Permutabilidade de quantidades aleatórias binárias e a falácia do apostador

Fernando Vieira Bonassi

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

Programa: **Estatística**

Orientador: **Prof. Dr. Sergio Wechsler**

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

- São Paulo, Março de 2009 -

Permutabilidade de quantidades aleatórias binárias e a falácia do apostador

Este exemplar corresponde à redação
final da dissertação/tese devidamente corrigida
e defendida por Fernando Vieira Bonassi
e aprovada pela Comissão Julgadora.

Banca Examinadora:

- Prof. Dr. Sergio Wechsler - IME-USP.
- Prof. Dr. Claudia Monteiro Peixoto - IME-USP.
- Prof. Dr. Dani Gamerman - UFRJ.

"The most important questions of life are indeed, for the most part, really only problems of probability."
— Pierre Simon Laplace

"Our world, our life, our destiny, are dominated by uncertainty; this is perhaps the only statement we may assert without uncertainty."
— Bruno de Finetti

Aos meus pais, irmã e irmãos,
afilhada e madrinha.

Agradecimentos

Primeiramente, agradeço a meus pais, Edna e Flavio, pelo apoio constante e incondicional para alcance das minhas realizações.

Agradeço a minha família como um todo, por formarem uma estrutura sólida, que orgulhosamente tenho como referência em minha vida pessoal e profissional. Além dos meus pais, fazem parte desse grupo: meus avós, tios, primos, afilhada Isabela, irmã Camila e madrinha Yone.

Registro meu sincero agradecimento aos amigos que estiveram próximos, e que me supriram com força e inspiração para superar os desafios dessa fase: irmãos jacobianos (Davi, Denis, Emerson, Hommenig, Luis e Paulo), a grande amiga Bia (TFPOTW), e os amigos do IME-USP (colegas de Graduação e Mestrado, pessoal da Seção de Alunos, CPG, Sec-Mae, Sandrão, Sylvia do CEA, ...).

Manifesto-me grato aos Professores do IME-USP que participaram mais proximamente da minha formação acadêmica na graduação. Neste grupo estão: Beti Kira, Fabio Prates, Denise Botter, Mônica Sandoval e Luis Gustavo. Pela formação matemática no ensino fundamental, agradeço à Professora Fátima.

Aos Professores da banca, Dani Gamerman e Claudia Peixoto, agradeço pelas correções sugeridas para este trabalho. O mesmo tipo de agradecimento se aplica aos Professores (e tios) Marcelo e Márcio.

É preciso ainda registrar meu agradecimento para três pessoas fundamentais na minha trajetória no Mestrado. São eles os Professores Carlinhos e Julio Stern, e o colega Rafael Bassi. Aos Professores, agradeço por orientações, amizade e apoio. Ao Rafael, agradeço pela grande amizade e pela contribuição brilhante nos trabalhos vinculados ao Mestrado.

Por fim, expresso meu agradecimento especial ao meu mentor, Professor Sergio Wechsler, que além de ter sido um excelente orientador para o projeto, foi um grande amigo que sempre me presenteou com valiosos aconselhamentos, ensinamentos e inspiração.

Resumo

O elemento central deste estudo é o problema de predição em seqüências de variáveis aleatórias binárias (0-1). Modelos são estudados para esse tipo de situação e então relacionados com a Falácia do Apostador - um famoso caso de estudo da Psicologia (também conhecida como Lei da Maturidade). Estudos estatísticos anteriores propõem tal modelagem sob a perspectiva bayesiana. Neles, tem-se a suposição de permutabilidade infinita e, como conseqüência, a maturidade é um comportamento inadmissível. Neste estudo, um novo modelo é apresentado, no qual a crença do apostador não é necessariamente uma falácia. Este é o modelo preditivista usual de população finita e, portanto, somente quantidades com significado operacional (parâmetros operacionais) são envolvidas. Uma classe de prioris para o parâmetro operacional que resulta em modelos não estendíveis é apresentada. Trata-se de uma classe de distribuições que definimos como mais estreitas que a Binomial. Maturidade é uma conseqüência da crença em prioris dessa classe. Apresenta-se ainda uma subclasse referente às distribuições mais estreitas de segunda ordem que a Binomial. Para prioris dessa subclasse tem-se taxa de falha preditiva crescente, que pode ser interpretado como o resultado mais extremo de maturidade. Os resultados deste estudo podem contribuir para o julgamento de quão razoável é a suposição de permutabilidade infinita em relação ao típico comportamento humano. Outra principal contribuição está associada ao estudo de condições de estendibilidade em processos binários.

Palavras-chave: Estatística Bayesiana, Preditivismo, Permutabilidade, Estendibilidade, Processos 0-1, Falácia do Apostador, Lei da Maturidade.

Abstract

We study the problem of prediction in sequences of binary random variables. Models are studied for this kind of situation and then considered vis-à-vis the Gambler's Fallacy - a famous case study in Psychology (also known as Law of Maturity). Previous statistical studies proposed such modeling under the bayesian perspective. In them there is the assumption of exchangeability and, as a result, maturity is a inadmissible behavior. In this study, a new model in which the Gambler's belief need not be a fallacy is presented. This one is the usual finite population model and, therefore, only operationally meaningful quantities (operational parameters) are involved. A class of prior distributions for the operational parameter which yield non-extendable models is presented. It is a class of distributions which we defined as tighter than the Binomial. Maturity is a consequence of the belief in the prior distributions of this class. Furthermore, a subclass which refers to the distributions that are second-order tighter than the Binomial is presented. For prior distributions of this subclass the predictive failure rate is increasing, which can be interpreted as the most extreme case of maturity. The results of this study may contribute on the judgment of how reasonable the assumption of infinite exchangeability is relative to typical human perception. Another major contribution is related to the study on extendibility conditions in binary processes.

Keywords: Bayesian Statistics, Predictivism, Exchangeability, Extendibility, 0-1 Processes, Gambler's Fallacy, Law of Maturity.

Conteúdo

Agradecimentos iii

Resumo v

Abstract vii

Lista de Figuras xi

1 Introdução 1

1.1 Inferência Estatística Bayesiana 1

1.2 Probabilidade Subjetiva 2

1.3 Permutabilidade 3

1.4 Teorema da Representação de De Finetti 3

1.4.1 Estudo do Teorema via Simulação Computacional 6

1.5 Preditivismo 8

1.5.1 Estendibilidade 9

2 Falácia do Apostador: Uma Abordagem Bayesiana 11

2.1 A Falácia do Apostador 11

2.2 Estudos Anteriores 12

2.2.1 O Estudo de O'Neill & Puza 12

2.2.2 O Estudo de Rodrigues & Wechsler 13

2.2.3 Uma Possível Demonstração do Resultado de O'Neill & Puza 14

2.2.4 Considerações sobre os Estudos 16

2.3 Um Novo Modelo 16

2.3.1 Distribuição Binomial como Priori 18

2.3.2 Mistura de Binomiais como Priori 19

2.3.3 Distribuições Mais Estreitas que a Binomial 20

2.3.4 Distribuição Hipergeométrica como Priori 25

2.3.5 Distribuições Mais Estreitas de Segunda Ordem que a Binomial 26

3 Conclusões **31**

3.1 Revisão dos Principais Pontos de Estudo 31

3.2 Considerações Finais 31

3.3 Pontos para Pesquisa Futura 33

A Cálculos Auxiliares **35**

A.1 Exemplo de Não-estendibilidade 35

A.2 Demonstração do resultado de Indiferença 36

A.3 Demonstração do resultado de Equivalência de Modelos 36

A.4 Identidades de Coeficientes Binomiais 37

A.5 Posteriori mais estreita que a Binomial 38

A.6 Distribuição Hipergeométrica simétrica - Mais estreita que a Binomial 39

A.7 Hipergeométrica simétrica - Mais estreita de segunda ordem 39

Referências Bibliográficas **41**

Referências Bibliográficas **41**

Lista de Figuras

1.1 Freqüências Relativas de Processos 0-1 Simulados Computacionalmente: N=10, n=1000 e Distribuição Beta(10,10) para Θ 6

1.2 Histograma das Freqüências Relativas de Processos 0-1 Simulados Computacionalmente: N=10000 e n=400 e Distribuição Beta(10,10) para Θ 7

2.1 Casos de Distribuição a Priori para γ (com N=10): 1 - Uniforme, 2 - Binomial e 3 - Degenerada em $N/2$ 20

2.2 Distribuições de Probabilidade: 1 - Binomial, 2 - Binomial Mais Estreita (I) e 3 - Binomial Mais Estreita(II) 21

2.3 Distribuições de Probabilidade: 1 - Binomial, 2 - Binomial Mais Estreita de Segunda Ordem (I) e 3 - Binomial Mais Estreita de Segunda Ordem (II) . . 27

Capítulo 1

Introdução

1.1 Inferência Estatística Bayesiana

Dá-se o nome de *inferência científica* ao processo de aprendizado sobre alguma característica desconhecida de interesse com base na observação de dados (amostra) vinculados a tal característica. Quando a conexão entre os dados e a característica é expressa em termos probabilísticos então o processo é chamado de *inferência estatística* (Johnson & Kotz [1]).

A abordagem dita bayesiana de inferência estatística é aquela na qual o mecanismo central do processo de inferência é dado pela utilização do Teorema de Bayes, que se descreve da forma a seguir:

$$P(H_i|D) = \frac{P(D|H_i)P(H_i)}{\sum_j P(D|H_j)P(H_j)} \quad (i = 1, 2, \ldots)$$

em que, H_i representa uma hipótese específica a respeito de um estado desconhecido. $P(D|H_i)$ é a probabilidade de se observar o conjunto de dados D caso H_i seja verdadeira. $P(H_i)$ é a probabilidade *a priori* de H_i ser verdadeira (determinada anteriormente à observação dos dados D). E por fim, $P(H_i|D)$, objeto resultante do procedimento, é chamada probabilidade *a posteriori* de H_i dado D.

O procedimento apresentado é apenas uma idéia geral do mecanismo de aprendizado obtido via Teorema de Bayes (em sua forma discreta). Em muitos casos, há uma estrutura relativamente mais complexa. Em geral, os dados são representados por observações de uma amostra $(x_1, x_2, \ldots, x_n)$ e as hipóteses por valores que a característica desconhecida (parâmetro θ) pode assumir.

1.2 Probabilidade Subjetiva

Como visto na seção anterior, dentro do procedimento de inferência estatística um componente que exerce um papel de fundamental importância é a *probabilidade*. Dessa forma, qualquer abordagem de inferência estatística está fortemente vinculada a esse conceito. Três das principais interpretações de probabilidade são: *clássica*, *freqüentista* e *subjetiva*.

O interpretação *clássica* de probabilidade de um evento baseia-se na suposição de que todas os casos são igualmente possíveis de ocorrer (Laplace [2]). Outra interpretação, a *freqüentista*, sustenta que a probabilidade de um evento ocorrer é medida pela frequência relativa de ocorrências desse evento após um grande número de ensaios independentes e nas mesmas condições (Venn [3]). Um ponto controverso de tal definição reside no fato de o conceito de probabilidade ser necessário para definir-se independência, resultando assim, em um problema de circularidade na definição.

A visão *subjetivista* é aquela que interpreta *probabilidade* como o *grau de credibilidade* que um indivíduo tem a respeito de uma proposição, como a ocorrência de um evento, por exemplo (De Finetti [4]). Dessa forma, duas pessoas podem ter probabilidades diferentes em relação a um determinado acontecimento. Em uma situação de lançamento de moeda, por exemplo, alguém poderia atribuir probabilidade 1/2 para *cara* e 1/2 para *coroa* enquanto que outra pessoa, em posse da informação de que aquela moeda é viciada, poderia designar probabilidade 1 para *coroa*.

Além de apresentar tal flexibilidade, a interpretação subjetivista é compatível com um conjunto vasto de situações de simetria de probabilidades sobre as quais pode existir consenso, como no caso em que há indiferença com respeito às faces de um dado e atribui-se probabilidade 1/6 para cada face em seu lançamento.

