sejam duas recompensas $r_1, r_2 \in R$. A relação do agente preferir a recompensa r_2 a r_1 é denotada por $r_1 \prec^* r_2$, enquanto se ele preferir r_2 ao menos tanto quanto r_1, $r_1 \precsim^* r_2$, ou ainda $r_2 \succsim^* r_1$. Caso ele prefira r_1 tanto quanto r_2, a notação será $r_1 \sim^* r_2$.

Definição 2.4.1 *Sejam as distribuições $P_1, P_2 \in \mathcal{P}$, $U(.)$ é dita função de utilidade se ela possuir a seguinte propriedade: $P_1, P_2 \in \mathcal{P}$ tais que existem $E(U|P_1)$ e $E(U|P_2)$. Então, $P_1 \precsim^* P_2 \iff E(U|P_1) \leq E(U|P_2)$.*

Definição 2.4.2 *Para qualquer distribuição $P \in \mathcal{P}$, a função $E(U|P)$, quando existir, será chamada utilidade de P, também conhecida como utilidade esperada.*

Definição 2.4.3 *A função de perda é definida por* $L : (D, \Theta) \to \mathcal{R}$*, tal que, para todo* $d \in D$*,* $L(., d)$ *é mensurável no espaço paramétrico* Θ*.*

Definição 2.4.4 *Para toda decisão* $d \in D$*, chamamos perda esperada ou risco, a função:*

$$\rho(P_\theta, d) = \int_{\theta \in \Theta} L(\theta, d)\, dP_\theta$$

DeGroot (2004) argumenta que, se possível, um estatístico deveria escolher, dado um conjunto de decisões D, a decisão $d \in D$ que maximizasse a utilidade de P_d. Além disso, ele mostra que maximizar a utilidade de P_d é o mesmo que minimizar a função de risco $\rho(P_\theta, d)$, quando ambas existem. As decisões derivadas da minimização da função de risco são tipicamente chamadas de **decisões ótimas**.

Uma outra formulação para a tomada de decisão é feita através do que chamamos de testes de hipóteses, o que será apresentado a seguir.

2.4.2 Testes de Hipóteses

Seja $d \in \mathcal{D}$, em que $\mathcal{D}$ é o espaço de decisões, testes de hipóteses são funções de decisão do tipo $D : \mathcal{X} \to d$ que nos permitem chegar a uma escolha entre hipóteses estatísticas para um determinado tipo de problema. Definimos testes de hipóteses conforme em Schervish (1995):

Definição 2.4.5 *Seja* Ω *o espaço paramétrico uma partição formada por* Ω_H *e* Ω_A*, tais que* $\Omega_H \cup \Omega_A = \Omega$ *e* $\Omega_H \cap \Omega_A = \emptyset$*. Definimos as hipóteses* $H : \theta \in \Omega_H$ *e* $A : \theta \in \Omega_A$*, denominando* H *como a hipótese nula e* A *a hipótese alternativa.*

Um problema de decisão é dito teste de hipóteses se, dado uma função de decisão $\mathcal{D} : \mathcal{X} \to d$*,* $d \in \{0, 1\}$ *e uma função de perda* $L(\theta, d)$*, são satisfeitas as condições que* $L(\theta, 1) > L(\theta, 0)$*,* $\forall \theta \in \Omega_H$ *e* $L(\theta, 1) < L(\theta, 0)$*,* $\forall \theta \in \Omega_A$*. Em que a decisão* $d = 1$ *é a rejeição da hipótese* H *e a decisão* $d = 0$ *é a aceitação ou não rejeição da mesma. Além disso, determinamos* $\mathcal{R}$ *a região de rejeição, que é o conjunto de todos os valores de* $\mathcal{X}$ *tais que a hipótese* H *é rejeitada.*

Podemos definir também a função teste conforme Casella e Berger (2016):

Definição 2.4.6 *Definimos como função teste uma função* $\varphi : \mathcal{X} \to \{0, 1\}$ *tal que* $\varphi(\boldsymbol{X}) = 1 \iff \boldsymbol{X} \in \mathcal{R}$*. Ou seja,* $\varphi(\boldsymbol{X}) = \mathbb{1}_{\mathcal{R}}(\boldsymbol{X})$*, em que* $\mathbb{1}_A(z)$ *é a função indicadora aplicada no ponto* z *sobre a região* A*.*

Assim, uma construção possível de teste é o teste bayesiano no qual tomamos as hipóteses: $H : \theta \in \Omega_0$ vs $A : \theta \in \Omega_1$, com $\Theta = \Omega_0 \cup \Omega_1$. Utilizando a perda 0-1-c (descrita na tabela **Tabela 2.2**), com uma amostra $\boldsymbol{X}$:

Decisão	Ω_0	Ω_1
d_0	0	1
d_1	c	0

Tabela 2.2: ***Tabela de decisão para a perda*** 0-1-c

Através dessa tabela de perda, temos os riscos por decisão:

$$\begin{aligned}\rho(P_{\theta|\boldsymbol{x}}, d_0) &= \mathbb{P}(\theta \in \Omega_1|\boldsymbol{x}) \\ \rho(P_{\theta|\boldsymbol{x}}, d_1) &= c\mathbb{P}(\theta \in \Omega_0|\boldsymbol{x})\end{aligned} \tag{2.2}$$

Como há apenas duas decisões possíveis, d_0 e d_1, o menor risco para decidir por d_0 será dada sempre que:

$$\mathbb{P}(\theta \in \Omega_1|\boldsymbol{x}) < c\mathbb{P}(\theta \in \Omega_0|\boldsymbol{x}) \iff 1 - \mathbb{P}(\theta \in \Omega_0|\boldsymbol{x}) < c\mathbb{P}(\theta \in \Omega_0|\boldsymbol{x}) \iff \mathbb{P}(\theta \in \Omega_0|\boldsymbol{x}) > \frac{1}{c+1}$$

Assim, decidimos por d_1 (rejeitamos H)

$$\mathbb{P}(\theta \in \Omega_0|\boldsymbol{x}) \leq \frac{1}{c+1} \tag{2.3}$$

e aceitamos H, caso contrário. Portanto, neste caso, a região de rejeição é determinada por $\mathcal{R} = \{\boldsymbol{x} \in \mathcal{X}|\mathbb{P}(\theta \in \Omega_0|\boldsymbol{x}) \leq \frac{1}{c+1}\}$.

Note que, para $\theta \in \Omega_0$, $L(\theta, 1) = 1, L(\theta, 0) = 0 \implies L(\theta, 1) => L(\theta, 0)$. Além disso, se $\theta \in \Omega_1$, $L(\theta, 1) = 0, L(\theta, 0) = 1 \implies L(\theta, 1) < L(\theta, 0)$.

Além disso, de acordo com a tabela, consideramos portanto que, nesse caso a perda é 0 quando a decisão correta é tomada, enquanto a perda é 1, caso a decisão d_0 seja tomada quando $\theta \in \Omega_1$ e é c quando a decisão d_1 é tomada quando $\theta \in \Omega_0$, penalidades associadas aos dois tipos de erros de decisão possíveis pela nossa formulação. Chamamos de erro do tipo I a rejeição da hipótese H quando ela é verdadeira e erro do tipo II a aceitação da hipótese H quando ela é falsa.

Note ainda que:

$$\mathbb{P}(\theta \in \Omega_0|\boldsymbol{x}) \leq \frac{1}{c+1} \iff c\mathbb{P}(\theta \in \Omega_0|\boldsymbol{x}) \leq \mathbb{P}(\theta \notin \Omega_0|\boldsymbol{x}) \tag{2.4}$$

Ou seja, c é a constante que dita até quantas vezes $P(\theta \notin \Omega_0)$ pode ser mais ou menos provável do que $P(\theta \in \Omega_0)$ para que se rejeite a hipótese. Por exemplo, se $c = 3$, isso significa que o teste vai rejeitar H se a probabilidade de $\theta \notin \Omega_0$ for maior ou igual ao triplo da probabilidade de $\theta \in \Omega_0$. Se, por exemplo, $c = 0.1$, então, o teste rejeitará H se a probabilidade de $\theta \in \Omega_0$ for menor ou igual ao décuplo da probabilidade de $\theta \notin \Omega_0$.

Capítulo 3

Modelos com adversários

Nos últimos 20 anos, métodos computacionais passaram a ficar cada vez mais populares no tratamento de problemas em áreas como reconhecimento de padrões (Bishop (2006)), finanças (Dixon *et al.* (2020)), aplicações médicas (Naqa *et al.* (2015)), sobretudo com o intuito de predição e automação.

No entanto, um dos problemas relacionados ao aumento da integração e automação de sistemas é o da segurança cibernética. Em 2017, cerca de 60% dos usuários de internet no Brasil foram vítimas de crimes cibernéticos, totalizando aproximadamente 62 milhões de pessoas e um montante estimado em 22 bilhões de dólares de prejuízo (Kshetri e DeFranco (2020)). Esse problema é mundial e estudos sugerem que ele tem crescido: de 2005 a 2009, a Romênia teve um aumento de 1500% na frequência desses ataques, enquanto países como a Indonésia, Tailândia, Bélgica e Colômbia apresentaram um aumento maior do que 550% desses ataques no mesmo período (Kim *et al.* (2012)).

Segundo Vorobeychik e Kantarcioglu (2018), com o intuito de lidar com os mais variados tipos de ataques que surgiram em cada área, pesquisadores passaram a contemplar o que são chamados **exemplos com adversários**. Esse nome é baseado na ideia de que, durante o processo de tomada de decisão, podem ser estabelecidas duas posições: a de defensor e a de atacante. Os defensores assumem o papel de indivíduos que tentam tomar uma decisão com base nos dados que têm em mãos, mas imaginando que esses dados podem ter sido comprometidos de algum modo pelo atacante (Vorobeychik e Kantarcioglu (2018)).

Dessa forma, a informação de interesse é portanto alterada por estratégias - as quais podem variar - do atacante. Em González-Ortega *et al.* (2019), são apresentados dois exemplos de cenários, um em que uma agente de segurança e um fraudador entram num confronto jurídico, mas que os únicos elementos observáveis (evidências) podem ter sido alterados pelo fraudador. Outro caso apresentado é o de que os dados em si não são alterados, mas sim o parâmetro relacionado à distribuição que os gerou.

Joseph *et al.* (2019) descreve que esse tema teve como um dos grandes motivadores o que chama de "aprendizado seguro", que seria uma base da inteligência artificial aplicada em segurança da informação. É uma área de pesquisa com uma base de aplicações em diversas áreas, tais como na construção de estratégias antiterrorismo, na solução de problemas de classificação e na segurança cibernética (Dalvi *et al.* (2004); Insua *et al.* (2012, 2019); Ríos e Insua (2012)).

Os primeiros trabalhos que fazem menção ao cenário com adversários propriamente dito foram modelos de classificação e de identificação de spam (Biggio e Roli (2018)). Uma quantidade substancial desses trabalhos dizem respeito à área de **aprendizado de máquina** (Biggio e Roli (2018);

Joseph *et al.* (2019)), sendo que há uma grande escassez de trabalhos que tenham se preocupado com propriedades de modelos probabilísticos ou estatísticos com adversários. Por exemplo, o trabalho mais antigo encontrado referente a testes de hipóteses com adversários foi Brandão *et al.* (2014) e apenas mais recentemente González-Ortega *et al.* (2019) apresentou mais formalmente os testes de hipóteses e a decisão ótima de forma geral sob a perspectiva bayesiana para um caso com adversários.

Apesar de existirem muitos estudos aplicados publicados na área, ainda são poucos os trabalhos abordam o tema com uma ênfase em elementos da estatística, como o comportamento da decisão ótima e dos testes de hipóteses para diferentes cenários, ou mesmo mais teóricas.

3.1 Teoria da Decisão com Adversários

Insua *et al.* (2018) é o primeiro trabalho a utilizar o termo "Teoria da Decisão com Adversários" (o qual abreviaremos por TDA). Nele, a teoria é apresentada como um confronto entre dois ou mais agentes, análogo ao que ocorre na Teoria dos Jogos não cooperativa.

Uma suposição bastante usual é a de que todos os agentes envolvidos possuem as mesmas informações a respeito do problema, porém ela pode não ser realista para muitas situações, a ponto de inviabilizar aplicações como a gestão de riscos de tomada de decisão (Insua *et al.* (2018), Insua *et al.* (2009)).

Desse modo, é proposta uma abordagem na qual se considera a perspectiva de um dos tomadores de decisão e, com base nessa perspectiva, estabelece-se que as decisões dos demais agentes são incertas para o tomador de decisão escolhido. Como existem essas incertezas, atribuem-se probabilidades a elas, considerando-se que as decisões de cada outro agente são também v.a. (Insua *et al.* (2018)).

As decisões ainda são tomadas da mesma forma que a abordagem em TD Bayesiana, de modo a maximizar (minimizar) a utilidade (perda) esperada. No entanto, as formas de interação entre atacante e defensor podem variar enormemente. Para facilitarmos a linguagem, iremos nos ater ao caso em que a agente decisora de interesse é uma defensora D e existe apenas um único outro agente A (atacante) envolvido. Utilizando essas personagens, apresentamos primeiro as ideias de Insua *et al.* (2018) em três tipos de estratégias de ataques diferentes:

- Não estratégica. A agente crê que o ataque será realizado sem considerar a ação de defesa. Essa categoria também engloba ataques de um adversário não inteligente, como desastres naturais

- Em equilíbrio de Nash ou Bayes-Nash. A defensora crê que o atacante assume que ambos têm os mesmos níveis de informação sobre os eventos

- "Pensamento em nível-$\mathcal{K}$". Quando $\mathcal{K} = 0$, temos o caso sem estratégia. Para $\mathcal{K} > 0$, a defensora considera em seu julgamento o que o atacante imagina da defesa que será realizada. Por exemplo, um pensamento de nível $\mathcal{K} = 1$ seria, pela defensora pensar que o atacante achar que ela não espera um ataque, ela crê que o ataque que será realizado é o mais vantajoso possível para ele nessa situação.

- Análise de equilíbrio em espelhamento. A defensora supõe que o atacante está modelando

sua decisão da mesma forma que ela está modelando a dele e ambos utilizam suas próprias distribuições subjetivas em todos os valores incertos.

Além dos tipos de estratégias, existem diversas categorias de ataques diferentes. Aqui, vamos nos ater aqueles listados em Vorobeychik e Kantarcioglu (2018), que diz que podemos pensar os ataques em pelo menos três dimensões: "momento", "informação" e "metas", que serão explicadas a seguir.

- Momento: relativo ao instante em que o ataque está sendo realizado. Usualmente, no aprendizado de máquina, temos a composição de modelos matemáticos de decisão e algoritmos. Ataques aos modelos (frequentemente chamados de "ataques de evasão") seriam referente a situações em que os modelos já foram ajustados e estão sendo utilizados para a tomada de decisão. Nessa situação, o atacante tenta explorar situações que levem o modelo a tomar predições erradas, o que pode ocorrer através de alguma mudança forçada de comportamento do algoritmo ou alterações nas observações. Ataques a algoritmos (também chamados "ataques de envenenamento") são ataques na amostra que será utilizada para ajustar o modelo (também conhecida como fase de fase de "treino" (aprendizagem), fazendo com que o modelo seja indevidamente ajustado, inviabilizando predições.

- Informação: trata-se da quantidade e qualidade de informação que o atacante tem do modelo ou algoritmo a ser atacado. Tipicamente, são chamados de "caixa-branca" quando o atacante possui total conhecimento sobre o funcionamento do modelo e de seu algoritmo e "caixa-preta" quando o atacante tem informações limitadas ou mesmo nenhuma informação sobre ele, mesmo que ele possa adquirir informações por alguma engenharia reversa.

- Metas: são o conjunto de razões para o ataque ocorrer. Tipicamente separados em duas categorias: "ataques direcionados" e "ataques de confiabilidade". Um ataque direcionado é aquele no qual o atacante tem como meta fazer com que o modelo falhe num cenário específico, enquanto os ataques de confiabilidade visam denegrir a confiabilidade do modelo de predição como um todo, maximizando o erro de predição.

Para cada uma dessas dimensões de ataque, há vários tipos de ataques possíveis, dos quais Insua *et al.* (2018) destaca três deles:

- ataque estrutural: A realiza um ataque de modo que perturba o parâmetro do modelo sob consideração (θ), gerando um processo baseado em um parâmetro modificado $a(\theta)$

- alteração de dados: A altera os dados $\boldsymbol{x}$ de modo que D tem como observável apenas $a(x) = y$

- ataque estrutural combinado com alteração de dados: A realiza o ataque estrutural e a alteração de dados concomitantemente

Usando a mesma configuração de agentes, diversos outros cenários de conflito entre essas duas entidades são apresentados em González-Ortega *et al.* (2019); Insua *et al.* (2009) com novas formas interativas entre atacante e defensor. Nelas, as decisões tanto de A quanto de B dependeriam das experiências passadas (ou seja, dos ataques desferidos por A e das decisões passadas de B). Neste trabalho, iremos estudar apenas o cenário de alteração de dados, explorando-o com algumas

situações que construímos: mais precisamente, consideraremos alguns mecanismos de alterações de dados binários (de Bernoulli).

A partir desse cenário apresentado nos trabalhos de González-Ortega *et al.* (2019); Insua *et al.* (2018), estabelerecemos primeiro as notações: os dados $\boldsymbol{Y}$ são as variáveis de Bernoulli observáveis para o agente decisor resultantes de perturbações do atacante (denotadas por a e b na sequência) sobre as variáveis originais $\boldsymbol{X}|$, que por sua vez não serão observáveis, em que $\mathbb{P}(Y_i = 1|X_i = 0, b) = b$ e $\mathbb{P}(Y_i = 0|X_i = 1, a) = 1 - a$. Além disso, vamos denotar por:

1. $P_\Theta(\theta)$, a priori de θ,
2. $P_A(a)$ e $P_B(b)$, as distribuições das respectivas ações a e b do atacante para determinados elementos de $\boldsymbol{X}$,
3. $P(\boldsymbol{x}|\theta)$, a distribuição do vetor aleatório $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ dado θ,
4. $P(\boldsymbol{y}|\boldsymbol{x}, a, b)$, a distribuição do vetor aleatório observável $\boldsymbol{Y} = (Y_1, Y_2, ..., Y_n)$ dados o vetor $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ (observável apenas pelo atacante) e as ações a e b (referentes, respectivamente, a $x_i = 1$ e $x_i = 0$, $i = 1, 2, ..., n$),
5. $l(d, \theta)$, a função de perda e
6. $d^*(\boldsymbol{y})$, a decisão ótima (estimação pontual) do defensor

O modelo acima acomoda as duas possíveis situações: uma em que o defensor não suspeita de fraude e outra em que ele suspeita. Para o caso em que o defensor não suspeita de fraude, suas estimativas e testes serão realizados da forma convencional, como se a amostra $\boldsymbol{y}$, adulterada, fosse uma realização legítima de $\boldsymbol{X}$ para fornecer informação de θ. Isto se dá considerando-se conhecidas as distribuições de a e b.

