

Cícero Carlos Felix de Oliveira

Uma *priori* Beta para distribuição Binomial Negativa

RECIFE-PE - JULHO/2011



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

Uma *priori* Beta para distribuição Binomial Negativa

Dissertação apresentada ao Programa de Pós-Graduação em Biometria e Estatística Aplicada como exigência parcial à obtenção do título de Mestre.

Área de Concentração: Modelagem Estatística e Computacional

Orientador: Prof. Dr. Eufrázio de Souza Santos

RECIFE-PE - JULHO/2011.

Ficha catalográfica

O48p Oliveira, Cícero Carlos Felix de
Uma priori Beta para distribuição Binominal Negativa / Cícero
Carlos Felix de Oliveira – 2011.
67 f.: il.

Orientador: Eufrázio de Souza Santos
Dissertação (Mestrado em Biometria e Estatística Aplicada) –
Universidade Federal Rural de Pernambuco, Departamento
de Estatística e Informática, Recife, 2011.
Inclui referências e anexo.

1. Inferência bayesiana 2. Distribuição binomial negativa
3. Distribuição beta 4. Piores não informativas 5. Estimadores
bayesiano 6. Estimador de máxima verossimilhança I. Santos,
Eufrázio de Souza, orientador II. Título

CDD 310


UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

Uma *priori* Beta para distribuição Binomial Negativa

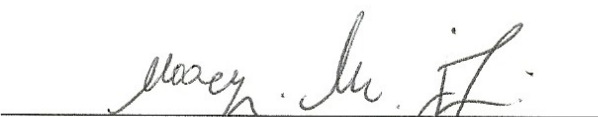
Cícero Carlos Felix de Oliveira


Dissertação julgada adequada para obtenção do título de mestre em Biometria e Estatística Aplicada, defendida e aprovada por unanimidade em 08/07/2011 pela Comissão Examinadora.

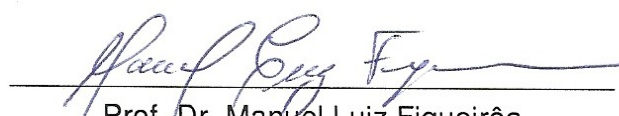
Orientador:


Prof. Dr. Eufrazio de Souza Santos
Universidade Federal Rural de Pernambuco

Banca Examinadora:


Prof. Dr. Moacyr Cunha Filho
Universidade Federal Rural de Pernambuco
DEINFO-UFRPE


Prof. Dr. Borko Stosic
Universidade Federal Rural de Pernambuco
DEINFO-UFRPE


Prof. Dr. Manuel Luiz Figueirôa
Universidade Federal de Sergipe
DECA-UFS

Dedico essa dissertação a toda minha família, como retribuição ao amor, carinho e cumplicidade durante este percurso.

Agradecimentos

Agradeço primeiramente a Deus, uma vez que sem sua vontade nada é possível e por acreditar que nenhum homem seja capaz de superar obstáculos sem fé.

Agradeço também aos meus pais, Antônia Freire Batista e Valmir Felix de Oliveira, cujos sonhos e objetivos sempre estiveram ligados à mim e a meus irmãos, demonstrando alegria nos momentos felizes, esperança e apoio em momentos difíceis.

A minha querida esposa Francisca que esteve presente no desenvolvimento deste trabalho, dando-me força e ajudando a corrigir os erros de português.

Ao professor Doutor Eufrázio de Souza Santos, pela orientação e sabedoria.

Ao meu amigo Professor M.sc. Pedro Ferreira Lima que me incentivou a fazer o mestrado em estatística e também a sua esposa Márcia.

A todos os professores e colaboradores do Departamento de Biometria e Estatística Aplicada da UFRPE, pelos conhecimentos, incentivos, amizade e carinho.

A todos os meus colegas que sempre estiveram carinhosamente presentes contribuindo com críticas, sugestões e paciente tolerância. Em especial aos colegas e amigos: Rosilda, Mariese, Rodrigo, Rita, Jáder, Rogério, Alvino, Aranildo, Darlon, David e Daniele.

Aos amigos Dâmocles, Josimar, Gabriel, Samuel, Vinícius, Katia, Carlos e Joseilme pelos bons momentos, vocês são amizades para uma vida inteira.

Aos funcionários Zuleide e Marco pela amizade e ajuda nestes dois anos.

Aos meus amigos do Instituto Federal do Ceará - Campus Crato onde trabalho, no qual eu tenho orgulho de ter amizade. Nesta Instituição agradeço em especial ao diretor e professor Eder Cardozo Gomes.

Ao meu tio Francisco Errasmo que faleceu este mês de fevereiro. Meu tio é para mim um grande exemplo de dignidade e coragem.

As pessoas que trabalham na empresa de ônibus Princesa do Agreste pelo bom serviço oferecido durante os dois anos do meu mestrado. Esta empresa foi responsável pelo meu transporte semanalmente no percurso Crato - CE × Recife - PE.

"O erro está em pensar neles (priors não informativas) como representante da ignorância"

Lindley, 1990

"O que é desconhecido é incerto, e toda incerteza deve ser quantificada em termos de probabilidade."

Thomas Bayes (1702 - 1761)

Resumo

Nesta dissertação está sendo abordado uma distribuição discreta baseada em ensaios de Bernoulli, que é a distribuição Binomial Negativa. O objetivo principal é propor uma nova distribuição *a priori* não informativa para o modelo Binomial Negativa, que está sendo denominado como uma possível distribuição *a priori* $Beta(0,0)$, que é uma distribuição imprópria. Essa distribuição também é conhecida para o modelo Binomial como *a priori* de Haldane, mas para o modelo Binomial Negativa não há nenhum estudo até o momento. O estudo do comportamento desta *a priori* foi baseada nos contextos bayesiano e clássico. A ideia da utilização de uma *a priori* não informativa é o desejo de fazer inferência estatística baseada no mínimo de informação subjetiva *a priori* quanto seja possível. Assim, torna possível a comparação com os resultados da inferência clássica que só usa informação amostral, como por exemplo, o estimador de máxima verossimilhança. Quando é comparado a distribuição $Beta(0,0)$ com *a priori* de Bayes - Laplace e *a priori* de Jeffreys, baseado-se nos estimadores bayesiano (média *a posteriori* e moda *a posteriori*) e no estimador de máxima verossimilhança, nota-se que a possível *a priori* $Beta(0,0)$ é menos informativa do que as outras *a priori*. É verificado também, que esta possível *a priori* é uma distribuição limitada no espaço paramétrico, sendo assim, uma característica importante para *a priori* não informativa. O principal argumento mostra que a possível *a priori* $Beta(0,0)$ é adequada, quando ela é aplicada numa distribuição *a posteriori* preditiva para modelo Binomial Negativa, levando a uma distribuição Beta Binomial Negativa (que corresponde a uma hipergeométrica multiplicada por uma probabilidade). Todas as observações citadas são fortalecidas por alguns estudos feitos, tais como: conceitos básicos associados à Inferência Bayesiana e conceitos das distribuições Binomial Negativa e Beta Binomial Negativa (que uma mistura da Beta com a Binomial Negativa).

Palavras-chave: Inferência Bayesiana; Distribuição Binomial Negativa; Distribuição Beta; Priors não Informativas; Estimador de Máxima Verossimilhança; Estimadores Bayesiano.

Abstract

This dissertation is being dealt with a discrete distribution based on Bernoulli trials, which is the Negative Binomial distribution. The main objective is to propose a new non-informative prior distribution for the Negative Binomial model, which is being termed as a possible prior distribution $Beta(0, 0)$, which is an improper distribution. This distribution is also known for the Binomial model as Haldane prior, but for the Negative Binomial model there are no studies to date. The study of the behavior of this prior was based on Bayesian and classical contexts. The idea of using a non-informative prior is the desire to make statistical inference based on the minimum of information prior subjective as possible. Well, makes it possible to compare the results of classical inference that uses only sample information, for example, the maximum likelihood estimator. When is compared the $Beta(0, 0)$ distribution with the Bayes-Laplace prior and Jeffreys prior, based on the Bayesian estimators (posterior mean and posterior mode) and the maximum likelihood estimator, note that the possible $Beta(0, 0)$ prior is less informative than the others prior. It is also verified that is prior possible is a limited distribution in parameter space, thus, an important feature for non-informative prior. The main argument shows that the possible $Beta(0, 0)$ prior is adequate, when it is applied in a predictive posterior distribution for Negative Binomial model, leading the a Beta-Negative Binomial distribution (which corresponds the a hypergeometric multiplied by a probability). All observations citas are strengthened by several studies, such as: basic concepts related to Bayesian Inference and concepts of the negative binomial distribution and Beta-Negative Binomial (a mixture of Beta with the negative binomial) distribution.

Key words: Bayesian inference; Negative Binomial distribution; Beta distribution; Non-informative priors, Maximum Likelihood Estimator; Bayesian estimators.

Lista de Figuras

1	O reverendo Thomas Bayes na única representação que existe	p.4
2	Distribuição $Beta(0,001;0,001)$	p.15
3	Distribuição $Beta(0,001;0,5)$	p.15
4	Distribuição $Beta(1;1)$	p.15
5	Distribuição $Beta(0,3;1,3)$	p.15
6	Distribuição $Beta(2;0,6)$	p.16
7	Distribuição $Beta(3;2)$	p.16
8	A densidade da distribuição Binomial Negativa com parâmetros $r = 8$ e $\theta = 0,15$	p.19
9	Distribuição <i>a priori</i> de Bayes - Laplace para a probabilidade de sucesso do modelo Binomial Negativa	p.22
10	Distribuição <i>a priori</i> de Jeffreys para a probabilidade de sucesso do modelo Binomial Negativa	p.24
11	A possível distribuição <i>a priori</i> $Beta(0;0)$ para a probabilidade de sucesso do modelo Binomial Negativa	p.28
12	A densidade da distribuição Beta Binomial Negativa com $a = 1,0; b = 1,0$ e $r = 40$	p.31
13	A densidade da distribuição Beta Binomial Negativa com $a = 1,1; b = 1,1$ e $r = 40$	p.31
14	A densidade da distribuição Beta Binomial Negativa com $a = 3, b = 5$ e $r = 10$	p.32
15	A densidade da distribuição Beta Binomial Negativa com $a = 3, b = 5$ e $r = 40$	p.32

- 16 Gráfico *das posterioris* de Bayes - Laplace, de Jeffreys e uma possível *a posteriori*. Para $n = 10$ e $r = 3$ p. 40
- 17 Gráfico *das posterioris* de Bayes - Laplace, de Jeffreys e uma possível *a posteriori*. Para $n = 100$ e $r = 30$ p. 40

Lista de Tabelas

- 1 Estimadores pontuais bayesiano para o parâmetro θ utilizando *as prioris* $Beta(1; 1)$, $Beta(0; 0, 5)$ e $Beta(0; 0)$. Sabendo que $n = 10$ e $r = 3$ p. 39
- 2 Estimadores pontuais bayesiano para o parâmetro θ utilizando *as prioris* $Beta(1; 1)$, $Beta(0; 0, 5)$ e $Beta(0; 0)$. Sabendo que $n = 100$ e $r = 30$. . . p. 39
- 3 Cenários *a posteriori* preditivas: $p(y|s, x, r)$ p. 44

Sumário

1	Introdução	p. 1
1.1	Organização dos capítulos	p. 2
2	Revisão da Literatura	p. 3
2.1	Inferência Bayesiana	p. 3
2.2	Teorema de Bayes	p. 5
2.3	Função de Verossimilhança e o Teorema de Bayes	p. 7
2.4	Distribuição <i>a priori</i> subjetiva	p. 9
2.5	Distribuições <i>a priori</i> conjugada	p. 9
2.6	Distribuições <i>a priori</i> não informativa	p. 9
2.6.1	Método de Bayes - Laplace	p. 10
2.6.2	Método de Jeffreys	p. 11
2.6.3	Distribuição <i>a priori</i> de referência	p. 12
2.7	Estimação Bayesiana Pontual	p. 12
2.8	Distribuição <i>a priori</i> Beta	p. 13
3	Materiais e Métodos	p. 17
3.1	Distribuição Binomial Negativa	p. 17
3.2	Estimador de Máxima Verossimilhança (EMV)	p. 19
3.3	Distribuições <i>a priori</i> e distribuições <i>a posteriori</i>	p. 20
3.3.1	Distribuição <i>a priori</i> de Bayes - Laplace	p. 22
3.3.2	Distribuição <i>a priori</i> de Jeffreys	p. 23
3.3.3	Uma possível distribuição <i>a priori</i> $Beta(0, 0)$	p. 25

3.4	Distribuição Beta Binomial Negativa	p. 29
3.5	Distribuições <i>a posteriori</i> preditiva para o modelo Binomial Negativa . . .	p. 33
4	Conclusões	p. 39
4.1	Os Principais Argumentos	p. 39
4.1.1	Primeiro Argumento	p. 39
4.1.2	Segundo Argumento	p. 41
4.1.3	Terceiro Argumento	p. 42
5	Considerações Finais	p. 46
	Referências	p. 48
	Anexo	p. 50

1 Introdução

Ao realizar um experimento o pesquisador pode ter interesse em respostas que assumam uma sequência enumerável de valores, por exemplo, o número de sementes germinadas numa certa platagem em pesquisa, numa plantação se define uma unidade espacial na qual é contado o número de insetos, etc. Dados que envolve esta sequência enumerável de valores surgem nos mais variados campos do conhecimento, tais como na biologia, medicina, engenharia, cursos agrários em geral, etc. Essa demanda fez com que surgisse uma variedade de métodos para análise deste tipo de dados, como por exemplo, as distribuições de Poisson, Binomial e Binomial Negativa. Neste trabalho será destacado a distribuição Binomial Negativa.

Ross e Preece (1985), disseram que quando a superdispersão ocorre, a distribuição Binomial Negativa é frequentemente apropriada para os dados.

Sabe-se que a distribuição Binomial Negativa é um modelo discreto de probabilidade baseada em ensaios de Bernoulli, cujo resultado é binário. Para esta distribuição a finalidade principal é analisar, nos contextos clássico e bayesiano, o desempenho das priores não informativas.

Os principais focos que nos levou a analisar se as priores são não informativas adequadas ou não adequadas na distribuição Binomial Negativa, esta baseado em: comparações entre os estimadores bayesiano pontuais e o estimador de máxima verossimilhança, verificando-se as priores não subjetivas são distribuições limitadas (*Bernardo, 1997*) e principalmente o comportamento da priori não informativa quando aplicada a uma distribuição *a posteriori* preditiva da distribuição em estudo.

O objetivo deste trabalho é fazer uma análise Bayesiana para a distribuição Binomial Negativa, e daí propor uma possível distribuição $Beta(0,0)$ como uma distribuição *a priori* não informativa adequada. O argumento principal para verificar esta distribuição é através de uma distribuição *a posteriori* preditiva para o modelo em estudo que coincidentemente, corresponde a uma distribuição Beta Binomial Negativa. Esta distribuição *a*

posteriori quando vista através de uma hipergeométrica equivalente multiplicada por uma probabilidade, é verificada um comportamento simétrico em relação ao número de fracassos nos ensaios de Bernoulli, quando aplicamos a possível *a priori* $Beta(0, 0)$. Esse resultado, que, tanto quanto sabemos, não tem aparecido antes na literatura. Outro fato observado, é que, esta hipergeométrica multiplicada por uma probabilidade dá uma visão interessante da Beta Binomial Negativa. Implica também que, para valores inteiros de a e b , pode ser avaliada facilmente em qualquer pacote de software que fornecem a distribuição hipergeométrica.