1.3 Permutabilidade

Dada uma situação em que se tem uma seqüência de variáveis aleatórias $X_1, \ldots, X_n$, um indivíduo pode julgar que os *rótulos* que identificam as variáveis não são informativos; isto é, o grau de credibilidade que é atribuído para a distribuição conjunta $P(x_1, \ldots, x_n)$ permanece o mesmo ao se permutar os valores de $x_1, \ldots, x_n$. A seguir, apresenta-se a descrição formal desse conceito (Paulino et al. [6]):

Definição 1.3.1 (Permutabilidade Finita) *A medida de probabilidade P para um seqüência de variáveis aleatórias $X_1, \ldots, X_n$ é julgada permutável quando a distribuição conjunta de probabilidade indicada satisfaz:*

$$P(x_1, \ldots, x_n) = P(x_{\pi(1)}, \ldots, x_{\pi(n)})$$

para todas permutações de π definidas no conjunto $\{1,\ldots,n\}$.

Um exemplo de medida de probabilidade permutável para duas variáveis aleatórias binárias ocorre quando $P(X_1 = 1, X_2 = 0) = P(X_1 = 0, X_2 = 1)$. Nota-se que independência implica em permutabilidade, porém a volta não é necessariamente válida. Disto, tem-se que a suposição de permutabilidade é mais fraca que a de independência.

Rigorosamente, o conceito de permutabilidade é vinculado à medida de probabilidade da seqüência de variáveis aleatórias. No entanto, é comum entre os autores da literatura estatística utilizarem a afirmação de que a seqüência de variáveis que é permutável.

Tal conceito pode ser estendido para seqüências infinitas. Uma seqüência infinita de variáveis aleatórias é infinitamente permutável se toda subseqüência finita dela for permutável.

1.4 Teorema da Representação de De Finetti

Consideramos como nosso objeto de interesse uma seqüência infinita de variáveis aleatórias binárias ($X_1, X_2, \ldots$, com $X_i = 0$ ou $X_i = 1$, para todo $i = 1, 2, \ldots$). Ao se supor que essa seqüência é infinitamente permutável, tem-se como conseqüência o Teorema da Representação de De Finetti [7]. Trata-se de um resultado que, dada a suposição mencionada, fornece uma representação matemática para distribuição conjunta de probabilidade $p(x_1, \ldots, x_n)$.

Teorema 1.4.1 (Teorema da Representação para variáveis 0-1) *Consideremos um processo de variáveis aleatórias binárias $X_1, X_2, \ldots$ com medida de probabilidade P. Então,*

(I) *A medida do processo é permutável se, e somente se, existe uma função distribuição Q tal que, para todo $n \geq 1$, a distribuição conjunta de probabilidade $p(x_1, \ldots, x_n)$ tem a forma:*

$$p(x_1, \ldots, x_n) = \int_0^1 \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} \, dQ(\theta).$$

(II)

Além disso, quando $n \to \infty$, $\dfrac{\sum^n X_i}{n}$ converge quase certamente.

(III)

E finalmente, chamando $lim_{n\to\infty}\dfrac{\sum^n X_i}{n}$ de Θ, tem-se que $Q(\theta)$ é a função distribuição de Θ.

Seguindo a idéia da demonstração dada em Heath & Sudderth [8], podemos fazer a prova da parte (I) do teorema por meio dos seguintes passos:

Devido à suposição de infinita permutabilidade, a probabilidade de ocorrência de qualquer seqüência de tamanho n é função do número de 1's que a mesma contém (chamamos esse número de k). Denomina-se tal probabilidade por $p_{k,n}$.

Sendo m o tamanho de uma seqüência que contém a seqüência estudada ($n < m$), verifica-se que dado o número de 1's na seqüência maior, então o número de 1's na seqüencia estudada tem distribuição Hipergeométrica.

Dessa forma, para todo $m > n$, é possível escrever que a probabilidade de ocorrer k 1's na seqüência estudada ($p_{k,n}$) é:

$$p_{k,n} = \sum_{r=0}^{m} \frac{\binom{r}{k}\binom{m-r}{n-k}}{\binom{m}{n}} \cdot P(\sum_{j=1}^{m} X_j = r), \quad 0 \leq k \leq n \leq m$$

Por intermédio de uma reparametrização conveniente, a probabilidade de ocorrer k 1's na seqüência estudada ($p_{k,n}$) pode ser reescrita como:

$$p_{k,n} = \binom{n}{k} \int_0^1 \frac{(\theta m)_k[(1-\theta)m]_{n-k}}{(m)_n} \, dQ_m(\theta)$$

* em que $(x)_k = \Pi_{j=0}^{k-1}(x-j)$

Como $m \to \infty$,

$$\frac{(\theta m)_k[(1-\theta)m]_{n-k}}{(m)_n} \to \theta^k(1-\theta)^{n-k}$$

uniformemente em θ. Além disso, tem-se pelo Teorema de Helly (Brunk et al. [9]) que existe uma subseqüência de naturais tal que Q_m converge para Q, que é uma função distribuição; concluindo assim a prova. ■

Um dos pontos de grande importância no Teorema da Representação reside em sua interpretação. Por intermédio dele, é possível traçar equivalência entre a suposição de processo 0-1 infinitamente permutável com uma situação que fosse constituída pelos seguintes aspectos (Bernardo & Smith [10]):

1. $X_1, X_2, \ldots$ fossem variáveis aleatórias Bernoulli, condicionalmente independentes dada uma variável aleatória Θ;

2. fosse atribuída uma distribuição $Q(\theta)$ para a variável aleatória Θ;

3. pela lei forte dos grandes números, Θ fosse o limite (quando $n \to \infty$) da freqüência relativa de sucessos, de forma que $Q(\theta)$ pudesse ser interpretada como a credidibilidade sobre tal limite.

Em outras palavras, no processo 0-1 infinitamente permutável existe uma equivalência com o caso em que as variáveis resultam de uma *amostra aleatória* de uma distribuição Bernoulli com *parâmetro* Θ, em que é atribuída a distribuição $Q(\theta)$ a priori para o parâmetro. Nesse caso, o sentido operacional para a distribuição a priori é representado pelo que se acredita esperar da freqüência relativa para um número grande de observações.

Ao realizarmos uma interpretação mais detalhada de cada parte do teorema, temos da parte (I), que θ executa apenas o papel de variável de integração (que usualmente é chamado de "parâmetro"). O Teorema da representação está, portanto, nos mostrando como surge o parâmetro, não arbitrariamente, mas sim como conseqüência matemática da suposição de permutabilidade (que está vinculada apenas a quantidades observáveis).

O resultado (II) é a lei dos grandes números para *v.a.* binárias em sua formulação geral, no sentido de valer para toda a função distribuição $Q(\theta)$. Ela garante a estabilidade da seqüência de freqüências relativas (com probabilidade 1) sem, no entanto, indicar o limite - que é fixo quando $Q(\theta)$ é degenerada, i.e., $X_1, X_2, \ldots$ são v.a.'s *i.i.d.*.

E por fim, da parte (III) tem-se que a incerteza a respeito do limite Θ, é dada justamente pela função distribuição $Q(\theta)$. Como foi visto, tal limite é interpretado como *limite da freqüência relativa de sucessos.*

1.4.1 Estudo do Teorema via Simulação Computacional

Para uma melhor ilustração dos aspectos que envolvem o Teorema da Representação, estudamos um caso simulado computacionalmente. Geramos seqüências de variáveis aleatórias binárias (0-1) utilizando o seguinte método (seqüencial):

1. Estabelece-se uma distribuição para a variável aleatória Θ;
2. Gera-se uma observação segundo uma distribuição Bernoulli(E(Θ)), em que E(Θ) é a esperança da *v.a.* Θ;
3. Dado o valor da nova observação, atualiza-se a distribuição a posteriori de Θ;
4. Para gerar a próxima observação da seqüência, volta-se ao passo 2.

Sendo n o tamanho da seqüência e N o número de seqüências geradas, obtiveram-se os seguintes resultados quando utilizada uma distribuição a priori Beta (10,10) para Θ:

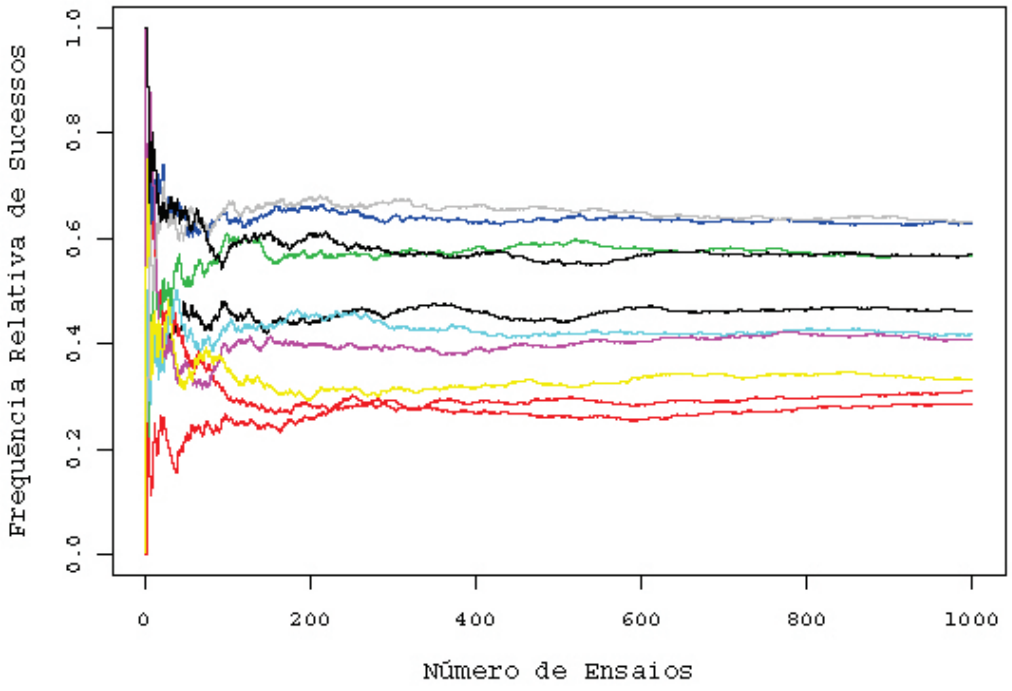

Figura 1.1: Resultados de simulações para N=10, n=1000 e distribuição Beta(10,10) para Θ

Observa-se nesse caso que a freqüência de sucessos converge para um limite conforme se aumenta o número de ensaios (n); entretanto, observa-se também que esse limite é aleatório - fato que confirma o que foi apontado na parte (II) do teorema. Para estudar a distribuição desse limite, fixou-se um valor de n razoável para a convergência da freqüência relativa de sucessos e gerou-se um número elevado de seqüências (N=10000), obtendo a distribuição que pode ser observada na figura 1.2.

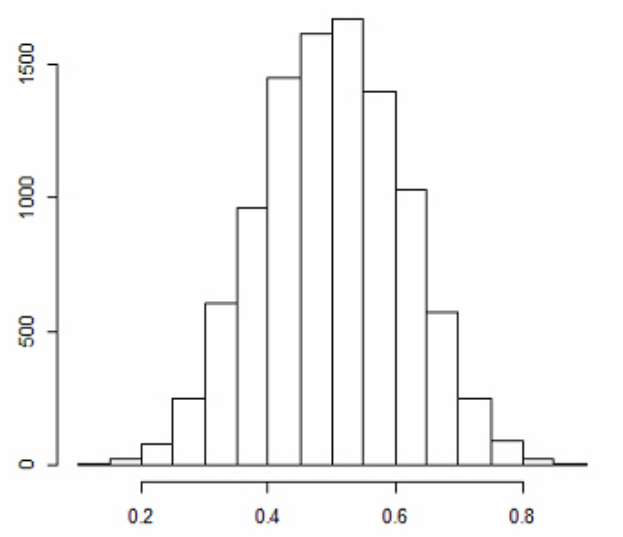

Figura 1.2: Distribuição das freqüências relativas (N=10000 e n=400)

Nota-se então que a distribuição das freqüências relativas de sucessos, obtida pela simulação, assemelha-se de forma significativa à da distribuição Beta(10,10). Fato que era esperado em função de ter sido a Beta(10,10) atribuída a priori para Θ, e também pelo que enuncia o próprio Teorema da Representação (parte III).

1.5 Preditivismo

Considerando problemas de interesse estatístico que se definem em termos de uma seqüência de observações $x_1, x_2, \ldots, x_n$, uma abordagem possível de inferência estatística é aquela cujo único objeto de manipulação é a distribuição conjunta $p(x_1, x_2, \ldots, x_n)$.