3.1.1 Estimação pontual para modelos permutáveis com adversários

Neste tópico, iremos lidar com o problema da decisão ótima na estimação pontual. Para melhor contextualização do problema, considere que temos interesse em estimar Θ. Vamos definir:

Definição 3.1.1 *Seja Ω o espaço paramétrico, um estimador pontual de Θ é uma estatística $d : \mathcal{X} \to \Omega$*

Com base nesse contexto, podemos ter interesse no valor mais provável de estar contido em Θ sob algum critério de otimalidade. Aos estimadores pontuais que satisfizerem o critério de otimalidade que iremos abordar (perda quadrática), chamaremos de decisões ótimas.

Para o caso em que o defensor suspeita de fraude, a decisão ótima é dada pelo seguinte:

$$d^*(\boldsymbol{y}) = \arg\min_d \int l(d, \theta) P(\boldsymbol{y}|\boldsymbol{x}, a, b)\, dP(\theta, \boldsymbol{x}, a, b) \tag{3.1}$$

Utilizando essa formulação geral e considerando a perda quadrática $l(d,\theta) = (d-\theta)^2$, temos,

$$d^*(\boldsymbol{y}) = \arg\min_d \Big[d^2 \int P(\boldsymbol{y}|\boldsymbol{x},a,b)\, dP(\theta,\boldsymbol{x},a,b) + \\ -2d \int \theta P(\boldsymbol{y}|\boldsymbol{x},a,b)\, dP(\theta,\boldsymbol{x},a,b) + \\ + \int \theta^2 P(\boldsymbol{y}|\boldsymbol{x},a,b)\, dP(\theta,\boldsymbol{x},a,b) \Big]$$

A obtenção de $d^*(\boldsymbol{y})$ é equivalente a achar o ponto máximo de uma função polinomial de segundo grau em d. Portanto,

$$d^*(\boldsymbol{y}) = \frac{\int \theta P(\boldsymbol{y}|x,a,b)\, dP(\theta,\boldsymbol{x},a,b)}{\int P(\boldsymbol{y}|\boldsymbol{x},a,b)\, dP(\theta,\boldsymbol{x},a,b)} \tag{3.2}$$

Podemos notar ainda que:

$$P(\theta|\boldsymbol{y}) = \frac{\int P(\boldsymbol{y}|\boldsymbol{x},a,b)\, dP(\boldsymbol{x},a,b)}{\int P(\boldsymbol{y}|\boldsymbol{x},a,b)\, dP(\theta,\boldsymbol{x},a,b)}, \tag{3.3}$$

e, portanto,

$$d^*(\boldsymbol{y}) = \int \theta P(\theta|\boldsymbol{y})\, dP\theta = \mathbb{E}(\theta|\boldsymbol{y}) \tag{3.4}$$

Para o caso multivariado, iremos utilizar a formulação encontrada em DeGroot (2004) para a função de perda. Considerando $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_k)^T$ $(dim(\boldsymbol{\theta}) \geq 2)$ o vetor de interesse a ser estimado e a decisão $\boldsymbol{d}$, para quaisquer pontos dados $\boldsymbol{\theta} \in \mathbb{R}^k$ e $\boldsymbol{d} \in \mathbb{R}^k$ temos a função de perda quadrática dada por:

$$L(\boldsymbol{\theta},\boldsymbol{d}) = (\boldsymbol{\theta}-\boldsymbol{d})^T \boldsymbol{A} (\boldsymbol{\theta}-\boldsymbol{d}), \tag{3.5}$$

em que $\boldsymbol{A}$ é uma matriz simétrica de dimensões k x k e que vamos assumir $\boldsymbol{A}$ positiva definida. Queremos, portanto, minimizar a perda esperada com respeito à decisão $\boldsymbol{d}$:

$$\begin{aligned} \boldsymbol{d^*}(\boldsymbol{y}) &= \arg\min_{\boldsymbol{d}} \int (\boldsymbol{\theta}-\boldsymbol{d})^T \boldsymbol{A} (\boldsymbol{\theta}-\boldsymbol{d}) P(\boldsymbol{y}|x,a,b)\, dP(\boldsymbol{\theta},\boldsymbol{x},a,b) = \\ &= \arg\min_{\boldsymbol{d}} \left[\int \left(\boldsymbol{\theta}^T \boldsymbol{A} \boldsymbol{\theta} - \boldsymbol{d}^T \boldsymbol{A} \boldsymbol{\theta} - \boldsymbol{\theta}^T \boldsymbol{A} \boldsymbol{\theta} + \boldsymbol{d}^T \boldsymbol{A} \boldsymbol{d} \right) P(\boldsymbol{y}|\boldsymbol{x},a,b)\, dP(\boldsymbol{\theta},\boldsymbol{x},a,b) \right] \end{aligned} \tag{3.6}$$

Uma vez que estamos fazendo a minimização em $\boldsymbol{d}$ e como podemos decompor a integral da soma como soma das integrais, o termo não $\boldsymbol{\theta}^T \boldsymbol{A} \boldsymbol{\theta}$ não depende de d. Além disso, podemos dividir a expressão inteira pela constante estritamente positiva $\int P(y|x,a,b) P(x|\boldsymbol{\theta})\, dP(\boldsymbol{\theta},x,a,b)$ sem alterarmos a otimização, obtendo

$$\boldsymbol{d^*}(\boldsymbol{y}) = \arg\min_{\boldsymbol{d}} [-\boldsymbol{d}^T \boldsymbol{A} \mathbb{E}(\boldsymbol{\theta}|y) - \mathbb{E}(\boldsymbol{\theta}^T|\boldsymbol{y}) \boldsymbol{A} \boldsymbol{d} + \boldsymbol{d}^T \boldsymbol{A} \boldsymbol{d}] =$$

Tomando o caso em que $\boldsymbol{A}$ é simétrica, temos que $\boldsymbol{d}^T \boldsymbol{A} \mathbb{E}(\boldsymbol{\theta}|\boldsymbol{y}) = \mathbb{E}(\boldsymbol{\theta}^T|\boldsymbol{y}) \boldsymbol{A} \boldsymbol{d}$, o que resulta que

$$\boldsymbol{d^*}(\boldsymbol{y}) = \arg\min_{\boldsymbol{d}} [\boldsymbol{d}^T \boldsymbol{A} \boldsymbol{d} - 2\boldsymbol{d}^T \boldsymbol{A} \mathbb{E}(\boldsymbol{\theta}|\boldsymbol{y})] \tag{3.7}$$

Calculando a derivada com respeito ao vetor $\boldsymbol{d}$ e igualando à zero e utilizando que $\boldsymbol{A}$ é positiva

definida (e, portanto, invertível), obtemos os pontos candidatos a mínimo:

$$2\boldsymbol{A}\boldsymbol{d} - 2\boldsymbol{A}\mathbb{E}(\boldsymbol{\theta}|\boldsymbol{y}) = \mathbf{0} \implies \boldsymbol{d} = \mathbb{E}(\boldsymbol{\theta}|\boldsymbol{y})$$

Tomando a segunda derivada em (3.7), podemos calcular a matriz Hessiana

$$\frac{\partial^2(\boldsymbol{d}^T\boldsymbol{A}\boldsymbol{d} - 2\boldsymbol{d}^T\boldsymbol{A}\mathbb{E}(\boldsymbol{\theta}|\boldsymbol{y}))}{\partial \boldsymbol{d}^T\boldsymbol{d}} = 2\boldsymbol{A} \tag{3.8}$$

Como $\boldsymbol{A}$ é definida positiva, a matriz Hessiana encontrada é também definida positiva e, portanto, $\mathbb{E}(\boldsymbol{\theta}|\boldsymbol{y})$ é decisão que minimiza a função de risco. Logo,

$$\boldsymbol{d^*}(\boldsymbol{y}) = \mathbb{E}(\boldsymbol{\theta}|\boldsymbol{y}) \tag{3.9}$$

A partir desse resultado, vemos que as decisões ótimas para o caso com adversários são análogas ao caso tradicional. No entanto, a distribuição à posteriori será também composta pela crença relativa às ações do agente atacante.

3.1.2 Teste de hipóteses para modelos com adversários

Os testes de hipóteses para os modelos com adversários são análogos aos testes de hipóteses para os modelos tradicionais. A alteração que temos é que, conforme já vimos, a distribuição à posteriori assume uma outra forma, por considerar as ações do atacante.

Assim, uma alternativa para testar as hipóteses: $H : \theta \in \Theta_0$ vs $A : \theta \in \Theta_1$, com $\Theta = \Theta_0 \cup \Theta_1$. Utilizando a perda 0-1-c, descrita na Tabela 2.4.2:

Assim, rejeitamos H se:

$$\mathbb{P}(\theta \in \Theta_0 | \boldsymbol{y}) \leq \frac{1}{c+1}, \tag{3.10}$$

e aceitamos H, caso contrário.

3.2 Modelo Bernoulli com adversários

Sejam $Y_1, Y_2, ..., Y_n$ uma sequência de variáveis originadas de uma sequência não observável $X_1, X_2..., X_n$ de $X|\theta \sim Bernoulli(\theta)$, com $\theta \sim$ Beta(α_0, β_0). Vamos considerar que a variável aleatória Y_i gerada do seguinte modo:

1. $Y_i|X_i = 1, a$ é v.a. com distribuição Bernoulli$(1-a)$,
2. $Y_i|X_i = 0, b$ é v.a. com distribuição Bernoulli(b),

Suponhamos, $a \sim$ Beta(α_a, β_a) e $b \sim$ Beta(α_b, β_b) independentes entre si e de θ e que α_a, α_b, β_a e β_b são hiperparâmetros (isto é, parâmetros definidos à priori, de acordo com a crença do estatístico).

Com essa formulação, inicialmente, podemos deduzir que a distribuição marginal de $Y_i|\theta, a, b$ é dada por:

$$\begin{aligned}
&\mathbb{P}(Y_i = y_i|\theta, a, b) = \sum_{x_i=0}^{1} \mathbb{P}(Y_i = y_i|X_i = x_i|\theta, a, b)\mathbb{P}(X_i = x_i|\theta, a, b) = \\
&= \mathbb{P}(Y_i = y_i|X_i = 0, \theta, a, b)\mathbb{P}(X_i = 0|\theta, a, b) + \mathbb{P}(Y_i = y_i|X_i = 1, \theta, a, b)\mathbb{P}(X_i = 1|\theta, a, b) = \\
&= b^{y_i}(1-b)^{1-y_i}(1-\theta) + (1-a)^{y_i}a^{1-y_i}\theta = \begin{cases} (1-b)(1-\theta) + a\theta, & \text{se } y_i = 0 \\ b(1-\theta) + (1-a)\theta, & \text{se } y_i = 1 \end{cases}
\end{aligned} \tag{3.11}$$

Assim, $Y_i|\theta, a, b \sim$ Bernoulli$(b(1-\theta)+(1-a)\theta)$ (note que $(1-b)(1-\theta)+a\theta+b(1-\theta)+(1-a)\theta = 1$).

É possível mostrar também (ver apêndice (A.11)) que a sequência de variáveis aleatórias $Y_1, Y_2, ...,$ dado θ, a, b é CIID Bernoulli$(b(1-\theta) + (1-a)\theta)$.

3.2.1 Suficiência para o modelo Bernoulli com adversários

Vamos utilizar a definição de estatística suficiente para o caso absolutamente contínuo apresentada na definição (2.3.1) para verificar que $k = \sum_{i=1}^{n} y_i$ é suficiente para (θ,a,b). Assim, considere a amostra $\boldsymbol{y} = (y_1, ..., y_n)$. Para uma dada priori genérica P_Θ e aplicando (A.11), a posteriori tem

densidade:

$$P(\theta|\boldsymbol{y}) = P(\theta|k(\boldsymbol{y})) = \\ = \frac{\int_0^1\int_0^1 \theta[(1-a)\theta + b(1-\theta)]^k[(1-b)(1-\theta)+a\theta]^{n-k}\,dPa\,dPb}{\int_0^1\int_0^1\int_0^1[(1-a)\theta + b(1-\theta)]^k[(1-b)(1-\theta)+a\theta]^{n-k}\,dP\theta\,dPa\,dPb} \tag{3.12}$$

Em que $k = \sum_{i=1}^{n} y_i$.

Definindo $K : \mathcal{X} \to \mathcal{T}$ com $K(\boldsymbol{Y}) = \sum_{i=1}^{n} Y_i$, temos que $K(\boldsymbol{Y})|\theta, a, b \sim$ Binomial(n,$(1-a)\theta + b(1-\theta)$). Logo, temos que:

$$P(\theta|K(\boldsymbol{y})) = \\ = \frac{\int_0^1\int_0^1 \theta\binom{n}{K(\boldsymbol{y})}[(1-a)\theta + b(1-\theta)]^{K(\boldsymbol{y})}[(1-b)(1-\theta)+a\theta]^{n-K(\boldsymbol{y})}\,dPa\,dPb}{\int_0^1\int_0^1\int_0^1\binom{n}{K(\boldsymbol{y})}[(1-a)\theta + b(1-\theta)]^{K(\boldsymbol{y})}[(1-b)(1-\theta)+a\theta]^{n-K(\boldsymbol{y})}\,dP\theta\,dPa\,dPb} \\ = \frac{\int_0^1\int_0^1 \theta[(1-a)\theta + b(1-\theta)]^{K(\boldsymbol{y})}[(1-b)(1-\theta)+a\theta]^{n-K(\boldsymbol{y})}\,dPa\,dPb}{\int_0^1\int_0^1\int_0^1[(1-a)\theta + b(1-\theta)]^{K(\boldsymbol{y})}[(1-b)(1-\theta)+a\theta]^{n-K(\boldsymbol{y})}\,dP\theta\,dPa\,dPb} = P(\theta|\boldsymbol{y})) \tag{3.13}$$

Logo, $K(\boldsymbol{Y}) = \sum_{i=1}^{n} Y_i$ é uma estatística suficiente para o modelo.

3.2.2 Decisão ótima para estimação do modelo Bernoulli com adversários

Para encontrar a decisão ótima $d^*(\boldsymbol{y})$ nesse contexto, inicialmente, notamos que:

$$d^*(\boldsymbol{y}) = \frac{\int_0^1\int_0^1\int_0^1 \theta \sum_{\boldsymbol{x}\in\{0,1\}^n} P(\boldsymbol{y}|\boldsymbol{x},b)P(\boldsymbol{x}|\theta)\,dP\theta\,dPa\,dPb}{\int_0^1\int_0^1\int_0^1 \sum_{\boldsymbol{x}\in\{0,1\}^n} P(\boldsymbol{y}|\boldsymbol{x},a,b)P(\boldsymbol{x}|\theta)\,dP\theta\,dPa\,dPb} \tag{3.14}$$

Essa expressão é calculada no apêndice e resulta que a decisão ótima $d^*(\boldsymbol{y})$ é dada por:

$$\frac{\sum_{t_{11}=0}^{k}\sum_{t_{00}=0}^{n-k}\binom{k}{t_{11}}\binom{n-k}{t_{00}}h_1h_2h_3}{\sum_{t_{11}=0}^{k}\sum_{t_{00}=0}^{n-k}\binom{k}{t_{11}}\binom{n-k}{t_{00}}h_4h_2h_3}, \tag{3.15}$$

em que:

$$k = \sum_{i=1}^{n} y_i$$

,

$$h_1 = \beta(\alpha_0 + n + t_{11} - k - t_{00} + 1, \beta_0 + k + t_{00} - t_{11})$$

$$h_2 = \beta(\alpha_a + n - k - t_{00}, \beta_a + t_{11})$$

$$h_3 = \beta(\alpha_b + t_{00}, \beta_b + k - t_{11})$$

$$h_4 = \beta(\alpha_0 + n + t_{11} - k - t_{00}, \beta_0 + k + t_{00} - t_{11})$$

Sendo $\beta(u,v) = \int_0^1 t^{u-1}(1-t)^{v-1}\,dt$ a função beta.

3.2.3 Testes de hipóteses para o modelo Bernoulli com adversários

Utilizando a expressão dada em 3.10 e a distribuição à posteriori geral para o caso Bernoulli com adversários calculada em A.14, para o caso em que $a \sim \beta(\alpha_a, \beta_a)$, $b \sim \beta(\alpha_b, \beta_b)$ e $\theta \sim (\alpha_0, \beta_0)$, rejeita-se H se:

$$\frac{\sum_{t_{11}=0}^{k}\sum_{t_{00}=0}^{n-k}\binom{k}{t_{11}}\binom{n-k}{t_{00}} g_1 g_2 \int_{\Theta_0} \theta^{\alpha_0-1+n-k+t_{11}-t_{00}}(1-\theta)^{\beta_0-1+k-t_{11}+t_{00}}\,d\theta}{\sum_{t_{11}=0}^{k}\sum_{t_{00}=0}^{n-k}\binom{k}{t_{11}}\binom{n-k}{t_{00}} g_1 g_2 g_3} \leq \frac{1}{c+1}, \tag{3.16}$$

em que

$$g_1 = \beta(\alpha_b + k - t_{11}, \beta_b + t_{00}),$$

$$g_2 = \beta(\alpha_a + n - k - t_{00}, \beta_a + t_{11}) \text{ e}$$

$$g_3 = \beta(\alpha_0 + n - k + t_{11} - t_{00}, \beta_0 + k - t_{11} + t_{00}),$$

e aceitamos H, caso contrário.

3.2.4 Decisões ótimas para diferentes cenários

Nesta seção, serão apresentados diversos cenários para diferentes incertezas expressas pelo agente A com relação à veracidade da amostra. Para todos os cenários, considere o conjunto de condições:

1. o conjunto de variáveis aleatórias de $\boldsymbol{X}|\Theta = \theta$ é CIID
2. $X_i|\Theta = \theta \sim$ Bernoulli(θ), $\forall i \in \{1, 2, ..., n\}$
3. A incerteza sobre θ não permite que algum valor seja mais provável. Portanto, para todos os casos, teremos então que $\Theta \sim$ Unif(0,1)
4. A amostra original $\{x_1, x_2, ..., x_n\}$ é observável apenas para o agente B
5. Os dados observáveis pelo agente A são $\{y_1, y_2, ..., y_n\}$
6. Serão designados, os valores $x_i = 1$ ou $y_i = 1$ por "sucesso" e $x_i = 0$ ou $y_i = 0$ por "fracasso", $\forall i \in \{1, 2, ..., n\}$, de modo que y_i, $\forall i \in \{1, 2, ..., n\}$ serão os resultados apresentados pelo agente B e x_i é não observável.

Os cálculos necessários para a decisão ótima, assim como as distribuições à posteriori são apresentados no apêndice.