1.1 Organização dos capítulos

O primeiro capítulo corresponde à parte introdutória da dissertação.

No segundo capítulo fazemos uma revisão da literatura abordando: os fatores históricos da bayesiana, o teorema de Bayes, o estimador de máxima verossimilhança, a distribuição *a priori* subjetiva, a distribuição *a priori* conjugada, as distribuições *a priori* não informativa, os estimadores bayesiano pontuais e a distribuição beta. Todos os pontos citados foram baseados nos livros: *Box - Tiao (1973)*, *Paulino et. al. (2003)*, *Berger, J. O. (1985)*, *Bernardo, J. M. and Smith, A. F. M. (1994)*, *Ehlers, R. S. (2007)*, *Lopes, H. L. e Migon, H. S. (2002)* e vários artigos que são citados no decorrer desse trabalho.

No terceiro capítulo são apresentados os materias e métodos que auxiliam na compreensão dos argumentos apresentados no capítulo quatro. Primeiro fazemos um estudo da distribuição Binomial Negativa verificando o comportamento da *a priori* não informativas e a possível *a priori* $Beta(0, 0)$. Em seguida fazemos um estudo da distribuição Beta Binomial Negativa com intuito de entender melhor o comportamento da distribuição *a posteriori* preditiva da Binomial Negativa. Neste momento, fazemos uma análise comparativa das *prior* não informativa e a possível *a priori* $Beta(0, 0)$, quando vista através de uma distribuição *a posteriori* preditiva que coincidentemente, corresponde a uma distribuição Beta Binomial Negativa.

No quarto capítulo, apresentamos os principais argumentos desenvolvido nesta dissertação com o objetivo de mostrar que a distribuição $Beta(0, 0)$ é uma possível *a priori* não informativa adequada para distribuição Binomial Negativa.

2 Revisão da Literatura

2.1 Inferência Bayesiana

Fazer inferência é uma das principais finalidades da estatística nos últimos séculos. Na abordagem clássica, os parâmetros desconhecidos utilizados são considerados fixos e todas as análises são baseadas nas informações contidas nos dados amostrais. Segundo alguns autores, esta abordagem foi adotada de forma quase unânime pelos estatísticos durante a primeira metade do século XX. No entanto, a abordagem Bayesiana renasceu nos últimos anos, depois de longos períodos esquecidos pelos pesquisadores.

Antes de continuarmos falando da Inferência Bayesiana nas últimas décadas, vamos lembrar como surgiu esta teoria a partir do seu principal criador. O reverendo Thomas Bayes descobriu o teorema que hoje leva seu nome. Considerando sua imensa importância na estatística, sabemos pouco sobre Thomas Bayes. Vamos citar alguns fatos que ficaram registrados durante sua vida:

- Ele foi um reverendo presbiteriano que viveu no início do século VIII na Inglaterra.
- Estudou teologia na Universidade de Edimburgo (Escócia), de onde saiu em 1722.
- Em 1731 assumiu a paróquia de Tunbridge Wells, no condado de Kent, a 58 km de Londres. Neste mesmo ano, apareceu na Inglaterra um livro anônimo, hoje creditado a Bayes, chamado *Benevolência divina*.
- Cinco anos depois, publicou seu primeiro e único livro de matemática, chamado “*The doctrine of fluxions*” (O nome fluxion foi dado pelo matemático e físico Isaac Newton (1642 - 1727) para a derivativa de uma função contínua, que Newton chamava de fluent).
- Com base nesse livro e em outras possíveis contribuições sobre as quais não temos dados precisos, Bayes foi eleito em 1752 para a *Real Sociedade*, entidade científica britânica criada em 1645.

- Dois anos após sua morte, um amigo, o filósofo Richard Price (1723 - 1791), publicou postumamente no *Philosophical Transactions of the Royal Society* em 1763, um artigo que aparentemente encontrou entre os papéis do reverendo, com o nome “*An essay towards solving a problem in the doctrine of chances*” (Ensaio buscando resolver um problema na doutrina das probabilidades). Neste artigo estava a demonstração do famoso teorema de Bayes. Após sua publicação, o trabalho caiu no esquecimento, o qual só foi resgatado pelo matemático francês Pierre-Simon de Laplace (1749 - 1827), que o revelou ao mundo. Neste artigo Thomas Bayes (*Figura 1*) mostrou como probabilidade inversa poderia ser usado para calcular a probabilidade de eventos antecedentes a partir da ocorrência do evento consequente.



Figura 1: O reverendo Thomas Bayes na única representação que existe

A ideia de probabilidade como grau de credibilidade (segundo Paulino et al. 2003), tão importante para entender a filosofia bayesiana, tem uma longa história (Kyburg e Smokler, 1964). Parece ter sido Bernoulli, na sua obra *Ars Conjectandi* (1713), um dos primeiros autores a definir probabilidade como grau de confiança numa dada proposição que não se sabe se é verdadeira ou falsa. Sabe-se também que De Morgan, na *Formal Logic* (1847), afirma:

- (1) a probabilidade identifica-se com um grau de credibilidade;
- (2) os graus de credibilidade podem medir-se;
- (3) os graus de credibilidade podem identificar-se com um certo complexo de sentimentos.

Os métodos de Bayes e de outros pesquisadores daquela época foram adotados por Laplace e outros cientistas do século 19, mas grande parte caiu em desgraça no começo do século 20. Em meados do século 20 interesse em métodos bayesianos foi renovado por De Finetti, Jeffreys, Savage, Press, Box e Tiao, Lindley, e análise de referência, cujo desenvolvimento começou no final de 70 (ver, por exemplo Bernardo S., 1994), fornece uma metodologia geral que inclui e generaliza as soluções pioneiras.

Eles desenvolveram um método completo de inferência estatística baseada no teorema de Bayes.

2.2 Teorema de Bayes

O teorema de Bayes é uma proposição sobre probabilidades condicionais indiscutível desde que se aceitem as leis tradicionais do cálculo de probabilidades ou axiomática de Kolmogorov (Box - Tiao, 1973). O que tem dado lugar a grande controvérsia é a sua interpretação e a sua aplicação a problemas de inferência estatística. Ocupa, como é óbvio, lugar crucial na inferência Bayesiana.

Seja θ uma quantidade de interesse desconhecida cujos possíveis valores são pertencentes ao conjunto Θ . O objetivo da inferência bayesiana pode ser a estimação de θ ou o teste de alguma hipótese envolvendo valores de θ . Um dos principais ingredientes para a realização de inferência bayesiana é a distribuição *a posteriori*, que representa o conhecimento a respeito de θ após a observação dos dados X .

A distribuição *a posteriori* é obtida através do teorema de Bayes, isto é, definido da seguinte forma (Box - Tiao, 1973):

Suponha que $\mathbf{x}' = (x_1, x_2, \dots, x_n)$ é um vetor de n observações cuja distribuição de probabilidade $p(\mathbf{x}|\theta)$ dependem dos valores de k parâmetros $\theta' = (\theta_1, \theta_2, \dots, \theta_k)$. Suponha também que θ em si tem uma distribuição de probabilidade $p(\theta)$. Então,

$$p(\mathbf{x}|\theta) \cdot p(\theta) = p(\mathbf{x}, \theta) = p(\theta|\mathbf{x}) \cdot p(\mathbf{x}). \quad (2.1)$$

Sendo os dados observados \mathbf{x} , a distribuição condicional de θ é

$$p(\theta|\mathbf{x}) = \frac{p(\mathbf{x}|\theta) \cdot p(\theta)}{p(\mathbf{x})} \quad (2.2)$$

Além disso, podemos escrever

$$p(\mathbf{x}) = E[p(\mathbf{x}|\theta)] = c^{-1} = \begin{cases} \int p(\mathbf{x}|\theta) \cdot p(\theta) d\theta & \theta \text{ contínuas} \\ \sum p(\mathbf{x}|\theta) \cdot p(\theta) & \theta \text{ discreta} \end{cases} \quad (2.3)$$

Onde o somatório ou a integral é tomada sobre o alcance admissível de θ , e onde $E[f(\theta)]$ é a esperança matemática de $f(\theta)$ com respeito a distribuição $p(\theta)$. Assim podemos escrever a expressão (2.2) alternativamente com

$$p(\theta|\mathbf{x}) = c \cdot p(\mathbf{x}|\theta)p(\theta). \quad (2.4)$$

A afirmação da expressão (2.2), ou seu equivalente (2.4), é geralmente denominado teorema de Bayes. Nesta expressão, $p(\theta)$, nos diz que é conhecido sobre θ sem conhecimento dos dados, é chamado distribuição *a priori* de θ , ou a distribuição de θ *a priori*. Correspondentemente, $p(\theta|\mathbf{x})$, que nos diz que é conhecido sobre θ o conhecimento dos dados, é chamado distribuição *a posteriori*. A quantidade c é simplesmente uma constante normalizadora necessária para garantir que a distribuição *a posteriori* $p(\theta|\mathbf{x})$ integra ou soma.

A distribuição *a posteriori* de um parâmetro θ contém toda a informação probabilística a respeito deste parâmetro e um gráfico da sua função de densidade *a posteriori* é a melhor descrição do processo de inferência. No entanto, algumas vezes é necessário resumir a informação contida nesta *a posteriori* através de alguns poucos valores numéricos. O caso mais simples é a estimação pontual de θ onde se resume a distribuição *a posteriori* através de um único número, $\hat{\theta}$.

As informações sobre o vetor de parâmetros de interesse, θ , que são obtidas antes da realização de um experimento podem ser representadas probabilisticamente através

das chamadas distribuições *a priori*. Portanto, essas distribuições expressam o estado de conhecimento ou ignorância sobre o vetor de parâmetros θ antes de os dados serem obtidos (Box - Tiao, 1973). Serão discutidas, nas seções (2.4), (2.5) e (2.6), algumas formas de especificação de distribuição *a priori*.

2.3 Função de Verossimilhança e o Teorema de Bayes

A função de verossimilhança de θ é a função que associa a cada θ a distribuição de probabilidade conjunta $p(\mathbf{x}|\theta)$, isto é, $L(\theta|\mathbf{x}) = p(\mathbf{x}|\theta)$. No caso particular em que X_1, X_2, \dots, X_n são variáveis aleatórias independentes e identicamente distribuídas, a função de verossimilhança de θ correspondente à amostra x_1, x_2, \dots, x_n é dada por

$$L(\theta|\mathbf{x}) = \prod_{i=1}^n p(x_i|\theta).$$

Note porém que a definição de verossimilhança não requer que os dados sejam observações de variáveis aleatórias independentes ou identicamente distribuídas. Além disso, fatores que dependem somente de \mathbf{x} e não dependem de θ podem ser ignorados quando se escreve a função de verossimilhança já que eles não fornecem informação sobre a plausibilidade relativa de diferentes valores de θ .

Agora, com base nos dados \mathbf{x} , $p(\mathbf{x}|\theta)$ em (2.4) pode ser considerada como uma função não de \mathbf{x} , mas de θ . Quando assim considerado, na sequência de Fisher (1922), ela é chamada a função de verossimilhança de θ para \mathbf{x} dado e pode ser escrito como $L(\theta|\mathbf{x})$. Deste modo, podemos escrever a fórmula de Bayes como

$$p(\theta|\mathbf{x}) = L(\theta|\mathbf{x}) \cdot p(\theta). \quad (2.5)$$

Em outras palavras, então, o teorema de Bayes diz que a distribuição de probabilidade para *a posteriori* θ dados \mathbf{x} é proporcional ao produto da distribuição *a priori* de θ ao da verossimilhança de θ dado \mathbf{x} . Isto é

$$\text{distribuição } a \text{ posteriori} \propto \text{verossimilhança} \times \text{distribuição } a \text{ priori}.$$

A função de verossimilhança $L(\theta|\mathbf{x})$ desempenha um papel muito importante no teorema de Bayes. É a função através da qual os dados \mathbf{x} modifica o conhecimento *a priori* de θ ; portanto, pode ser considerado como o representante das informações sobre θ prove-

nientes de dados.

A função de verossimilhança é definida até como constante multiplicativa, isto é, a multiplicação por uma constante deixa a verossimilhança inalterado. Isto está de acordo com o papel que desempenha na fórmula de Bayes, pois multiplicando a função de verossimilhança por uma constante arbitrária não terá nenhum efeito sobre a distribuição *a posteriori* de θ . A constante quer cancelar sobre o produto normalizado no lado direito de (2.1). É apenas o valor relativo do verossimilhança que é de grande importância (Box - Tiao, 1973).

Quanto à natureza sequencial do teorema de Bayes, suponhamos que temos uma amostra inicial x_1 , então o teorema de Bayes é dado por:

$$p(\theta|x_1) \propto p(\theta) \cdot L(\theta|x_1). \quad (2.6)$$

Agora, suponha que temos uma segunda amostra de observação x_2 distribuídos de forma independente da primeira amostra, então

$$p(\theta|x_2, x_1) \propto p(\theta) \cdot L(\theta|x_1) \cdot L(\theta|x_2) \propto p(\theta|x_1) \cdot L(\theta|x_2). \quad (2.7)$$

A expressão (2.7) é precisamente a mesma forma (2.6), exceto que em $p(\theta|x_1)$, a distribuição *a posteriori* para θ dado x_1 , desempenha o papel da distribuição *a priori* para a segunda amostra. Obviamente, este processo pode ser repetido várias vezes. Em particular, se temos n observações independentes, a distribuição *a posteriori* pode, se desejar, ser recalculado a cada nova observação, de modo que, no m -ésimo estágio a verossimilhança associada com a m -ésimo observação que é combinada com a distribuição *a posteriori* de θ após $m - 1$ observações para dar a nova distribuição *a posteriori*:

$$p(\theta|x_1, x_2, \dots, x_m) \propto p(\theta|x_1, x_2, \dots, x_{m-1})L(\theta|x_m), \quad m = 2, \dots, n \quad (2.8)$$

onde

$$p(\theta|x_1) \propto p(\theta)L(\theta|x_1).$$

Assim, o teorema de Bayes descreve, em uma era fundamental, o processo de aprendizagem por experiência, e mostra como o conhecimento sobre o estado de natureza representada por θ é continuamente modificado a medida que novos dados estiverem disponíveis.

2.4 Distribuição *a priori* subjetiva

A distribuição *a priori* subjetiva é usada quando não existe nenhuma base teórica objetiva para o ajuste do modelo. Nessas situações, a distribuição *a priori* representa a expectativa subjetiva do pesquisador a respeito do parâmetro de interesse. Com isso, a distribuição *a posteriori* para o parâmetro, obtida combinando essa distribuição *a priori* e a função de verossimilhança, mostra como o conhecimento pessoal do pesquisador é modificado pela informação dos dados (Box - Tiao, 1973). O problema de como quantificar essa informação subjetiva e transformá-la em uma distribuição *a priori* que possa ser usada na abordagem Bayesiana, tem sido discutido abundantemente na literatura. Uma interação entre o pesquisador e o estatístico é fundamental para a determinação de tal *a priori*.