Um exemplo de problema de interesse refere-se à obtenção da probabilidade preditiva de uma próxima observação dado um conjunto de dados observados. Tem-se então um processo de aprendizado sobre o que foi observado:

$$p(x_{n+1}|x_1, \ldots, x_n) = p(x_1, \ldots, x_{n+1})/p(x_1, \ldots, x_n)$$

Dessa forma, supõe-se nesse tipo de abordagem que a especificação de graus de credibilidade se faz apenas sobre a forma da distribuição conjunta para todo subconjunto de $x_1, x_2, \ldots$. A essa especificação dá-se o nome de *modelo de probabilidade preditiva* (Bernardo & Smith [10]).

É neste contexto de inferência estatística que se apresenta a grande importância do Teorema da Representação de De Finetti. Como foi visto, a única suposição presente nele refere-se à de permutabilidade infinita do processo 0-1, e como conseqüencia matemática tem-se equivalência com o modelo paramétrico de distribuições Bernoulli(Θ) condicionalmente independentes. Nesse caso, o problema de predição se apresentar-se-ia assim:

$$p(x_{n+1}|x_1, \ldots, x_n) = \int_0^1 \theta \cdot f(\theta|x_1, \ldots, x_n) d\theta$$

A força do Teorema da Representação apresenta-se, portanto, no fato de que parte-se de uma suposição de simetria que restringe-se apenas aos valores observáveis. A construção paramétrica surge por uma equivalência matemática, e não por razões obscuras. Além disso, a utilização do modelo paramétrico pode dar-se apenas por uma questão de facilidade operacional e não necessariamente crença em entidades chamadas parâmetros (Wechsler [11]).

Essa interpretação do Teorema da Representação de De Finetti é a essência de abordagem de inferência estatística conhecida como *preditivismo* ou *abordagem bayesiana operacional* (De Finetti [4], Mendel [12]). Nesse tipo de abordagem, as suposições restrigem-se somente às quantidades observáveis. Em geral, trata-se de suposições de simetria diretamente sobre as observações ou também sobre medidas resumo sobre elas - os chamados *parâmetros operacionais*.

1.5.1 Estendibilidade

Um ponto importante de estudo relacionado à visão preditivista de inferência refere-se à propriedade de *estendibilidade*. Em muitos casos, o problema caracteriza-se por não haver a possibilidade de observar infinitas realizações. Dessa forma, a suposição de processo infinitamente permutável não é imediatamente razoável, e com isso pode não se ter a validade de importantes resultados inferenciais como o Teorema da Representação.

Uma alternativa para lidar com esse tipo de situação é supor que a seqüência finita de quantidades observáveis advém de um processo infinito, ou seja, pode ser *estendida* para uma população infinita. De forma menos direta, é possível também especificar modelos de população finita que, sob certas condições, são equivalentes a especificações em população infinita (Iglesias et al. [13]). Por intermédio desse artifício é possível facilitar o problema inferencial.

Entretanto, nem sempre é possível estabelecer condições de estendibilidade. No caso em que a seqüência é formada por duas variáveis aleatórias binárias (X_1, X_2), poderiam ser atribuídas, por exemplo, as seguintes probabilidades:

$$P(X_1 = 0, X_2 = 0) = 0 \qquad P(X_1 = 1, X_2 = 1) = 0$$
$$P(X_1 = 1, X_2 = 0) = P(X_1 = 0, X_2 = 1) = 1/2$$

Neste caso verifica-se que não existe nenhuma priori $Q(\theta)$ no modelo de Bernoullis(Θ) *c.i.i.d.* para a qual tem-se equivalência com as probabilidades preditivas descritas acima. Sendo assim, esse caso é uma exemplificação de um modelo não *estendível* (as contas para chegar a essa conclusão são apresentadas no Apêndice A.1).

Um outro exemplo possível refere-se ao caso com seguinte modelo de probabilidade preditiva:

$$P(X_1 = 0, X_2 = 0) = 1/4 \qquad P(X_1 = 1, X_2 = 1) = 1/4$$
$$P(X_1 = 1, X_2 = 0) = P(X_1 = 0, X_2 = 1) = 1/4$$

Neste caso verifica-se equivalência dessas probabilidades com aquelas resultantes do modelo de distribuições independentes Bernoulli(1/2). Logo, podemos declarar tal modelo como *estendível* (para um processo infinitamente permutável em que a distribuição para Θ é degenerada em 1/2).

Capítulo 2

Falácia do Apostador: Uma Abordagem Bayesiana

2.1 A Falácia do Apostador

Há uma série de estudos de psicologia experimental [14] que apontam vieses sistemáticos no comportamento humano diante de situações que envolvem incertezas. Nesse contexto apresenta-se um tipo de crença que é vastamente conhecido como a *falácia do apostador* - ao se lançar uma moeda e observar uma seqüência longa de *caras*, acredita-se que é mais provável observar *coroa* no próximo lançamento ao invés de outra *cara*. Trata-se de uma espécie de lei da compensação e que, em âmbito mais geral, apresenta-se quando as pessoas predizem que uma tendência recente terá de se reverter para o equilíbrio geral prevalecer. Pode também ser chamada como *lei da maturidade*.

Nessa linha, um aspecto relevante é abordado por O'Neill & Puza [15] seguindo as idéias apresentadas por Cowan [16] - para caracterizar-se a falácia é necessário encontrar um conjunto de condições que implicam em uma conclusão lógica contrária à afirmação potencialmente falaciosa. Diante disso, há então a necessidade de analisar de forma lógica quais são as premissas que levam à conclusão de que a crença na *lei da maturidade* é um comportamento inadmissível.

Nessa perspectiva, a abordagem estatística bayesiana apresenta-se como ferramenta adeqüada para tratar o problema, determinando as conseqüências lógicas que se têm para cada conjunto de premissas (suposições). Dois estudos da literatura estatística que seguem essa proposta são: O'Neill & Puza [15] e Rodrigues e Wechsler [17]. Nas seções seguintes apresentam-se suposições, definições, resultados e a forma como cada um desses estudos se relaciona com a *falácia do apostador*.

2.2 Estudos Anteriores

2.2.1 O Estudo de O'Neill & Puza

O'Neill & Puza [15] têm como proposta principal estudar os aspectos que fornecem um suporte lógico à *falácia do apostador*. Para tanto, propõe-se modelar uma situação em que se observa uma subseqüência de variáveis aleatórias discretas ($\mathtt{x}_n = (x_1, x_2, \ldots, x_n)$) que assumem valores em 1,2,...,k (faces de um dado, por exemplo). Além disso, supõe-se que a seqüência $\mathtt{x} \equiv (x_1, x_2, \ldots)$ é infinitamente permutável.

Nessas condições, os autores definem a *gambler's belief* como entender que a probabilidade preditiva $p(x_{n+1}|\mathtt{x}_n)$ de cada característica deve ser ordenada segundo relação inversa do número de ocorrências observadas das mesmas; ou seja, a característica que mais ocorreu até então teria a menor probabilidade preditiva, a segunda mais freqüente seria a com segunda menor probabilidade preditiva, e assim sucessivamente.

Dessa forma, um indivíduo que segue essa crença (*gambler's belief*) acredita que eventos que já ocorreram muitas vezes no passado tendem a ocorrer menos no futuro, e no caso de eventos que ocorreram poucas vezes, ele acredita que ocorrerão com maior probabilidade no futuro.

Entretanto, o principal resultado de O'Neill & Puza [15] apresenta-se no sentido contrário da *gambler's belief*. Para introduzir tal resultado é preciso, no entanto, elucidar alguns aspectos do modelo.

Uma vez que a principal suposição do modelo é a de que o processo seja infinitamente permutável, tem-se como conseqüência do Teorema de De Finetti, um vetor de *v.a.*'s $\mathbf{\Theta} = (\Theta_1, \ldots, \Theta_k)$ referente ao limite da freqüência relativa de ocorrências de cada uma das k possíveis características.

Além disso, supõe-se nesse modelo que a distribuição a priori de $\mathbf{\Theta}$ é permutável. Trata-se de uma suposição adicional de simetria, que demonstra que a priori existe indiferença na comparação entre as faces do dado.

Sendo assim, o modelo desse estudo é constituído por duas suposições de permutabilidade - a primeira sobre o processo $\mathtt{x} \equiv (x_1, x_2, \ldots)$ e a segunda sobre o vetor $\mathbf{\Theta}$. O principal resultado de O'Neill & Puza [15] declara que, sob essas condições, a probabilidade preditiva $p(x_{n+1}|\mathtt{x}_n)$ é ordenada segundo o número de ocorrências de cada característica já observada; ou seja, quanto mais freqüente nas observações mais se acredita que certa característica possa surgir numa próxima observação. A esse resultado é atribuído o título de *reverse gambler's belief*. A seguir, apresenta-se a descrição formal de tal resultado:

Resultado 2.2.1 (Reverse Gambler's Belief) *Sejam n_a e n_b as freqüências observadas das faces a e b, respectivamente; Dadas as suposições anteriormente descritas ($\mathbf{x}$ infinitamente permutável e priori permutável para $\mathbf{\Theta}$), para quaisquer a, b em $\{1, \ldots, k\}$, tem-se que:*

1. *se $n_a \geq n_b$ então $p(a|\mathbf{x}_n) \geq p(b|\mathbf{x}_n)$; e*
2. *se $n_a > n_b$ e $\mathbf{\Theta}$ não-degenerado então $p(a|\mathbf{x}_n) > p(b|\mathbf{x}_n)$.*

É possível ver então que, sob as condições descritas, a conclusão lógica é **contrária** à *gambler's belief*, visto portanto, como uma falácia (supostas as condições de permutabilidade).

2.2.2 O Estudo de Rodrigues & Wechsler

O modelo apresentado em Rodrigues & Wechsler [17] descreve a situação em que se observa uma subseqüência de variáveis aleatórias discretas ($\mathtt{x}_n = (x_1, x_2, \ldots, x_n)$) que assumem valores 0 ou 1 (fracasso ou sucesso), e nele também se supõe que a seqüência $\mathtt{x} = (x_1, x_2, ...)$ é infinitamente permutável. Entende-se que a *lei da maturidade* apresenta-se nesse modelo quando se acredita que, com a observação de uma seqüência crescente de fracassos consecutivos, a probabilidade preditiva de sucesso em uma próxima observação torna-se cada vez maior (o sucesso vai "amadurecendo").

O principal objeto de manipulação deste estudo refere-se à taxa de falha preditiva, que é definida da seguinte forma:

Definição 2.2.1 *Seja M a variável aleatória referente ao ensaio em que se observa o primeiro sucesso, então a Taxa de Falha Preditiva é dada por:*

$$r(m) = Pr(M = m) \ / \ Pr(M \geq m), \ \textit{para } m = 1, 2, \ldots$$

O resultado que é obtido apresenta-se formalmente da seguinte maneira:

Resultado 2.2.2 (Taxa de Falha Decrescente) *Para qualquer escolha de priori não degenerada para Θ, tem-se que $r(m)$ é **decrescente** em m.*

Prova:

Primeiramente, nota-se que:

$$r(m) = \left[\int_0^1 \theta \cdot (1-\theta)^{m-1} p(\theta)\ d\theta\right] / \left[\int_0^1 (1-\theta)^{m-1} p(\theta)\ d\theta\right].$$

Sendo assim, é preciso demonstrar que $E[\Theta \cdot (1-\Theta)^{m-1}]/E[(1-\Theta)^{m-1}]$ decresce em $m = 1, 2, \ldots$ para qualquer *v.a.* Θ satisfazendo $Pr[0 < \Theta < 1] = 1$.

Define-se então $Z \equiv 1 - \Theta$. E para todo $m = 1, 2, \ldots$, buscamos provar que:

$$E[(1-Z) \cdot Z^m]/E[Z^m] \geq E[(1-Z) \cdot Z^{m+1}]/E[Z^{m+1}]$$

Mas, definindo $T \equiv Z^{m/2}$ e $V \equiv Z^{(m/2)+1}$, podemos aplicar a desigualdade de Cauchy-Schwarz para obter (para todo $m \geq 1$):

$$[E(Z^{m+1})]^2 \leq E[Z^m] \cdot E[Z^{m+2}],$$

que é equivalente à desigualdade anterior, concluindo assim a prova. □

Uma outra relação interessante, que é simples de se verificar, é a seguinte:

$$r(m) = E(\theta \mid (m-1)\ fracassos) = p(x_m = 1 | x_1 = 0, \ldots, x_{m-1} = 0)$$

Disto, tem-se que $r(m)$ é equivalente à probabilidade preditiva de sucesso após observar-se (m-1) fracassos consecutivos.