Cenário 3.2.1 *Caso I: o agente A não acredita que o agente B adultere os dados. Um meio de representar essa ideia seria: $Y_i|x_i \sim$ Degenerada(x_i), $\forall x_i \in \{0, 1\}$. Em outras palavras, este caso*

reflete aquele em que a amostra obtida $\boldsymbol{y}$, aos olhos de A, reflete exatamente a amostra $\boldsymbol{x}$. Nesse caso, $\mathbb{P}(a=0)=1$ e $\mathbb{P}(b=0)=1$.

Assim, pelas condições apresentadas em temos a decisão ótima:

$$d_I^*(k) = \frac{k+1}{n+2} \tag{3.17}$$

Cenário 3.2.2 *Caso II: o agente A é tão incerto sobre os eventos de "sucesso" quanto de "fracasso" apresentados por B. Um meio de representar essa ideia seria considerando: $Y_i|x_i = 1, a \sim$ Bernoulli(1-a), $Y_i|X_i = 0, b \sim$ Bernoulli(b), $\forall x_i \in \{0,1\}$, $i \in \{1,2,...,n\}$. Vamos considerar ainda que o agente A não faz ideia de como B modificaria os resultados de "sucesso" ou "fracasso" e, portanto, entende que essa mudança poderia ser feita com quaisquer probabilidades, não havendo assim quaisquer valores para a ou b que fossem mais prováveis. Isso resulta que $a \sim$ Unif(0,1) e $b \sim$ Unif(0,1).*

Assim, pelas condições apresentadas na seção A.2.1 temos a decisão ótima (equação 3.15),

$$d_{II}^*(k) = \frac{\sum_{t_{11}=0}^{k}\sum_{t_{00}=0}^{n-k}\frac{1}{(k+t_{00}-t_{11}+1)}}{(n+2)\sum_{t_{11}=0}^{k}\sum_{t_{00}=0}^{n-k}\frac{1}{(n-k-t_{00}+t_{11}+1)(k+t_{00}-t_{11}+1)}} = \frac{1}{2} \tag{3.18}$$

Podemos interpretar esse resultado de modo que, se A é tão incerto sobre os casos de sucesso quanto sobre os casos de fracasso, a decisão ótima considerará ambos os lados tendo o mesmo peso, para quaisquer que sejam os valores apresentados na amostra.

Cenário 3.2.3 *Caso III: o agente A é incerto de que B alegaria "sucesso" quando ele ocorreu, mas tem certeza de que alegaria "fracasso" se ele tivesse ocorrido. Um meio de representar essa ideia seria considerando: $Y_i|x_i = 1, a \sim$ Bernoulli$(1-a)$ e $Y_i|x_i = 0 \sim$ Degenerada(0) (isto é, $\mathbb{P}(b=0)=1$). Novamente, consideraremos que o agente A não faz ideia de com que probabilidade B modificaria os resultados de "sucesso", o que leva à suposição de que $a \sim$ Unif(0,1).*

Por tais condições, temos a seguinte decisão ótima:

$$d_{III}^*(k) = \frac{n-k+1}{(n+2)(\psi(n+2)-\psi(k+1)}, \tag{3.19}$$

em que $\psi(z) := \frac{d}{dz}(log(\Gamma(z))) = \frac{\Gamma'(z)}{\Gamma(z)}$ é a função digama aplicada no ponto z, $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}\,dt$ e $\Gamma'(z)$ é sua respectiva derivada.

Note ainda que $d_{III}^*(k) \geq d_I^*(k)$:

$$d_{III}^*(k) \geq d_I^*(k) \iff \frac{n-k+1}{\psi(n+2)-\psi(k+1)} \geq k+1$$

Utilizando que H_k é o (k-ésimo) número harmônico (ver definição A.1.1), podemos aplicar o resultado A.1, resultando que $\psi(n+2)-\psi(k+1) = H_{n+1} - H_k = \sum_{s=k+1}^{n+1}\frac{1}{s}$,

$$d_{III}^*(k) \geq d_I^*(k) \iff \frac{n-k+1}{k+1} \geq \sum_{s=k+1}^{n+1}\frac{1}{s}$$

Mas,

$$\sum_{s=k+1}^{n+1} \frac{1}{s} = \frac{1}{k+1} + \frac{1}{k+2} + ... + \frac{1}{n+1} \leq \frac{n-k+1}{k+1} \implies d^*_{III}(k) \geq d^*_I(k)$$

Esse resultado é condizente com a ideia de que, se o agente tem suspeita de que alguns "sucessos" foram apresentados como "fracassos", o peso dado a cada um deveria ser menor do que no cenário em que eles foram considerados certezas.

Cenário 3.2.4 *Caso IV: o agente A é incerto de que B alegaria "fracasso" quando ele ocorreu, mas tem certeza de que alegaria "sucesso" se ele tivesse ocorrido. Um meio de representar essa ideia seria considerando:* $Y_i|x_i = 1, \sim$ *Degenerada*(1) *(isto é,* $\mathbb{P}(a = 0) = 1$*) e* $Y_i|x_i = 0, b \sim$ *Bernoulli*(b)*. Novamente, consideraremos que o agente A não faz ideia de com que probabilidade B modificaria os resultados de "fracasso", o que leva à suposição de que* $b \sim$ *Unif(0,1).*

Por tais condições, temos a seguinte decisão ótima:

$$d^*_{IV}(k) = 1 - \frac{k+1}{(n+2)(\psi(n+2) - \psi(n-k+1))} \tag{3.20}$$

Note que $d^*_{IV}(k) \leq d^*_I(k)$*:*

$$d^*_{IV}(k) \leq d^*_I(k) \iff 1 \leq \frac{k+1}{n+2}\left[1 + \frac{1}{\psi(n+2) - \psi(n-k+1)}\right]$$

Como $\psi(n+2) - \psi(n-k+1) = H_{n+1} - H_{n-k} = \sum_{s=n-k+1}^{n+1} \frac{1}{s}$*,*

$$d^*_{IV}(k) \leq d^*_I(k) \iff \sum_{s=n-k+1}^{n+1} \frac{n+2}{s} \leq \sum_{s=n-k+1}^{n+1} \frac{k+1}{s} + (k+1) \iff \sum_{s=n-k+1}^{n+1} \frac{1}{s} \leq \frac{k+1}{n-k+1}$$

Mas,

$$\sum_{s=n-k+1}^{n+1} \frac{1}{s} = \frac{1}{n-k+1} + \frac{1}{n-k+2} + ... + \frac{1}{n+1} \leq \frac{k+1}{n-k+1} \implies$$

$$\implies \sum_{s=n-k+1}^{n+1} \frac{1}{s} \leq \frac{k+1}{n-k+1} \implies d^*_{IV}(k) \leq d^*_I(k)$$

Este resultado é condizente com a ideia de que, se o agente suspeita de que alguns casos de "fracasso" foram apresentados como "sucessos", então, o peso dado a cada "sucesso" deveria ser menor do que quando o agente tem certeza de que o "sucesso" de fato ocorreu.

Cenário 3.2.5 *Caso V: o agente A crê que B trocará todos os resultados para apenas "sucessos" ou apenas "fracassos". Logo,* $Y_i|x_i \sim$ *Degenerada(s), em que* $s = 1$*, se ele crê que todos os dados foram alterados para sucesso e* $s = 0$*, se os dados foram alterados para fracasso.*

Pelo resultado dado em A.2.4, temos que, considerando $\Theta \sim$ *Unif*$(0, 1)$*, tanto para* $\sum_{i=1}^{n} y_i = n$ *(com* $s = 1$*) quanto para* $\sum_{i=1}^{n} y_i = 0$ *com* $(s = 0)$*,*

$$\mathbb{E}_V(\Theta|\boldsymbol{y}) = \mathbb{E}_V(\Theta) = \frac{1}{2} \tag{3.21}$$

Note que essa decisão coincide com a do caso II. Assim, a decisão é $\frac{1}{2}$ para qualquer tamanho de amostra, desde que satisfeitas as condições do cenário 3.2.2 ou do cenário 3.2.5.

Para os casos I, II, III e IV, foi criado um gráfico do comportamento das decisões ótimas, fixado $n = 1000$ em função de $k \in \{0, 1, 2, ..., 1000\}$, apresentado a seguir.

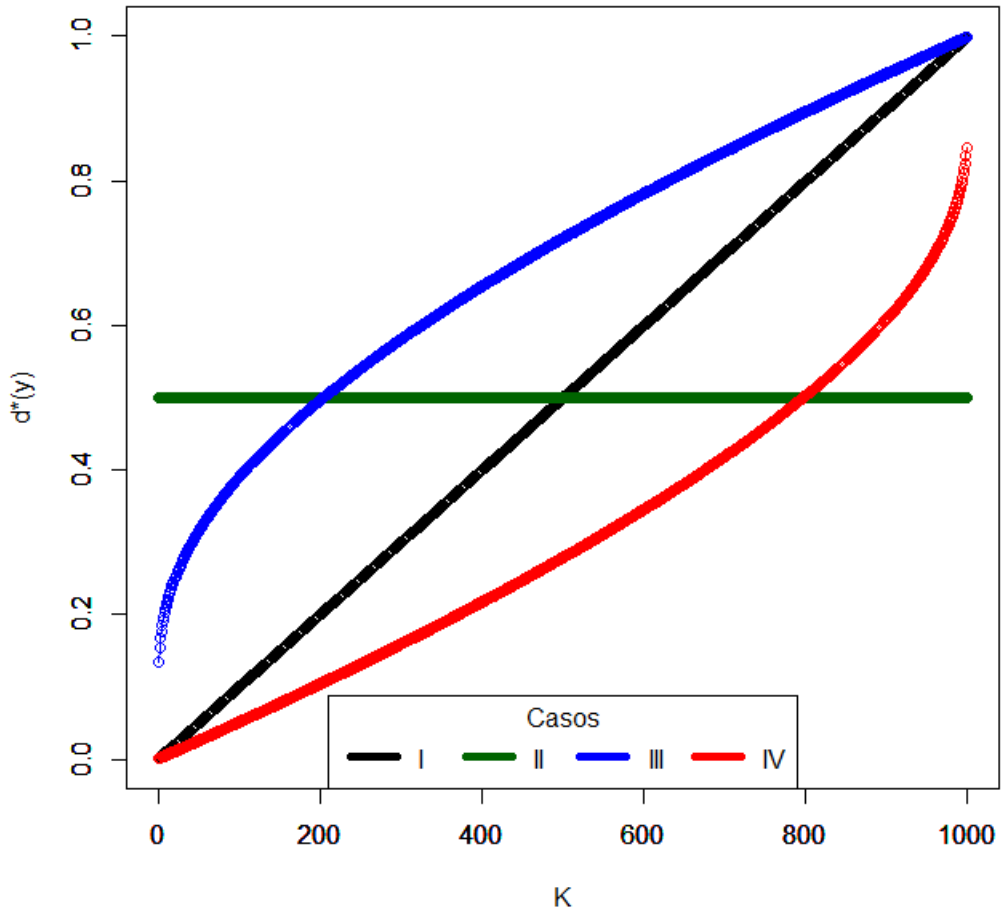

Figura 3.1: *Comparação de decisões ótimas - casos I a IV, para n = 1000*

Considerando que, para cada valor k fixado, as decisões ótimas se alteram bastante de cenário para cenário apresentado, uma questão natural seria se o risco de Bayes teria também um comportamento diferente em cada situação.

Calculamos o risco de Bayes para todos os casos utilizando $\theta \sim$ Unif(0,1). As contas para os casos III e IV utilizaram resultados de funções hipergeométricas e podem ser encontradas no apêndice.

Considerando $d_q^*(k)$ a decisão ótima para o caso q com $K = k$, $q \in \{I, II, III, IV\}$ e a função de perda L (quadrática), temos os riscos de Bayes $r_q(d_q^*, L)$:

$$r_I(d_I^*, L) = \frac{1}{6(n+2)} \tag{3.22}$$

$$r_{II}(d_{II}^*, L) = \frac{1}{12} \tag{3.23}$$

$$r_{III}(d^*_{III}, L) = \frac{1}{3} + \sum_{k=0}^{n} \frac{\binom{n}{k} d^*_{III}(k)}{k+1} \Big[-2\frac{F_{3,2}(k-n, k+1, k+2, k+2, k+3, 1)}{k+2} + \\ + d^*_{III}(k) \frac{F_{3,2}(k-n, k+1, k+1, k+2, k+2, 1)}{k+1} \Big], \tag{3.24}$$

em que $F_{p,q}(a_1, ..., a_p, b_1, ..., b_q, t)$ é a função hipergeométrica generalizada (ver A.7) nos parâmetros $(a_1, ..., a_p, b_1, ..., b_q)$ e ponto t.

$$r_{IV}(d^*_{IV}, L) = \frac{1}{3} + \sum_{k=0}^{n} \frac{\binom{n}{k} d^*_{IV}(k)}{(n-k+1)^2} \Big[-2\frac{F_{3,2}(-k, n-k+1, n-k+1, n-k+2, n-k+3, 1)}{n-k+2} + \\ + d^*_{IV}(k) F_{3,2}(-k, n-k+1, n-k+1, n-k+2, n-k+2, 1) \Big] \tag{3.25}$$

Com base nessas fórmulas apresentadas, foi feito um gráfico para ilustrar como variam os riscos de Bayes para diferentes tamanhos amostrais.

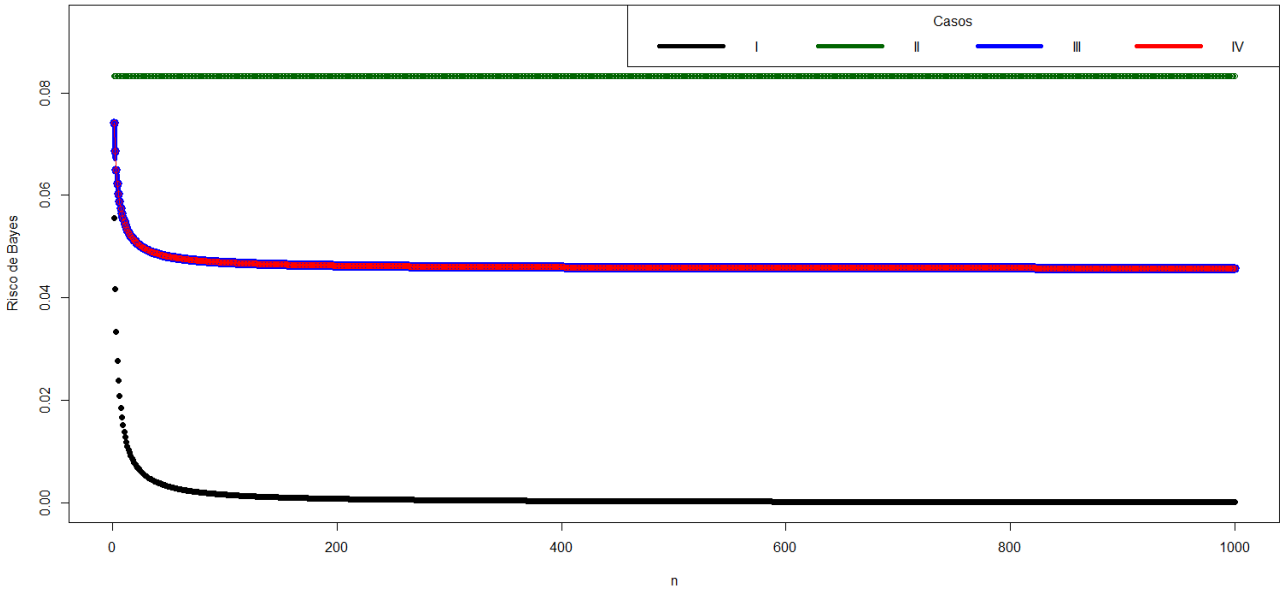

Figura 3.2: *Comparação de riscos de Bayes - casos I a IV, para n de 1 a 1000*

Comparando os riscos de Bayes em cada cenário, vemos que o risco da decisão $d^*_{II}(k)$ é sempre superior aos demais, indicando que essa forma de ataque se mostra a mais efetiva das listadas, pois a melhor decisão possível incorrerá num risco maior do que em qualquer outra situação. Também observamos que, numericamente, os riscos de Bayes das decisões $d^*_{III}(k)$ e $d^*_{IV}(k)$ são muito próximas, o que é esperado pela forma que construímos o modelo. Além disso, o risco da decisão $d^*_I(k)$ é o menor de todos, o que é condizente com a ideia do estatístico não contemplar que tenha havido algum ataque.

3.2.5 Distribuições marginais de $K|\theta$

Foram calculadas as distribuições dos valores observados $Y_i|\theta, a, b$, conforme na seção A.2, para $i \in \{1, ..., n\}$:

I - $Y_i|\theta, a, b \sim Y_i|\theta \sim$ CIID Bernoulli(θ)

II - $Y_i|\theta, a, b \sim$ CIID Bernoulli$(\theta - a\theta + b - b\theta)$

III - $Y_i|\theta, a, b \sim Y_i|\theta, a \sim$ CIID Bernoulli$(\theta(1-a))$

IV - $Y_i|\theta, a, b \sim Y_i|\theta, b \sim$ CIID Bernoulli$(\theta + b - b\theta)$

Dada a crença do estatístico sobre a e b, podemos calcular as marginais correspondentes para cada caso:

I - $Y_i|\theta \sim$ Bernoulli(θ)

II - $Y_i|\theta \sim$ Bernoulli$(\frac{1}{2})$

III - $Y_i|\theta \sim$ Bernoulli$(\frac{\theta}{2})$

IV - $Y_i|\theta \sim$ Bernoulli$(\frac{1+\theta}{2})$

Verificamos que $k = \sum_{i=1}^{n} y_i$ é uma estatística suficiente para o modelo e que $Y_i|\theta, a, b$, $i \in \{1, ..., n\}$ é CIID para todos os casos. Porém podemos verificar que o mesmo não vale para $Y_i|\theta$ nos casos II, III e IV, uma vez que a distribuição de $K|\theta = \sum_{i=1}^{n} Y_i|\theta$ claramente não é binomial (ver resultados A.21,A.30 e A.39).

Para ilustrarmos como a distribuição se altera, construímos o que seria a distribuição de $K|\theta$ em diferentes situações, com $\theta \sim$ Degenerada$(0, 3)$, $\theta \sim$ Degenerada$(0, 5)$ e $\theta \sim$ Degenerada$(0, 7)$, respectivamente.