2.5 Distribuições *a priori* conjugada

A idéia de distribuição conjugada é que as distribuições *a priori* e *a posteriori* pertençam a mesma classe de distribuições e assim a atualização do conhecimento que se tem sobre o parâmetro θ envolve apenas uma mudança nos parâmetros indexadores da família de distribuições *a priori*, denominados hiperparâmetros, que diferem do parâmetro θ . Neste caso, o aspecto sequencial do método Bayesiano pode ser explorado definindo-se apenas a regra de atualização dos hiperparâmetros já que as distribuições permanecem as mesmas. Podemos definir a distribuição *a priori* conjugada da seguinte forma (Ehlers, 2007):

- Se $F = \{p(x|\theta), \theta \in \Theta\}$ é uma classe de distribuições amostrais então uma classe de distribuições P é conjugada a F se

$$\forall p(x|\theta) \in F \quad e \quad p(\theta) \in P \Rightarrow p(\theta|x) \in P.$$

Gamerman e Lopes (2006) ressalta que a distribuição *a priori* conjugada deve ser usada com cuidado, pois sua utilização está muitas vezes associadas às facilidades analíticas e nem sempre é uma representação adequada do conhecimento *a priori* do parâmetro.

2.6 Distribuições *a priori* não informativa

A especificação de distribuição *a priori* não informativa é usada quando se espera que a informação dos dados seja dominante, no sentido que a informação *a priori* é vaga, ou seja, não existe informação *a priori* palpável, tanto de natureza subjetiva quanto objetiva.

Os conceitos de “conhecimento vago”, “não informação”, ou “ignorância *a priori*” claramente não são únicos e o problema de caracterizar *a priori* com tais características pode se tornar bastante complexo. Por outro lado, reconhece-se a necessidade de alguma forma de análise que, em algum sentido, consiga captar esta noção de uma *a priori* que tenha um efeito mínimo, relativamente aos dados, na inferência final. Tal análise pode ser pensada como um ponto de partida quando não se consegue fazer uma elicitação detalhada do “verdadeiro” conhecimento *a priori*.

Esta distribuição *a priori* não informativa começou por ser predominantemente interpretadas como representações formais de ignorância, mas há hoje uma tendência (motivada pela não aceitação de representações objetivas únicas da ignorância) para encará-la como opções convencionais de defeito a que se recorre em caso de informação *a priori* insuficiente que torne difícil elicitar uma distribuição subjetiva considerada adequada (Paulino et al., 2003).

Uma *priori* não informativa pode ser obtida a partir de uma *priori* conjugada definindo-se o hiperparâmetro de escala tendendo a zero e mantendo os outros constantes. A seguir são apresentados alguns métodos para obtenção de distribuição *a priori* não informativa.

2.6.1 Método de Bayes - Laplace

O argumento primeiramente representado para gerar distribuição *a priori* não informativa foi o *Princípio da Razão Insuficiente* devido a Bayes - Laplace. De acordo com este princípio, na ausência de razão suficiente para privilegiar umas possibilidades em detrimento de outras, decorrente da escassez informativa *a priori* deve-se adaptar a equiprobabilidade.

A primeira idéia de *a priori* não informativa que se pode ter é pensar em todos os possíveis valores de θ como igualmente prováveis, isto é, como uma distribuição *a priori* uniforme. Essa idéia motivou Bayes e Laplace a sugerir $p(\theta) \propto k$ para θ variando em um subconjunto da reta, significando que nenhum valor particular tem preferência. Porém esta escolha da *priori* pode trazer algumas dificuldades técnicas (Ehlers, 2007).

- Se o intervalo de variação de θ for ilimitado então a distribuição *a priori* é imprópria, isto é,

$$\int p(\theta) d\theta = \infty.$$

- Se $\phi = g(\theta)$ é uma reparametrização não linear monótona de θ então $p(\phi)$ é não

uniforme já que pelo teorema de transformação de variáveis

$$p(\phi) = p(\theta(\phi)) \left| \frac{d\theta}{d\phi} \right| \propto \left| \frac{d\theta}{d\phi} \right|.$$

2.6.2 Método de Jeffreys

Jeffreys crítica a distribuição *a priori* de Bayes - Laplace, pois o primeiro método considera uma constante para representar a ignorância *a priori* e não é invariante sob transformação 1 - 1. Por conta disso Harold Jeffreys propôs, em 1961, uma forma de obtenção de distribuições *a priori* que assegura invariância sob transformação injetora (Paulino et al. 2003). Antes de definir esta *a priori*, precisamos primeiro definir a medida de informação de Fisher.

Seja X uma única observação com função de (densidade) de probabilidade $p(x|\theta)$. A medida de informação esperada de Fisher de θ através de X é definida como

$$I(\theta) = E \left[-\frac{\partial^2 \log p(x|\theta)}{\partial \theta^2} \right] \quad (2.9)$$

Se θ for um vetor paramétrico define-se então a matriz de informação esperada de Fisher de θ através de X como

$$I(\theta) = E \left[-\frac{\partial^2 \log p(x|\theta)}{\partial \theta \partial \theta'} \right].$$

(Paulino et al. 2003; Berger, 1985) Seja X uma variável aleatória com função (densidade) de probabilidade $p(x|\theta)$. A *a priori* de Jeffreys para o caso uniparamétrico é dado por

$$p(\theta) = [I(\theta)]^{\frac{1}{2}};$$

e no caso multiparamétrico é dado por

$$p(\theta) = [\det I(\theta)]^{\frac{1}{2}}.$$

Vale resaltar, que a *a priori* de Jeffreys viola o princípio da verossimilhança, já que a informação de Fisher depende da distribuição amostral. Mas um importante ganho desta distribuição *a priori* em relação à proposta de Bayes - Laplace é que o parâmetro θ não precisa necessariamente pertencer a um espaço limitado.

Bernardo (1989) chama a atenção para alguns aspectos da distribuição *a priori* de Jeffreys:

- A principal motivação intuitiva da distribuição *a priori* de Jeffreys é ser invariante, a qual é uma condição necessária mas não suficiente, para determinar uma referência sensível de distribuição *a priori*.
- A existência da distribuição *a priori* de Jeffreys requer condições de regularidade fortes, como a normalidade assintótica da distribuição *a posteriori* de θ , conforme lembra *Bernardo & Smith (2000)*.

Jeffreys acreditava que uma distribuição *a priori* deveria ser convencionalizada, um padrão de referência, da mesma maneira que existem padrões de natureza científica. Baseado na invariância sugeriu um modelo de obtenção da distribuição *a priori*. Bernardo na tentativa de superar as dificuldades encontradas por Jeffreys propôs, em 1979, a distribuição *a priori* de referência.

2.6.3 Distribuição *a priori* de referência

A distribuição *a priori* de referência surgiu como uma forma objetiva para formular uma distribuição *a priori* não informativa que maximiza a falta de informação (*Bernardo, 2005; Bernardo & Smith, 2000*).

A especificação da distribuição *a priori* de referência é complexa computacionalmente, mas para a função de verossimilhança abordada neste trabalho, a mesma coincide com o método de Jeffreys (segundo Yang & Berger, 1997). Na verdade, a distribuição *a priori* de Jeffreys se tornou um caso particular desta distribuição *a priori* em situações na qual a densidade *a posteriori* de θ possui normalidade assintótica.

2.7 Estimação Bayesiana Pontual

A distribuição *a posterior* sintetiza a nossa crença sobre o parâmetro depois de observamos os dados. Leva em conta a nossa crença *a priori* (a distribuição *a priori*) e os dados (verossimilhança). Um gráfico da distribuição *a posteriori* nos mostra tudo o que podemos saber sobre o parâmetro, após observação dos dados. A distribuição é difícil de interpretar. Muitas vezes queremos encontrar alguns números que a caracterizam. Estas incluem medidas de posição que determinam onde a maior probabilidade está no conjunto de dados, e as medidas de dispersão que determinam como a probabilidade é amplamente espalhado. Nós podemos querer determinar um intervalo que tem uma alta probabilidade de conter o

parâmetro. Estes são conhecidos como intervalos de credibilidade bayesianos e são um pouco análoga à intervalos de confiança. Esta ultima parte não será estudo neste trabalho.

A escolha das estimativas bayesianas de θ depende naturalmente da forma de $p(\theta|x)$, bem como dos objetivos do seu uso. As estimativas mais usadas são: a média *a posteriori*, a moda *a posteriori* e a mediana *a posteriori* (de parâmetros escalares), cuja a definição se apresenta a seguir onde $\theta = (\theta_1, \theta_2, \dots, \theta_k)$ (Paulino et. al. 2003):

- Média *a posteriori*: é uma das medidas mais utilizadas entre as medidas de posição.

$$\hat{\theta} = E(\theta_i|X) = \int_{\Theta} \theta_i p(\theta|x) d\theta, \quad i = 1, 2, \dots, k;$$

- Moda *a posteriori*: este é o valor que maximiza a distribuição *a posteriori*.

$$\hat{\theta} = M(\theta|X) = \max_{\theta \in \Theta} p(\theta|x) = \max_{\theta \in \Theta} \{L(\theta|x)p(\theta)\};$$

- Mediana *a posteriori*: este é um valor que está no meio de um conjunto de dados ordenados. Se $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k)$, então

$$p\{\theta_i \geq \hat{\theta}_i|x\} \geq \frac{1}{2} \quad \text{e} \quad p\{\theta_i \leq \hat{\theta}_i|x\} \geq \frac{1}{2}, \quad i = 1, 2, \dots, k.$$

Outra estimativa bayesiana que iremos destacar será as medidas de dispersão. Depois que encontramos a média, a segunda coisa que queremos saber sobre a distribuição *a posteriori* como ela se comporta em torno da média. Se ela apresentar uma grande dispersão, então o nosso conhecimento sobre o parâmetro, mesmo depois de analisar os dados observados, ainda é imprecisa. Vamos definir apenas a variância e o desvio padrão *a posteriori*. Isto é:

- Variância *a posteriori*:

$$Var(\theta|X) = E(\theta|X)^2 - [E(\theta|X)]^2;$$

- Desvio padrão *a posteriori*: é a raiz quadrada da variância *a posteriori*, ou seja,

$$DP(\theta|X) = \sqrt{Var(\theta|X)}.$$

2.8 Distribuição *a priori* Beta

Na Inferência Bayesiana, ao se utilizar diferentes distribuições *a priori*, são requeridos diferentes desenvolvimentos do teorema de Bayes,

$a \text{ posteriori} \propto a \text{ priori} \times \text{Verossimilhança},$

isto é, para cada distribuição *a priori* considerada é necessário obter uma distribuição *a posteriori*. Uma forma viável de facilitar este processo é utilizar uma mesma distribuição *a priori*, e adotar diferentes valores para seus parâmetros, os quais são definidos como hiperparâmetros, de forma que esta passe a representar diferentes formas, e assim caracterizar outras distribuições de probabilidade. Dentre as distribuições que apresentam esta qualidade, destaca-se a distribuição Beta, pois sua função densidade de probabilidade é relativamente simples e dispõe de grande flexibilidade (Casela & Berger, 1994). Em outras palavras, a distribuição Beta pode adquirir facilmente a forma de distribuições simétricas, assimétricas e constantes mediante simples transformação nos valores de seus parâmetros.

A partir do conhecimento sobre o parâmetro que descreve a probabilidade de sucesso na distribuição Binomial Negativa, pode-se definir uma família paramétrica de densidade. Neste caso, a distribuição *a priori* é representada de forma funcional, cujos parâmetros devem ser especificados de acordo com este conhecimento. Estes parâmetros indexadores da família de distribuição *a priori* são chamados de hiperparâmetros para distinguí-los do parâmetro de interesse θ .

Essa abordagem facilita a análise e o caso mais importante é o de distribuição *a priori* conjugadas. A idéia é que as distribuições *a priori* e *a posteriori* pertençam à mesma classe de distribuições e assim a atualização do conhecimento sobre θ envolve apenas uma mudança nos hiperparâmetros.

Como foi dito acima e como está presente em várias literaturas bayesiana, que a família de distribuições Beta é conjugada ao modelo Bernoulli e conseqüentemente ao modelo Binomial. Não é difícil mostrar que o mesmo vale para as distribuições amostrais Binomial Negativa e conseqüentemente para a distribuição Geométrica.

A distribuição *a priori* conjugada para o parâmetro θ é a distribuição Beta. Logo, a distribuição *a posteriori* tem densidade Beta com atualização nos hiperparâmetros. A função de densidade da distribuição $Beta(a, b)$ para quaisquer hiperparâmetros $a > 0$ e $b > 0$ é dado por

$$p(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1} (1-\theta)^{b-1} \quad 0 < \theta < 1.$$

A distribuição uniforme é um caso especial da distribuição Beta, com $a = b = 1$, que é também chamado de *a priori* de Bayes - Laplace (será visto adiante). A razão pela qual se

poderia considerar o uso da distribuição Beta como *a priori*, é porque a distribuição Beta e a distribuição de Bernoulli forma um par conjugado, de modo que a distribuição *a posterior* ainda é uma versão Beta (por exemplo, *DeGroot, 1970*). Isto dá-nos alguma conveniência analítica.

As seguintes propriedades da distribuição beta são úteis. Se $\theta \sim \text{Beta}(a, b)$, então

$$\mathbb{E}(\theta) = \frac{a}{a+b}, \quad \mathbb{M}_0 = \frac{a-1}{a+b-2} \quad \text{e} \quad \mathbb{V}(\theta) = \frac{ab}{(a+b+1)(a+b)^2}.$$