Dessa forma, uma vez que o principal resultado do estudo declara que $r(m)$ é decrescente para qualquer escolha de priori para Θ, sua interpretação é que conforme se aumenta o número de fracassos consecutivamente observados, então se diminui a probabilidade preditiva de sucesso numa próxima observação, comportamento esse que é completamente oposto à *lei da maturidade*, e que pode ser definido como a *crença na maturidade reversa*.

2.2.3 Uma Possível Demonstração do Resultado de O'Neill & Puza

Apesar da existência de algumas diferenças entre os dois estudos mencionados anteriormente, mostraremos que é possível utilizar o resultado de Rodrigues & Wechsler para obter uma demonstração simples e intuitiva do resultado de O'Neill & Puza. Para tanto, precisamos fazer uso da seguinte propriedade:

Propriedade 2.2.1 *Nas condições descritas no modelo de O'Neill e Puza, se em n ensaios, for obtido $n_a = n_b$, então Θ_a e Θ_b são permutáveis a posteriori.*

Essa propriedade pode ser tirada do fato de que a posteriori é proporcional ao produto da verossimilhança e priori, e uma vez que a troca entre θ_a e θ_b não altera nem a verossimilhança e nem a priori, então são permutáveis a posteriori.

Tendo conhecimento de tal propriedade e do resultado de Rodrigues & Wechsler, já é possível partir para a demonstração de interesse, de que se tivermos permutabilidade do processo (infinito) $\mathtt{x}$ e permutabilidade em $\mathbf{\Theta}$, então teremos *reverse gambler's belief* (O'Neill & Puza).

A idéia da prova de tal resultado é utilizar o fato de que x_1, x_2,..., x_n são permutáveis, e sendo assim o processo de atualização da priori de $\mathbf{\Theta}$ permanece a mesmo para o vetor $(x_1, x_2,\ldots, x_n)$ rearranjado de forma conveniente. Rearranja-se então tal vetor de forma que $n_a - n_b$ observações de característica a fiquem em seu final.

Particionamos então o novo vetor em dois subvetores: o segundo com as $n_a - n_b$ observações mencionadas acima, e o primeiro com as observações restantes. Tem-se então que a atualização da priori é equivalente ao processo seqüencial de atualizá-la utilizando o primeiro vetor e então reatualizá-la utilizando o segundo vetor.

Considerando tal processo, temos que ao final da primeira etapa a priori atualizada se encaixa nas condições da *Propriedade 2.2.1* e sendo assim, ao final dessa etapa as probabilidades preditivas $p(a)$ e $p(b)$ serão iguais.

Como na segunda etapa têm-se $n_a - n_b$ observações consecutivas de a, utilizando o resultado da *crença na maturidade reversa* é possível concluir que a reatualização da priori resulta num decrescimento de $p(b)$ e crescimento de $p(a)$, implicando que $p(a) > p(b)$.

O procedimento sugerido acima é válido para quaisquer a e b tais que $n_a > n_b$. Quando $n_a = n_b$ implica que $p(a) = p(b)$ (pela propriedade enunciada), e por fim, quando $\mathbf{\Theta}$ degenerado (e permutável), então $p(a) = p(b) = 1/k$ (para quaisquer n_a e n_b). Assim mostramos que o resultado da *reverse gambler's belief* pode ser demonstrado com utilização imediata da *crença na maturidade reversa* (Rodrigues & Wechsler).

2.2.4 Considerações sobre os Estudos

A seguir apresenta-se uma tabela que contém os aspectos gerais dos dois estudos apresentados nas seções anteriores.

Tabela 2.2.1. Aspectos Gerais dos Estudos

	O'Neill & Puza	***Rodrigues & Wechsler***
Tipo de Variável	$1, 2, \ldots, k$	0 ou 1
Suposição - Observações	Estendibilidade	Estendibilidade
Suposição - Priori para Θ	Permutável	Qualquer
Resultado	*reverse gambler's belief*	*crença na maturidade reversa*

Os dois trabalhos anteriormente mencionados são semelhantes em alguns aspectos de suas suposições e distintos em outros. Entretanto, ambos têm dois pontos importantes em comum. Primeiro, os dois estudam modelos em que a principal suposição é a de que o processo é infinitamente permutável. Sendo assim, são compatíveis com a visão preditivista de inferência estatística.

O segundo ponto refere-se ao fato de que ambos indicam que dadas suas respectivas suposições e seguindo a abordagem bayesiana, então o comportamento coerente seria contrário ao encontrado na *falácia do apostador*. Mesmo assim, não é possível afirmar de modo **geral** que o apostador tem um comportamento equivocado. Na verdade, isso depende do fato de ele ver ou não razoabilidade nas suposições dos modelos em questão.

2.3 Um Novo Modelo

Como a *maturidade* é um comportamento comprovadamente integrante da natureza humana, entendemos que há indicações de que as suposições utilizadas nos estudos anteriores podem nem sempre serem condizentes com a percepção humana.

A principal suposição que está presente nos dois estudos é a de permutabilidade infinita do processo do qual se obtém as observações. Trata-se de uma suposição forte e freqüentemente feita em estudos bayesianos. Porém, no sentido da percepção humana é uma premissa muitas vezes questionável.

Com o intuito de variar a suposição de permutabilidade infinita, propomos então um novo modelo para a situação estudada de predição em uma seqüência de variáveis aleatórias binárias. A idéia central desse modelo é supor que o que se observa (0 ou 1 - fracasso ou

sucesso) advém de uma população finita permutável não necessariamente estendível. Trata-se então do modelo usual de população finita estudado sob o ponto de vista preditivista (Iglesias et al. [13]).

A seguir apresentamos uma descrição formal dos primeiros aspectos do modelo:

DEFINIÇÕES INICIAIS: Seja $\mathtt{X}_N = (X_1, X_2, \ldots, X_N)$ um vetor de valores desconhecidos e permutáveis do tipo 0 ou 1 ("população"). Definimos a função da população $\gamma\ (X)$ como a contagem de sucessos em $\mathtt{X}_N$, isto é, uma quantidade com significado essencialmente operacional (*parâmetro operacional*). Uma amostra de tamanho n refere-se a um subvetor da população que se torna disponível. Por simplicidade, uma vez que se tem permutabilidade em $\mathtt{X}_N$, podemos considerar a amostra como o subvetor $\mathtt{x}_n = (x_1, x_2, \ldots, x_n)$. Considere n_1 e n_0 as contagens de sucesso e fracasso em $\mathtt{x}_n$, respectivamente.

A suposição base do modelo é que a população é permutável. Dessa forma, a propriedade de permutabilidade que se apresenta nos estudos anteriores é mantida. No entanto, no atual modelo permutabilidade pode **não** ser estendível para um processo infinito. Portanto, as suposições desse modelo são mais fracas do que as dos anteriores.

É possível notar que não há suposição a respeito de processo amostral algum. Entretanto, para clarificar o modelo, pode-se interpretá-lo como a situação típica de extração de bolas de uma urna. A idéia é que existe uma urna com N bolas, γ_0 pretas ("1"s) e $N - \gamma_0$ brancas ("0"s), e n bolas são retiradas da urna sem reposição. Mesmo que pensemos neste exemplo de bolas em urnas, $\sum_{i=1}^{n} X_i|\gamma = \gamma_0$ será Hipergeométrica (N, γ_0, n) em virtude da permutabilidade de $\mathtt{X}_n$, não havendo necessariamente sorteio ou plano amostral algum envolvido.

A partir desta equivalência, a distribuição de probabilidade de $\sum_{i=1}^{n} X_i|\gamma = \gamma_0$ é:

$$P(\sum_{i=1}^{n} X_i = s|\gamma = \gamma_0) = \frac{\binom{\gamma_0}{s}\binom{N-\gamma_0}{n-s}}{\binom{N}{n}}$$

Como as retiradas são feitas sem reposição, nós sabemos que qualquer par de ensaios tem associação negativa. Esse fato parece apresentar-se no sentido favorável ao da *lei da maturidade*. No entanto, a abordagem bayesiana incorpora o aprendizado sobre γ, e às vezes direciona a predição no sentido inverso ao da *maturidade* como apontado nos estudos anteriormente mencionados.

Dessa forma, têm-se a princípio, diferentes possibilidades de resultados esperados no novo modelo. Pode-se tornar complexo então estabelecer resultados bem comportados que possam ser interpretados como favoráveis ou não à *maturidade* (um resultado "bem com-

portado" seria, por exemplo, taxa de falha preditiva decrescente como em Rodrigues & Wechsler [17]). Inclusive, exemplos de resultados não comportados que podem surgir no novo modelo referem-se a situações em que a taxa de falha preditiva não é monótona (se alterna entre momentos de crescimento e decrescimento).

Sendo assim, o caminho seguido neste trabalho é de estudar classes de distribuições a priori para γ que impliquem em resultados interpretáveis no sentido favorável ou reverso da *maturidade*.

Outro ponto a ser estudado refere-se à estendibilidade nos modelos considerados, isto é, a possibilidade de a medida de probabilidade do processo envolvido ser equivalente a de uma seqüência finita advinda de um processo infinitamente permutável.

A base para a investigação sobre estendibilidade é o conhecimento adquirido pelos estudos anteriores, de que *maturidade reversa* é uma propriedade implicada por todos os processos infinitamente permutáveis. Dessa forma, se conseguirmos traçar equivalência com algum processo infinitamente permutável, saberemos que há *maturidade reversa* em tal situação. Da mesma forma, se mostrarmos que há *maturidade*, podemos descartar equivalência com algum processo infinitamente permutável e, portanto, declarar não-estendibilidade em tal situação.

2.3.1 Distribuição Binomial como Priori

Um primeiro caso de distribuição a priori para γ com resultado interessante de ser estudado é o da distribuição Binomial:

Resultado 2.3.1 (Indiferença) *Se a distribuição a priori escolhida para γ for Binomial(N, π), então:*

$$p(x_{n+1} = 1|\boldsymbol{x}_n) = \pi, \textit{ para quaisquer } \boldsymbol{x}_n \in \{0,1\}^n \textit{ e } n \leq N$$

Prova. Os detalhes dessa prova são apresentados no Apêndice A.2. A demonstração desse resultado é feita marginalizando-se a probabilidade $p(x_1, x_2, \ldots, x_n|\gamma)$ com respeito a γ e após algumas operações algébricas chega-se à expressão $p(x_1, x_2, \ldots, x_n) = \pi^{\Sigma x_i} \cdot (1-\pi)^{n-\Sigma x_i}$, a mesma expressão do modelo independente com probabilidade π de sucesso. □

Um ponto que chama a atenção é a probabilidade preditiva de sucesso não depender das observações. Tal comportamento chamamos de *indiferença*, pois não segue nem na direção da maturidade e nem mesmo em sua direção reversa.

Em um primeiro momento, esse resultado pode ser tido como inesperado. Entretanto, a idéia de que a priori binomial resulta na *indiferença* torna-se mais razoável de ser entendida

quando traçamos uma relação de equivalência com uma situação experimental - atribuir priori binomial para γ é equivalente a dizer que cada bola da urna tem probabilidade π de ser preta independentemente das demais, e dessa forma fica razoável visualizar que a probabilidade preditiva de a próxima bola a ser observada ser preta é π independentemente do que foi observado anteriormente.

2.3.2 Mistura de Binomiais como Priori

Uma outra distribuição interessante é aquela que pode ser representada como mistura de Binomias. Se essa for utilizada como priori para γ, obtemos resultados comparáveis à *falácia do apostador*. Esses resultados são obtidos relacionando o novo modelo (com a priori descrita) àquele de permutabilidade infinita abordado nos estudos anteriores. Trata-se, portanto, de uma forma de explorar a propriedade de estendibilidade do modelo para obtenção de resultados imediatos de estudos anteriores.

Teorema 2.3.1 (Equivalência de Modelos) *Seja $Q(\pi)$ uma função distribuição em $[0,1]$; se a distribuição a priori escolhida para γ puder ser expressa como*

$$p(\gamma) = \int_0^1 \binom{N}{\gamma} \cdot \pi^{\gamma} \cdot (1-\pi)^{N-\gamma} \; dQ(\pi),$$

então:

$$p(x_1, x_2, \ldots, x_n) = \int_0^1 \pi^{\Sigma x_i} \cdot (1-\pi)^{n-\Sigma x_i} \; dQ(\pi), \; \forall \; n \leq N$$

Prova. Detalhes dessa demonstração estão no Apêndice A.3. Para provar este resultado, vê-se que $p(x_1, x_2, \ldots, x_n) = \sum_{\gamma} p(x_1, x_2, \ldots, x_n|\gamma)p(\gamma)$. Utilizando o Teorema de Fubini é possível trocar a ordem da somatória com a integral na equação e após alguns passos algébricos chega-se ao lado direito da igualdade do teorema. □

A probabilidade preditiva $p(x_{n+1} = 1|\mathtt{x}_n)$ no novo modelo, que é nosso objeto de interesse, é obtida diretamente pela probabilidade conjunta das observações (Teorema 2.3.1). É possível ver que essa probabilidade é a mesma que aquela do modelo de permutabilidade infinita, como provado pelo Teorema de De Finetti. Dessa forma, essa classe de prioris tem as mesmas implicações dos modelos de estudos anteriores (de permutabilidade infinita).