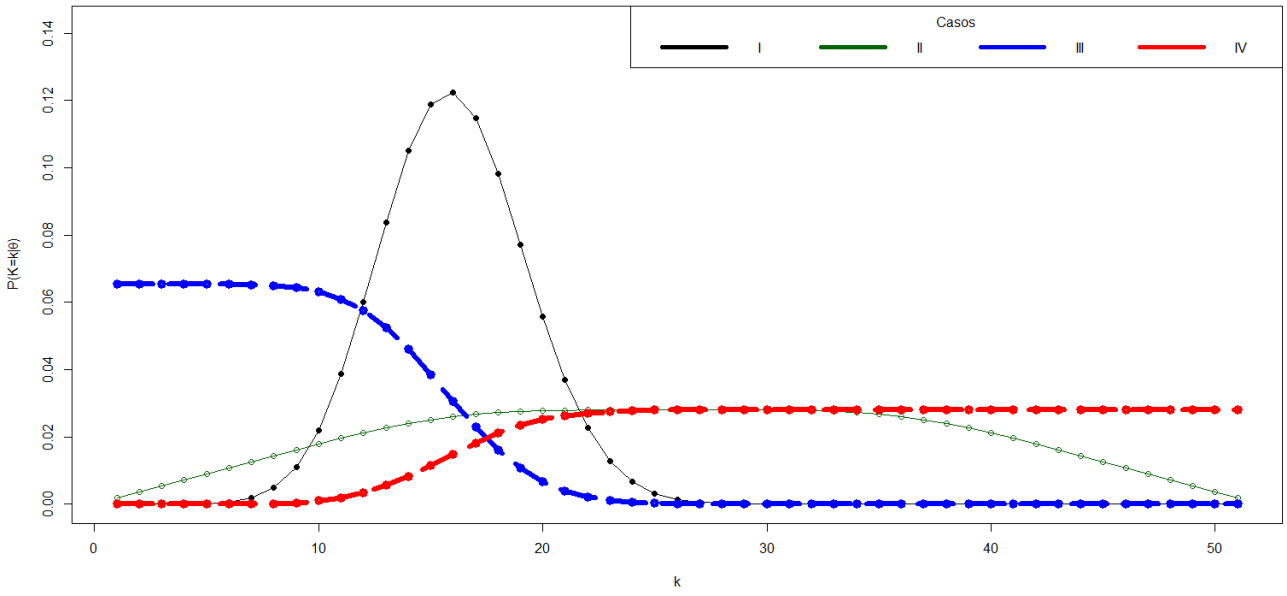

Figura 3.3: *Comparação de distribuições condicionais marginais - casos I a IV - para* $n = 50$ *e* $\theta \sim$ *Degenerada*$(0, 3)$

Com base nesse gráfico, podemos ver que a distribuição de $K = \sum_{i=1}^{n} Y_i$ em que não se crê em ataque (caso I) contemplada para o caso de que $\theta \sim$ Degenerada$(0, 3)$ (portanto, uma Binomial$(n,0, 3)$) é bastante modificada pela crença nas ações a e b. Os valores da distribuição relativa ao caso III concentra muito mais valores menores do que 12, enquanto a distribuição do caso IV se distribui mais em valores maiores do que 20. Ao mesmo tempo, para o caso II, a distribuição tende a se centralizar na média 25 (ponto médio do suporte da distribuição) e decrescer de forma simétrica até os extremos 0 e 50.

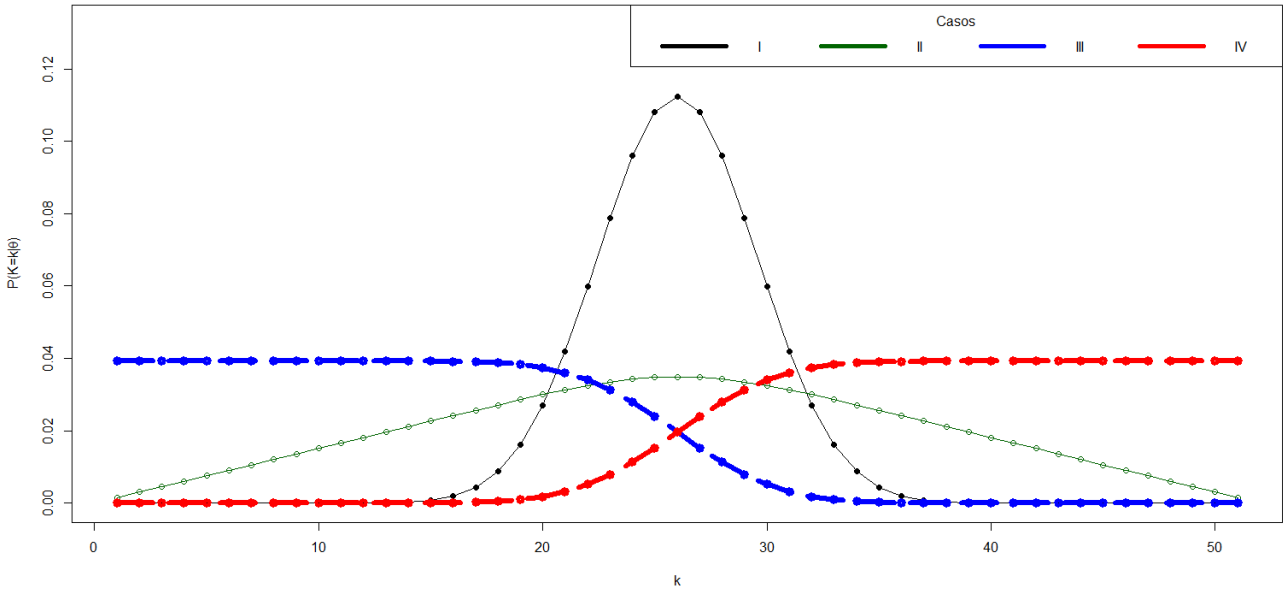

Figura 3.4: *Comparação de distribuições condicionais marginais - casos I a IV - para $n = 50$ e $\theta \sim$ Degenerada$(0, 5)$*

Para o caso em que $\theta \sim$ Degenerada$(0, 5)$, vemos que as distribuições correspondentes às amostras atacadas pelos adversários dos casos III e IV são simétricas, sendo que no caso III a distribuição se concentra mais a valores menores do que 20 e, no caso IV, a valores maiores do que 30. Novamente, o ataque a ambos os valores '1' e '0' tende a nos dar uma distribuição simétrica em torno do ponto médio do suporte da distribuição.

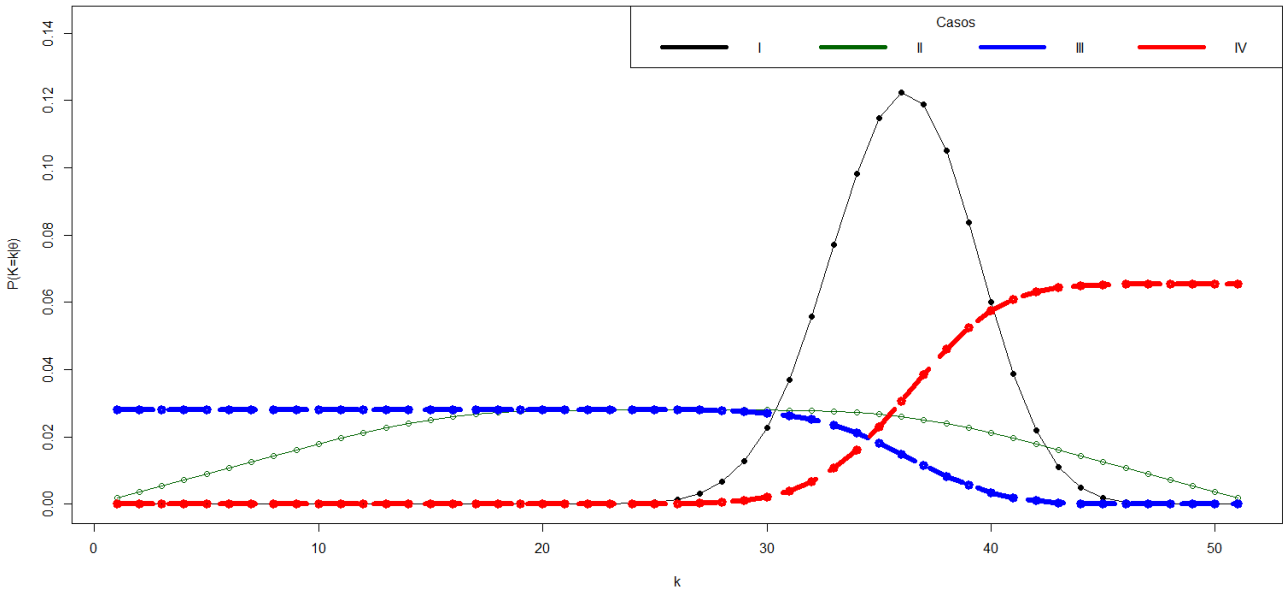

Figura 3.5: *Comparação de distribuições condicionais marginais - casos I a IV - para* $n = 50$ *e* $\theta \sim$ *Degenerada*$(0, 7)$

Para o caso em que $\theta \sim$ Degenerada$(0, 7)$, podemos notar que as distribuições dos casos III e IV apresentam uma simetria: o comportamento da distribuição de $K|\theta = 0,3$ no caso III é similar em relação ao comportamento da distribuição de $K|\theta = 0,7$ no caso IV. O mesmo vale para a distribuição no caso IV de $K|\theta = 0,3$ em relação à distribuição de $K|\theta = 0,7$ no caso III, enquanto a distribuição do caso II é idêntica tanto para $\theta \sim$ Degenerada$(0, 3)$ quanto para $\theta \sim$ Degenerada$(0, 7)$.

3.2.6 Testes de hipóteses para diferentes cenários

Nesta seção, serão apresentados e comparados resultados de testes bayesianos utilizando a perda 0-1-c aplicados aos cenários já apresentados: Casos I (3.2.1), II (3.2.2), III (3.2.3) e IV (3.2.4).

Para todos os cenários, vamos supor que o interesse seja testar as hipóteses: $H : \theta \leq \theta_0$ vs $A : \theta > \theta_0$, $\theta_0 \in \Theta$. Assim, o teste de Bayes, para todos os casos, consiste em rejeitar H, se e somente se,

$$\int_0^{\theta_0} f(\theta|\boldsymbol{y})\, d\theta \leq \frac{1}{c+1}$$

Conforme vimos na seção 2.4.2, a constante c está relacionada com até quantas vezes a probabilidade de $\theta \notin \Theta_0$ pode ser grande em relação à probabilidade de $\theta \in \Theta_0$. Assim, calcularemos as respectivas constantes $\bar{c}$, que majoram c que rejeita a hipótese H para cada caso.

Caso I: Para esse caso, temos a posteriori

$$f(\theta|\boldsymbol{y}) = \frac{\theta^{k+1-1}(1-\theta)^{n-k+1-1}}{\beta(k+1, n-k+1)} \mathbb{1}_{(0,1)}(\theta) \tag{3.26}$$

Definindo por $B(z, a, b) = \int_0^z t^{a-1}(1-t)^{b-1}\, dt$ a função beta incompleta (Osborn e Madey (1967)),

com $z \in [0,1]$, $a > 0$, $b > 0$, a região de rejeição é dada por

$$\int_0^{\theta_0} \frac{\theta^{k+1-1}(1-\theta)^{n-k+1-1}}{\beta(k+1, n-k+1)}\, d\theta \leq \frac{1}{c+1} \iff \frac{B(\theta_0, k+1, n-k+1)}{\beta(k+1, n-k+1)} \leq \frac{1}{c+1} \tag{3.27}$$

Aplicando o resultado A.3 a 3.27, com $t = \theta_0$, $P = k+1$ e $N = n+1$ e $Z \sim \text{Bin}(n+1, \theta_0)$, dizemos que o teste rejeita H se, e somente se,

$$\mathbb{P}_{\theta_0}(Z \geq k+1) \leq \frac{1}{c+1} \tag{3.28}$$

Vamos definir $\bar{c}_i$ como sendo o maior valor do custo c para uma dada amostra $k = \sum_{i=1}^{n} y_i$ de tamanho n, sob o cenário i, $i \in \{I, II, III, IV\}$ em que o teste rejeita a hipótese H.

Dessa forma, podemos expressar a região de rejeição de H em função de $\bar{c}_I$:

$$c \leq \frac{\mathbb{P}_{\theta_0}(Z \leq k)}{\mathbb{P}_{\theta_0}(Z > k)} \implies \bar{c}_I = \frac{\mathbb{P}_{\theta_0}(Z \leq k)}{\mathbb{P}_{\theta_0}(Z > k)} \tag{3.29}$$

Em outras palavras, para uma dada amostra de resultado k, tamanho amostral n e um determinado θ_0, no caso I, a hipótese seria rejeitada se, e somente se, $\bar{c}_I$ for maior ou igual ao custo c.

Em outras palavras, podemos interpretar que o teste considerando a priori $\theta \sim \text{Unif}(0,1)$ e a distribuição de $Y_i | y_i = x_i, \theta \sim \text{Bernoulli}(\theta)$, com as hipóteses H e K e a perda 0-1-c, é equivalente ao teste em que se rejeita H se e somente se, a probabilidade de uma variável aleatória $\text{Bin}(n+1, \theta_0)$ é inferior à quantidade estipulada $\frac{1}{c+1}$.

Caso II: Para este caso, a posteriori é (ver (A.15) para detalhes):

$$f_{II}(\theta|\boldsymbol{y}) = \frac{\sum_{t_{11}=0}^{k} \sum_{t_{00}=0}^{n-k} \binom{n+2}{k-t_{11}+t_{00}+1} \theta^{n-k+t_{11}-t_{00}} (1-\theta)^{k-t_{11}+t_{00}}}{(n+2) \sum_{t_{11}=0}^{k} \sum_{t_{00}=0}^{n-k} \frac{1}{(k-t_{11}+t_{00}+1)(n-k+t_{11}-t_{00})+1}} \tag{3.30}$$

Portanto, a região de rejeição e $\bar{c}_{II}$ são dados por

$$c \leq \frac{(n+2) \sum_{t_{11}=0}^{k} \sum_{t_{00}=0}^{n-k} \frac{1}{(k-t_{11}+t_{00}+1)(n-k+t_{11}-t_{00})+1}}{\sum_{t_{11}=0}^{k} \sum_{t_{00}=0}^{n-k} \binom{n+2}{k-t_{11}+t_{00}+1} \int_0^{\theta_0} \theta^{n-k+t_{11}-t_{00}} (1-\theta)^{k-t_{11}+t_{00}}\, d\theta} - 1 \tag{3.31}$$

Utilizando o resultado A.3, o teste rejeita H se, e somente se,

$$c \leq \frac{(n+2)\sum_{t_{11}=0}^{k}\sum_{t_{00}=0}^{n-k}\frac{1}{(k-t_{11}+t_{00}+1)(n-k+t_{11}-t_{00}+1)}}{\sum_{t_{11}=0}^{k}\sum_{t_{00}=0}^{n-k}\binom{n+2}{k-t_{11}+t_{00}+1}\int_0^{\theta_0}\theta^{n-k+t_{11}-t_{00}}(1-\theta)^{k-t_{11}+t_{00}}\,d\theta} - 1 =$$

$$\frac{(n+2)\sum_{t_{11}=0}^{k}\sum_{t_{00}=0}^{n-k}\frac{1}{(k-t_{11}+t_{00}+1)(n-k+t_{11}-t_{00}+1)}}{\sum_{t_{11}=0}^{k}\sum_{t_{00}=0}^{n-k}\frac{(n+2)}{(k-t_{11}+t_{00}+1)(n-k+t_{11}-t_{00}+1)}\mathbb{P}_{\theta_0}(Z \geq n-k+t_{11}-t_{00}+1)} - 1$$

$$\therefore c \leq \frac{\sum_{t_{11}=0}^{k}\sum_{t_{00}=0}^{n-k}\frac{1}{(k-t_{11}+t_{00}+1)(n-k+t_{11}-t_{00}+1)}\mathbb{P}_{\theta_0}(Z \leq n-k+t_{11}-t_{00})}{\sum_{t_{11}=0}^{k}\sum_{t_{00}=0}^{n-k}\frac{1}{(k-t_{11}+t_{00}+1)(n-k+t_{11}-t_{00}+1)}\mathbb{P}_{\theta_0}(Z > n-k+t_{11}-t_{00})} \tag{3.32}$$

Portanto,

$$\bar{c}_{II} = \frac{\sum_{t_{11}=0}^{k}\sum_{t_{00}=0}^{n-k}\frac{1}{(k-t_{11}+t_{00}+1)(n-k+t_{11}-t_{00}+1)}\mathbb{P}_{\theta_0}(Z \leq n-k+t_{11}-t_{00})}{\sum_{t_{11}=0}^{k}\sum_{t_{00}=0}^{n-k}\frac{1}{(k-t_{11}+t_{00}+1)(n-k+t_{11}-t_{00}+1)}\mathbb{P}_{\theta_0}(Z > n-k+t_{11}-t_{00})} \tag{3.33}$$

Caso III: Para este caso, a posteriori é dada por (ver (A.2.2) para detalhes):

$$f_{III}(\theta|\boldsymbol{y}) = \frac{\sum_{t_{00}=0}^{n-k}\binom{n+1}{t_{00}}\theta^{n-t_{00}}(1-\theta)^{t_{00}}}{\psi(n+2)-\psi(k+1)} \tag{3.34}$$

Portanto, a região de rejeição é dada por

$$\frac{\sum_{t_{00}=0}^{n-k}\binom{n+1}{t_{00}}\int_0^{\theta_0}\theta^{n-t_{00}}(1-\theta)^{t_{00}}\,d\theta}{\psi(n+2)-\psi(k+1)} \leq \frac{1}{c+1} \iff c \leq \frac{\psi(n+2)-\psi(k+1)}{\sum_{t_{00}=0}^{n-k}\binom{n+1}{t_{00}}B(\theta_0, n-t_{00}+1, t_{00}+1)} - 1 \tag{3.35}$$

Novamente, utilizando o resultado A.3, o teste rejeita H se, e somente se,

$$c_{III} \leq \frac{\sum_{t_{00}}^{n-k} \frac{1}{n-t_{00}+1}}{\sum_{t_{00}=0}^{n-k} \binom{n+1}{t_{00}} \beta(n-t_{00}+1, t_{00}+1) \sum_{s=n-t00+1}^{n+1} \binom{n+1}{s} \theta_0^s (1-\theta_0)^{n+1-s}} - 1 =$$

$$= \frac{\sum_{t_{00}}^{n-k} \frac{1}{n-t_{00}+1} - \sum_{t_{00}=0}^{n-k} \frac{1}{n-t_{00}+1} \mathbb{P}_{\theta_0}(Z \geq n-t_{00}+1)}{\sum_{t_{00}=0}^{n-k} \frac{1}{n-t_{00}+1} \mathbb{P}_{\theta_0}(Z \geq n-t_{00}+1)} = \frac{\sum_{t_{00}}^{n-k} \frac{1}{n-t_{00}+1} \mathbb{P}_{\theta_0}(Z < n-t_{00}+1)}{\sum_{t_{00}=0}^{n-k} \frac{1}{n-t_{00}+1} \mathbb{P}_{\theta_0}(Z \geq n-t_{00}+1)} \tag{3.36}$$

$$\therefore c \leq \frac{\sum_{t_{00}=0}^{n-k} \frac{1}{n-t_{00}+1} \mathbb{P}_{\theta_0}(Z \leq n-t_{00})}{\sum_{t_{00}=0}^{n-k} \frac{1}{n-t_{00}+1} \mathbb{P}_{\theta_0}(Z > n-t_{00})}$$