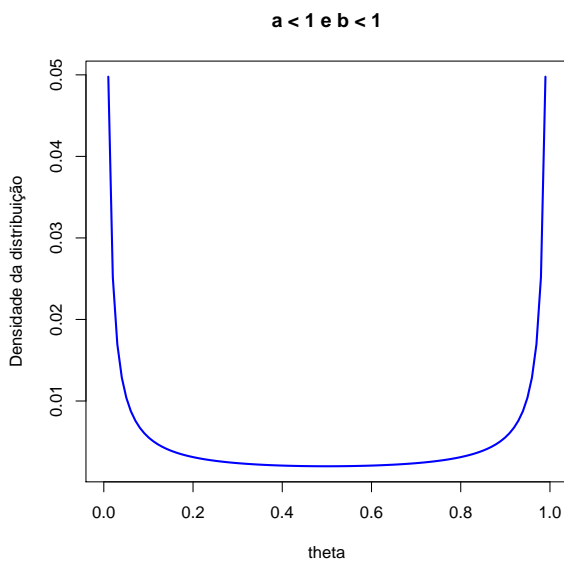


Figura 2: Distribuição $\text{Beta}(0,001; 0,001)$

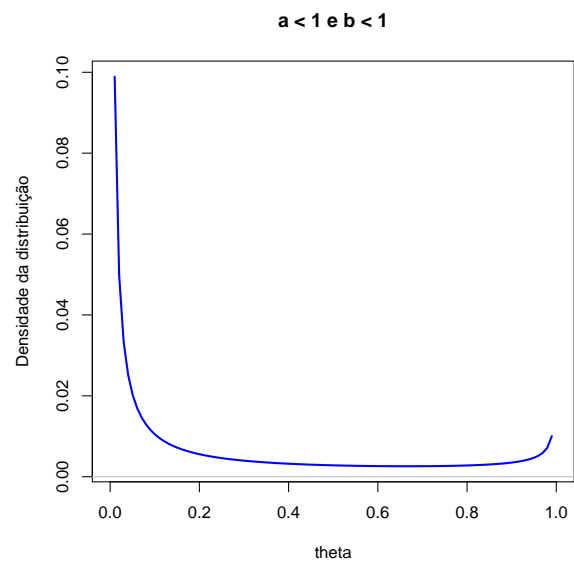


Figura 3: Distribuição $\text{Beta}(0,001; 0,5)$

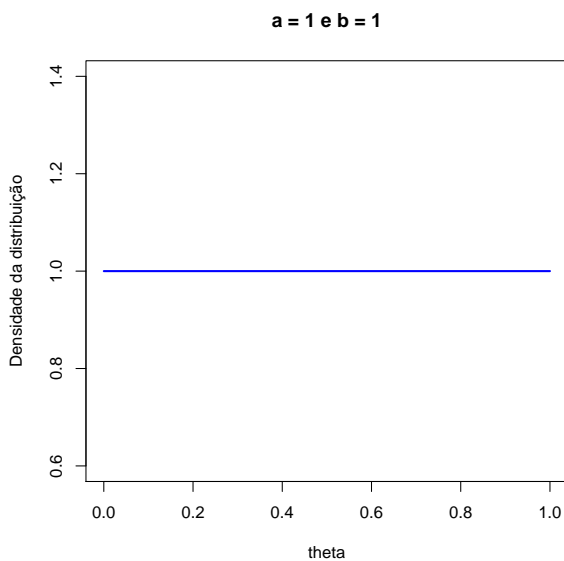


Figura 4: Distribuição $\text{Beta}(1; 1)$

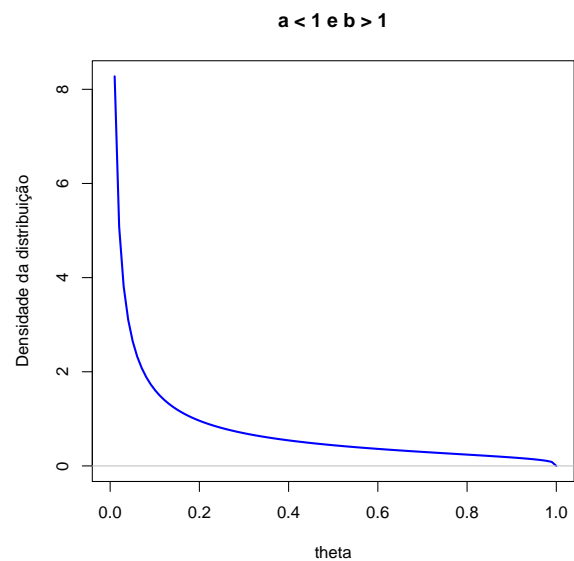


Figura 5: Distribuição $\text{Beta}(0,3; 1,3)$

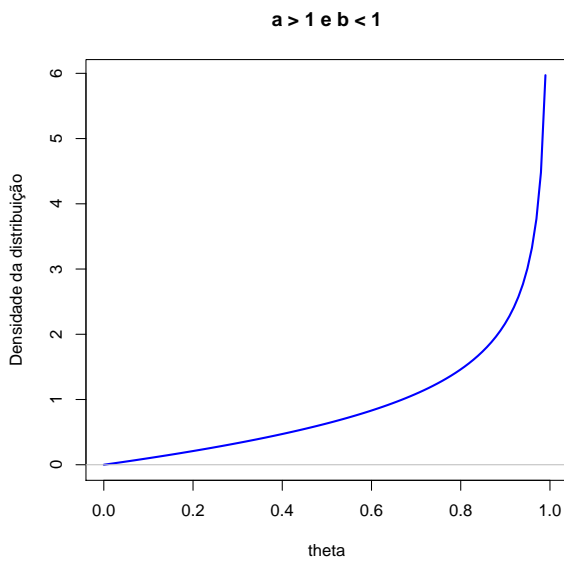


Figura 6: Distribuição $Beta(2; 0, 6)$

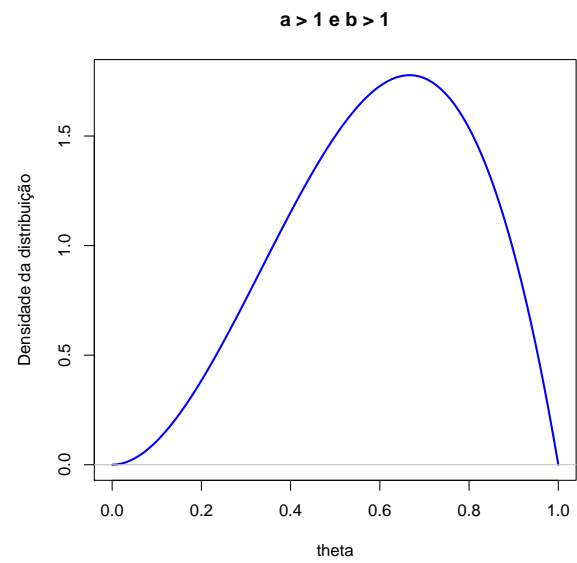


Figura 7: Distribuição $Beta(3; 2)$

Observe que nas *figuradas 2, 3, 4, 5 e 6* as distribuições beta não apresentam moda. Porém, na *figura 7* a distribuição beta é unimodal e a moda ocorre em

$$M_0 = \frac{3 - 1}{3 + 2 - 2} = \frac{2}{3} \approx 0,667$$

3 Materiais e Métodos

3.1 Distribuição Binomial Negativa

Novas distribuições de probabilidade são obtidas quando é admitido que o parâmetro da distribuição também varie segundo uma distribuição de probabilidade, isto é, o parâmetro de interesse passa a ser considerado também uma variável aleatória com sua própria distribuição. Esse mecanismo de obtenção de distribuições de probabilidade é definido como mistura. Uma distribuição conhecida pelo processo de mistura é a Binomial Negativa, muito usada no ajuste de dados biológicos. Essa distribuição é resultado da mistura das distribuições Poisson e Gama.

Outro processo similar para se obter novas distribuições é dado pelo conceito de distribuições generalizadas, resultado da soma de um número aleatório de variáveis aleatórias independentes e identicamente distribuídas. Novamente, obtém-se a distribuição Binomial Negativa como distribuição generalizada de uma distribuição de Poisson com uma distribuição Logarítmica.

Assim, tem-se que, a distribuição Binomial Negativa pode ser obtida por dois processos distintos, com pressupostos diferentes, mistura e distribuição generalizada. Mas não mostraremos como encontrar a distribuição Binomial Negativa, vamos dar, somente um conceito intuitivo.

Suponha que haja uma sequência de ensaios de Bernoulli independentes, cada ensaio com dois resultados possíveis, chamado “SUCESSO” e “FRACASSO”. Em cada ensaio, a probabilidade de sucesso é θ e de fracasso $1 - \theta$. Suponha agora que o interesse seja contar o número de ensaios de Bernoulli necessários para se obter um número de sucesso fixo. Essa última afirmação conduz à distribuição Binomial Negativa (*Casella & Berger, 2002*).

Considere uma sequência de ensaios de Bernoulli realizada de forma independente com probabilidade θ e defina X como o número de fracassos anteriores ao r -ésimo sucesso.

A variável aleatória X segue uma distribuição Binomial Negativo com parâmetros r e θ , em que $0 < \theta < 1$ e $r > 0$ (Haldane, 1945).

Assim, $X|\theta$ tem distribuição condicional Binomial Negativa dada por:

$$P(x|r, \theta) = \binom{x+r-1}{x} \theta^r (1-\theta)^x, \quad x = 0, 1, 2, \dots \quad (3.1)$$

Aqui a quantidade entre parênteses é chamado de coeficiente binomial, e é igual a:

$$\binom{x+r-1}{x} = \frac{(x+r-1)!}{x!(r-1)!} = \frac{\Gamma(x+r)}{\Gamma(x+1)\Gamma(r)} = (-1)^x \binom{-r}{x}$$

O valor esperado e a variância da variável aleatória X são respectivamente

$$E(X) = \frac{r(1-\theta)}{\theta} \quad \text{e} \quad Var(X) = \frac{r(1-\theta)}{\theta^2}$$

Quando r é um número inteiro positivo, a expressão (3.1) pode ser interpretada como a distribuição do tempo de espera para a ocorrência do r -ésimo sucessos, sendo também conhecida por distribuição Pascal (Feller, 1968). Para $r = 1$ ela se reduz a distribuição geométrica. A distribuição binomial negativa (BN) apresenta uma propriedade importante, a variância excede a média. Essa característica é, às vezes, referida como superdispersão.

O parâmetro θ , que está presente nas distribuições Binomial e Binomial Negativa, tem a mesma interpretação teórica. A diferença está no planejamento amostral.

Podemos verificar facilmente que a distribuição Binomial Negativa é unimodal, ou seja, apresenta uma única moda. Uma maneira de se obter a moda desta distribuição é utilizar o fato de que a função gama generaliza o fatorial permitindo assim uma extensão contínua. Sabemos por definição que a moda é calculada pela primeira derivada do logaritmo da expressão (3.1) e depois igualando a zero, isto é:

$$\begin{aligned} \frac{\partial \log P(x|r, \theta)}{\partial x} &= \frac{\partial}{\partial x} \log \left(\frac{\Gamma(x+r)}{\Gamma(x+1)\Gamma(r)} \theta^r (1-\theta)^x \right) \\ &= \frac{\partial}{\partial x} (\log \Gamma(x+r) - \log \Gamma(x+1) - \log \Gamma(r) + r \log \theta + x \log (1-\theta)) \end{aligned}$$

Pelo programa de Computação Algébrico MAPLE ou de acordo com (Madeira, 2009), temos que:

$$\frac{\partial}{\partial x} (\log \Gamma(x+r)) = \Psi(x+r) \quad \frac{\partial}{\partial x} (\log \Gamma(x+1)) = \Psi(x+1)$$

onde $\Psi(x) = \frac{\Gamma'(x)}{\Gamma(x)}$, então

$$\frac{\partial \log P(x|r, \theta)}{\partial x} = \Psi(x+r) - \Psi(x+1) + \log(1-\theta) = 0 \quad (3.2)$$

Conhecendo-se os valores dos parâmetros r e θ é possível, com o auxílio de um software, obter uma aproximação para o valor da moda.

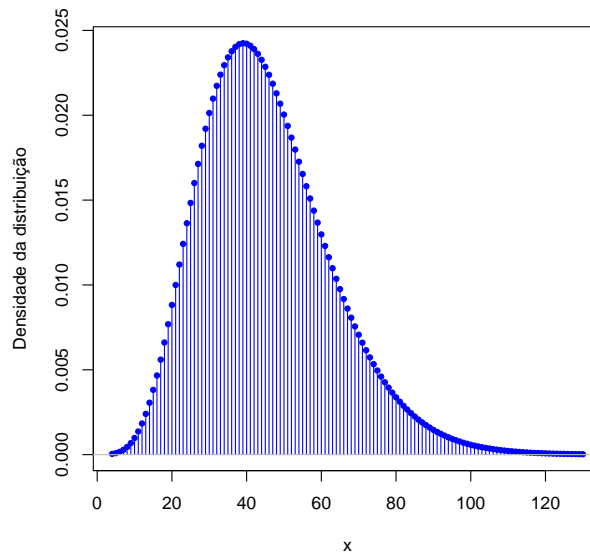


Figura 8: A densidade da distribuição Binomial Negativa com parâmetros $r = 8$ e $\theta = 0,15$

Na distribuição Binomial Negativa correspondente a *figura 8*, o valor da moda é aproximadamente $x_{M_0} = 39,165$. Esta moda foi calculada usando a *expressão (3.2)* no programa de Computação Algébrico MAPLE.

3.2 Estimador de Máxima Verossimilhança (EMV)

Sem considerar qualquer opinião *a priori*, uma abordagem comum para estimar θ é o método de máxima verossimilhança. Sendo X_1, X_2, \dots, X_n uma amostra aleatória da distribuição Binomial Negativa de parâmetros r e θ . A função de *log-verossimilhança* da distribuição Binomial Negativa é

$$\log L(\theta|X) = \log \prod_{i=1}^n \binom{x_i + r - 1}{x_i} + nr \log(\theta) + \log(1-\theta) \sum_{i=1}^n x_i.$$

Vamos escrever $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ ao longo da dissertação. Para maximizar essa *log-verossimilhança*, vamos calcular a primeira derivada e depois igualar a zero, isto é:

$$\frac{nr}{\theta} - \frac{\sum x_i}{1 - \theta} = 0$$

logo, temos que

$$\hat{\theta}_{EMV} = \frac{r}{r + \bar{x}} \quad (3.3)$$

3.3 Distribuições *a priori* e distribuições *a posteriori*

A distribuição Beta é a distribuição *a priori* conjugada para a probabilidade de sucesso da distribuição Binomial Negativa, pois o núcleo da distribuição Binomial Negativa é proporcional ao núcleo da distribuição Beta com parâmetros a e b , isto é:

$$p(\theta) = \theta^{a-1}(1 - \theta)^{b-1}$$

Assim, a distribuição *a posteriori* também será uma distribuição Beta com parâmetro a e b .

Considerando a função de verossimilhança $L(\theta|x) = p(x|\theta)$, a densidade *a posteriori* $p(\theta|x)$ é obtida pelo teorema de Bayes, isto é,

$$\begin{aligned} p(\theta|x) &= \frac{L(\theta|x)p(\theta)}{\int_{\Theta} L(\theta|x)p(\theta)d\theta} \\ &\propto L(\theta|x)p(\theta) \\ &\propto [\theta^r(1 - \theta)^x][\theta^{a-1}(1 - \theta)^{b-1}] \\ &\propto \theta^{r+a-1}(1 - \theta)^{x+b-1} \end{aligned}$$

segue uma densidade da distribuição *Beta*($r + a, x + b$)

Daí, temos que a distribuição preditiva $p(x)$ é dado por:

$$\begin{aligned} p(x) &= \binom{x+r-1}{x} \frac{1}{\beta(a,b)} \int_0^1 \theta^{r+a-1}(1 - \theta)^{x+b-1} d\theta \\ &= \binom{x+r-1}{x} \frac{\beta(r+a, x+b)}{\beta(a,b)} \end{aligned}$$

A partir deste resultado, podemos observar que $p(x)$ segue uma distribuição Beta Binomial Negativa, que mostraremos com mais detalhe na seção 3.4.