Este resultado pode ser interpretado observando que o fator de mistura $Q(\pi)$ pode ser visto como a priori $Q(\theta)$ no modelo de permutabilidade infinita. Sendo assim, se $Q(\pi)$ é não

degenerada, teremos *crença na maturidade reversa* (Rodrigues & Wechsler [17]). Se $Q(\pi)$ for simétrica em torno de 1/2, então vale a *reverse gambler's belief* (O'Neill & Puza [15]).

Um exemplo de distribuição deste tipo é dada pelo modelo Beta-Binomial. Trata-se do modelo obtido quando $Q(\pi)$ corresponde à distribuição da família Beta. Tal modelo é amplamente utilizado e tem propriedades interessantes como solução analítica simples para a distribuição a posteriori.

2.3.3 Distribuições Mais Estreitas que a Binomial

Nas seções anteriores estudamos distribuições a priori para γ que levam a *maturidade reversa* ou *indiferença*. Trabalhos anteriores também provaram que em outros modelos a *falácia do apostador* é incoerente. Guiado por esses resultados pode-se questionar se todas as distribuições a priori para γ têm essa propriedade. Nesta seção, apresentamos uma classe de prioris para γ para as quais a crença na *maturidade* é coerente.

Para chegar a essa classe, começamos estudando casos que podem servir de referência. Primeiro, tomamos como base a distribuição Uniforme, que é caso particular da Beta-Binomial - logo, ela implica em *maturidade reversa*. Um segundo caso a ser observado refere-se à distribuição Binomial(N, 1/2), que resulta em *indiferença*. E por fim, tem-se a distribuição degenerada em $N/2$, que é claramente um caso extremo de *maturidade*. A figura 2.1 ilustra essa linha de comparação.

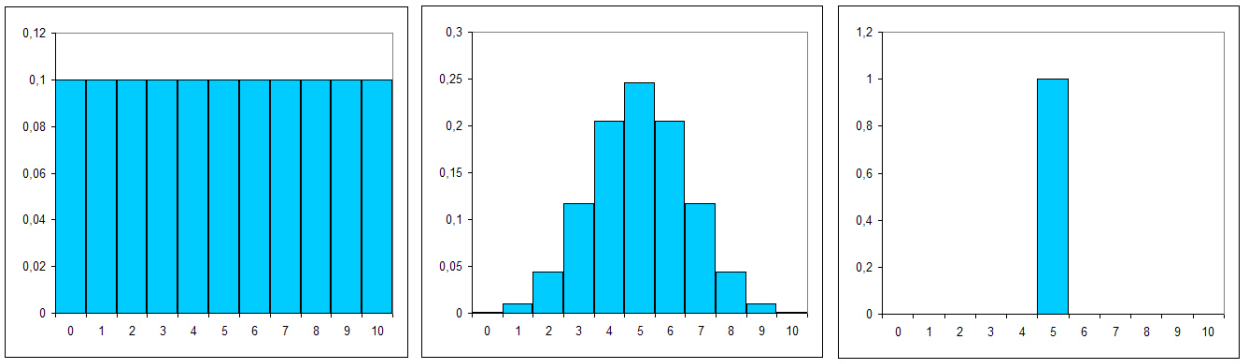

Figura 2.1: Casos de distribuição a priori com N=10: 1 - Uniforme, 2 - Binomial e 3 - Degenerada

Com base nessa análise, é possível suspeitar sobre a existência de uma classe de distribuições "entre" os dois últimos casos que compartilham a propriedade de *maturidade*. Para examinar esse ponto, definimos uma classe de distribuições cuja forma em certo sen-

tido é mais *afunilada* que a da Binomial. A seguir, apresenta-se a descrição formal das distribuições pertencentes a essa classe:

Definição 2.3.1 *Uma v.a. Y discreta tem distribuição de probabilidade* ***mais estreita que a Binomial(N, 1/2)*** *se satisfaz as seguintes condições:*

1. *Seu suporte é o mesmo que o da Binomal(N,1/2);*
2. *Sua distribuição também é simétrica em torno de* $N/2$*;*
3. *P(Y=y) / P(Y=y−1) > (N−y+1) / y ,* $\quad y = 1, \ldots, \lfloor N/2 \rfloor$

Como pode ser visto, a essência dessa definição está na restrição de que cada valor da razão P(Y=y) / P(Y=y−1) tem que ser maior do que a respectiva razão no caso de distribuição Binomial para os valores anteriores à $N/2$ (e pela simetria, menor para os valores posteriores). Dessa forma,o gráfico de distribuições desse tipo tem crescimento mais acentuado do que a Binomial até $N/2$, e decrescimento mais acentuado para os valores da segunda metade.

A figura 2.2 apresenta dois casos de distribuições mais estreitas que a Binomial em comparação com a própria Binomial.

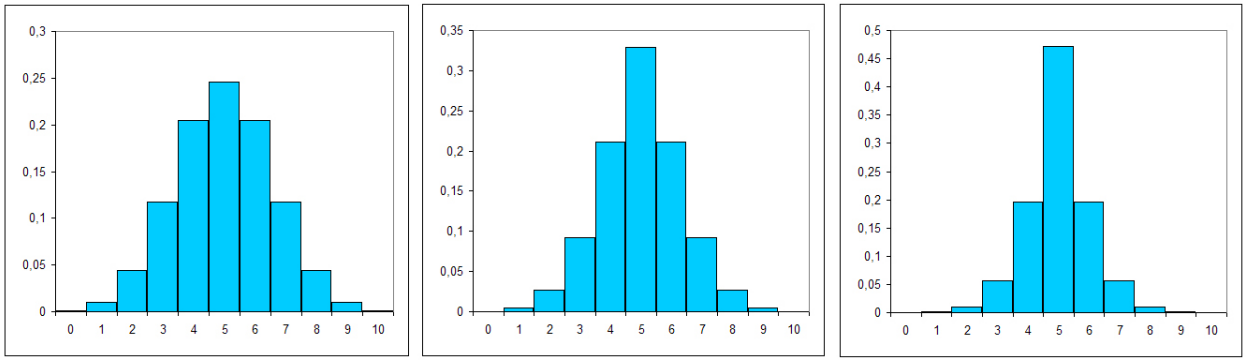

Figura 2.2: Distribuições de Probabilidade: 1- Binomial, 2- Mais Estreita (I) e 3- Mais Estreita(II)

Para realizar o estudo da propriedade de *maturidade*, faremos uso da função taxa de falha preditiva. Sua definição foi apresentada na seção 2.2.2 ($r(m) := \Pr(M{=}m)\ /\ \Pr(M \geq m)$ sendo M o primeiro ensaio em que ocorre sucesso). Essa função tem como valor a probabilidade preditiva de sucesso condicionada a uma sequência apenas de fracassos com $(m{-}1)$ observações.

Teorema 2.3.2 *Se para γ for escolhida uma distribuição priori $p(\gamma)$ mais estreita que a Binomial(N, 1/2), então:*

$$r(m) = Pr(M = m)/Pr(M \geq m) \ > \ 1/2 \qquad (m = 2, \ldots, N)$$

Prova. Alguns dos passos iniciais envolvem operações com identidades de coeficientes binomiais, que são apresentados no Apêndice A.4.

$$P(M = m) = \sum_{k=1}^{N-m+1} \binom{N-m}{k-1} p(k)/\binom{N}{k}, \quad \text{e chamamos } p(k)/\binom{N}{k} \text{ de } t(k)$$

$$r(m) = P(M = m)/[t(0) + \sum_{i=m}^{N} P(M = i)] = \frac{\displaystyle\sum_{k=1}^{N-m+1} \binom{N-m}{k-1} t(k)}{t(0) + \displaystyle\sum_{k=1}^{N-m+1} \sum_{i=m}^{N-k+1} \binom{N-i}{k-1} t(k)} =$$

$$= \frac{\displaystyle\sum_{k=1}^{N-m+1} \binom{N-m}{k-1} t(k)}{t(0) + \displaystyle\sum_{k=1}^{N-m+1} \binom{N-m+1}{k} t(k)} = \frac{\displaystyle\sum_{k=1}^{N-m+1} \binom{N-m}{k-1} t(k)}{\displaystyle\sum_{k=1}^{N-m+1} \binom{N-m}{k-1} (t(k) + t(k{-}1))} =$$

$$= \frac{\displaystyle\sum_{k=1}^{min(m-1,N-m+1)} \binom{N-m}{k-1} t(k) + \left[\sum_{k=m}^{N-m+1} \binom{N-m}{k-1} t(k) \right] \mathbf{I}_{(m \leq\ N/2+1)}}{\displaystyle\sum_{k=1}^{min(m-1,N-m+1)} \binom{N-m}{k-1} (t(k) + t(k{-}1)) + \left[\sum_{k=m}^{N-m+1} \binom{N-m}{k-1} (t(k) + t(k{-}1)) \right] \mathbf{I}_{(m \leq\ N/2+1)}}$$

A estratégia para continuidade da prova é mostrar que: (I) a primeira somatória do numerador sobre a primeira do denominador é maior do que 1/2; e (II) a segunda somatória do numerador sobre a segunda do denominador é maior do que 1/2. Além disso, nota-se que se $m > N/2 + 1$ somente existe a parte (I).

Provamos (I) notando que se k menor que $(N{+}1)/2$ então $t(k)/(t(k){+}t(k{-}1)) > 1/2$, o

que é demonstrado imediatamente do fato de $p(\gamma)$ ser mais estreita que a Binomial(N,1/2):

$$\frac{t(k)}{t(k)+t(k-1)} > 1/2 \Leftrightarrow \frac{1}{1+\frac{t(k+1)}{t(k)}} > 1/2 \Leftrightarrow \frac{t(k-1)}{t(k)} < 1 \Leftrightarrow \frac{p(k)}{p(k-1)} > \frac{(N-k+1)}{k}$$

Para provar (II):

$$\frac{\sum_{k=m}^{N-m+1} \binom{N-m}{k-1} t(k)}{\sum_{k=m}^{N-m+1} \binom{N-m}{k-1} (t(k)+t(k-1))} > 1/2 \ ,$$

$$\text{se} \quad \frac{\sum_{k=m}^{\lfloor N/2 \rfloor} [\binom{N-m}{k-1} t(k) + \binom{N-m}{N-k} t(N-k+1)]}{\sum_{k=m}^{\lfloor N/2 \rfloor} [\binom{N-m}{k-1} (t(k)+t(k-1)) + \binom{N-m}{N-k} (t(N-k+1)+t(N-k))]} > 1/2 \ ,$$

Notamos que $\binom{N-m}{k-1} = \binom{N-m}{N-m-k+1}$, e $\binom{N-m}{N-k-m+1} > \binom{N-m}{N-k}$ sempre que $m \leq N/2$ e $k < (N+1)/2$.

Além disso, $[t(k) + t(N-k+1)]/[t(k) + t(k-1) + t(N-k+1) + t(N-k)] = 1/2$ e $t(k)/(t(k)+t(k-1)) > 1/2$, e assim valem as desigualdades acima. ■

Com base no teorema acima, identifica-se que ao se escolher uma priori mais estreita que a Binomial(N,1/2) e observando uma sequência de consecutivos fracassos passa-se a acreditar mais na ocorrência de sucesso do que de fracasso no próximo ensaio. A seguir mostramos como esse resultado compara-se com aqueles obtidos por O'Neill & Puza [15].

Corolário 2.3.1 (Gambler's Belief) *: Se para γ for escolhida uma distribuição priori mais estreita que a Binomial(N, 1/2), então:*

$$\textit{se } n_1 < n_0 \textit{ então } p_{n+1}(1|\mathbf{x}_n) > p_{n+1}(0|\mathbf{x}_n)$$

Prova. A demonstração do resultado acima decorre de um artifício de manipulação na ordem das observações. A idéia é que devido à propriedade de permutabilidade entre as observações, o aprendizado sobre o vetor de observações mantém-se o mesmo para qualquer permutação de seus valores. Dessa forma, construímos os seguintes subvetores: o final ($x^{(2)}$) - com $(n_0 - n_1)$ fracassos; e o inicial ($x^{(1)}$) - com as observações restantes de $\mathbf{x}_n$ (com o mesmo número de sucessos e fracassos).