Então,

$$\bar{c}_{III} = \frac{\sum_{t_{00}=0}^{n-k} \frac{1}{n-t_{00}+1} \mathbb{P}_{\theta_0}(Z \leq n-t_{00})}{\sum_{t_{00}=0}^{n-k} \frac{1}{n-t_{00}+1} \mathbb{P}_{\theta_0}(Z > n-t_{00})} \tag{3.37}$$

Caso IV: Para este caso, a posteriori é dada por

$$f_{IV}(\theta|\boldsymbol{y}) = \frac{\sum_{t_{11}=0}^{k} \binom{n+1}{t_{11}} \theta^{t_{11}} (1-\theta)^{n-t_{11}}}{\psi(n+2) - \psi(n-k+1)} \tag{3.38}$$

Portanto, a região de rejeição é dada por

$$\frac{\sum_{t_{11}=0}^{k} \binom{n+1}{t_{11}} \int_0^{\theta_0} \theta^{t_{11}} (1-\theta)^{n-t_{11}}}{\psi(n+2) - \psi(n-k+1)} \leq \frac{1}{c+1} \iff c \leq \frac{\psi(n+2) - \psi(n-k+1)}{\sum_{t_{11}=0}^{k} \binom{n+1}{t_{11}} B(\theta_0, t_{11}+1, n-t_{11}+1)} - 1 \tag{3.39}$$

Aplicando o resultado A.3, o teste rejeita H se, e somente se,

$$c_{IV} \leq \frac{\sum_{t_{11}=0}^{k} \frac{1}{n-t_{11}+1} \mathbb{P}_{\theta_0}(Z \leq t_{11})}{\sum_{t_{11}=0}^{k} \frac{1}{n-t_{11}+1} \mathbb{P}_{\theta_0}(Z > t_{11})}$$

$$\therefore \bar{c}_{IV} = \frac{\sum_{t_{11}=0}^{k} \frac{1}{n-t_{11}+1} \mathbb{P}_{\theta_0}(Z \leq t_{11})}{\sum_{t_{11}=0}^{k} \frac{1}{n-t_{11}+1} \mathbb{P}_{\theta_0}(Z > t_{11})} \tag{3.40}$$

Tendo $\bar{c}_I$,$\bar{c}_{II}$,$\bar{c}_{III}$ e $\bar{c}_{IV}$, podemos comparar as constantes para alguns valores de θ_0 e $n = 50$. A escolha desse tamanho amostral foi porque os comportamentos foram bastante similares para

diferentes tamanhos amostrais testados e um tamanho muito acima desse dificultaria a análise.

Através da tabela B.1, podemos ver que o comportamento de $\bar{c}_I$ e $\bar{c}_{IV}$ são bastante similares, para quando $\theta_0 = 0,3$ enquanto $\bar{c}_{III}$ cresce numa escala muito maior. Também é possível verificar que $\bar{c}_{II}$ tem uma variação muito menor conforme k varia.

Vamos considerar, por exemplo, o teste bayesiano com a perda 0-1-c com $c = 1,5$. Tomemos três valores distintos para θ_0 e consideremos as hipóteses nulas: $H_1 : \theta \leq 0,3$, $H_2 : \theta \leq 0,5$ e $H_3 : \theta \leq 0,7$. A região de rejeição RC_{ij} para o teste de hipótese H_j sob o caso i, $i \in \{I, II, III, IV\}$, $j \in \{1, 2, 3\}$ são destinadas a seguir.

Para a hipótese H_1, temos as regiões de rejeição:

- $RC_{I1} = \{k \in \mathcal{X} : k \geq 16\}$
- $RC_{II1} = \{k \in \mathcal{X} : 1 \leq k \leq 49\}$
- $RC_{III1} = \{k \in \mathcal{X} : k \geq 7\}$
- $RC_{IV1} = \{k \in \mathcal{X} : k \geq 30\}$

Assim, vemos que essa hipótese seria mais facilmente rejeitada mais facilmente no caso III do que nos casos I e II, o que condiz com a crença adotada pelo estatístico que contempla o caso III (ele crê que, com alguma probabilidade, os resultados '1' podem ter sido trocados por '0', mas não o contrário). Verificamos também que, para a perda definida, o teste aplicado ao caso II rejeita a hipótese para quase qualquer valor de k e que, diferentemente dos três outros casos, H_1 é rejeitada para os valores extremos 0 e 50, havendo uma simetria em torno do valor 25. Isso condiz com a crença que constituiu o modelo, em que o indivíduo contempla a possibilidade de ter tanto valores '0' quanto '1' fraudados pelo adversário.

Para a hipótese H_2 (tabela B.2), temos as regiões de rejeição:

- $RC_{I2} = \{k \in \mathcal{X} : k \geq 26\}$
- $RC_{II2} = \emptyset$
- $RC_{III2} = \{k \in \mathcal{X} : k \geq 16\}$
- $RC_{IV2} = \{k \in \mathcal{X} : k \geq 42\}$

Em H_2, novamente vemos um comportamento similar ao da hipótese H_1, porém com a região crítica sendo menor. Uma observação é que, para essa hipótese, haveria rejeição no caso II se, e somente se, a perda fosse 0-1-c com $c \leq 1$. No caso de $c \leq 1$, o teste rejeitaria H_2 para qualquer valor de k.

Para a hipótese H_3 (tabela B.3), temos as regiões de rejeição:

- $RC_{I3} = \{k \in \mathcal{X} : k \geq 37\}$
- $RC_{II3} = \emptyset$
- $RC_{III3} = \{k \in \mathcal{X} : k \geq 28\}$
- $RC_{IV3} = \{k \in \mathcal{X} : k \geq 48\}$

Novamente, para H_3, observamos que os comportamentos para os casos I, III e IV são análogos aos das hipóteses H_1 e H_2. Porém, para o caso II, os maiores valores $\bar{c}_{II}$ se encontram nos extremos ($k = 0$ e $k = 50$), decrescendo até o mínimo em $k \in \{22, 23, 24, 25, 26, 27, 28\}$

.

Capítulo 4

Aplicação

Neste capítulo, ilustraremos, com uma aplicação, os modelos com adversários discutidos no capítulo anterior. Para isso, basear-nos-emos em conceitos apresentados em 3.1.

4.1 Ataque e defesa

Um sistema avaliador de fontes de informação está sendo construído. Para que ele seja ajustado, foi disponibilizado para que pessoas pudessem utilizá-lo e, de forma colaborativa, avaliassem se as notícias recebidas foram devidamente classificadas como falsas ou não. Uma fonte $\mathcal{I}_j$ de informação é considerada confiável se a fração de notícias falsas divulgadas por ela, θ_j, não excede 10%. Vamos supor que, para cada fonte de informação $\mathcal{I}_j$, a i-ésima notícia i recebida por cada pessoa siga uma distribuição dado $\Theta_j = \theta_j \sim \text{Bernoulli}(\theta_j)$ (isto é, $X_{ij}|\theta_j \sim \text{Bernoulli}(\theta_j)$), em que $X_{ij} = 1$ significa que a notícia i da fonte j é falsa e $X_{ij} = 0$ significa que a notícia i é verdadeira.

Foram avaliadas quatro diferentes fontes de informação, $\mathcal{I}_1$, $\mathcal{I}_2$, $\mathcal{I}_3$, $\mathcal{I}_4$.

No meio do processo de revisão, o estatístico percebe que as amostras das quatro fontes parecem ter comportamentos bastante distintos:

1. Para a fonte $\mathcal{I}_1$, foi observado que todas as amostras foram devidamente classificadas
2. Para a fonte $\mathcal{I}_2$, foi observado que cerca de metade das amostras que eram falsas foram classificadas como verdadeiras e vice-e-versa.
3. Para a fonte $\mathcal{I}_3$, as notícias verdadeiras foram classificadas corretamente, porém, cerca de metade das notícias falsas foram classificadas como verdadeiras.
4. Para a fonte $\mathcal{I}_4$, as notícias falsas foram classificadas corretamente, porém, cerca de metade das notícias verdadeiras foram classificadas como falsas.

Como esse processo de revisão é muito custoso, o estatístico deve então pensar em utilizar os modelos estudados para o cenário com adversários para fazer estimativas mais próximas da realidade das frações de notícias falsas divulgadas pelas fontes. Repare que, oportunamente, escolhemos esses cenários para que se encaixassem nos exemplos apresentados em 3.2.1, 3.2.2, 3.2.3 e 3.2.4 para montar uma defesa. Podiam haver diversos outros cenários, cada um descrito pelas distribuições correspondentes com as incertezas contempladas pelo estatístico ao criar o modelo.

Vamos supor que ele opte por usar prioris uniformes "não informativas" para todos os parâmetros em questão. Devido às características das fontes, observa-se, de fato, variáveis Y_{ij}, conforme descritas nos exemplos citados. Deseja-se estimar θ_j e avaliar se a j-ésima fonte é confiável, $j = 1, ..., 4$.

Foram coletadas $n_j = 1000$ notícias de cada fonte de informação $\mathcal{I}_j$ com as classificações dadas pelos usuários, $j = 1, 2, 3, 4$. Seja $k_j = \sum_{i=1}^{n_j} y_{ij}$, $i = 1, 2, ..., 1000$, a quantidade de notícias classificadas como falsas para cada fonte $\mathcal{I}_j$, $j = 1, 2, 3, 4$, obteve-se $k_1 = 105$, $k_2 = 78$, $k_3 = 70$, $k_4 = 328$

Utilizando os resultados apresentados para cada caso, chegamos nas estimativas ótimas aproximadas para a fração de notícias falsas para cada fonte de informação:

- $d_1^*(k_1) \approx 10,58\%$
- $d_2^*(k_2) = 50\%$
- $d_3^*(k_3) \approx 35,01\%$
- $d_4^*(k_4) \approx 17,55\%$

Essas são as estimativas pontuais para cada proporção de notícias falsas em cada um dos cenários, ou seja, os valores que minimizam o risco de erro quanto à fração real de notícias falsas sob as suposições de cada modelo e perda quadrática. Note que as médias observadas de notícias falsas, para os quatro casos são, respectivamente, $\bar{y}_1 = 10,5\%$, $\bar{y}_2 = 7,8\%$,$\bar{y}_3 = 7\%$ e $\bar{y}_4 = 32,8\%$ e que o único valor que é aparentemente muito próximo é para a fonte de informação $\mathcal{I}_1$, o que é esperado. Podemos reparar ainda que, para as fontes de informação $\mathcal{I}_2$ e $\mathcal{I}_3$, as proporções amostrais de notícias falsas foram consideradas subestimadas em relação à estimação pontual, enquanto, para $\mathcal{I}_4$, esse valor foi considerado superestimado por esse modelo.

Agora, vamos supor que o sistema avaliador classifica a fonte de informação $\mathcal{I}_j$ como "confiável", se a probabilidade $\theta_{\mathcal{I}_j}$ dela divulgar uma notícia falsa for menor do que 10% e não confiável se essa probabilidade for superior.

Formulamos então as hipóteses:

- H_j: a fonte de informação $\mathcal{I}_j$ é confiável ($\theta_j \in \Theta_{0j}$, com $\Theta_{0j} = \{\theta_j \in \Theta_{0j} : \theta_j \leq 10\%\}$
- A_j: a fonte de informação $\mathcal{I}_j$ não é confiável ($\theta_j \in \Theta_{1j}$, com $\Theta_{1j} = \{\theta_j \in \Theta_{1j} : \theta_j > 10\%\}$

Também vamos assumir que o erro do tipo I tenha o custo $c_I = 3$ e o erro do tipo II tenha o custo $c_{II} = 1$. Esses custos seriam compatíveis com o uso da perda 0-1-c, com $c = c_I = 3$.

Utilizando o teste bayesiano para os cenários correspondentes, temos então as decisões:

- $\varphi_1(\boldsymbol{x}_1) = 0 \implies$ não rejeitamos H_1
- $\varphi_2(\boldsymbol{x}_2) = 1 \implies$ rejeitamos H_2
- $\varphi_3(\boldsymbol{x}_3) = 1 \implies$ rejeitamos H_3
- $\varphi_4(\boldsymbol{x}_4) = 0 \implies$ não rejeitamos H_4

Portanto, temos que as fontes de informação $\mathcal{I}_1$ e $\mathcal{I}_4$ seriam classificadas como confiáveis e as fontes $\mathcal{I}_2$ e $\mathcal{I}_3$ não.

Capítulo 5

Conclusões

Nesse trabalho, estudamos alguns cenários com adversidade para o modelo Bernoulli. Sob esses cenários, desenvolvemos algumas soluções para os problemas de estimação e testes de hipóteses. Pudemos estabelecer relações entre algumas diferentes formas de ataques estipulados. Sendo um modelo recentemente apresentado, buscou-se descrever resultados analíticos sempre que possível para que fossem construídos meios de se realizar análise qualitativas sobre os resultados.

Através dos casos construídos (I, II, III e IV), pôde-se observar o comportamento de decisões ótimas, as quais foram condizentes com as ideias propostas: decisões baseadas num modelo que atribui incerteza para os elementos de "sucesso", mas não de "fracasso" tendem a atribuir uma estimativa de valor menor do que quando não são consideradas incertezas sobre a amostra apresentada. Em contrapartida, o modelo que apresentou incerteza para as amostras relativas ao "fracasso", mas não para aquelas relativas ao "sucesso" estima valores maiores para a decisão quando comparado ao modelo sem essas incertezas.

Além disso, verificamos que um modelo que contempla a mesma chance de permutar "fracasso" por "sucesso" e vice-e-versa, incorre em decisões que parecem depender muito pouco da amostra, o que é condizente com a crença do indivíduo que utiliza esse modelo.

Verificamos ainda que os riscos de Bayes associados às decisões ótimas tendem a aumentar na ocorrência de ataques, para todos os casos de ataques elencados. Além disso, para ataques como o do caso II, como a decisão ótima é a mesma para qualquer amostra, o risco de Bayes é constante para qualquer tamanho amostral.

Pudemos ainda ilustrar como as distribuições amostrais podem variar de acordo com os casos contemplados, além de estabelecermos uma relação entre a crença individual de cada estatístico que vislumbra cada cenário e a forma da distribuição correspondente.

Ao trabalharmos com os testes de hipóteses, considerando hiperparâmetros inteiros, pudemos encontrar uma relação entre a região de rejeição do teste e a função acumulada de uma variável aleatória Z com parâmetros $n+1$ e θ_0, relativo à hipótese H. Também verificamos como a crença em cada ataque facilita ou dificulta a rejeição da hipótese no teste de Bayes com hipóteses do tipo $\theta \geq \theta_0$ versus $\theta \leq \theta_0$.

Por fim, foi feito um exemplo numérico de como os casos apresentados poderiam ser utilizados para se obter a melhor inferência sob esses modelos.

5.1 Considerações Finais

Ao longo deste trabalho, foram desenvolvidos os cálculos para a estimação pontual e o teste de hipóteses bayesiano em diferentes situações com adversários. Também foram analisadas as distribuições dos dados observados perturbados pela ação do adversário, assim como calculados os respectivos risco de bayes dos estimadores pontuais e regiões críticas dos testes de hipótese bayesianos para cada situação descrita.

Durante a construção desse trabalho, pudemos observar que existem inúmeros tópicos que ainda podem ser mais explorados. Questões como o comportamento de decisões minimax, comportamento assintótico do risco de Bayes para esses estimadores, verificação de princípios, relação com identificabilidade, mecanismos de perturbação que preservem a permutabilidade, aplicações a outros modelos e analisar o modelo sob a perspectiva do atacante são algumas possibilidades.

Existe ainda a possibilidade de se explorar outros tipos de interações de ataque e defesa que não foram considerados, como a perturbação aplicada ao próprio parâmetro da distribuição e não mais à amostra em si, o que é seria uma mudança de paradigma na forma que contemplamos as ações dos adversários neste trabalho.

Apêndice A

Apêndice

A.1 Resultados auxiliares

Nesse apêndice, serão apresentados alguns resultados e definições, que foram aplicados aos cálculos dos exemplos.

A.1.1 Números harmônicos e função digama

Definição A.1.1 *Segundo Choi e Srivastava (2011), número harmônico clássico é definido por:*

$$H_n = \begin{cases} \sum_{k=1}^{n} \frac{1}{k}, & se\ n \in \mathbb{N}^* \\ 0, & se\ n = 0 \end{cases}$$

Em Wu *et al.* (2000), é demonstrado o seguinte teorema:

Teorema A.1.1 *Seja H_k um número harmônico, tal como na definição A.1.1 e $n \in \mathbb{N}^*$. Então,*

$$\sum_{k=1}^{n} H_k = (n+1)H_n - n$$

Além disso, temos o resultado, apresentado pelo mesmo autor:

$$H_n = \psi(n+1) - \psi(1), \tag{A.1}$$

em que $\psi(z) := \frac{d}{dz}(log(\Gamma(z))) = \frac{\Gamma'(z)}{\Gamma(z)}$ é a função digama aplicada no ponto z, $z \in \mathbb{C} \diagdown \mathbb{Z}_0^-$.

A.1.2 Função beta incompleta

Para $a > 0, b > 0$ no ponto $z \in [0,1]$, a função beta incompleta, denotada por $B(z,a,b)$ é dada por:

$$B(z,a,b) = \int_0^z t^{a-1}(1-t)^{b-1}\,dt \tag{A.2}$$

O seguinte resultado apresentado em Osborn e Madey (1967) nos diz que, para $t \in (0,1)$, $P \in \mathbb{N}^*$, $N \in \mathbb{N}^*$ e $N \geq P$, temos:

$$\frac{B(t,P,N-P+1)}{\beta(P,N-P+1)} = \sum_{s=P}^{N} \binom{N}{s} t^s (1-t)^{N-s} \tag{A.3}$$

A.1.3 Função (série) hipergeométrica

O conteúdo desta sessão se encontra de forma mais detalhada em Vaz Jr. e Oliveira (2016). Aqui, apresentaremos de forma resumida os resultados de interesse. Para definirmos a função hipergeométrica, primeiro, apresentamos a equação hipergeométrica (ou de Gauss):

$$z(1-z)y'' + [\gamma - (\alpha+\beta+1)z]y' - \alpha\beta y = 0 \tag{A.4}$$

Seja $(t)_n = \frac{\Gamma(t+n)}{\Gamma(n)}$ a notação para o símbolo de *Pochhammer* (também conhecido como "fatorial crescente"), a solução da equação A.4 em torno dos pontos $z_0 = 0$ e $z_0 = 1$ é dada pela série hipergeométrica:

$$F_{2,1}(\alpha,\beta,\gamma,z) = \sum_{n=0}^{\infty} \frac{(\alpha)_n(\beta)_n z^n}{(\gamma)_n n!}, \tag{A.5}$$

em que $y = y(z)$, α, β, γ parâmetros reais e $y' = \frac{\partial y}{\partial z}$ e $y'' = \frac{\partial^2 y}{\partial^2 z}$ as respectivas derivadas parciais de primeira e segunda ordem de y em função de z, $z, y \in \mathbb{C}$. definido nos

Para os pontos α,β,γ, definidos no conjunto dos reais e z definido nos complexos, em que há convergência da série, ela é chamada função hipergeométrica. Assim, a função hipergeométrica está definida no conjunto dos números reais para $|z| = 1$, se $\gamma - \alpha - \beta > 0$. Existem outras combinações de pontos em que a função hipergeométrica pode ser definida, sobretudo no conjunto dos números complexos, mas eles não serão de nosso interesse neste trabalho.