Vamos obter alguns estimadores pontuais para a probabilidade de sucesso da distribuição Binomial Negativa, isto é:

* Média *a posteriori*:

$$\begin{aligned}
 \mathbb{E}(\theta|X) &= \int_{\Theta} \theta p(\theta|x) d\theta \\
 &= \int_0^1 \theta \left[\frac{\theta^{r+a-1} (1-\theta)^{x+b-1}}{\beta(r+a, x+b)} \right] d\theta \\
 &= \frac{1}{\beta(r+a, x+b)} \int_0^1 [\theta^{r+a} (1-\theta)^{x+b-1}] d\theta \\
 &= \frac{\beta(r+a+1, x+b)}{\beta(r+a, x+b)} \\
 &= \frac{\Gamma(x+r+a+b)\Gamma(r+a+1)\Gamma(x+b)}{\Gamma(r+a)\Gamma(x+b)\Gamma(x+r+a+b+1)} \\
 &= \frac{(x+r+a+b-1)!(r+a)!}{(r+a-1)!(x+r+a+b)!} \\
 &= \frac{r+a}{x+r+a+b}
 \end{aligned}$$

* Moda *a posteriori*:

$$\begin{aligned}
 \mathbb{M}(\theta|X) &= \max_{\theta \in \Theta} \{L(\theta|x)p(\theta)\} \\
 &= \max_{\theta \in (0,1)} \{[\theta^r (1-\theta)^x][\theta^{a-1} (1-\theta)^{b-1}]\} \\
 &= \max_{\theta \in (0,1)} \{\theta^{r+a-1} (1-\theta)^{x+b-1}\} \\
 &= \frac{r+a-1}{x+r+a+b-2}
 \end{aligned}$$

* Variância *a posteriori*:

$$\mathbb{V}(\theta|X) = \mathbb{E}(\theta|X)^2 - [\mathbb{E}(\theta|X)]^2$$

Vamos primeiro calcular o valor de $\mathbb{E}(\theta|X)^2$, isto é:

$$\begin{aligned}\mathbb{E}(\theta|X)^2 &= \int_0^1 \theta^2 \left[\frac{\theta^{r+a-1}(1-\theta)^{x+b-1}}{\beta(r+a, x+b)} \right] d\theta \\ &= \frac{\beta(r+a+2, x+b)}{\beta(r+a, x+b)} \\ &= \frac{\Gamma(x+r+a+b)\Gamma(r+a+2)\Gamma(x+b)}{\Gamma(r+a)\Gamma(x+b)\Gamma(x+r+a+b+2)} \\ &= \frac{(r+a+1)(r+a)}{(x+r+a+b+1)(x+r+a+b)}\end{aligned}$$

Assim, temos que

$$\begin{aligned}\mathbb{V}(\theta|X) &= \frac{(r+a+1)(r+a)}{(x+r+a+b+1)(x+r+a+b)} - \frac{(r+a)^2}{(x+r+a+b)^2} \\ &= \frac{(x+r+a+b)(r+a+1)(r+a) - (r+a)^2(x+r+a+b+1)}{(x+r+a+b+1)(x+r+a+b)^2} \\ &= \frac{(r+a)(x+b)}{(x+r+a+b+1)(x+r+a+b)^2}\end{aligned}$$

Podemos observar facilmente que a esperança dos estimadores média e moda da distribuição *a posteriori* de θ são diferentes de θ , ou seja, eles são viciados.

3.3.1 Distribuição *a priori* de Bayes - Laplace

A distribuição *a priori* de Bayes - Laplace foi inicialmente proposta como uma distribuição *a priori* não informativa para a probabilidade de sucesso do modelo Binomial. Esta distribuição pode ser escrita em função da distribuição Beta com parâmetros ($a = 1$; $b = 1$) e é constante em relação ao espaço paramétrico (*Figura 9*).

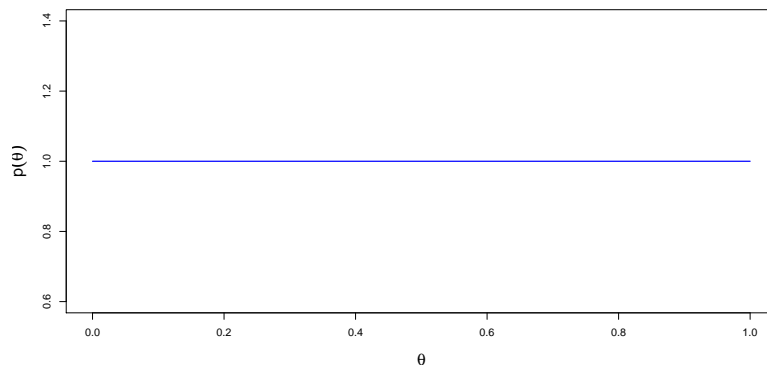


Figura 9: Distribuição *a priori* de Bayes - Laplace para a probabilidade de sucesso do modelo Binomial Negativa

Na *figura 9* pode-se ver que a distribuição *a priori* de Bayes - Laplace para a probabilidade de sucesso do modelo Binomial Negativa segue uma distribuição uniforme, como foi dita na *subsecção 2.6.1*, e logo não tem diferença em relação ao modelo Binomial.

Com a distribuição *a priori* $p(\theta) \propto 1$, a distribuição *a posteriori* é proporcional à função de verossimilhança do modelo Binomial Negativa. A distribuição *a posteriori*, pelo teorema de Bayes é dado por

$$\begin{aligned} p(\theta|x) &\propto L(\theta|x) \cdot p(\theta) \\ &\propto [\theta^r(1-\theta)^x] \cdot [\theta^{1-1}(1-\theta)^{1-1}] \\ &\propto \theta^r(1-\theta)^x, \quad 0 < \theta < 1 \end{aligned}$$

A distribuição *a posteriori* $\theta|x \sim \text{Beta}(r+1, x+1)$, utilizando a distribuição *a priori* de Bayes - Laplace fornece os estimadores bayesiano:

- Média *a posteriori*

$$\mathbb{E}(\theta|X) = \frac{r+1}{x+r+2}; \quad (3.4)$$

- Moda *a posteriori*

$$\mathbb{M}(\theta|X) = \frac{r}{x+r}; \quad (3.5)$$

- Variância *a posteriori*

$$\mathbb{V}(\theta|X) = \frac{(r+1)(x+1)}{(x+r+3)(x+r+2)^2}.$$

A moda desta distribuição *a posteriori*, ao considerar a distribuição *a priori* Uniforme, coincide com o estimador de máxima verossimilhança para o parâmetro θ da distribuição Binomial Negativa.

3.3.2 Distribuição *a priori* de Jeffreys

O geofísico Harold Jeffreys, propôs uma distribuição *a priori* não informativa para a probabilidade de sucesso do modelo Binomial Negativa. Esta distribuição *a priori* é proporcional à raiz quadrada da informação de Fisher para o parâmetro de interesse, conforme visto na *subsecção 2.6.2*.

Assim, a distribuição *a priori* de Jeffreys é calculado da seguinte forma: se x_1, x_2, \dots, x_n segue uma distribuição Binomial Negativa de parâmetros r e θ , então o logaritmo da função

de probabilidade conjunta é dado por

$$\log p(x|\theta) = \log \left[\prod_{i=1}^n \binom{x_i + r - 1}{x_i} \right] + nr \log \theta + \sum_{i=1}^n x_i \cdot \log (1 - \theta)$$

Fazendo a primeira derivada, teremos que

$$\frac{\partial \log p(x|\theta)}{\partial \theta} = \frac{nr}{\theta} - \frac{1}{(1-\theta)} \sum_{i=1}^n x_i$$

e tomando a segunda derivada, segue que

$$\frac{\partial^2 \log p(x|\theta)}{\partial \theta^2} = -\frac{nr}{\theta^2} - \frac{1}{(1-\theta)^2} \sum_{i=1}^n x_i$$

De acordo com a *expressão (2.9)*, chegamos a seguinte informação de Fisher

$$\begin{aligned} I(\theta) &= \mathbb{E} \left[\frac{nr}{\theta^2} + \frac{1}{(1-\theta)^2} \sum_{i=1}^n x_i \right] \\ &= \frac{nr}{\theta^2} + \frac{1}{(1-\theta)^2} \sum_{i=1}^n \mathbb{E}(x_i) \\ &= \frac{nr}{\theta^2} + \frac{nr}{\theta(1-\theta)} \\ &= \frac{nr}{\theta^2(1-\theta)} \propto \frac{1}{\theta^2(1-\theta)} \end{aligned}$$

Portanto, a *a priori* não informativa de Jeffreys para θ é na distribuição Binomial Negativa é

$$p(\theta) \propto \theta^{-1}(1-\theta)^{-\frac{1}{2}}$$

Lembramos que a distribuição *a priori* de referência neste caso coincide com a de Jeffreys.

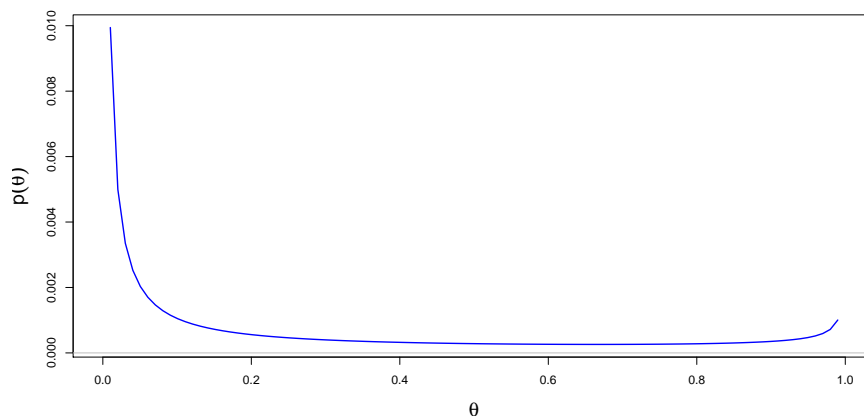


Figura 10: Distribuição *a priori* de Jeffreys para a probabilidade de sucesso do modelo Binomial Negativa

Na *figura 10* pode-se ver que a distribuição *a priori* de Jeffreys para a probabilidade de sucesso do modelo Binomial Negativa não é simétrica, diferente do que ocorre com o modelo Binomial.

Apesar da distribuição *a priori* de Jeffreys para a probabilidade de sucesso não ser própria, a distribuição *a posteriori* será própria. Utilizando a função de verossimilhança do modelo Binomial Negativa apresentada na *seção 3.3*, tem-se a distribuição *a posteriori* para $\theta|x$:

$$\begin{aligned} p(\theta|x) &\propto L(\theta|x) \cdot p(\theta) \\ &\propto [\theta^r(1-\theta)^x] \cdot [\theta^{-1}(1-\theta)^{-\frac{1}{2}}] \\ &\propto \theta^{r-1}(1-\theta)^{x-\frac{1}{2}} \quad 0 < \theta < 1 \end{aligned}$$

A distribuição *a posteriori* $\theta|x \sim \text{Beta}\left(r, x + \frac{1}{2}\right)$, utilizando a distribuição *a priori* de Jeffreys fornece os estimadores bayesiano:

- Média *a posteriori*

$$\mathbb{E}(\theta|X) = \frac{r}{x + r + \frac{1}{2}}; \quad (3.6)$$

- Moda *a posteriori*

$$\mathbb{M}(\theta|X) = \frac{r-1}{x + r - \frac{1}{2}}; \quad (3.7)$$

- Variância *a posteriori*

$$\mathbb{V}(\theta|X) = \frac{r \left(x + \frac{1}{2}\right)}{\left(x + r + \frac{3}{2}\right) \left(x + r + \frac{1}{2}\right)^2}.$$

3.3.3 Uma possível distribuição *a priori* $\text{Beta}(0, 0)$

O geneticista Haldane, em 1948, propôs uma distribuição *a priori* $\text{Beta}(0, 0)$ imprópria para o modelo Binomial. Vamos primeiro calcular esta distribuição baseado no modelo Binomial e depois vamos calcular para o modelo Binomial Negativa com intuito de compararmos os resultados.

Considerando uma variável aleatória X tendo distribuição Binomial com parâmetros n

e θ , denotando-se $X \sim \text{Bin}(n, \theta)$, se sua função de probabilidade é dada por

$$p(x|n, \theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}, \quad x = 0, 1, 2, \dots, n$$

para $n \geq 0$ e $0 < \theta < 1$.

Então, podemos calcular *a priori* de Haldane da seguinte forma:

$$\begin{aligned} p(\theta) &= - \sum_{i=0}^n p(x|n, \theta) \frac{\partial^2}{\partial \theta^2} \log p(x|n, \theta) \\ &= - \sum_{i=0}^n \left[\binom{n}{x} \theta^x (1 - \theta)^{n-x} \right] \left[\frac{\partial^2}{\partial \theta^2} \left(\log \prod_{i=0}^n \binom{n}{x} + x \log \theta + (n - x) \log (1 - \theta) \right) \right] \\ &= - \sum_{i=0}^n \left[\binom{n}{x} \theta^x (1 - \theta)^{n-x} \right] \left[\frac{\partial}{\partial \theta} \left(\frac{x}{\theta} - \frac{n - x}{(1 - \theta)} \right) \right] \\ &= - \sum_{i=0}^n \left[\binom{n}{x} \theta^x (1 - \theta)^{n-x} \right] \left[-\frac{x}{\theta^2} - \frac{n - x}{(1 - \theta)^2} \right] \\ &= \sum_{i=0}^n \left[\frac{\Gamma(n + 1)}{\Gamma(x + 1) \Gamma(n - x + 1)} \cdot \theta^{x-2} (1 - \theta)^{n-x-2} \right] \cdot [x - 2x\theta + n\theta^2] \end{aligned}$$

De acordo com programa de Computação Algébrico MAPLE, encontramos o seguinte resultado:

$$\begin{aligned} p(\theta) &= -\frac{2n\theta^2 - n\theta - n\theta^2}{\theta^2(1 - 2\theta + \theta^2)} \\ &= \frac{n\theta(1 - \theta)}{\theta^2(1 - \theta)^2} \\ &= n\theta^{-1}(1 - \theta)^{-1} \\ &\propto \theta^{-1}(1 - \theta)^{-1} \end{aligned}$$

Podemos ver este resultado também em (Zellner, 1996)

Considere agora, a variável aleatória X tendo distribuição Binomial Negativa com parâmetros r e θ . Vamos fazer os mesmos cálculos feitos anteriormente, isto é,

$$\begin{aligned}
p(\theta) &= - \sum_{i=0}^{\infty} p(x|r, \theta) \frac{\partial^2}{\partial \theta^2} \log p(x|r, \theta) \\
&= - \sum_{i=0}^{\infty} \left[\binom{x+r-1}{x} \theta^r (1-\theta)^x \right] \frac{\partial^2}{\partial \theta^2} \left(\log \prod_{i=1}^{\infty} \binom{x+r-1}{x} + r \log \theta + x \log (1-\theta) \right) \\
&= - \sum_{i=0}^{\infty} \left[\binom{x+r-1}{x} \theta^r (1-\theta)^x \right] \left[\frac{\partial}{\partial \theta} \left(\frac{r}{\theta} - \frac{x}{(1-\theta)} \right) \right] \\
&= - \sum_{i=0}^{\infty} \left[\binom{x+r-1}{x} \theta^r (1-\theta)^x \right] \left[-\frac{r}{\theta^2} - \frac{x}{(1-\theta)^2} \right] \\
&= \sum_{i=0}^{\infty} \left[\frac{\Gamma(x+r)}{\Gamma(x+1)\Gamma(r)} \cdot \theta^{r-2} (1-\theta)^{x-2} \right] \cdot [r(1-\theta)^2 + x\theta^2] \\
&= \sum_{i=0}^n \left[\frac{\Gamma(x+r)}{\Gamma(x+1)\Gamma(r)} \cdot (r\theta^{r-2}(1-\theta)^x + x\theta^r(1-\theta)^{x-2}) \right]
\end{aligned}$$

De acordo com programa de Computação Algébrico MAPLE, encontramos o seguinte resultado:

$$\begin{aligned}
p(\theta) &= -\frac{\Gamma(r+1)}{(-1+\theta)\Gamma(r)} \\
&= \frac{r}{\theta^2(1-\theta)} \\
&= r\theta^{-2}(1-\theta)^{-1} \\
&\propto \theta^{-2}(1-\theta)^{-1}
\end{aligned}$$

Observe que o resultado encontrado não corresponde a uma distribuição $Beta(0,0)$ como foi encontrado para o modelo Binomial, mas comparando os resultados se pode notar uma aproximação que tal vez seja significativa. Daí pergunto, será que a distribuição $Beta(0,0)$ (*Figura 11*) pode ser uma *a priori* não informativa para o modelo Binomial Negativa? A resposta para esta pergunta vai ser respondida no decorrer deste trabalho. Talvez os resultados que serão apresentados para propor esta *a priori* não informativa não seja respondido com tanto rigor matemático, mas tem lógica o que está sendo proposto.