Fazemos então o processo de atualização da priori em duas etapas: na primeira, baseada no subvetor inicial. Como o número de sucessos é igual ao de fracassos temos que $P(\gamma|x^{(1)}) \propto P(\gamma)\binom{\gamma}{n_1}\binom{(N-\gamma)}{n_1}$. Verifica-se no Apêndice A.5 que a priori atualizada manterá a propriedade de ser mais estreita que a Binomial; na segunda etapa, baseada no subvetor final, como temos $(n_0 - n_1)$ fracassos consecutivos, pelo Teorema 2.3.2, temos que a probabilidade preditiva de sucesso ao final do processo é maior que 1/2, o que conclui a demonstração. □

A interpretação deste resultado pode ser feita da seguinte maneira: uma vez que especifica-se uma distribuição a priori para γ mais estreita que a Binomial, e tendo observado mais fracassos do que sucessos na amostra, acredita-se que o próxima observação tem maior probabilidade de ser sucesso do que fracasso. O mesmo padrão aplica-se no caso em que se observa mais sucessos (acredita-se mais em um próximo fracasso). Trata-se então de um resultado exatamente contrário ao obtido no estudo de O'Neill & Puza [15] (*reverse gambler's belief*).

Pelo Teorema 2.3.2., tem-se que para a classe de distribuições mais estreitas que a Binomial a taxa de falha preditiva $r(m)$ é sempre maior que 1/2 para $m = 2, \ldots, N$ (tem-se sempre $r(1) = 1/2$ pela simetria da priori). Sabe-se também que em todos os modelos estendíveis com distribuição não-degenerada para o limite da freqüência relativa de sucessos Θ temos a *crença na maturidade reversa* - ou seja, taxa de falha preditiva decrescente. A partir disso, é possível concluir que qualquer especificação de modelo gerada por uma distribuição a priori para γ na classe de prioris mais estreitas que a Binomial não é estendível.

2.3.4 Distribuição Hipergeométrica como Priori

Uma outra classe de distribuições a priori que compartilha da propriedade de implicação em *maturidade* é a Hipergeométrica. Apesar de se tratar de uma classe menos ampla que a anterior, não há nela a restrição de simetria.

Formalmente, tem-se que, se a distribuição a priori para γ puder ser expressa como:

$$p(\gamma) = \frac{\binom{T_1}{\gamma}\binom{T_0}{N-\gamma}}{\binom{T_0+T_1}{N}}, \quad T_0 + T_1 \geq N$$

então a taxa de falha preditiva $r(m)$ do modelo especificado é **crescente** em m.

Em outras palavras, tem-se que a partir de tal resultado conclui-se que quanto maior o número de fracassos consecutivamente observados, mais passa-se a acreditar na ocorrência de um sucesso no próximo ensaio, o que leva à idéia de que o sucesso está "amadurecendo". A este resultado dá-se o nome de *crença na maturidade*.

Para ilustrar tal resultado, realiza-se novamente uma comparação com a situação de retirada de bolas de uma urna. A distribuição considerada pode ser obtida retirando sem reposição N bolas de uma urna com T_0 bolas brancas e T_1 pretas e, em seguida, escolher aleatoriamente uma combinação de n destas bolas retiradas. Contudo, qualquer uma destas combinações de bolas tem mesma distribuição. Em particular, a combinação $(1,\ldots,n)$ tem essa distribuição, que é aquela obtida retirando exatamente n bolas da mesma urna. Portanto, a distribuição de Σx é uma Hipergeométrica, cuja taxa de falha é conhecidamente decrescente.

Outra maneira de ilustrar esta situação é imaginar que em um primeiro momento N bolas são retiradas de uma *super urna* e colocadas em uma urna menor. E no passo seguinte, n bolas são retiradas da urna menor. Nota-se que tal processo equivale a retirar diretamente as n bolas da *super urna*. Dessa forma, a retirada de consecutivas bolas brancas sempre levará a um aumento na probabilidade preditiva de retirar uma bola preta na próxima tentativa.

No que diz respeito à estendibilidade, sabe-se que para qualquer seqüência estendível a crença na *maturidade* é incoerente. Disto, pode se concluir que o modelo resultante da especificação Hipergeométrica a priori não é estendível.

Um fato interessante de se observar é que, caso $T_0 = T_1$, a priori será mais estreita que a Binomial(N,1/2) (conforme verificado no Apêndice A.6). Dessa forma, tem-se pelo Corolário 2.3.1 que valerá a *gambler's belief*. Tal fato não é surpreendente uma vez que a *crença na maturidade* (mostrada nesta seção) é um resultado mais forte que a *gambler's belief*. No entanto, tal fato leva à suspeita de que seja possível estabelecer uma subclasse dentro das

distribuições mais estreitas que a Binomial de modo que haja a implicação no resultado mais forte - a *crença na maturidade*.

2.3.5 Distribuições Mais Estreitas de Segunda Ordem que a Binomial

Por intermédio do Corolário 2.3.1 foi possível estabelecer resultados favoráveis à *maturidade* e comparáveis aos obtidos por O'Neill & Puza [15] (*reverse gambler's belief*). Nessa seção estudaremos prioris em uma nova classe de distribuições e mostraremos que vale a *crença na maturidade* - um resultado comparável ao obtido por Rodrigues & Wechsler [17].

Este resultado apresenta-se de uma forma mais aguda que o resultado obtido para as prioris mais estreitas que a Binomial (*gambler's belief*) uma vez que diz respeito ao comportamento monótono da probabilidade preditiva (crescente) com o passar das observações de fracassos. Ainda assim, o sentido de ambos os resultados é o mesmo - favorável à *maturidade*. Espera-se então que seja possível definir um subclasse dentro daquela das distribuições mais estreitas que a Binomial de forma que os referentes modelos especificados tenham a *crença na maturidade* como conseqüência.

Foi observado na seção anterior que a priori Hipergeométrica simétrica resulta em *crença na maturidade*, e além disso, ela verifica-se como caso de distribuição mais estreita que a Binomial. Tal fato sugere a existência de alguma característica na distribuição Hipergeométrica simétrica que pode ser utilizada para determinar uma subclasse dentro das distribuições mais estreitas que a Binomial de modo que nela seja válida a *crença na maturidade*.

Baseando-se em tal investigação, chegou-se à conclusão que a característica de interesse poderia estar associada à razão "de segunda ordem"da distribuição Binomial.

A seguir, apresenta-se formalmente a descrição de distribuições pertencentes a uma classe com essa característica:

Definição 2.3.2 *Uma v.a. Y discreta tem distribuição de probabilidade* ***mais estreita de segunda ordem que a Binomial(N, 1/2)*** *se satisfaz as seguintes condições:*

1. *Sua distribuição é mais estreita que a Binomal(N,1/2);*
2. $\dfrac{P(Y=y+1)/P(Y=y)}{P(Y=y)/P(Y=y-1)} < \dfrac{(N-y)/(y+1)}{(N-y+1)/y}\ , \qquad y = 1, \ldots, \lfloor N/2 \rfloor - 1$

Dessa forma, as distribuições dessa classes têm forma mais estreita que a Binomial e, além disso, apresentam uma restrição na variação de seu crescimento quando comparadas à Binomial. É possível verificar a presença desta característica em qualquer distribuição Hipergeométrica simétrica (as contas são apresentadas no Apêndice A.7).

Na figura 2.3, são apresentados dois exemplos de distribuições mais estreitas de segunda ordem que a Binomial em comparação com a própria Binomial (N, 1/2).

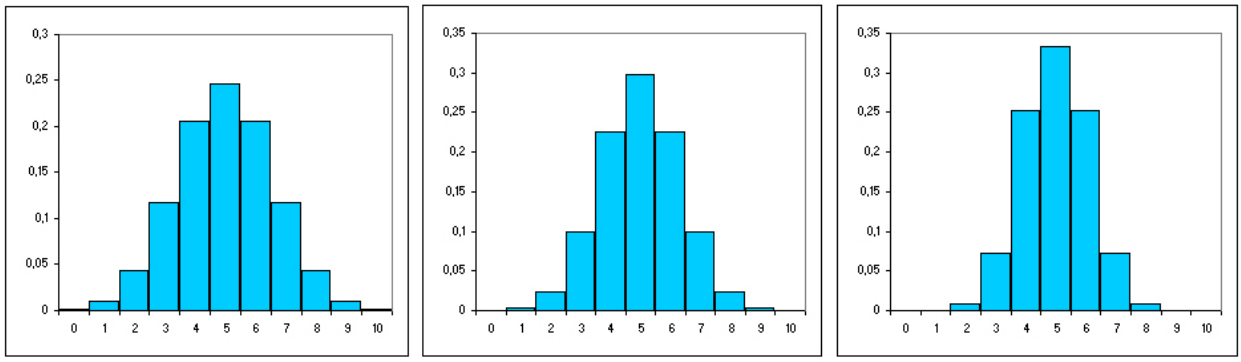

Figura 2.3: Distribuições de Probabilidade: 1- Binomial, 2- Mais Estreita de Segunda Ordem (I) e 3- Mais Estreita de Segunda Ordem (II)

Teorema 2.3.3 (Crença na Maturidade) *Se para γ for escolhida uma distribuição priori mais estreita de segunda ordem que a Binomial(N, 1/2), então:*

$$r(m) = Pr(M{=}m) \ / \ Pr(M \geq m) \text{ é } \textbf{\textit{crescente}} \text{ em } m.$$

Prova.

$$P(M=m) = \sum_{k=1}^{N-m+1} \binom{N-m}{k-1} p(k) / \binom{N}{k}, \text{ e novamente chamamos } p(k)/\binom{N}{k} \text{ de } t(k),$$

seguindo os mesmos passos iniciais da demonstração do teorema anterior, temos:

$$r(m) = \frac{\sum_{k=1}^{N-m+1} \binom{N-m}{k-1} t(k)}{\sum_{k=0}^{N-m+1} \binom{N-m+1}{k} t(k)} \quad \text{e} \quad r(m+1) = \frac{\sum_{k=1}^{N-m} \binom{N-m-1}{k-1} t(k)}{\sum_{k=0}^{N-m} \binom{N-m}{k} t(k)}$$

Podemos reescrever $r(m)$ como:

$$r(m) = \frac{\sum_{k=1}^{N-m+1} \binom{N-m}{k-1} t(k)}{\sum_{k=0}^{N-m+1} \binom{N-m+1}{k} t(k)} = \frac{\sum_{k=1}^{N-m} \binom{N-m-1}{k-1} t(k) + \sum_{k=2}^{N-m+1} \binom{N-m-1}{k-2} t(k)}{\sum_{k=0}^{N-m} \binom{N-m}{k} t(k) + \sum_{k=1}^{N-m+1} \binom{N-m}{k-1} t(k)}$$

Na última expressão obtida, nota-se que a primeira somatória do numerador juntamente com a primeira somatória do denominador são as mesmas que aquelas que aparecem na expressão de $r(m+1)$.

Portanto, para que $r(m+1) > r(m)$, tem-se a seguinte condição suficiente:

$$\frac{\sum_{k=1}^{N-m} \binom{N-m-1}{k-1} t(k)}{\sum_{k=0}^{N-m} \binom{N-m}{k} t(k)} > \frac{\sum_{k=2}^{N-m+1} \binom{N-m-1}{k-2} t(k)}{\sum_{k=1}^{N-m+1} \binom{N-m}{k-1} t(k)}$$

que é equivalente a:

$$\frac{\sum_{k=1}^{N-m} \binom{N-m-1}{k-1} t(k)}{\sum_{k=1}^{N-m} \binom{N-m-1}{k-1} (t(k)+t(k-1))} > \frac{\sum_{k=2}^{N-m+1} \binom{N-m-1}{k-2} t(k)}{\sum_{k=2}^{N-m+1} \binom{N-m-1}{k-2} (t(k)+t(k-1))}$$

Em particular, tal desigualdade será válida quando valer o seguinte:

$$\frac{t(y)}{t(y)+t(y-1)} > \frac{t(y+1)}{t(y+1)+t(y)}, \quad y = 1, \ldots, N-m$$

$$\Leftrightarrow \quad \frac{1}{1+t(y-1)/t(y)} > \frac{1}{1+t(y)/t(y+1)}, \quad y = 1, \ldots, N-m$$

$$\Leftrightarrow \quad \frac{t(y-1)}{t(y)} < \frac{t(y)}{t(y+1)}, \quad y = 1, \ldots, N-m$$

$$\Leftrightarrow \quad \frac{p(y+1)/p(y)}{p(y)/p(y-1)} < \frac{(N-y)/(y+1)}{(N-y+1)/y}, \quad y = 1, \ldots, N-m$$

Nota-se que tal desigualdade é a mesma que aquela vista na definição de distribuição mais estreita de segunda ordem que a Binomial para $y = 1, \ldots, \lfloor N/2 \rfloor - 1$.