Utilizando a notação $\Re(x)$ como sendo a parte real de x, temos o resultado, para $\Re(\gamma) > \Re(\beta)$ com $\beta > 0$:

$$\int_0^1 t^{\beta-1}(1-t)^{\gamma-\beta-1}(1-tz)^{-\alpha}\,dt = F_{2,1}(\alpha,\beta,\gamma,z)\beta(\beta,\gamma-\beta) \tag{A.6}$$

Podemos definir também a função (série) hipergeométrica generalizada, conforme em Hannah (2013):

$$F_{p,q}(1,...p,b_1,...,b_q,z) = \sum_{s=0}^{\infty} \frac{\prod_{i=1}^{p}(a_i)_s}{\prod_{i=1}^{q}(b_i)_s} \frac{z^s}{s!} \tag{A.7}$$

Ainda segundo a mesma autora, temos um teorema que assegura a relação integral, para

$\Re(b_{q+1}) > \Re(a_{p+1}) > 0$:

$$\begin{aligned}&\int_0^1 t^{a_{p+1}-1}(1-t)^{b_{q+1}-a_{p+1}-1}F_{p,q}(a_1,...,a_p,b_1,...,b_q,tz)\,dt = \\ &= F_{p+1,q+1}(a_1,...,a_{p+1},b_1,...,b_{q+1},z)\frac{\Gamma(a_{p+1})\Gamma(b_{q+1}-a_{p+1})}{\Gamma(b_{q+1})}\end{aligned} \tag{A.8}$$

A.2 Cálculos para os casos II, III, IV e V

A.2.1 Caso II (exemplo 3.2.2)

Para esse caso, consideramos que $Y_i|x_i = 1, a \sim$ Bernoulli(1-a), $Y_i|X_i = 0, a \sim$ Bernoulli(b), $\forall x_i \in \{0,1\}$, $i \in \{1,2,...,n\}$.

Queremos, primeiramente, calcular $p(\boldsymbol{y}|\theta,a,b) = \sum_{\boldsymbol{x}\in\{0,1\}^n} p(\boldsymbol{y}|\boldsymbol{x},a,b)p(\boldsymbol{x}|\theta)$. Como estamos realizando a soma em todas as combinações possíveis de $\boldsymbol{x}$, como x_i pode assumir apenas os valores 0 ou 1, temos um total de 2^n diferentes vetores $\boldsymbol{x}$ a serem aplicados no somatório.

Inicialmente, vamos definir que, para um dado $\boldsymbol{x}$ e $\boldsymbol{y}$ fixados:

- k = número de elementos do vetor $\boldsymbol{y}$ que assumem valor 1, isto é, $k = \sum_{i=1}^{n} y_i$
- t_{11} = número de valores $y_i = 1$ associados a $x_i = 1$, com $i = 1,...,n$,
- t_{00} = número de valores $y_i = 0$ associados a $x_i = 0$, com $i = 1,...,n$

A partir desses números, podemos pensar que, de k componentes de y que se apresentam como 1, t_{11} deles se "originaram" de valores de x que assumiam 1 e de $k - t_{11}$ valores de x que se assumiam como 0. Utilizando as especificações do exemplo, $Y_i|X_i = 1 \sim Bernoulli(1-a)$ e $Y_i|X_i = 0 \sim Bernoulli(b)$ e que, dados a e b, cada Y_i depende exclusivamente de X_i temos que:

$$\begin{aligned}\mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}|\boldsymbol{X} = \boldsymbol{x}) &= \prod_{i=1}^{n}\mathbb{P}(Y_i = y_i|X_i = x_i) = \prod_{i:x_i=0}\mathbb{P}(Y_i = y_i|X_i = x_i)\prod_{j:x_j=1}\mathbb{P}(Y_j = y_j|X_j = x_j) = \\ &= \prod_{i:x_i=0} b^{y_i}(1-b)^{1-y_i}\prod_{j:x_j=1}(1-a)^{y_j}a^{1-y_j} = (1-b)^{t_{00}}b^{k-t_{11}}(1-a)^{t_{11}}a^{n-k-t_{00}}\end{aligned} \tag{A.9}$$

Ao mesmo tempo, para esses dados x e y fixados,

$$\mathbb{P}(\boldsymbol{X} = \boldsymbol{x}|\Theta = \theta) = \theta^{t_{11}+n-k-t_{00}}(1-\theta)^{k-t_{11}+t_{00}} \tag{A.10}$$

Nosso interesse é, para um dado $\boldsymbol{y}$, calcularmos $\mathbb{P}(\boldsymbol{y}|\theta,a,b)$, que envolve somarmos as probabilidades possíveis em $\boldsymbol{x}$. Iremos então fazer a contagem de todas as possibilidades.

Para um dado $\boldsymbol{y}$ e um dado $\boldsymbol{x}$, o valor k nos indica o número de elementos no vetor $\boldsymbol{y}$ que assumem valor "1". Então, vamos primeiro contar que, desses k valores "1", existem t_{11} elementos de $\boldsymbol{x}$ que assumem valor "1"também. Para cada uma dessas escolhas, restam $n-k$ elementos de $\boldsymbol{y}$ que assumirão valor "0"e, desses, t_{00} estarão associados a um valor "0"do vetor $\boldsymbol{x}$. Assim, temos $\binom{k}{t_{11}}$ modos de escolher os elementos $i : x_i = 1, i \in \{1,...,n\}$ e $y_i = 1$ e $\binom{n-k}{t_{00}}$ modos de escolher os elementos $j : x_j = 0, j \in \{1,...,n\}$ e $y_j = 0$.

No entanto, como estamos somando para todas as possíveis combinações de $\boldsymbol{x}$, teremos situações em que $t_{00} = 0$, $t_{00} = 1$,...,$t_{00} = n-k$ ($n-k$ é o número de elementos "0"em $\boldsymbol{y}$). Ao mesmo tempo, temos os possíveis valores $t_{11} = 0$,$t_{11} = 1$,...,$t_{11} = k$.

Assim, fixado y, temos $\binom{n-k}{t_{00}}$ modos de escolher $j : x_j = 0, j \in \{1, ..., n\}$ e $y_j = 0$. Como queremos somar em todas as combinações de x, temos todos os casos possíveis de valores t_{00} e t_{11}, para um k fixo, pois ele depende apenas de y (fixado). Assim, sem perda de generalidade:

$$\mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}|\theta, a, b) = \sum_{\boldsymbol{x}\in\{0,1\}^n} p(\boldsymbol{y}|\boldsymbol{x}, a, b)p(\boldsymbol{x}|\theta) =$$

$$= \sum_{t_{11}=0}^{k}\sum_{t_{00}=0}^{n-k}\binom{k}{t_{11}}\binom{n-k}{t_{00}}(1-b)^{t_{00}}b^{k-t_{11}}(1-a)^{t_{11}}a^{n-k-t_{00}}\theta^{n+t_{11}-k-t_{00}}(1-\theta)^{k+t_{00}-t_{11}} =$$

$$= \underbrace{\sum_{t_{11}=0}^{k}\binom{k}{t_{11}}[(1-a)\theta]^{t_{11}}[(1-\theta)b]^{k-t_{11}}}_{\text{(I)}}\overbrace{\sum_{t_{00}=0}^{n-k}\binom{n-k}{t_{00}}[(1-b)(1-\theta)]^{t_{00}}(a\theta)^{n-k-t_{00}}}^{\text{(II)}} =$$

$$= [(1-a)\theta + (1-\theta)b]^k[(1-b)(1-\theta) + a\theta]^{n-k} = \prod_{i=1}^{n}[(\theta - a\theta + b - b\theta)^{y_i}(1-b-\theta+b\theta+a\theta)^{1-y_i}] \tag{A.11}$$

Note que (I) e (II) são binômios de Newton. Logo, como a densidade conjunta de $\boldsymbol{Y}|\theta, a, b$ pode ser escrita como o produtório da densidade de VA CIID Bernoulli$(\theta(1-a)+(1-\theta)b)$, $i = 1, 2, ..., n$ para qualquer $n \in \mathbb{N}^*$, temos que a sequência $(Y_k)_{k\geq 1}$ é permutável.

Substituindo (A.11) em (3.14) e aplicando o Teorema de Fubini, temos o resultado (3.15).

Usando (A.11), podemos calcular a função densidade de probabilidade da distribuição à posteriori de θ:

$$f_{II}(\theta|\boldsymbol{y}) \propto \int_{\{0,1\}^n}\int_0^1\int_0^1 p(\boldsymbol{y}|\boldsymbol{x}, a, b)p(\boldsymbol{x}|\theta)p(\theta)\,dP\boldsymbol{x}\,dPa\,dPb \propto \int_0^1\int_0^1\sum_{\boldsymbol{x}\in\{0,1\}^n} p(\boldsymbol{y}|\boldsymbol{x}, a, b)p(\boldsymbol{x}|\theta)p(\theta)\,dPa\,dPb \propto$$

$$\int_0^1\int_0^1\sum_{t_{11}=0}^{k}\sum_{t_{00}=0}^{n-k}\binom{k}{t_{11}}\binom{n-k}{t_{00}}(1-b)^{t_{00}}b^{k-t_{11}}(1-a)^{t_{11}}a^{n-k-t_{00}}\theta^{\alpha_0-1+n-k+t_{11}-t_{00}}(1-\theta)^{\beta_0-1+k-t_{11}+t_{00}}\,dPa\,dPb \tag{A.12}$$

Então, temos, admitindo $a \sim$ Beta(α_a,β_a) e $b \sim$ Beta(α_b,β_b), a e b independentes, que

$$f_{II}(\theta|\boldsymbol{y}) \propto$$
$$\sum_{t_{11}=0}^{k}\sum_{t_{00}=0}^{n-k}\binom{k}{t_{11}}\binom{n-k}{t_{00}}\beta(\alpha_a+n-k-t_{00}, \beta_a+t_{11})\beta(\alpha_b+k-t_{11}, \beta_b+t_{00})\theta^{\alpha_0-1+n+t_{11}-k-t_{00}}(1-\theta)^{\beta_0-1+k+t_{00}-t_{11}} \tag{A.13}$$

$$\therefore \quad f_{II}(\theta|\boldsymbol{y}) \quad = \quad \frac{\sum_{t_{11}=0}^{k}\sum_{t_{00}=0}^{n-k}\binom{k}{t_{11}}\binom{n-k}{t_{00}}g_1g_2\theta^{\alpha_0-1+n-k+t_{11}-t_{00}}(1-\theta)^{\beta_0-1+k-t_{11}+t_{00}}}{\sum_{t_{11}=0}^{k}\sum_{t_{00}=0}^{n-k}\binom{k}{t_{11}}\binom{n-k}{t_{00}}g_1g_2g_3}, \quad \text{(A.14)}$$

em que:

$$g_1 = \beta(\alpha_b + k - t_{11}, \beta_b + t_{00})$$

$$g_2 = \beta(\alpha_a + n - k - t_{00}, \beta_a + t_{11})$$

$$g_3 = \beta(\alpha_0 + n - k + t_{11} - t_{00}, \beta_0 + k - t_{11} + t_{00})$$

Fazendo $\alpha_0 = \alpha_a = \alpha_b = \beta_0 = \beta_a = \beta_b = 1$, resulta que

$$f_{II}(\theta|\boldsymbol{y}) = \frac{\sum_{t_{11}=0}^{k}\sum_{t_{00}=0}^{n-k}\binom{n+2}{k-t_{11}+t_{00}+1}\theta^{n-k+t_{11}-t_{00}}(1-\theta)^{k-t_{11}+t_{00}}}{(n+2)\sum_{t_{11}=0}^{k}\sum_{t_{00}=0}^{n-k}\frac{1}{(k-t_{11}+t_{00}+1)(n-k+t_{11}-t_{00})+1}} \quad \text{(A.15)}$$

A decisão ótima é dada por

$$E_{II}(\Theta|\boldsymbol{y}) = \frac{\sum_{t_{11}=0}^{k}\sum_{t_{00}=0}^{n-k}\frac{1}{(k+t_{00}-t_{11}+1)}}{(n+2)\sum_{t_{11}=0}^{k}\sum_{t_{00}=0}^{n-k}\frac{1}{(n-k-t_{00}+t_{11}+1)(k+t_{00}-t_{11}+1)}} \quad \text{(A.16)}$$

Assim, utilizando A.1.1 e denotando $E(\Theta|\boldsymbol{y}) = \frac{U}{V}$, temos:

$$\begin{aligned}
U &= \sum_{t_{11}=0}^{k}\sum_{t_{00}=0}^{n-k}\frac{1}{(k+t_{00}-t_{11}+1)} = \sum_{t_{11}=0}^{k}\sum_{u=k-t_{11}+1}^{n-t_{11}+1}\frac{1}{u} = \sum_{t_{11}=0}^{k}\left(\sum_{u=1}^{n-t_{11}+1}\frac{1}{u} - \sum_{u=1}^{k-t_{11}}\frac{1}{u}\right) = \\
&= \sum_{t_{11}=0}^{k}\left(H_{n-t_{11}+1} - H_{k-t_{11}}\right) = \sum_{v=n-k+1}^{n+1}H_v - \sum_{v=0}^{k}H_v = \sum_{v=1}^{n+1}H_v - \sum_{v=1}^{k}H_v - \sum_{v=1}^{n-k}H_v = \\
&= [(n+2)H_{n+1} - (n+1)] - [(k+1)H_k - k] - [(n-k+1)H_{n-k} - (n-k)] = \\
&= (n+2)H_{n+1} - (k+1)H_k - (n-k+1)H_{n-k} - 1 = \\
&= (n+2)[\psi(n+2) - \psi(1)] - (k+1)[\psi(k+1) - \psi(1)] - (n-k+1)[\psi(n-k+1) - \psi(1)] - 1 = \\
&= (n+2)\psi(n+2) - (n-k+1)\psi(n-k+1) - (k+1)\psi(k+1) - 1
\end{aligned} \quad \text{(A.17)}$$

$$V = (n+2)\sum_{t_{11}=0}^{k}\sum_{t_{00}=0}^{n-k}\frac{1}{(n-k-t_{00}+t_{11}+1)(k+t_{00}-t_{11}+1)} \quad \text{(A.18)}$$

Para utilizarmos as propriedas dos números harmônicos clássicos, queremos encontrar J e L que satisfazem:

$$\frac{1}{(n-k-t_{00}+t_{11}+1)(k+t_{00}-t_{11}+1)} = \frac{J}{(n-k-t_{00}+t_{11}+1)} + \frac{L}{(k+t_{00}-t_{11}+1)}$$

Ou seja,

$$\frac{1}{(n-k-t_{00}+t_{11}+1)(k+t_{00}-t_{11}+1)} = \frac{J}{(n-k-t_{00}+t_{11}+1)} + \frac{L}{(k+t_{00}-t_{11}+1)} =$$
$$= \frac{J(k+t_{00}-t_{11}+1)+L(n-k-t_{00}+t_{11}+1)}{(n-k-t_{00}+t_{11}+1)(k+t_{00}-t_{11}+1)} \implies kJ-kL+nL+(J-L)t_{00}-(J-L)t_{11}+J+L=1$$
$$\therefore \begin{cases} J = L \\ nJ+J+L=1 \iff J=\frac{1}{2+n} \end{cases} \tag{A.19}$$

Portanto, usando A.19, A.17 e o teorema A.1.1

$$V = \sum_{t_{11}=0}^{k}\sum_{t_{00}=0}^{n-k}\frac{1}{(n-k-t_{00}+t_{11}+1)} + \sum_{t_{11}=0}^{k}\sum_{t_{00}=0}^{n-k}\frac{1}{(k+t_{00}-t_{11}+1)} =$$
$$= \sum_{t_{11}=0}^{k}\sum_{t_{00}=0}^{n-k}\frac{1}{(n-k-t_{00}+t_{11}+1)} + U = \sum_{t_{00}=0}^{n-k}\sum_{t_{11}=0}^{k}\frac{1}{(n-k-t_{00}+t_{11}+1)} + U = \tag{A.20}$$
$$= \sum_{t_{00}=0}^{n-k}(H_{n-t_{00}+1}-H_{n-k-t_{00}}) + U = \sum_{j=1}^{n+1}H_j - \sum_{j=1}^{k}H_j - \sum_{j=1}^{n-k}H_j + U = 2U$$

Logo, $V = 2U \implies \frac{U}{V} = \frac{1}{2} \implies d^*_{II}(k) = \frac{1}{2}$

Podemos também obter $\mathbb{P}(K=k|\theta)$ utilizando a equação A.11:

$$\mathbb{P}(K=k|\theta) = \int_0^1\int_0^1 \mathbb{P}(K=k|\theta,a,b)\,dPa\,dPb =$$
$$= \sum_{t_{11}=0}^{k}\sum_{t_{00}=0}^{n-k}\binom{k}{t_{11}}\binom{n-k}{t_{00}}\beta(k-t_{11}+1,t_{00}+1)\beta(n-k-t_{00}+1,t_{11}+1)\theta^{n+t_{11}-k-t_{00}}(1-\theta)^{k+t_{00}-t_{11}} =$$
$$= n!\sum_{t_{11}=0}^{k}\sum_{t_{00}=0}^{n-k}\frac{\theta^{n+t_{11}-t_{00}-k}(1-\theta)^{k+t_{00}-t_{11}}}{(k-t_{11}+t_{00}+1)!(n-k-t_{00}+t_{11}+1)!}\mathbb{1}_{0,\dots,n}(k) \tag{A.21}$$

Além disso, podemos calcular o risco de Bayes para o estimador $d_{II}(k) = \frac{1}{2}$, considerando, para simplificar, $\theta \sim$ Unif(0,1):

$$r_{II} = r(d^*_{II}(k),L) = \mathbb{E}[\mathbb{E}_\theta(L(\theta,d^*_{II}(k))] = \int_0^1\int_0^1\int_0^1\sum_{k=0}^{n}\big(\theta-\frac{1}{2}\big)^2\mathbb{P}(K=k|\theta,a,b)p(a)p(b)p_\Theta(\theta)\,da\,db\,d\theta =$$
$$= \int_0^1\big(\theta^2-\theta+\frac{1}{4}\big)\int_0^1\int_0^1\sum_{k=0}^{n}\mathbb{P}(K=k|\theta,a,b)p(a)p(b)p_\Theta(\theta)\,da\,db\,d\theta = \int_0^1\big(\theta^2-\theta+\frac{1}{4}\big)p_\Theta(\theta)\,d\theta = \frac{1}{12} \tag{A.22}$$

A.2.2 Caso III (exemplo 3.2.3)

Para esse caso, consideramos que $Y_i|x_i = 1, a \sim$ Bernoulli$(1-a)$ e $Y_i|x_i = 0 \sim$ Degenerada(0).