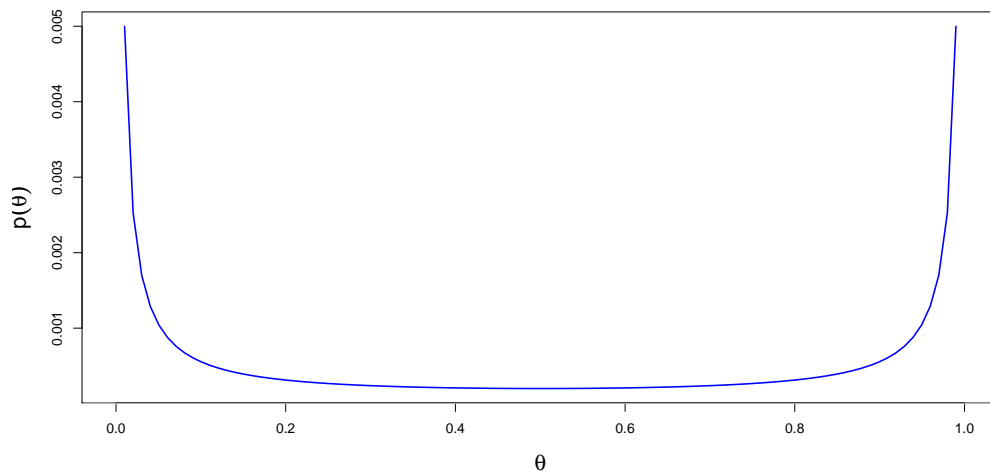


Figura 11: A possível distribuição *a priori* $Beta(0;0)$ para a probabilidade de sucesso do modelo Binomial Negativa

Suponha que a distribuição $Beta(0,0)$ seja uma *a priori* não informativa para o modelo Binomial Negativa, então a distribuição *a posteriori* para $\theta|x$ é dado por:

$$\begin{aligned} p(\theta|x) &\propto L(\theta|x)p(\theta) \\ &\propto [\theta^r(1-\theta)^x][\theta^{-1}(1-\theta)^{-1}] \\ &\propto \theta^{r-1}(1-\theta)^{x-1} \quad 0 < \theta < 1 \end{aligned}$$

A distribuição *a posteriori* $\theta|x \sim Beta(r,x)$, utilizando a possível distribuição *a priori* $Beta(0,0)$ fornece os seguintes estimadores bayesiano:

- Média *a posteriori*

$$\mathbb{E}(\theta|X) = \frac{r}{x+r}; \quad (3.8)$$

- Moda *a posteriori*

$$\mathbb{M}(\theta|X) = \frac{r-1}{x+r-2}; \quad (3.9)$$

- Variância *a posteriori*

$$\mathbb{V}(\theta|X) = \frac{rx}{(x+r+1)(x+r)^2}.$$

A média *a posteriori*, ao considerar a possível distribuição *a priori* $Beta(0,0)$, coincide com o estimador de máxima verossimilhança para o parâmetro θ da distribuição Binomial Negativa.

3.4 Distribuição Beta Binomial Negativa

Considere uma sequência de ensaios de Bernoulli realizada de forma independente com probabilidade θ constante e defina X como o número de fracassos anteriores ao r -ésimo sucesso. A variável aleatória X segue uma distribuição Binomial Negativo com parâmetros r e θ , em que $0 < \theta < 1$ e $r > 0$, e tem função densidade de probabilidade dada por:

$$p(x|r, \theta) = \binom{x+r-1}{x} \theta^r (1-\theta)^x, \quad x = 0, 1, 2, 3, \dots$$

Uma forma de contornar o problema da variabilidade de θ é supor que θ varia segundo uma distribuição Beta, isto é,

$$p(\theta) = \frac{1}{\beta(a, b)} \theta^{a-1} (1-\theta)^{b-1} \quad 0 < \theta < 1.$$

Neste caso, obtém-se como mistura a distribuição discreta Beta Binomial Negativa (BBN) dada por

$$\begin{aligned} p(x|a, b, r, \theta) &= \int_0^1 p(x|r, \theta) p(\theta) d\theta \\ &= \int_0^1 \left[\binom{x+r-1}{x} \theta^r (1-\theta)^x \right] \left[\frac{1}{\beta(a, b)} \theta^{a-1} (1-\theta)^{b-1} \right] d\theta \\ &= \binom{x+r-1}{x} \frac{1}{\beta(a, b)} \int_0^1 \theta^{r+a-1} (1-\theta)^{x+b-1} d\theta \\ &= \binom{x+r-1}{x} \frac{\beta(r+a, x+b)}{\beta(a, b)} \quad x = 0, 1, 2, 3, \dots \end{aligned}$$

Utilizando-se a função Beta em termos da função Gama, isto é,

$$\beta(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \quad a > 0 \quad e \quad b > 0,$$

a expressão acima pode ser expressa em termos da função Gama, ou seja,

$$p(x|a, b, r, \theta) = \binom{x+r-1}{x} \frac{\Gamma(r+a)\Gamma(x+b)\Gamma(a+b)}{\Gamma(x+r+a+b)\Gamma(a)\Gamma(b)}$$

Deste modo, podemos definir esta distribuição como segue.

A distribuição Beta Binomial Negativa é uma distribuição Binomial Negativa, cuja probabilidade de sucesso do parâmetro θ segue uma distribuição Beta com parâmetros a e b . Em outras palavras, se a probabilidade de sucesso do parâmetro θ de uma distribuição

Binomial Negativa (com parâmetros r e θ) tem uma distribuição Beta com parâmetros a e b , então a distribuição resultante é referido como a distribuição Beta Binomial Negativa com parâmetros a, b, r e θ , denotada por $BBN(a, b, r, \theta)$. Para uma distribuição Binomial Negativa Padrão, θ é geralmente assumido como fixo para os ensaios sucessivos, mas o valor de θ muda para cada ensaio, para a distribuição Beta Binomial Negativa. A distribuição Beta Binomial Negativa é muitas vezes referida como uma distribuição de Markov Pólya Inversa.

Considere uma variável aleatória X tendo uma distribuição Beta Binomial Negativo. Então, para $x = 0, 1, 2, 3, \dots$, esta distribuição de probabilidade (*Johnson et al. 2005 e Teerapabolarn, 2008*) é definida por

$$p(x|a, b, r, \theta) = \binom{x+r-1}{x} \frac{\Gamma(r+a)\Gamma(x+b)\Gamma(a+b)}{\Gamma(x+r+a+b)\Gamma(a)\Gamma(b)} \quad (3.10)$$

Para esta distribuição podemos observar as seguintes fatos:

- No caso de $r = 1$, a distribuição Beta Binomial Negativa é a distribuição Beta Geométrica com parâmetros a, b e θ ;
- Se $b = \alpha$ e $a = c - \alpha$, então a distribuição é denominada distribuição Generalizada Waring (*Johnson et al. 2005 e Teerapabolarn, 2008*).

A expressão (3.10) pode ser representada também pela seguinte forma:

$$p(x|a, b, r, \theta) = \frac{r}{x+r} \frac{\binom{x+b-1}{x} \binom{r+a-1}{r}}{\binom{x+r+a+b-1}{x+r}} \quad (3.11)$$

Essa representação, que, tanto quanto sabemos, não tem aparecido antes na literatura, dá uma visão interessante da Beta Binomial Negativa. Implica também que, para valores reais de $a > 0$ e $b > 0$, pode ser avaliada facilmente em qualquer pacote de software que fornecem comandos para combinação.

O valor esperado (ou esperança matemática) e a variância de uma variável aleatória $X \sim BBN(a, b, r, \theta)$ são respectivamente,

$$E(X) = \frac{rb}{a-1} \quad \text{Var}(X) = \frac{rb(r+a-1)(a+b-1)}{(a-2)(a-1)^2}$$

A unimodalidade da distribuição Beta Binomial Negativa é estudada por *Hassan & Bilal (2008)* e logo depois estudada por *Madeira (2009)*. Os autores justificam a unimodalidade

da distribuição com base nos resultados obtidos por *Holgate (1970)*. Em *Madeira (2009)*, tem o seguinte teorema: (*Hassan & Bilal, 2008*) a distribuição Beta Binomial Negativa é unimodal para todos os valores de (a, b, r) e a moda ocorre em $x = 0$ se $rb < 1$ e para $rb > 1$ a moda é algum outro ponto $x = x_M$ tal que

$$\frac{r(b-1) - (a+b)}{a+1} < x_M < \frac{(r-1)(b+1)}{a+1} \quad (3.12)$$

Agora, obtem-se uma melhor aproximação para a moda da distribuição Beta Binomial Negativa utilizar o fato de que a função gama generaliza o fatorial permitindo assim uma extensão contínua. Assim, podemos calcular a moda fazendo primeira derivada do logaritmo da expressão (3.10) e depois igualando a zero, isto é:

$$\frac{\partial}{\partial x} \log p(x|a, b, r, \theta) = \frac{\partial}{\partial x} \log \left(\frac{\Gamma(x+r)\Gamma(r+a)\Gamma(x+b)\Gamma(a+b)}{\Gamma(x+1)\Gamma(r)\Gamma(x+r+a+b)\Gamma(a)\Gamma(b)} \right).$$

De acordo com o programa de Computação Algébrico MAPLE, temos:

$$\Psi(x+r) + \Psi(x+b) - \Psi(x+1) - \Psi(x+r+a+b) = 0 \quad (3.13)$$

A construção dos gráficos abaixo foram baseados na distribuição Beta Binomial Negativa que corresponde a equação (3.11).

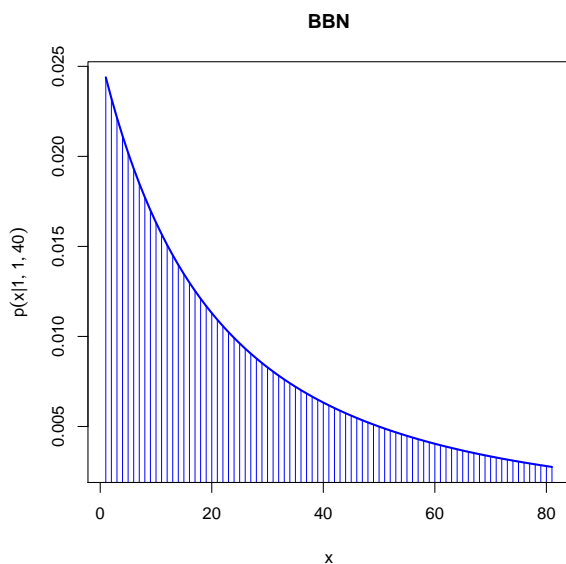


Figura 12: A densidade da distribuição Beta Binomial Negativa com $a = 1,0$; $b = 1,0$ e $r = 40$

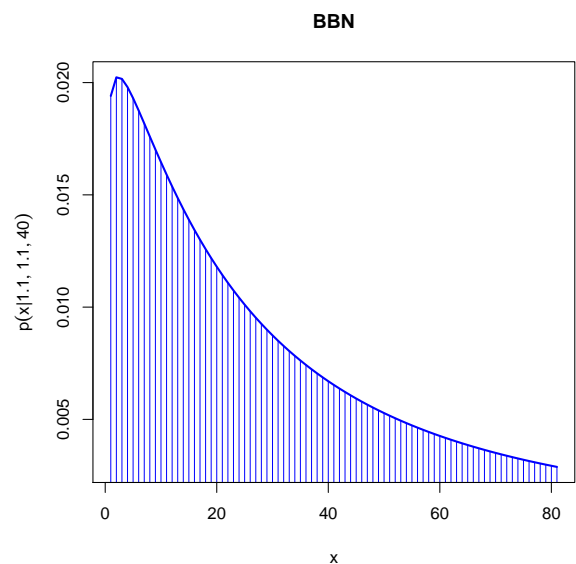


Figura 13: A densidade da distribuição Beta Binomial Negativa com $a = 1,1$; $b = 1,1$ e $r = 40$

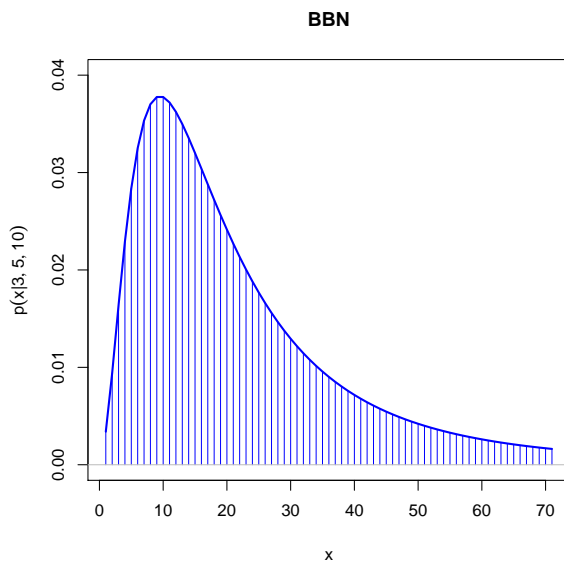


Figura 14: A densidade da distribuição Beta Binomial Negativa com $a = 3$, $b = 5$ e $r = 10$

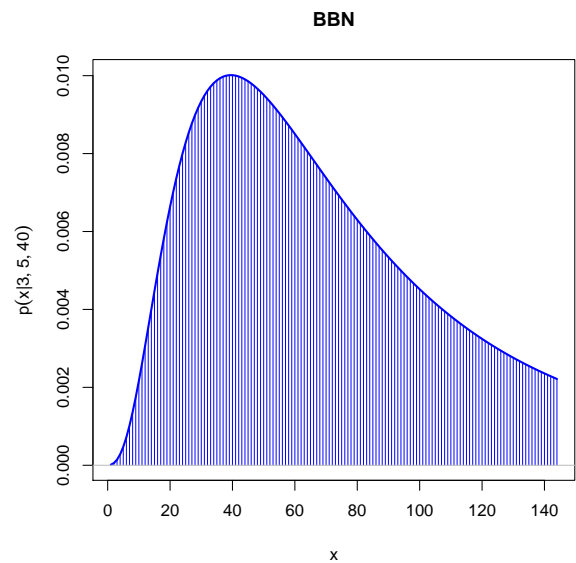


Figura 15: A densidade da distribuição Beta Binomial Negativa com $a = 3$, $b = 5$ e $r = 40$

Para a *Figura 12*, a moda calculada usando a expressão (3.13) é aproximadamente $-40,5$; para expressão (3.12), a moda está no intervalo aberto $(-1; 39)$ e observando na *Figura 12*, a moda se aproxima a zero. Daí, podemos concluir que a expressão (3.12) tem uma aproximação melhor, mas com um intervalo muito amplo.