Para $y = \lceil N/2 \rceil + 1, \ldots, N$; verifica-se o seguinte:

$$\frac{p(y+1)/p(y)}{p(y)/p(y-1)} = \frac{p(N-y-1)/p(N-y)}{p(N-y)/p(N-y+1)} = \frac{p(N-y+1)/p(N-y)}{p(N-y)/p(N-y-1)}$$

Como $N - y$ é um número pertencente a $\{1, \ldots, \lfloor N/2 \rfloor - 1\}$, então pela definição de distribuição mais estreita de segunda ordem que a Binomial verifica-se também a última desigualdade para esses valores de y.

Resta, então, analisar os casos $y = \lfloor N/2 \rfloor$ *e* $y = \lceil N/2 \rceil$ quando N for ímpar e o caso $N/2$ quando N for par. Primeiramente, para N ímpar verificamos que:

$$\frac{p(\lfloor N/2 \rfloor + 1)/p(\lfloor N/2 \rfloor)}{p(\lfloor N/2 \rfloor)/p(\lfloor N/2 \rfloor - 1)} = \frac{1}{p(\lfloor N/2 \rfloor)/p(\lfloor N/2 \rfloor - 1)}$$

e nesse caso, a desigualdade fica equivalente a:

$$\frac{1}{p(\lfloor N/2 \rfloor)/p(\lfloor N/2 \rfloor - 1)} < \frac{1}{(N - \lfloor N/2 \rfloor + 1)/\lfloor N/2 \rfloor}$$

$$\Leftrightarrow \quad p(\lfloor N/2 \rfloor)/p(\lfloor N/2 \rfloor - 1) > (N - \lfloor N/2 \rfloor + 1)/\lfloor N/2 \rfloor$$

que é a mesma desigualdade verificada na definição de distribuição mais estreita que a Binomial. O caso em que $y = \lceil N/2 \rceil$ se apresenta da mesma maneira.

O último caso a para verificação da desigualdade se refere a $y = N/2$ quando N for par. Nesse caso, tem-se:

$$\frac{p(N/2 + 1)/p(N/2)}{p(N/2)/p(N/2 - 1)} = \frac{[p(N/2 - 1)]^2}{[p(N/2)]^2}$$

e sendo assim, a desigualdade fica equivalente a:

$$\frac{[p(N/2 - 1)]^2}{[p(N/2)]^2} < \frac{(N/2)^2}{(N/2 + 1)^2} \Leftrightarrow \frac{p(N/2)}{p(N/2 - 1)} > \frac{(N/2 + 1)}{(N/2)}$$

que é a mesma desigualdade verificada na definição de distribuição mais estreita que a Binomial. Dessa forma, verifica-se que para o caso de distribuição mais estreita de segunda ordem que a Binomial tem-se a validade de desigualdades que implicam que $r(m + 1) > r(m)$ para $m = 1, \ldots, N - 1$. ■

Trata-se, portanto, do resultado mais forte a favor da *maturidade* que é possível de se obter no novo modelo. A partir desse resultado conclui-se que a probabilidade preditiva de sucesso somente aumenta à medida que consecutivos fracassos são observados.

No que diz respeito a estendibilidade, tem-se que os modelos resultantes da especificação de priori mais estreita de segunda ordem que a Binomial não são estendíveis. Isso se deve ao fato de que correspondem a uma subclasse de uma classe mais geral (prioris mais estreitas que a Binomial) em que os modelos não são estendíveis.

Capítulo 3

Conclusões

3.1 Revisão dos Principais Pontos de Estudo

A seguir, apresentamos de forma sumarizada os principais pontos de cada estudo estatístico relacionado à *falácia do apostador*. Dentre os aspectos ressaltados, são apresentadas definições, suposições e os principais resultados obtidos no estudo em questão.

O estudo de **O'Neill & Puza**: Baseia-se em um processo x_1, x_2, x_3,... de v.a.'s em $\{1,2,\ldots,k\}$. Seu objeto de principal interesse é a probabilidade preditiva $p(x_{n+1} = x|\mathtt{x}_n)$. Suas principais suposições são de que $\mathtt{x}$ é (infinitamente) permutável, e $\boldsymbol{\Theta} = (\Theta_1, \Theta_2, \ldots, \Theta_k)$ também é permutável. Seu principal resultado é a *Reverse Gambler's Belief*.

O estudo de **Rodrigues & Wechsler**: Baseia-se em um processo x_1, x_2, x_3,... v.a.'s do tipo 0 ou 1. Seu objeto de principal interesse é a *Função Taxa de Falha Preditiva* $r(m)$. Sua principal suposição é a de que $\mathtt{x}$ é (infinitamente) permutável. Admite-se qualquer priori não degenerada para Θ. Seu principal resultado é a *Crença na Maturidade Reversa*.

Este estudo: Se baseia em um vetor $(X_1, X_2, \ldots, X_N)$ de valores desconhecidos **0 ou 1** (populacionais). Seu objeto de principal interesse é a probabilidade preditiva $p(x_{n+1} = x|\mathtt{x}_n)$. Sua principal suposição é a de que o vetor $(X_1, X_2, \ldots, X_N)$ é permutável. Seus principais resultados são: *Indiferença, Equivalência de Modelos, Gambler's Belief e Crença na Maturidade*

3.2 Considerações Finais

Como foi mostrado, estudos anteriores provaram que a *reverse gambler's belief* e a *maturidade reversa* são conseqüências lógicas da suposição de permutabilidade infinita. Neste trabalho apresentamos um modelo no qual uma população finita é considerada e, portanto, a principal suposição dos estudos anteriores não é necessariamente verificada.

No Teorema 2.3.1, estudamos o uso de priori dada por mistura de Binomiais para o total de sucessos na população. Pela equivalência de modelos, obtemos as mesmas conclusões lógicas dos estudos anteriores. No Teorema 2.3.2, estudamos a utilização de prioris mais estreitas que a Binomial. Foi possível concluir que essa especificação leva à *gambler's belief*. Finalmente, no Teorema 2.3.3 foram consideradas distribuições mais estreitas de segunda ordem que a Binomial. Para essa classe, tem-se *crença na maturidade*.

Outro ponto principal de nosso interesse foi a estendibilidade, que é, a existência de algum processo inifinitamente permutável equivalente ao modelo especificado no que diz respeito à distribuição de probabilidade das observações. Como corolário dos resultados obtidos, foi possível concluir que os modelos gerados por especificação de prioris mais estreitas ou mais estreitas de segunda ordem que a Binomial não são estendíveis. Isso acontece porque a suposição de permutabilidade infinita sempre leva à conclusão lógica de *maturidade*. Portanto, resultados opostos a essa crença são indicadores de não-estendibilidade do modelo.

Estes resultados também levam a direções interessantes no entendimento do comportamento humano. Como visto neste trabalho, o novo modelo proposto explicaria não somente o comportamento prescrito pela *falácia do apostador* como também o reverso. Em função disso, quando uma pessoa acredita nas premissas adotados no novo modelo, não é necessariamente irracional se comportar de acordo com a *maturidade*.

Tal fato de o novo modelo ser compatível com *maturidade* é um dos principais pontos que o diferencia em relação aos modelos estudados anteriormente. A outra grande diferença diz respeito a suposição de permutabilidade infinita presente nos estudos anteriores. Trata-se possivelmente de uma indicação de que tal suposição seja inadequada para descrever situações gerais que envolvam a percepção humana.

Acreditamos também que os resultados apresentados podem passar alguma mensagem referente à modelagem estatística de processos 0-1. Embora prioris conjugadas como a Beta-Binomial sejam bastante flexíveis e cubram uma ampla variedade de opiniões, elas sempre levam a conclusões opostas à *maturidade*. Dessa forma, essa classe de prioris exclui distribuições que são consistentes com *maturidade* e, portanto, podem levar a conclusões inferenciais significantemente diferentes.

Por fim, considerando aspectos da teoria estatística, um outro ponto de potencial contribuição deste trabalho refere-se às duas classes de distribuições de probabilidade que foram definidas e estudadas: as mais estreitas e também as mais estreitas de segunda ordem que a Binomial. É possível que distribuições dessas classes apresentam implicações em outros tipos de modelos estatísticos, sendo assim um aspecto com margem para estudos e desenvolvimento.

3.3 Pontos para Pesquisa Futura

Um ponto potencial de pesquisa refere-se ao estudo mais aprofundado das relações lógicas entre *maturidade* e estendibilidade, ou seja, das condições e implicações de ida e volta entre tais propriedades. Poderia se investigar, por exemplo, se todos os modelos que resultem em *maturidade reversa* são estendíveis (sabemos que a implicação da volta é sempre válida).

Outro campo com margem para estudo futuro surge com a definição de novos conceitos de *maturidade*. Neste trabalho, os resultados foram estabelecidos com base em duas definições (*gambler's belief* e *crença na maturidade*), entretanto, outras definições são possíveis para esse conceito, como por exemplo, basear-se apenas nas últimas k observações ($k < n$).

Por fim, outra linha aberta para pesquisa diz respeito à determinação de novas classes de prioris no novo modelo para as quais seja possível estabelecer resultados comportados reversos ou favoráveis à *maturidade*. Uma possibilidade seria, por exemplo, tentar relaxar a suposição de simetria da classe das distribuições mais estreitas que a Binomial, obtendo assim, uma classe mais geral.

Apêndice A

Cálculos Auxiliares

Neste apêndice são apresentadas contas auxiliares utilizadas em algumas passagens deste trabalho. Tais passagens referem-se a exemplos explorados, propriedades e demonstrações, como por exemplo, os resultados de *Indiferença* e *Equivalência de Modelos*.

A.1 Exemplo de Não-estendibilidade

Modelo de probabilidade preditiva de duas variáveis aleatórias binárias (X_1, X_2):

$$P(X_1 = 0, X_2 = 0) = 0 \qquad P(X_1 = 1, X_2 = 1) = 0$$
$$P(X_1 = 1, X_2 = 0) = P(X_1 = 0, X_2 = 1) = 1/2$$

Buscando a equivalência desse modelo com algum processo infinitamente permutável:

$$P(X_1 = 0, X_2 = 0) = 0 \Rightarrow \int_0^1 \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}\, dQ(\theta) = 0 \Rightarrow$$

$$\Rightarrow E[(1-\Theta)^2] = 0 \Rightarrow 1 - 2E(\Theta) + E(\Theta^2) = 0$$

Além disso,

$$P(X_1 = 1, X_2 = 1) = 0 \Rightarrow E[\Theta^2] = 0$$

A solução para esse sistema é $E(\Theta) = 1/2$ e $E(\Theta^2) = 0$, que implicaria que a v.a. Θ tivesse variância negativa, tornando evidente que não é possível encontrar uma distribuição para Θ que satisfaça tais condições.