Usando a formulação dada em A.9, temos que $t_{01} = 0$ pois $Y_i|x_i = 0 \sim$ Degenerada(0). Como

$t_{00}+t_{01}+t_{11}+t_{10}=t_{00}+t_{11}+t_{10}=n$ e $t_{11}+t_{01}=k$, temos que $t_{11}=k$ e $t_{10}=n-k-t_{00}$. Logo,

$$\begin{aligned}\mathbb{P}(\boldsymbol{Y}=\boldsymbol{y}|\theta,a)&=\sum_{t_{00}=0}^{n-k}\binom{n-k}{t_{00}}(1-a)^k a^{n-k-t_{00}}\theta^{n-t_{00}}(1-\theta)^{t_{00}}=\\&=[\theta(1-a)]^k\sum_{t_{00}=0}^{n-k}\binom{n-k}{t_{00}}(a\theta)^{n-k-t_{00}}(1-\theta)^{t_{00}}=[\theta(1-a)]^k[1-\theta(1-a)]^{n-k}[\theta(1-a)]^k=\\&=\prod_{i=1}^{n}[1-(\theta-a\theta)]^{1-y_i}[\theta(1-a)]^{y_i}=\prod_{i=1}^{n}\mathbb{P}(Y_i=y_i|\theta,a)\end{aligned}\tag{A.23}$$

Então, para o caso III, $Y_i|\theta,a\sim$ CIID Bernoulli$(\theta(1-a))$, $i=\{1,...,n\}$ (e, portanto, permutáveis).

Assim, para $a\sim\beta(\alpha_a,\beta_a)$, temos a densidade de probabilidade da distribuição à posteriori:

$$\begin{aligned}f_{III}(\theta|\boldsymbol{y})&\propto\int_0^1\sum_{\boldsymbol{x}\in\{0,1\}^n}p(\boldsymbol{y}|\boldsymbol{x},a)p(\boldsymbol{x}|\theta)p(\theta)p(a)\,da\propto\\&\propto\int_0^1\sum_{t_{00}=0}^{n-k}\binom{n-k}{t_{00}}(1-a)^k a^{n-k-t_{00}}\theta^{\alpha_0-1+n-t_{00}}(1-\theta)^{\beta_0-1+t_{00}}p(a)\,da\propto\\&\propto\sum_{t_{00}=0}^{n-k}\binom{n-k}{t_{00}}\beta(\alpha_a+n-k-t_{00},\beta_a+k)\theta^{\alpha_0-1+n-t_{00}}(1-\theta)^{\beta_0-1+t_{00}}\end{aligned}\tag{A.24}$$

Então,

$$f_{III}(\theta|\boldsymbol{y})=\frac{\sum_{t_{00}=0}^{n-k}\binom{n-k}{t_{00}}\beta(\alpha_a+n-k-t_{00},\beta_a+k)\theta^{\alpha_0-1+n-t_{00}}(1-\theta)^{\beta_0-1+t_{00}}}{\sum_{t_{00}=0}^{n-k}\binom{n-k}{t_{00}}\beta(\alpha_a+n-k-t_{00},\beta_a+k)\beta(\alpha_0+n-t_{00},\beta_0+t_{00})}\tag{A.25}$$

Fazendo $\alpha_0=\alpha_a=\beta_0=\beta_a=1$,

$$f_{III}(\theta|\boldsymbol{y})=\frac{\sum_{t_{00}=0}^{n-k}\binom{n+1}{t_{00}}\theta^{n-t_{00}}(1-\theta)^{t_{00}}}{\sum_{t_{00}=0}^{n-k}\frac{1}{n-t_{00}+1}}\tag{A.26}$$

Note que, usando A.1.1,

$$\sum_{t_{00}=0}^{n-k}\frac{1}{n-t_{00}+1}=\sum_{s=k+1}^{n+1}\frac{1}{s}=H_{n+1}-H_k=\psi(n+2)-\psi(k+1)\tag{A.27}$$

Temos que:

$$f_{III}(\theta|\boldsymbol{y})=\frac{\sum_{t_{00}=0}^{n-k}\binom{n+1}{t_{00}}\theta^{n-t_{00}}(1-\theta)^{t_{00}}}{\psi(n+2)-\psi(k+1)}\tag{A.28}$$

Assim,

$$\mathbb{E}_{III}(\Theta|\boldsymbol{y}) = \frac{\sum_{t_{00}=0}^{n-k}\binom{n+1}{t_{00}}\beta(n+2-t_{00},t_{00}+1)}{\psi(n+2)-\psi(k+1)} = \frac{\sum_{t_{00}=0}^{n-k}\frac{(n+1)!(n-t_{00}+1)!(t_{00})!}{(n+2)!(t_{00})!(n-t_{00}+1)!}}{\psi(n+2)-\psi(k+1)} =$$
$$= \frac{n+1-\sum_{i=1}^{n} y_i}{(n+2)(\psi(n+2)-\psi(1+\sum_{i=1}^{n} y_i))} \tag{A.29}$$
$$\therefore d^*_{III}(k) = \frac{n-k+1}{(n+2)[\psi(n+2)-\psi(k+1)]}$$

Podemos ainda calcular a distribuição marginal $K|\theta$ usando A.23 ($K|\theta, a \sim \text{Bin}(n,\theta(1-a))$). Então, com $a \sim \text{Unif}(0,1)$,

$$\mathbb{P}(K=k|\theta) = \int_0^1 \mathbb{P}(K=k|\theta,a)\,dPa = \binom{n}{k}\int_0^1 [\theta(1-a)]^k[1-\theta(1-a)]^{n-k}\mathbb{1}_{\{0,1,\dots,n\}}(k)\;\;dPa =$$
$$= \binom{n}{k}\theta^k\int_0^1 (1-a)^k[1-\theta(1-a)]^{n-k}\mathbb{1}_{\{0,1,\dots,n\}}(k)\,dPa = \binom{n}{k}\frac{\theta^k F_{2,1}(k-n,k+1,k+2,\theta)}{k+1}\mathbb{1}_{\{0,1,\dots,n\}}(k) \tag{A.30}$$

em que, a última igualdade decorre do resultado A.6. Assim, o risco de bayes para a decisão $d_{III}(k)$ pode ser dado por:

$$r_{III} = r(\theta, d^*_{III}(k), L) = \mathbb{E}[\mathbb{E}_\theta(L(\theta, d^*_{III}(K)))] =$$
$$= \int_0^1 \sum_{k=0}^{n}(\theta - d^*_{III}(k))^2\binom{n}{k}\frac{\theta^k F_{2,1}(k-n,k+1,k+2,\theta)}{k+1}\,dP\theta =$$
$$= \int_0^1 \theta^2 + \sum_{k=0}^{n}(-2\theta d^*_{III}(k) + d^*_{III}(k)^2)\binom{n}{k}\frac{\theta^k F_{2,1}(k-n,k+1,k+2,\theta)}{k+1}\,dP\theta = \tag{A.31}$$
$$\therefore r_{III} \overset{**}{=} \frac{1}{3} + \sum_{k=0}^{n}\frac{\binom{n}{k}d^*_{III}(k)}{k+1}\Big[\frac{-2F_{3,2}(k-n,k+1,k+2,k+2,k+3,1)}{k+2} +$$
$$+ d^*_{III}(k)\frac{F_{3,2}(k-n,k+1,k+1,k+2,k+2,1)}{k+1}\Big]$$

Em que para a passagem $**$, utilizamos o resultado A.8.

A.2.3 Cálculos para o caso IV (exemplo 3.2.4)

Para esse caso, consideramos que $Y_i|x_i = 1, \sim$ Degenerada(1) e $Y_i|x_i = 0, b \sim$ Bernoulli(b). Usando novamente A.9, temos que $t_{10} = 0 \implies t_{00} = n - k$.

$$\begin{aligned}
&\mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}|\theta, b) = \sum_{t_{11}=0}^{k} \binom{k}{t_{11}} b^{k-t_{11}}(1-b)^{n-k}\theta^{t_{11}}(1-\theta)^{n-t_{11}} = \\
&= [(1-b)(1-\theta)]^{n-k} \sum_{t_{11}=0}^{k} \binom{k}{t_{11}} [b(1-\theta)]^{k-t_{11}}\theta^{t_{11}} = [(1-b)(1-\theta)]^{n-k}[\theta + b(1-\theta)]^{k} = \\
&= \prod_{i=1}^{n} [1-(\theta+b-b\theta)]^{1-y_i}(\theta+b-b\theta)^{y_i} = \prod_{i=1}^{n} \mathbb{P}(Y_i|\theta, b)
\end{aligned} \tag{A.32}$$

Então, para o caso IV, $Y_i|\theta, b \sim$ Bernoulli($\theta + b - b\theta$) e $Y_1, ..., Y_n$ são independentes (e, portanto, permutáveis).

Assim, com $b \sim \beta(\alpha_b, \beta_b)$, temos a densidade de probabilidade da distribuição à posteriori:

$$\begin{aligned}
f_{IV}(\theta|\boldsymbol{y}) &\propto \int_0^1 \sum_{\boldsymbol{x}\in\{0,1\}^n} p(\boldsymbol{y}|\boldsymbol{x}, b)p(\boldsymbol{x}|\theta)p(\theta)\, dPb \propto \\
&\propto \int_0^1 \sum_{t_{11}=0}^{k} \binom{k}{t_{11}} b^{k-t_{11}}(1-b)^{n-k}\theta^{\alpha_0+t_{11}-1}(1-\theta)^{\beta_0+n-t_{11}-1} p(b)\, db \propto \\
&\propto \sum_{t_{11}=0}^{k} \binom{k}{t_{11}} \beta(\alpha_b + k - t_{11}, \beta_b + n - k)\theta^{\alpha_0+t_{11}-1}(1-\theta)^{\beta_0+n-t_{11}-1} \\
\therefore f_{IV}(\theta|\boldsymbol{y}) &= \frac{\sum_{t_{11}=0}^{k} \binom{k}{t_{11}} \beta(\alpha_b + k - t_{11}, \beta_b + n - k)\theta^{\alpha_0+t_{11}-1}(1-\theta)^{\beta_0+n-t_{11}-1}}{\sum_{t_{11}=0}^{k} \binom{k}{t_{11}} \beta(\alpha_b + k - t_{11}, \beta_b + n - k)\beta(\alpha_0 + t_{11}, \beta_0 + n - t_{11})}
\end{aligned} \tag{A.33}$$

Fazendo $\alpha_0 = \alpha_b = \beta_0 = \beta_b = 1$, temos que

$$f_{IV}(\theta|\boldsymbol{y}) = \frac{\sum_{t_{11}=0}^{k} \frac{k!(k-t_{11})!(n-k)!\theta^{\alpha_0+t_{11}-1}(1-\theta)^{\beta_0+n-t_{11}-1}}{t_{11}!(k-t_{11})!(n-t_{11}+1)!}}{\sum_{t_{11}=0}^{k} \frac{k!(k-t_{11})!(n-k)!t_{11}!(n-t_{11})!}{t_{11}!(k-t_{11})!(n-t_{11}+1)!(n+1)!}} = \frac{\sum_{t_{11}=0}^{k} \binom{n+1}{t_{11}}\theta^{t_{11}}(1-\theta)^{n-t_{11}}}{\sum_{t_{11}=0}^{k} \frac{1}{n-t_{11}+1}} \tag{A.34}$$

Note que $\sum_{t_{11}=0}^{k} \frac{1}{n-t_{11}+1} = \sum_{s=n-k+1}^{n+1} \frac{1}{s} = H_{n+1} - H_{n-k} = \psi(n+2) - \psi(n-k+1)$, assim

$$f_{IV}(\theta|\boldsymbol{y}) = \frac{\sum_{t_{11}=0}^{k} \binom{n+1}{t_{11}}\theta^{t_{11}}(1-\theta)^{n-t_{11}}}{\psi(n+2) - \psi(n-k+1)} \tag{A.35}$$

Assim,

$$\mathbb{E}_{IV}(\Theta|\boldsymbol{y}) = \frac{\sum_{t_{11}=0}^{k} \binom{n+1}{t_{11}} \beta(t_{11}+2, n-t_{11}+1)}{\psi(n+2)-\psi(n-k+1)} = \frac{\sum_{t_{11}=0}^{k} \frac{t_{11}+1}{(n-t_{11}+1)}}{(n+2)[\psi(n+2)-\psi(n-k+1)]} \tag{A.36}$$

Note que

$$\begin{aligned}\sum_{t_{11}=0}^{k} \frac{t_{11}+1}{(n-t_{11}+1)} &= -\sum_{t_{11}=0}^{k} [\frac{n-t_{11}+1}{(n-t_{11}+1)} - \frac{n+2}{(n-t_{11}+1)}] = -(k+1) + \sum_{t_{11}=0}^{k} \frac{n+2}{n-t_{11}+1} = \\ &= (n+2)[\psi(n+2)-\psi(n-k+1)] - (k+1)\end{aligned} \tag{A.37}$$

Aplicando esse resultado em A.36,

$$\begin{aligned}\mathbb{E}_{IV}(\Theta|\boldsymbol{y}) &= 1 - \frac{1+\sum_{i=1}^{n} y_i}{(n+2)[\psi(n+2)-\psi(n+1-\sum_{i=1}^{n} y_i)]} \\ \therefore d_{IV}(k) &= 1 - \frac{1+k}{(n+2)[\psi(n+2)-\psi(n-k+1)]}\end{aligned} \tag{A.38}$$

Para calcularmos o risco de Bayes, primeiro, vamos obter a distribuição marginal de $K|\theta$. Usando A.32, considerando $b \sim$ Unif(0,1),

$$\begin{aligned}\mathbb{P}(K=k|\theta) &= \int_0^1 \mathbb{P}(K=k|\theta,b)\, dPb = \int_0^1 \binom{n}{k} (\theta+b-b\theta)^k (1-\theta-b+b\theta)^{n-k} \mathbb{1}_{\{0,1,\dots,n\}}(k)\, dPb = \\ &= \binom{n}{k}(1-\theta)^{n-k} \int_0^1 [1-(1-b)(1-\theta)]^k (1-b)^{n-k} \mathbb{1}_{\{0,1,\dots,n\}}(k)\, dPb = \\ &= \binom{n}{k}(1-\theta)^{n-k} \int_0^1 [1-u(1-\theta)]^k u^{n-k} \mathbb{1}_{\{0,1,\dots,n\}}(k)\, du \\ &\overset{*}{=} \frac{\binom{n}{k}(1-\theta)^{n-k} F_{2,1}(-k, n-k+1, n-k+2, 1-\theta)}{n-k+1} \mathbb{1}_{\{0,1,\dots,n\}}(k)\end{aligned} \tag{A.39}$$

Em que, para a passagem $*$, utilizamos o resultado A.6. Assim, o risco de bayes para a decisão $d_{IV}(k)$ pode ser dado por:

$$
\begin{aligned}
r_{IV} = r(\theta, d^*_{IV}(k), L) = \mathbb{E}[\mathbb{E}_\theta(L(\theta, d^*_{IV}(K)))] = \\
= \int_0^1 \sum_{k=0}^{n} (\theta - d^*_{IV}(k))^2 \frac{\binom{n}{k}(1-\theta)^{n-k} F_{2,1}(-k, n-k+1, n-k+2, 1-\theta)}{n-k+1} \, dP\theta = \\
= \int_0^1 \theta^2 + \sum_{k=0}^{n} (-2\theta d^*_{IV}(k) + d^*_{IV}(k)^2) \frac{\binom{n}{k}(1-\theta)^{n-k} F_{2,1}(-k, n-k+1, n-k+2, 1-\theta)}{n-k+1} \, dP\theta = \\
\therefore r_{IV} \overset{**}{=} \frac{1}{3} + \sum_{k=0}^{n} \frac{\binom{n}{k} d^*_{IV}(k)}{(n-k+1)^2} \Big[\frac{-2F_{3,2}(-k, n-k+1, n-k+1, n-k+2, n-k+3, 1)}{n-k+2} + \\
+ d^*_{IV}(k) F_{3,2}(-k, n-k+1, n-k+1, n-k+2, n-k+2, 1) \Big]
\end{aligned}
\tag{A.40}
$$

Em que para a passagem $**$, utilizamos o resultado A.8.

A.2.4 Caso V (3.2.5)

Para esse caso, consideramos $Y_i|x_i \sim$ Degenerada(s), $s \in \{0, 1\}$, s fixado.

$$
\begin{aligned}
\mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}|\theta) = \sum_{\boldsymbol{x} \in \{0,1\}^n} \mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}|\boldsymbol{x}) \mathbb{P}(\boldsymbol{X} = \boldsymbol{x}|\theta) = \\
= \sum_{\boldsymbol{x} \in \{0,1\}^n} \prod_{i=1}^{n} \mathbb{1}_s(y_i) P(\boldsymbol{X} = \boldsymbol{x}|\theta) = \prod_{i=1}^{n} \mathbb{1}_s(y_i)
\end{aligned}
\tag{A.41}
$$

Logo,

$$
\begin{cases}
\mathrm{P}(\boldsymbol{Y}{=}\boldsymbol{y}|\theta) = 1 \iff \sum_{i=1}^{n} y_i = n, \text{ se } s = 1 \\
\mathrm{P}(\boldsymbol{Y}{=}\boldsymbol{y}|\theta) = 1 \iff \sum_{i=1}^{n} y_i = 0, \text{ se } s = 0
\end{cases}
$$

Para $\boldsymbol{y} = \{1\}^n$ e $s = 1$,

$$
f_V(\theta|\boldsymbol{y}) \propto \sum_{\boldsymbol{x} \in \{0,1\}^n} \mathbb{P}(\boldsymbol{X} = \boldsymbol{x}|\theta) p(\theta) \propto p(\theta) \tag{A.42}
$$

Logo, a posteriori será a própria priori. Um resultado análogo é obtido para o caso em que $\boldsymbol{y} = \{0\}^n$ e $s = 0$.