Para *Figura 13*, a moda calculada usando a expressão (3.13) é aproximadamente $1,31$; para expressão (3.12), a moda está no intervalo aberto $(0, 86; 39)$ e observado na *Figura 13*, a moda se aproxima do valor encontrado na expressão (3.13). Daí, podemos concluir que a moda encontrada na expressão (3.13) é melhor do que a encontrada na expressão (3.12), apesar da moda esta num intervalo muito amplo.

Para a *Figura 14*, a moda calculada usando a expressão (3.13) é aproximadamente $8,49$; para expressão (3.12), a moda está no intervalo aberto $(8; 13,5)$. Neste caso, a expressão (3.12) tem uma boa previsão da moda, mas a expressão (3.13) também tem uma boa aproximação.

Para a *Figura 15*, a moda calculada usando a expressão (3.13) é aproximadamente $38,50$; para expressão (3.12), a moda está no intervalo aberto $(38; 58,5)$. Neste caso, a expressão (3.13) tem uma aproximação da moda melhor do que na *Figura 14*, mas a expressão (3.12) também tem uma boa aproximação, mesmo tendo um intervalo amplo.

Desta análise, podemos concluir dois fatos importantes:

- de acordo com as *Figuras 12 e 13*, podemos observar que a distribuição Beta Bino-

mial Negativa é unimodal somente para $a > 1$ e $b > 1$. Quando a e b se aproxima de 1, o valor de r tem que crescer para garantir a afirmação anterior.

- de acordo com as Figuras 14 e 15, podemos observar que o resultado da moda na expressão (3.13) se aproxima muito do resultado real da moda, quando o valor de r cresce.

3.5 Distribuições *a posteriori* preditiva para o modelo Binomial Negativa

Quando fazemos inferencia sobre dados futuros Y , com base em dados passados X e um modelo paramétrico Θ , estamos tratando de inferencia preditiva. Então, definimos a distribuição *a posteriori* preditiva da seguinte forma:

$$p(y|x) = \int p(y|\theta)p(\theta|x)d\theta,$$

que é uma ferramenta poderosa Bayesiana para efeitos de previsões.

De acordo com (Kinas, P. & Andrade, H. (2010)), a distribuição de probabilidade preditiva permite que se faça previsões em modelos de regressão e em estudos que envolvem um componente temporal. Porém, essa forma de inferência também é importante para estudar a validade dos modelos probabilísticos utilizados nas análises bayesiana. Como novos valores y devem ser observados com probabilidades $p(y|x)$ pode-se inferir quais são os valores prováveis. Se isso não se confirmar na prática, há indícios de que pelo menos um dos componentes do modelo (verossimilhança ou *a priori*) podem estar especificados inadequadamente.

Para distribuição Binomial Negativa a distribuição *a posteriori* preditiva descreve a probabilidade de y fracassos e s sucessos dados x fracassos e r sucessos. Ou seja,

$$\begin{aligned}
p(y|s, x, r) &= \int p(y|\theta)p(\theta|x)d\theta \\
&= \int_0^1 \left[\binom{y+s-1}{y} \theta^s (1-\theta)^y \right] \cdot \left[\frac{\theta^{r+a-1} (1-\theta)^{x+b-1}}{\beta(r+a; x+b)} \right] d\theta \\
&= \binom{y+s-1}{y} \frac{1}{\beta(r+a; x+b)} \int_0^1 \theta^{s+r+a-1} (1-\theta)^{y+x+b-1} d\theta \\
&= \binom{y+s-1}{y} \frac{\beta(s+r+a; y+x+b)}{\beta(r+a; x+b)} \\
&= \binom{y+s-1}{y} \frac{\Gamma(x+r+a+b)\Gamma(y+x+b)\Gamma(s+r+a)}{\Gamma(r+a)\Gamma(x+b)\Gamma(y+x+s+r+a+b)} \quad (3.14)
\end{aligned}$$

Observa-se que esta última equação segue uma distribuição Beta Binomial Negativa, isto é, segue a mesma distribuição correspondente a expressão (3.10). Agora, fazendo um re-arranjo na expressão (3.14), teremos:

$$\begin{aligned}
p(y|s, x, r) &= \frac{(y+s-1)!(x+r+a+b-1)!(y+x+b-1)!(s+r+a-1)!}{y!(s-1)!(r+a-1)!(x+b-1)!(y+x+s+r+a+b-1)!} \\
&= \frac{(y+x+b-1)!}{y!(x+b-1)!} \frac{(r+a)(s+r+a-1)!}{(s-1)!(r+a)!} \frac{(y+s-1)!(x+r+a+b)!}{(x+r+a+b)(y+x+s+r+a+b-1)!} \\
&= \frac{r+a}{x+r+a+b} \cdot \frac{\binom{y+x+b-1}{y} \binom{s+r+a-1}{s-1}}{\binom{y+x+s+r+a+b-1}{y+s-1}}
\end{aligned}$$

Sabemos que $m = y + s \Rightarrow s = m - y$ e $n = x + r \Rightarrow r = n - x$, onde m é o tamanho da amostra da variável Y e n é o tamanho da amostra da variável X . Logo,

$$p(y|s, x, r) = \frac{r+a}{n+a+b} \cdot \frac{\binom{y+x+b-1}{y} \binom{s+r+a-1}{m-1-y}}{\binom{m+n+a+b-1}{m-1}} \quad (3.15)$$

Daí, observa-se que a varável aleatória Y segue uma distribuição hipergeométrica negativa multiplicado por uma certa probabilidade (média *a posteriori* de X). Fazendo um re-arranjo na expressão (3.15), teremos:

$$\begin{aligned}
p(y|s, x, r) &= \frac{(x+b)(y+x+b)!}{(y+x+b)y!(x+b)!} \frac{s(r+a)(s+r+a)!}{(s+r+a)s!(r+a)!} \frac{(y+x+s+r+a+b)(y+s)!(x+r+a+b)!}{(y+s)(x+r+a+b)(y+x+s+r+a+b)!} \\
&= \frac{s(x+b)(r+a)(y+x+s+r+a+b)}{(y+x+b)(s+r+a)(y+s)(x+r+a+b)} \cdot \frac{\binom{y+x+b}{y} \binom{s+r+a}{s}}{\binom{y+x+s+r+a+b}{y+s}} \\
&= \frac{s(x+b)(r+a)(m+n+a+b)}{m(y+x+b)(s+r+a)(n+a+b)} \cdot \frac{\binom{y+x+b}{y} \binom{s+r+a}{s}}{\binom{m+n+a+b}{m}} \tag{3.16}
\end{aligned}$$

Da mesma forma, vamos calcular para distribuição binomial negativa a distribuição *a posteriori* preditiva que descreve a probabilidade de x fracassos e r sucessos dados y fracassos e s sucessos. Ou seja,

$$\begin{aligned}
p(x|r, y, s) &= \int p(x|\theta)p(\theta|y)d\theta \\
&= \int_0^1 \left[\binom{x+r-1}{x} \theta^r (1-\theta)^x \right] \cdot \left[\frac{\theta^{s+a-1} (1-\theta)^{y+b-1}}{\beta(s+a; y+b)} \right] d\theta \\
&= \binom{x+r-1}{x} \frac{1}{\beta(s+a; y+b)} \int_0^1 \theta^{s+r+a-1} (1-\theta)^{y+x+b-1} d\theta \\
&= \binom{x+r-1}{x} \frac{\beta(s+r+a; y+x+b)}{\beta(s+a; y+b)} \\
&= \binom{x+r-1}{x} \frac{\Gamma(y+s+a+b)\Gamma(y+x+b)\Gamma(s+r+a)}{\Gamma(s+a)\Gamma(y+b)\Gamma(y+x+s+r+a+b)} \tag{3.17}
\end{aligned}$$

Observa-se que esta última equação segue uma distribuição Beta Binomial Negativa, isto é, segue a mesma distribuição correspondente a expressão (3.10). Agora, fazendo um

re-arranjo na expressão (3.17), teremos:

$$\begin{aligned}
 p(x|r, y, s) &= \frac{(x+r-1)!(y+s+a+b-1)!(y+x+b-1)!(s+r+a-1)!}{x!(r-1)!(s+a-1)!(y+b-1)!(y+x+s+r+a+b-1)!} \\
 &= \frac{(y+x+b-1)!}{x!(y+b-1)!} \frac{(s+a)(s+r+a-1)!}{(r-1)!(s+a)!} \frac{(x+r-1)!(y+s+a+b)!}{(y+s+a+b)(y+x+s+r+a+b-1)!} \\
 &= \frac{s+a}{y+s+a+b} \cdot \frac{\binom{y+x+b-1}{x} \binom{s+r+a-1}{r-1}}{\binom{y+x+s+r+a+b-1}{x+r-1}}
 \end{aligned}$$

Sabemos que $m = y + s \Rightarrow s = m - y$ e $n = x + r \Rightarrow r = n - x$, onde m é o tamanho da amostra da variável Y e n é o tamanho da amostra da variável X . Logo,

$$p(x|r, y, s) = \frac{s+a}{m+a+b} \cdot \frac{\binom{y+x+b-1}{x} \binom{s+r+a-1}{n-1-x}}{\binom{m+n+a+b-1}{n-1}} \quad (3.18)$$

Daí, observa-se que a varável aleatória X segue uma distribuição hipergeométrica negativa multiplicado por uma certa probabilidade (média *a posteriori* de Y). Fazendo um re-arranjo na expressão (3.18), teremos:

$$\begin{aligned}
 p(x|r, y, s) &= \frac{(y+b)(y+x+b)!}{(y+x+b)x!(x+b)!} \frac{r(s+a)(s+r+a)!}{(s+r+a)r!(s+a)!} \frac{(y+x+s+r+a+b)(x+r)!(y+s+a+b)!}{(x+r)(y+s+a+b)(y+x+s+r+a+b)!} \\
 &= \frac{r(y+b)(s+a)(y+x+s+r+a+b)}{(y+x+b)(s+r+a)(x+r)(y+s+a+b)} \cdot \frac{\binom{y+x+b}{x} \binom{s+r+a}{r}}{\binom{y+x+s+r+a+b}{x+r}} \\
 &= \frac{r(y+b)(s+a)(m+n+a+b)}{n(y+x+b)(s+r+a)(m+a+b)} \cdot \frac{\binom{y+x+b}{x} \binom{s+r+a}{r}}{\binom{m+n+a+b}{n}} \quad (3.19)
 \end{aligned}$$

Vamos observar como se comporta a razão entre as distribuições *a posteriori* preditiva calculada anteriormente, isto é, as expressões correspondente a (3.16) e (3.19), quando aplicarmos as prioris de Bayes - Laplace $Beta(1, 1)$, a de Jeffreys $Beta(0, \frac{1}{2})$ e uma possível *a priori* $Beta(0, 0)$.

- Para *a priori* de Bayes - Laplace $Beta(1, 1)$, temos:

$$p(y|s, x, r) = \frac{s(x+1)(r+1)(m+n+2)}{m(y+x+1)(s+r+1)(n+2)} \frac{\binom{y+x+1}{y} \binom{s+r+1}{s}}{\binom{m+n+2}{m}}$$

$$p(x|r, y, s) = \frac{r(y+1)(s+1)(m+n+2)}{n(y+x+1)(s+r+1)(m+2)} \frac{\binom{y+x+1}{x} \binom{s+r+1}{r}}{\binom{m+n+2}{n}}$$

A razão entre as probabilidades acima é dada por:

$$\frac{p(y|s, x, r)}{p(x|r, y, s)} = \frac{sn(n+1)}{rm(m+1)}$$

- Para *a priori* de Jeffreys $Beta(0, \frac{1}{2})$, temos:

$$p(y|s, x, r) = \frac{sr(x+\frac{1}{2})(m+n+1)}{m(y+x+\frac{1}{2})(s+r)(n+1)} \frac{\binom{y+x+\frac{1}{2}}{y} \binom{s+r}{s}}{\binom{m+n+1}{m}}$$

$$p(x|r, y, s) = \frac{rs(y+\frac{1}{2})(m+n+1)}{n(y+x+\frac{1}{2})(s+r)(m+1)} \frac{\binom{y+x+\frac{1}{2}}{x} \binom{s+r}{r}}{\binom{m+n+1}{n}}$$

A razão entre as probabilidades acima é dada por:

$$\frac{p(y|s, x, r)}{p(x|r, y, s)} = \frac{x!(m-1)! \left(n - \frac{1}{2}\right)! \left(y - \frac{1}{2}\right)!}{y!(n-1)! \left(m - \frac{1}{2}\right)! \left(x - \frac{1}{2}\right)!}$$

- Para uma possível *a priori* $Beta(0, 0)$, temos:

$$p(y|s, x, r) = \frac{sxr(m+n)}{m(y+x)(s+r)n} \frac{\binom{y+x}{y} \binom{s+r}{s}}{\binom{m+n}{m}}$$

$$p(x|r, y, s) = \frac{rys(m+n)}{n(y+x)(s+r)m} \frac{\binom{y+x}{x} \binom{s+r}{r}}{\binom{m+n}{n}}$$

A razão entre as probabilidades acima é dada por:

$$\frac{p(y|s, x, r)}{p(x|r, y, s)} = \frac{x}{y}$$

Esta última expressão, ocorre pelo fato de que as variáveis X e Y serem simétricas.

4 Conclusões

4.1 Os Principais Argumentos

4.1.1 Primeiro Argumento

A *Tabela 1* e a *Tabela 2* mostram os resultados para os estimadores bayesiano pontual *a posteriori* para o parâmetro de interesse θ , juntamente com suas respectivas distribuições *a priori* (*a priori* de Bayes - Laplace, *a priori* de Jeffreys e uma possível *a priori* $Beta(0; 0)$) e as distribuições *a posteriori* resultantes. Como esperado, todas as distribuições *a priori* resultaram em uma distribuição *a posteriori* Beta, mas com formas diferentes e parâmetros escalares. Para construção destas tabelas foram propostas, uma amostra aleatória de tamanho $n = 10$ onde ocorreu exatamente $r = 3$ sucessos (*Tabela 1*) e outra amostra aleatória de tamanho $n = 100$, onde ocorreu exatamente $r = 30$ sucessos (*Tabela 2*).

Tabela 1: Estimadores pontuais bayesiano para o parâmetro θ utilizando *as prioris* $Beta(1; 1)$, $Beta(0; 0, 5)$ e $Beta(0; 0)$. Sabendo que $n = 10$ e $r = 3$.

<i>a priori</i>	$Beta(1; 1)$	$Beta(0; 0, 5)$	$Beta(0; 0)$
<i>a posteriori</i>	$Beta(4; 8)$	$Beta(3; 7, 5)$	$Beta(3; 7)$
média <i>a posteriori</i>	0,3333	0,2857	0,3000
moda <i>a posteriori</i>	0,3000	0,2353	0,2500
variância <i>a posteriori</i>	0,0171	0,0177	0,0191

Tabela 2: Estimadores pontuais bayesiano para o parâmetro θ utilizando *as prioris* $Beta(1; 1)$, $Beta(0; 0, 5)$ e $Beta(0; 0)$. Sabendo que $n = 100$ e $r = 30$.