A.2 Demonstração do resultado de Indiferença

$$p(x_1, x_2, \ldots, x_n|\gamma) = \frac{\binom{\gamma}{\Sigma x}\binom{N-\gamma}{n-\Sigma x}}{\binom{N}{n}\binom{n}{\Sigma x}}$$

Marginalizando-se com respeito a γ:

$$p(x_1, x_2, \ldots, x_n) = \sum_{\gamma=\Sigma x}^{N-n+\Sigma x} \frac{\binom{\gamma}{\Sigma x}\binom{N-\gamma}{n-\Sigma x}}{\binom{N}{n}\binom{n}{\Sigma x}} \cdot \binom{N}{\gamma}\pi^{\gamma}(1-\pi)^{N-\gamma}$$

Após abrir as expressões dos coeficientes binomiais e eliminando aqueles termos que aparecem simultaneamente no numerador e denominador, chega-se em:

$$p(x_1, x_2, \ldots, x_n) = \sum_{\gamma=\Sigma x}^{N-n+\Sigma x} \binom{N-n}{\gamma-\Sigma x}\pi^{\gamma}(1-\pi)^{N-\gamma}$$

$$p(x_1, x_2, \ldots, x_n) = \pi^{\Sigma x}(1-\pi)^{n-\Sigma x} \cdot \sum_{\gamma=\Sigma x}^{N-n+\Sigma x} \binom{N-n}{\gamma-\Sigma x}\pi^{\gamma-\Sigma x}(1-\pi)^{N-n-\gamma+\Sigma x}$$

Ao fazer a transformação $z = \gamma - \Sigma x$, verifica-se que a somatória na expressão é equivalente à soma de todas as probabilidades de uma distribuição Binomial$(N-n, \pi)$, que por sua vez é igual a 1. Portanto:

$$p(x_1, x_2, \ldots, x_n) = \pi^{\Sigma x}(1-\pi)^{n-\Sigma x} \ \square$$

A.3 Demonstração do resultado de Equivalência de Modelos

$$p(x_1, x_2, \ldots, x_n|\gamma) = \frac{\binom{\gamma}{\Sigma x}\binom{N-\gamma}{n-\Sigma x}}{\binom{N}{n}\binom{n}{\Sigma x}}$$

Marginalizando-se com respeito a γ:

$$p(x_1, x_2, \ldots, x_n) = \sum_{\gamma=\Sigma x}^{N-n+\Sigma x} \frac{\binom{\gamma}{\Sigma x}\binom{N-\gamma}{n-\Sigma x}}{\binom{N}{n}\binom{n}{\Sigma x}} \int_0^1 \binom{N}{\gamma}\pi^{\gamma}(1-\pi)^{N-\gamma} dQ(\pi)$$

Utilizando o Teorema de Fubini para inverter a ordem entre somatório e integral:

$$p(x_1, x_2, \ldots, x_n) = \int_0^1 \sum_{\gamma=\Sigma x}^{N-n+\Sigma x} \frac{\binom{\gamma}{\Sigma x}\binom{N-\gamma}{n-\Sigma x}}{\binom{N}{n}\binom{n}{\Sigma x}} \binom{N}{\gamma} \pi^{\gamma}(1-\pi)^{N-\gamma} dQ(\pi)$$

Após abrir as expressões dos coeficientes binomiais e eliminando aqueles termos que aparecem simultaneamente no numerador e denominador, chega-se a:

$$p(x_1, x_2, \ldots, x_n) = \int_0^1 \sum_{\gamma=\Sigma x}^{N-n+\Sigma x} \binom{N-n}{\gamma-\Sigma x} \pi^{\gamma}(1-\pi)^{N-\gamma} dQ(\pi)$$

E por fim, seguindo os mesmos passos da demonstração anterior, tem-se que:

$$p(x_1, x_2, \ldots, x_n) = \int_0^1 \pi^{\Sigma x}(1-\pi)^{n-\Sigma x} dQ(\pi) \ \square$$

A.4 Identidades de Coeficientes Binomiais

A seguir são apresentadas duas identidades envolvendo coeficientes binomiais que foram utilizadas nos primeiros passos de demonstração do Teorema 2.3.2:

1. $$\binom{k-1}{k-1} + \binom{k}{k-1} + \binom{k+1}{k-1} + \cdots + \binom{N-m}{k-1} = \binom{N-m+1}{k}$$

2. $$\binom{N-m+1}{k} = \binom{N-m}{k} + \binom{N-m}{k-1}$$

A.5 Posteriori mais estreita que a Binomial

Verificaremos que ao adotar-se priori mais estreita que a Binomial e observar-se mesmo número de sucessos e fracassos, então a priori atualizada nesse processo manterá a propriedade de ser mais estreita que a Binomial.

Primeiro, chamando de $\mathtt{x}^{(1)}$ a amostra observada com n_1 sucessos e $n_0 = n_1$ fracassos, então observamos que o cálculo da posteriori de γ dá-se por:

$$p(\gamma = y|\mathtt{x}^{(1)}) \propto p(\gamma = y)\binom{y}{n_1}\binom{N-y}{n_1}$$

Dessa forma, verifica-se que a razão das probabilidades a posteriori entre dois valores consecutivos é:

$$\frac{p(\gamma = y|\mathtt{x}^{(1)})}{p(\gamma = y-1|\mathtt{x}^{(1)})} = \frac{p(\gamma = y)\cdot y}{p(\gamma = y-1)\cdot(N-y+1)} \cdot \frac{(N-y-n_1+1)}{(y-n_1)}$$

Uma vez que já foram observados n_1 sucessos (bolas pretas), então já se sabe que $n_1 \leq \gamma \leq N - n_1$. Dessa forma, é preciso fazer uma reparametrização: $\gamma* = \gamma - n_1$ $(0 \leq \gamma* \leq N - 2n_1)$. Tem-se então que:

$$\frac{p(\gamma* = y|\mathtt{x}^{(1)})}{p(\gamma* = y-1|\mathtt{x}^{(1)})} = \frac{p(\gamma = y+n_1|\mathtt{x}^{(1)})}{p(\gamma = y+n_1-1|\mathtt{x}^{(1)})} =$$

$$= \left[\frac{p(\gamma = y+n_1)\cdot(y+n_1)}{p(\gamma = y+n_1-1)(N-y-n_1+1)}\right] \cdot \left[\frac{(N-y-2n_1+1)}{y}\right]$$

Sabemos que o primeiro termo da expressão acima é maior do que 1 até $(N-2n_1)/2$ e menor do que 1 para os valores seguintes. Isso decorre das desigualdades conhecidas sobre $p(\gamma = y+n_1)/p(\gamma = y+n_1-1)$ uma vez que γ tem distribuição mais estreita que a Binomial. Dessa forma, verifica-se que:

- $p(\gamma* = y|\mathtt{x}^{(1)})/p(\gamma* = y-1|\mathtt{x}^{(1)}) > (N-2n_1-y+1)/y$, *se* $y < (N-2n_1)/2$
- $p(\gamma* = y|\mathtt{x}^{(1)})/p(\gamma* = y-1|\mathtt{x}^{(1)}) < (N-2n_1-y+1)/y$, *se* $y > (N-2n_1)/2$

Tais desigualdades se encaixam perfeitamente na terceira condição da definição de distribuição mais estreita que a Binomial$(N-2n_1, 1/2)$. E por fim, verifica-se também que a priori atualizada $(\gamma*)$ satisfaz as duas primeiras condições de tal definição (suporte e simetria), concluindo a averiguação da propriedade. □

A.6 Distribuição Hipergeométrica simétrica - Mais estreita que a Binomial

Nesta seção verificamos que uma v.a. X com distribuição Hipergeométrica$(2T, T, N)$ (simétrica) satisfaz as propriedades que a define como tendo uma distribuição mais estreita que a Binomial$(N, 1/2)$. Para tanto, estudamos a razão entre as probabilidades de dois valores consecutivos nesta distribuição:

$$\frac{P(X=x)}{P(X=x-1)} = \frac{\dfrac{\binom{T}{x}\binom{T}{N-x}}{\binom{2T}{N}}}{\dfrac{\binom{T}{x-1}\binom{T}{N-x+1}}{\binom{2T}{N}}} = \frac{N-x+1}{x} \cdot \frac{T-x+1}{T-N+x}$$

Estudando as situações em que o segundo termo da expressão acima $((T-x+1)/(T-N+x))$ é maior, menor ou igual a 1, então verificamos que:

- $P(X=x)/P(X=x-1) > (N-x+1)/x \ , \ se \ x < (N+1)/2$
- $P(X=x)/P(X=x-1) < (N-x+1)/x \ , \ se \ x > (N+1)/2$
- $P(X=x)/P(X=x-1) = 1 \ , \ se \ x = (N+1)/2$

Tais desigualdades se encaixam perfeitamente na terceira condição da definição de distribuição mais estreita que a Binomial. E por fim, verifica-se também que a Hipergeométrica simétrica satisfaz as duas primeiras condições de tal definição (suporte e simetria), concluindo a averiguação da propriedade. □

A.7 Distribuição Hipergeométrica simétrica - Mais estreita de segunda ordem que a Binomial

A proposta agora é verificar que uma v.a. X com distribuição Hipergeométrica$(2T, T, N)$ (simétrica) satisfaz as propriedades que a define como tendo uma distribuição mais estreita de segunda ordem que a Binomial$(N, 1/2)$. A primeira condição para isso já foi verificada - trata-se de uma distribuição mais estreita que a Binomial. A seguir, estudamos a razão de

segunda ordem dessa distribuição:

$$\frac{P(X=x+1)/P(X=x)}{P(X=x)/P(X=x-1)} = \frac{\left(\dfrac{N-x}{y+1}\right)}{\left(\dfrac{N-x+1}{y}\right)} \cdot \frac{\left(\dfrac{N-x}{T-N+x+1}\right)}{\left(\dfrac{N-x+1}{T-N+x}\right)}$$

Verifica-se a o segundo termo da expressão acima sempre será maior do que 1 se $T > (N-1)/2$, desigualdade essa que é sempre válida pela própria definição da Hipergeométrica simétrica. Dessa forma, tem-se que:

$$\frac{P(X=x+1)/P(X=x)}{P(X=x)/P(X=x-1)} < \frac{\left(\dfrac{N-x}{y+1}\right)}{\left(\dfrac{N-x+1}{y}\right)}$$

Sendo assim, verificamos a validade da segunda condição da definição de mais estreita de segunda ordem que a Binomial, concluindo a averiguação da propriedade nas distribuições Hipergeométricas simétricas. $\square$

Referências Bibliográficas

[1] Johnson, N. & Kotz, S. (1982-1989) *Encyclopedia of Statistical Sciences, Vol. 3.* Wiley, New York.

[2] Laplace, P.S. 1814, English edition 1951, *A Philosophical Essay on Probabilities.* New York: Dover Publications Inc.

[3] Venn, J. (1876) *The Logic of Chance.* 2nd ed., Macmillan and co; reprinted, New York, 1962.

[4] De Finetti, B. (1974) *Theory of probability, Vol 1.*. Wiley, New York.

[5] Bernando, J.M. (1998) Bruno de Finetti en la Estadística Contemporánea *Historia de la Matemática en el Siglo XX*, Madrid: Real Academia de Ciencias,**63-80**.

[6] Paulino, C. D., Amaral Turkman, A. e Murteira, B. (2003) *Estatística Bayesiana.* Fundação Calouste Gulbenkian, Lisboa.

[7] De Finetti, B. (1937) La Prévision: Ses Lois Logiques, Ses Sources Subjectives. *Ann. Inst.Henri Poincaré 7, 1-68. Translation reprinted in H.E. KYBURG and H.E. SMOKLER, Eds.(1980), Studies in Subjective Probability, 2nd ed*, Robert Krieger, New York,**53-118**.

[8] Heath, D., and Sudderth, W. (1976) De Finetti's Theorem on Exchangeable Variables *The American Statistician*, 3,**188-189**.

[9] Brunk, H. D., Ewing, G. M. e Utz, W. R. (1956) Some Helly Theorems for monotone functions. *Proc. Amer. Math. Soc.*, **7**.

[10] Bernardo, J.M. and Smith, A.F.M. (1994) *Bayesian Theory*. John Wiley, Chichester.

[11] Wechsler, S. (1993) Exchangeability and Predictivism. *Erkenntnis - International Journal of Analytic Philosophy*, **38**, n.3, 343:350

[12] Mendel, M. (1994) Operational Parameters in. Bayesian Models. *Test*, **3**(2), 195-206.

[13] Iglesias, P., Loschi, R.H, Pereira, C.A.B. and Wechsler, S. (2008) A Note on Extendibility and Predictivistic Inference in Finite Populations *Brazilian Journal of Probability and Statistics*, **32**(1), 1-9.

[14] Kahneman, D., Slovic, P. and Tversky, A. (1982) *Judgement under uncertainty: Heuristics and biases*. Cambridge University Press, Cambridge, pp 3-22.

[15] O'Neill, P. and Puza, B. (2005) In Defence of the Reverse Gambler's Belief. *The Mathematical Scientist*, **30**.

[16] Cowan, J.L. (1969) The Gambler's Fallacy. *Philosophy and Phenomenological Research*, **30(2)**, pp 238-251.

[17] Rodrigues, F.W. and Wechsler, S. (1993) A Discrete Bayes Explanation of a Failure-Rate Paradox *IEEE Transactions on Reliability*, **42(1)**, pp 132-133.