Portanto, para esse caso, considerando $\Theta \sim \text{Unif}(0, 1)$, tanto para $\sum_{i=1}^{n} y_i = n$ com $s = 1$ quanto para $\sum_{i=1}^{n} y_i = 0$ com $s = 0$,

$$
\mathbb{E}_V(\Theta|\boldsymbol{y}) = \mathbb{E}_V(\Theta) = \frac{1}{2} \tag{A.43}
$$

Apêndice B

Apêndice

k	$\bar{c}_I$	$\bar{c}_{II}$	$\bar{c}_{III}$	$\bar{c}_{IV}$
0	1.25-08	1.46	3.63e-01	1.25e-08
1	2.87-07	1.53	5.20e-01	1.51e-07
2	3.23-06	1.60	6.63e-01	1.20e-06
3	2.38-05	1.66	8.12e-01	7.04e-06
4	1.30-04	1.72	9.77e-01	3.26e-05
5	5.57-04	1.78	1.16	1.24e-04
6	1.96-03	1.83	1.39	4.04e-04
7	5.86-03	1.90	1.66	1.13e-03
8	1.51-02	1.96	2.01	2.81e-03
9	3.48-02	2.02	2.45	6.24e-03
10	7.21-02	2.09	3.05	1.25e-02
11	1.37-01	2.16	3.87	2.31e-02
12	2.46-01	2.23	5.04	3.95e-02
13	4.21-01	2.30	6.77	6.32e-02
14	6.98-01	2.37	9.38	9.54e-02
15	1.13	2.44	1.35e+01	1.36e-01
16	1.85	2.51	2.02e+01	1.87e-01
17	3.04	2.57	3.17e+01	2.48e-01
18	5.10	2.63	5.23e+01	3.17e-01
19	8.84	2.68	9.09e+01	3.93e-01
20	1.59+01	2.73	1.66e+02	4.77e-01
21	3.03+01	2.76	3.24e+02	5.66e-01
22	6.10+01	2.79	6.71e+02	6.61e-01
23	1.30+02	2.81	1.47e+03	7.59e-01
24	2.98+02	2.82	3.46e+03	8.63e-01
25	7.32+02	2.83	8.66e+03	9.70e-01
26	1.92+03	2.82	2.31e+04	1.08
27	5.45+03	2.81	6.63e+04	1.19
28	1.66+04	2.79	2.03e+05	1.32
29	5.45+04	2.76	6.69e+05	1.44
30	1.93+05	2.73	2.36e+06	1.58
31	7.41+05	2.68	9.00e+06	1.72
32	3.08+06	2.63	3.69e+07	1.87
33	1.38+07	2.57	1.63e+08	2.02
34	6.81+07	2.51	7.85e+08	2.19
35	3.65+08	2.44	4.09e+09	2.36
36	2.13+09	2.37	2.32e+10	2.55
37	1.37+10	2.30	1.43e+11	2.75
38	9.80+10	2.23	9.77e+11	2.96
39	7.74+11	2.16	7.32e+12	3.20
40	6.83+12	2.09	6.07e+13	3.45
41	6.78+13	2.02	5.62e+14	3.73
42	7.65+14	1.96	5.85e+15	4.04
43	9.94+15	1.90	6.91e+16	4.39
44	1.50+17	1.83	9.39e+17	4.80
45	2.72+18	1.78	1.48e+19	5.26
46	6.04+19	1.72	2.80e+20	5.82
47	1.70+21	1.66	6.46e+21	6.52
48	6.57+22	1.60	1.90e+23	7.46
49	3.86+24	1.53	7.60e+24	8.86
50	4.64+26	1.46	4.64e+26	1.16e+01

Tabela B.1: ***Majorantes de constantes*** $\bar{c}$ ***para*** $\theta_0 = 0,3$***,*** $n = 50$ ***para teste de hipótese bayesiano com perda*** 0-1-c

k	$\bar{c}_I$	$\bar{c}_{II}$	$\bar{c}_{III}$	$\bar{c}_{IV}$
0	4.44e-16	1	1.81e-01	4.44e-16
1	2.30e-14	1	2.45e-01	1.18e-14
2	5.89e-13	1	2.98e-01	2.08e-13
3	9.83e-12	1	3.47e-01	2.68e-12
4	1.20e-10	1	3.97e-01	2.72e-11
5	1.16e-09	1	4.49e-01	2.26e-10
6	9.16e-09	1	5.03e-01	1.58e-09
7	6.05e-08	1	5.62e-01	9.52e-09
8	3.43e-07	1	6.25e-01	4.99e-08
9	1.69e-06	1	6.95e-01	2.31e-07
10	7.36e-06	1	7.72e-01	9.55e-07
11	2.85e-05	1	8.60e-01	3.55e-06
12	9.90e-05	1	9.59e-01	1.19e-05
13	3.10e-04	1	1.07	3.67e-05
14	8.85e-04	1	1.20	1.03e-04
15	2.30e-03	1	1.36	2.67e-04
16	5.51e-03	1	1.55	6.40e-04
17	1.21e-02	1	1.79	1.42e-03
18	2.50e-02	1	2.08	2.94e-03
19	4.81e-02	1	2.46	5.70e-03
20	8.74e-02	1	2.95	1.03e-02
21	1.51e-01	1	3.62	1.78e-02
22	2.50e-01	1	4.54	2.89e-02
23	4.04e-01	1	5.83	4.48e-02
24	6.39e-01	1	7.73	6.64e-02
25	1.00	1	1.05e+01	9.43e-02
26	1.56	1	1.50e+01	1.29e-01
27	2.47	1	2.22e+01	1.71e-01
28	3.98	1	3.44e+01	2.20e-01
29	6.62	1	5.61e+01	2.76e-01
30	1.14e+01	1	9.63e+01	3.38e-01
31	2.07e+01	1	1.75e+02	4.06e-01
32	3.99e+01	1	3.39e+02	4.79e-01
33	8.20e+01	1	7.03e+02	5.58e-01
34	1.81e+02	1	1.56e+03	6.42e-01
35	4.33e+02	1	3.73e+03	7.32e-01
36	1.12e+03	1	9.67e+03	8.28e-01
37	3.21e+03	1	2.72e+04	9.31e-01
38	1.00e+04	1	8.35e+04	1.04
39	3.50e+04	1	2.81e+05	1.16
40	1.35e+05	1	1.04e+06	1.29
41	5.90e+05	1	4.32e+06	1.43
42	2.91e+06	1	2.00e+07	1.59
43	1.65e+07	1	1.04e+08	1.77
44	1.09e+08	1	6.30e+08	1.98
45	8.59e+08	1	4.40e+09	2.22
46	8.27e+09	1	3.66e+10	2.51
47	1.01e+11	1	3.71e+11	2.87
48	1.69e+12	1	4.80e+12	3.35
49	4.33e+13	1	8.41e+13	4.07
50	2.25e+15	1	2.25e+15	5.51

Tabela B.2: ***Majorantes de constantes*** $\bar{c}$ ***para*** $\theta_0 = 0,5$***,*** $n = 50$ ***para teste de hipótese bayesiano com perda*** 0-1-c

k	$\bar{c}_I$	$\bar{c}_{II}$	$\bar{c}_{III}$	$\bar{c}_{IV}$
0	2.15e-27	0.68	8.56e-02	2.15e-27
1	2.58e-25	0.65	1.12e-01	1.31e-25
2	1.52e-23	0.62	1.33e-01	5.25e-24
3	5.84e-22	0.60	1.53e-01	1.54e-22
4	1.65e-20	0.58	1.71e-01	3.56e-21
5	3.66e-19	0.56	1.89e-01	6.72e-20
6	6.62e-18	0.54	2.08e-01	1.06e-18
7	1.00e-16	0.52	2.27e-01	1.44e-17
8	1.30e-15	0.50	2.46e-01	1.70e-16
9	1.47e-14	0.49	2.67e-01	1.77e-15
10	1.46e-13	0.47	2.89e-01	1.64e-14
11	1.29e-12	0.46	3.12e-01	1.36e-13
12	1.01e-11	0.44	3.36e-01	1.02e-12
13	7.25e-11	0.43	3.63e-01	6.95e-12
14	4.67e-10	0.42	3.91e-01	4.30e-11
15	2.73e-09	0.40	4.22e-01	2.44e-10
16	1.46e-08	0.39	4.56e-01	1.27e-09
17	7.19e-08	0.38	4.93e-01	6.10e-09
18	3.24e-07	0.37	5.34e-01	2.70e-08
19	1.34e-06	0.37	5.80e-01	1.11e-07
20	5.16e-06	0.36	6.31e-01	4.22e-07
21	1.83e-05	0.36	6.90e-01	1.49e-06
22	6.02e-05	0.35	7.56e-01	4.91e-06
23	1.83e-04	0.35	8.33e-01	1.50e-05
24	5.19e-04	0.35	9.23e-01	4.31e-05
25	1.36e-03	0.35	1.03	1.15e-04
26	3.34e-03	0.35	1.15	2.88e-04
27	7.65e-03	0.35	1.31	6.77e-04
28	1.63e-02	0.35	1.51	1.49e-03
29	3.29e-02	0.36	1.76	3.07e-03
30	6.25e-02	0.36	2.09	5.99e-03
31	1.13e-01	0.37	2.53	1.09e-02
32	1.95e-01	0.37	3.15	1.90e-02
33	3.28e-01	0.38	4.03	3.14e-02
34	5.39e-01	0.39	5.32	4.93e-02
35	8.78e-01	0.40	7.30	7.40e-02
36	1.43	0.42	1.04e+01	1.06e-01
37	2.37	0.43	1.58e+01	1.47e-01
38	4.05	0.44	2.52e+01	1.98e-01
39	7.26	0.46	4.31e+01	2.57e-01
40	1.38e+01	0.47	7.96e+01	3.27e-01
41	2.87e+01	0.49	1.60e+02	4.06e-01
42	6.58e+01	0.50	3.55e+02	4.97e-01
43	1.70e+02	0.52	8.79e+02	6.00e-01
44	5.08e+02	0.54	2.47e+03	7.18e-01
45	1.79e+03	0.56	8.01e+03	8.56e-01
46	7.69e+03	0.58	3.06e+04	1.02
47	4.18e+04	0.60	1.41e+05	1.23
48	3.09e+05	0.62	8.33e+05	1.50
49	3.47e+06	0.65	6.59e+06	1.92
50	7.94e+07	0.68	7.94e+07	2.75

Tabela B.3: ***Majorantes de constantes*** $\bar{c}$ ***para*** $\theta_0 = 0,7$**,** $n = 50$ ***para teste de hipótese bayesiano com perda*** 0-1-c

Referências Bibliográficas

Bentham(1780) Jeremy Bentham. *An Introduction to the Principles of Morals and Legislation*. T.PAYNE, and SON, 15th edição. Citado na pág. 9

Biggio e Roli(2018) B. Biggio e F. Roli. Wild Patterns: Ten Years After the Rise of Adversarial Machine Learning. *Pattern Recognition*, 84:317–331. Citado na pág. 1, 13

Biggio *et al.***(2015)** B. Biggio, G. Fumera, P. Russu, L. Didaci e F. Roli. Adversarial Biometric Recognition. *IEEE SIGNAL PROCESSING MAGAZINE*, 15:31–41. Citado na pág. 1

Bishop(2006) C.M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 1st edição. Citado na pág. 13

Brandão *et al.***(2014)** FGSL Brandão, AW Harrow, JR Lee e Y Peres. Adversarial hypothesis testing and a quantum stein's lemma for restricted measurements. *5th conference on Innovations in Theoretical Computer Science*. Citado na pág. 14

Casella e Berger(2016) G. Casella e R.L. Berger. *Inferência Estatística*. CENGAGE Learning, 2nd edição. Citado na pág. 10

Choi e Srivastava(2011) Junesang Choi e HM Srivastava. Some summation formulas involving harmonic numbers and generalized harmonic numbers. *Mathematical and Computer Modelling*, 54:2220–2234. Citado na pág. 39

Dalvi *et al.***(2004)** N. Dalvi, P. Domingos, Mausam, Sumit Sanghai e Deepak Verma. Adversarial classification. páginas 99–108. Citado na pág. 13

DeGroot(2004) M.H. DeGroot. *Optimal Statistical Decisions*. John Wiley and Sons, wiley classics library edition edição. Citado na pág. 3, 4, 8, 9, 10, 17

DeGroot e Schervish() M.H. DeGroot e M.J. Schervish. *Probability and Statistics*, páginas 69,370–371. Addilson Wesley, 3rd edição. Citado na pág. 7, 8

Dixon *et al.***(2020)** M.F. Dixon, I. Halperin e P. Bilokon. *Machine Learning in Finance*. Springer, 1st edição. Citado na pág. 13

Fisher(1956) R.A. Fisher. *Statistical Methods and Scientific Inference. Oliver and Boyd*. Edinburgh. Citado na pág. 6

González-Ortega *et al.***(2019)** J. González-Ortega, D. Ríos Insua, F. Ruggeri e R. Soyer. Hypothesis testing in presence of adversaries. *The American Statistician*, 0(0):1–18. doi: 10.1080/00031305.2019.1630001. URL https://doi.org/10.1080/00031305.2019.1630001. Citado na pág. 1, 2, 13, 14, 15, 16

Greenland e Robins(1986) S. Greenland e J.M. Robins. Identifiability, Exchangeability, and Epidemiological Confounding. *International Journal of Epidemiology*, 15:413–419. Citado na pág. 6

Hannah(2013) J.P. Hannah. *Identities for the gamma and hypergeometric functions: an overview from Euler to the present*. Dissertação de Mestrado, School of Mathematics University of Witwatersrand, Johannesburg, South Africa. Citado na pág. 40

Insua *et al.***(2009)** D. Ríos Insua, J. Ríos e D. L. Banks. Adversarial risk analysis. *Journal of the American Statistical Association*, (104):841–854. Citado na pág. 14, 15

Insua *et al.***(2012)** D. Ríos Insua, D. L. Banks, J. Ríos e J. González-Ortega. Adversarial risk analysis for structured expert judgement modelling. *Risk Analysis*, (32):894–915. Citado na pág. 13

Insua *et al.***(2018)** David Ríos Insua, Jorge González-Ortega, David Banks e Jesús Ríos. *The Mathematics of the Uncertain - A Tribute to Pedro Gil - Concept Uncertainty in Adversarial Statistical Decision Theory*. Springer, 1st edição. Citado na pág. 1, 14, 15, 16

Insua *et al.***(2019)** David Ríos Insua, Aitor Couce Vieira, José Antonio Rubio, Wolter Pieters, Katsiaryna Labunets e Daniel Garcia Rasines. An adversarial risk analysis framework for cybersecurity. *Risk Analysis*, páginas 1–21. doi: 10.1111/risa.13331. Citado na pág. 1, 13

Joseph *et al.***(2019)** Anthony D. Joseph, Blaine Nelson, Benjamin I. P. Rubinstein e J. D. Tygar. *Adversarial Machine Learning*, páginas 4–9. Cambridge University Press. doi: 10.1017/9781107338548. Citado na pág. 13, 14

Julious e Mulee(1994) S.A. Julious e M.A. Mulee. Confounding and Simpson's paradox. *BMJ*, 309:1480–1481. Citado na pág. 6

Kim *et al.***(2012)** S.H. Kim, Q-H. Wang e J.B. Ullrich. A Comparative Study of Cyberattacks. *Communications of the ACM*, 55:66–73. Citado na pág. 13

Kshetri e DeFranco(2020) N. Kshetri e J.F. DeFranco. The Economiics of Cyberattacks on Brazil. *Computer*, 53:85–90. Citado na pág. 13

Lima(2019) E.L. Lima. *Curso de análise*, volume 1, página 58. PROJETO EUCLIDES, 15th edição. Citado na pág. 5

Lindley(2014) D.V. Lindley. *Understanding Uncertainty*. WILEY, 2nd edição. Citado na pág. 3, 4

Lindley e Novick(1981) D.V. Lindley e M.R. Novick. The Role of Exchangeability in Inference. *The Annals of Statistics*, 9:45–48. Citado na pág. 5, 6

Mendoza e Gutiérrez-Peña(2010) M. Mendoza e E. Gutiérrez-Peña. *International Encyclopedia of Education*. Elsevier Science, 3rd edição. Citado na pág. 8

Mohri *et al.***(2012)** M. Mohri, A. Rostamizadeh e A. Talwalkar. *Foundations of Machine Learning*. The MIT Press, 1st edição. Citado na pág. 1

Moscati(2020) Ivan Moscati. History of Utility Theory. *BAFFI CAREFIN Centre Research Paper*, 129:1–20. Citado na pág. 9

Myerson(1991) R.B. Myerson. *Game Theory - Analysis of Conflict*. Harvard University Press, 1st edição. Citado na pág. 8, 9

Naqa *et al.***(2015)** I.E. Naqa, R. Li e Murphy M.J. *Machine Learning in Radiation Oncology*. Springer, 1st edição. Citado na pág. 13

Neumann e Oskar(1955) John Von Neumann e Morgenstern Oskar. *Theory of Games and Economic Behavior*. Princenton University Press, 6th edição. Citado na pág. 9

Osborn e Madey(1967) David Osborn e Richard Madey. The Incomplete Beta Function and its Ratio to the Complete Beta Function. *Mathematics of Computation*, 22:159–162. Citado na pág. 28, 40

Pearl(2009) J. Pearl. *Causality*. Cambridge University Press, 2nd edição. Citado na pág. 6

Ríos e Insua(2012) J. Ríos e D. R. Insua. Adversarial risk analysis for counterterrorism modeling. *Risk Analysis*, (32):894–915. Citado na pág. 1, 13

Schervish(1995) M.J. Schervish. *Theory of Statistics*, páginas 26–52,82–84. Springer-Verlag New York, 2nd edição. Citado na pág. 3, 5, 7, 8, 10

Singpurwalla *et al.***(2016)** N.D. Singpurwalla, B.C. Arnold, J.L. Gastwirth, A.S. Gordon e H.K.T. Ng. Adversarial and Amiable Inference in Medical Diagnosis, Reliability, and Survival Analysis. *International Journal of Epidemiology*, 15:413–419. Citado na pág. 1

Stigler(1950) George J. Stigler. The Development of Utility Theory. I. *Journal of Political Economy*, 58:307–327. Citado na pág. 9

Vaz Jr. e Oliveira(2016) J. Vaz Jr. e E.C. Oliveira. *Métodos Matemáticos - Volume I*. EDITORA UNICAMP, 1st edição. Citado na pág. 40

Vorobeychik e Kantarcioglu(2018) Y. Vorobeychik e M. Kantarcioglu. *Adversarial Machine Learning*. Springer, 1st edição. Citado na pág. 13, 15

Wu *et al.***(2000)** Tsu-Chen Wu, Shih-Tong Tu e HM Srivastava. Some Combinatorial Series Identities Associated with the Digamma Function and Harmonic Numbers. *Applied Mathematics Letters*, 13:101–106. Citado na pág. 39