<i>a priori</i>	$Beta(1; 1)$	$Beta(0; 0, 5)$	$Beta(0; 0)$
<i>a posteriori</i>	$Beta(31; 71)$	$Beta(30; 70, 5)$	$Beta(30; 70)$
média <i>a posteriori</i>	0,3039	0,2985	0,3000
moda <i>a posteriori</i>	0,3000	0,2944	0,2959
variância <i>a posteriori</i>	0,0021	0,0021	0,0021

Por questão de comparação dos estimadores bayesiano, vamos construir os seguintes gráficos:

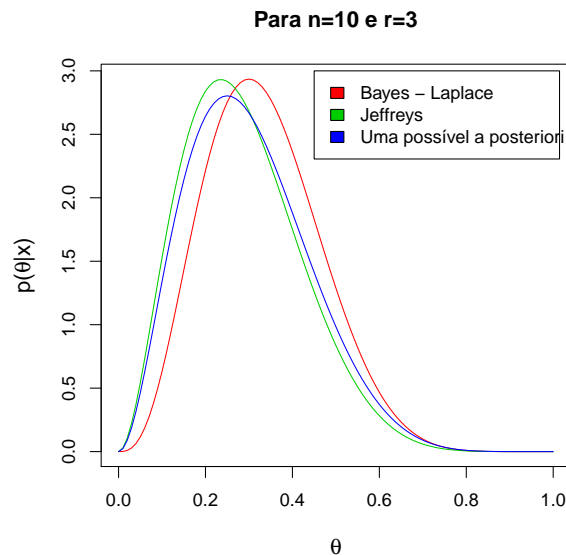


Figura 16: Gráfico das posteriores de Bayes - Laplace, de Jeffreys e uma possível a posteriori. Para $n = 10$ e $r = 3$.

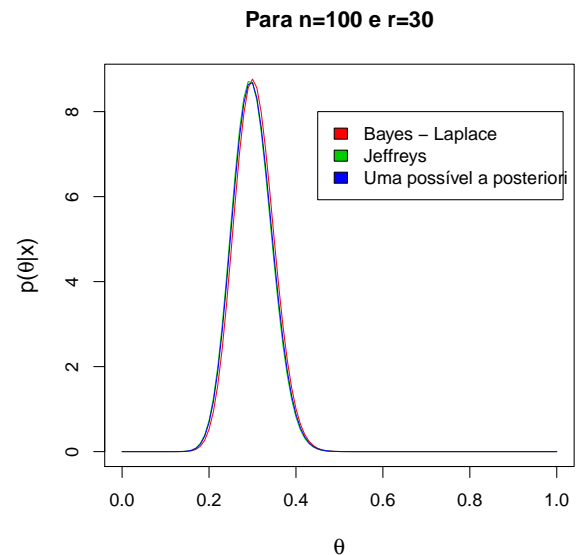


Figura 17: Gráfico das posteriores de Bayes - Laplace, de Jeffreys e uma possível a posteriori. Para $n = 100$ e $r = 30$.

Pela Tabela 1 e Figura 16, podemos observar que as três *prioris* (Bayes - Laplace, Jeffreys e uma possível a priori $Beta(0, 0)$) apresentaram variâncias a posteriori diferentes, sendo a priori de Bayes - Laplace a de menor dispersão, enquanto a possível a priori $Beta(0, 0)$ a de maior dispersão. Isto ocorre, pelo fato da média e da moda a posteriori em relação a Bayes - Laplace estarem mais próximas, isto é, está ocorrendo uma maior probabilidade a posteriori em torno destas medidas, que possivelmente represente bem a distribuição.

Agora, pela Tabela 2 e Figura 17, podemos observar que as três *prioris* em estudo apresentaram a mesma variância a posteriori, que foi de 0,0021. Isso, entretanto, torna impossível avaliar e comparar os resultados a partir das distribuições a priori com base na variância a posteriori como medida de dispersão. Timidamente, o fato da amostra em estudo não apresentar dispersão em relação a distribuição a posteriori pode ser explicado pelo número de ensaios em relação ao número de sucessos ocorridos nesse experimento.

Mesmo a priori de Bayes - Laplace está tendo um comportamento melhor em relação as outras a priori, podemos afirmar que a possível distribuição a priori $Beta(0, 0)$ tem o mesmo comportamento quando o número de ensaios for grande. Assim, para este tipo de análise a priori que estamos estudando é adequada.

4.1.2 Segundo Argumento

Como o estimador de máxima verossimilhança (EMV) não é afetado por qualquer opinião *a priori*, simplesmente perguntamos: será que existe uma distribuição *a priori*, que produziria uma estimativa Bayesiana (por exemplo, a média *a posteriori* ou moda *a posteriori*) que coincida com o estimador de máxima verossimilhança (EMV)? Em caso afirmativo, a distribuição *a priori* poderia ser considerada como não informativa, desde que a opinião *a priori* exerça qualquer influência sobre qualquer estimativa pontual Bayesiana (Mu Zhu & Arthur Y. Lu, 2004).

Usando essa heurística simples, podemos observar o seguinte comportamento:

- em relação a moda *a posteriori*, a *a priori* de Bayes - Laplace é realmente menos informativa do que a *a priori* de Jeffreys, pois a moda *a posteriori* (veja a expressão (3.5)) em relação a Bayes - Laplace coincide com o (EMV) (veja a expressão (3.3));
- em relação a média *a posteriori*, a *a priori* de Jeffreys é menos influente do que a *a priori* de Bayes - Laplace, pois a média *a posteriori* (veja a expressão (3.6)) em relação a *a priori* Jeffreys está mais próximo do (EMV) (veja a expressão (3.3)).

Levando em conta que a média *a posteriori* tem uma maior aplicabilidade nos trabalhos científicos, por ser uma medida compreendida entre as variáveis de maior probabilidade, a afirmativa a seguir se torna ainda mais forte:

- em relação a média *a posteriori*, a distribuição $Beta(0,0)$ (que é uma distribuição imprópria) é menos informativa do que a *a priori* de Bayes - Laplace e do que a *a priori* de Jeffreys, pois a média *a posteriori* (veja a expressão (3.8)) em relação a $Beta(0,0)$ coincide com o (EMV) (veja a expressão (3.3)).

A partir daí, considera-se que a menos informativa *a priori* é, surpreendentemente, uma distribuição extremamente “opinativa” abordando duas massas pontuais em 0 e 1, ou seja, uma distribuição *a priori* $Beta(\varepsilon, \varepsilon)$ com $\varepsilon > 0$ e ε pequeno aborda duas massa pontuais em 0 e 1. Tal crença *a priori* parece ser extremamente forte, pois diz que o parâmetro θ é essencialmente 0 ou 1. Com esta intuição, pode-se considerar tal crença *a priori* forte ser extremamente irrazoável, mas esta é *a priori* que resultaria em uma média *a posteriori* mais próximo possível de EMV , como foi visto anteriormente. Assim, a *a priori* $Beta(\varepsilon, \varepsilon)$ com $\varepsilon > 0$ e pequeno, realmente pode ser considerado como a menos influente *a priori* neste modelo.

Ressaltamos aqui que não é nossa intenção dar a entender que a nossa heurística simples é o melhor ou mesmo um ponto de vista adequado para avaliar a não informatividade das diferentes priores na análise bayesiana. E nem, afirmar diretamente que a distribuição $Beta(0, 0)$ é uma distribuição *a priori* não informativa para o parâmetro de sucesso, θ , da distribuição Binomial Negativa, mas esta distribuição pode ser uma possível *a priori* não informativa para este modelo.

De acordo com (Bernardo, 1997), as priores não subjetivas adequadas são geralmente encontradas sempre que o espaço paramétrico é limitado, embora isto não seja uma regra geral; *a priori* não subjetiva tipicamente recomendada para o parâmetro θ de uma distribuição Binomial Negativa é $Beta(0, \frac{1}{2})$, que é imprópria, embora o espaço de parâmetros seja limitado.

Ainda de acordo com ele, pode-se trabalhar sempre com a *priori* não subjetiva adequada se o espaço paramétrico for levado para ser escolhido em conjunto limitado. Por exemplo, a *priori* padrão não subjetiva de parâmetro pertencente aos reais \mathbb{R} é uniforme em \mathbb{R} , que é imprópria. No entanto, se for dados algumas medidas experimentais, inferências são feitas sobre o verdadeiro valor de algumas quantidades físicas, mas o espaço paramétrico realista seria $[0, c]$, para alguma constante c . Consequentemente, a *priori* não subjetiva de parâmetro uniforme em $[0, c]$, seria de fato adequada. Da mesma forma, no cenário Binomial Negativa, onde a probabilidade de sucesso deve ser estritamente positiva, um espaço paramétrico na forma $[\varepsilon, \varepsilon]$, para algum pequeno $\varepsilon > 0$ levaria a *priori* não subjetiva adequada. Este fato fica mais forte pelo teorema que diz (Mu Zhu & Arthur Y. Lu, 2004): a limitação de distribuição $Beta(0, 0)$ consiste de duas massas pontuais iguais a 0 e 1.

4.1.3 Terceiro Argumento

Em relação à Binomial Negativa a distribuição *a posteriori* preditiva (veja a seção 2.5) descreve a probabilidade de y fracassos e s sucessos dados x fracassos e r sucessos, onde ocorre $m = y + s$ ensaios para a variável Y e $n = x + r$ ensaios para variável X . Ou seja,

$$p(y|s, x, r) = \binom{y+s-1}{y} \frac{\Gamma(x+r+a+b)\Gamma(y+x+b)\Gamma(s+r+a)}{\Gamma(r+a)\Gamma(x+b)\Gamma(y+x+s+r+a+b)} \quad (\text{p.34 - 3.14}).$$

Observe que a *priori* generalizada $Beta(a; b)$ leva à distribuição Beta Binomial Negativa (Johnson et al. 2005; Teerapabolarn, 2008 e Madeira, 2009).

Usando fatoriais, em vez de funções gama, o pesquisador *Arthur Roger Thatcher* (em 1964), refere-se a uma variável aleatória com modelo Beta-Binomial como tendo uma distribuição Hipergeométrica Negativa, enquanto os pesquisadores *Frank Tuyl, Richard Gerlachy e Kerrie Mengersenz* (2009) em seu artigo, propôs um re-arranjo, obtendo um componente que chamou de hypergeometric-like.

Da mesma forma, observa-se que a variável aleatória y segue também uma distribuição hipergeométrica negativa multiplicado por uma certa probabilidade (média *a posteriori* de X). Fazendo um re-arranjo da expressão anterior, obtemos:

$$p(y|s, x, r) = \frac{r + a}{x + r + a + b} \frac{\binom{y + x + b - 1}{y} \binom{s + r + a - 1}{s - 1}}{\binom{y + x + s + r + a + b - 1}{y + s - 1}} \quad (\text{p.34 - 3.15})$$

Semelhante a expressão (2) apresentado por (Tuyl F., Gerlach R. e Mengersen K., 2009), a expressão acima pode ser vista como o produto das duas probabilidades seguinte:

$$p(y|s, x, r) = \frac{s(x + b)(r + a)(y + x + s + r + a + b)}{(y + x + b)(s + r + a)(y + s)(x + r + a + b)} \frac{\binom{y + x + b}{y} \binom{s + r + a}{s}}{\binom{y + x + s + r + a + b}{y + s}} \quad (\text{p.35 - 3.16})$$

Essa representação, que, tanto quanto sabemos, não tem aparecido antes na literatura, dá uma visão interessante da Beta Binomial Negativa. Implica também que, para valores inteiros de a e b , pode ser avaliada facilmente em qualquer pacote de software que fornecem a distribuição hipergeométrica.

O componente da hipergeométrica equivalente multiplicada por uma certa probabilidade, será assim que chamaremos a expressão (3.16), sugere uma possível distribuição *a priori* $Beta(0, 0)$, levando a simetria em y e x . Aplicando esta *a priori*, o multiplicador resultante será:

$$\frac{xsr(m + n)}{nm(y + x)(s + r)}$$

é razoável também: por exemplo, a

$$\frac{p(y|s, x, r)}{p(x|r, y, s)} = \frac{x}{y}, \quad (4.1)$$

que simplesmente reflete o fato de que, nos m e n ensaios se tem y e x fracassos possíveis, respectivamente. Esta proporção é 1, quando $x = y$, de modo que $p(y|s, x, r)$ é igual a $p(x|r, y, s)$, como ilustraremos adiante, isso é perfeitamente razoável, mas segue a partir de uma possível distribuição *a priori* $Beta(0, 0)$ somente.

Estamos apoiando este último resultado (expressão (4.1)), somente baseado no comportamento da possível *a priori* $Beta(0;0)$ que vem sendo analisada desde do princípio deste trabalho.

Para uma maior aceitação deste resultado, vamos analisar o comportamento da probabilidade *a posteriori* preditiva apresentadas nas expressões (3.16) e (3.19). Para isso, vamos utilizar as *priors* não informativas: *a priori* de Bayes - Laplace $Beta(1;1)$, *a priori* de Jeffreys $Beta(0; \frac{1}{2})$ e a possível *a priori* $Beta(0;0)$, com o intuito também de comparar essas *priors*.

Na Tabela 3, as probabilidades *a posteriori* preditiva (expressões (3.16) e (3.19)) induzidas pelas *priors* citadas anteriormente são comparadas por 11 cenários.

Tabela 3: Cenários *a posteriori* preditivos: $p(y|s, x, r)$

Caso	y	s	x	r	$Beta(1;1)$	$Beta(0; \frac{1}{2})$	$Beta(0;0)$
					Bayes-Laplace	Jeffreys	
1a	0	1	0	1	0,666667	0,666667	1,000000
1b	1	0	1	0	0,000000	0,000000	0,000000
2a	20	14	20	3	0,003008	0,001428	0,001589
2b	20	3	20	14	0,001389	0,001739	0,001589
3a	1	1	20	20	0,244186	0,243939	0,243902
3b	20	20	1	1	0,017867	0,012896	0,012195
3c	18	20	1	1	0,019700	0,013903	0,013495
4a	0	1	20	20	0,500000	0,493827	0,500000
4b	20	20	0	1	0,012195	0,008702	0,000000
5a	4	4	4	2	0,081585	0,063176	0,067988
5b	4	2	4	4	0,069930	0,073329	0,067988

Naturalmente, nenhuma afirmação geral pode ser feita sobre a correção das *priors* com base somente neste cenário, mas podemos fazer as seguintes observações:

- De acordo com o caso 1 e principalmente em 1a, fortalece ainda mais a afirmação feita em (4.1), pois os resultados obtidos pela possível *a priori* $Beta(0;0)$ são mais lógicos de acordo com a realidade do problema. Ou seja, no caso 1a queremos a probabilidade de um experimento aleatório ocorrer exatamente $s = 1$ sucesso entre $m = 1$ ensaio dado que já ocorreu um experimento aleatório com exatamente $r = 1$ sucesso entre $n = 1$ ensaio. Se observarmos logicamente, podemos dizer que está

probabilidade tende a ser 100%, como ocorre com a possível *a priori* $Beta(0; 0)$.

- Como esperado, as probabilidades em relação a possível *a priori* $Beta(0, 0)$ são as mesmas para os casos $2a$ e $2b$, e nós argumentamos que $s = 14$ sucessos fixos entre 34 ensaios dado $r = 3$ sucessos fixo entre 23 ensaios deveria ser realmente tão inesperado quanto $s = 3$ sucessos fixos entre 23 ensaios dado $r = 14$ sucessos fixo entre 34 ensaios. Chegamos a mesma conclusão nos casos $5a$ e $5b$. Podemos observar que a probabilidade em $2a$ é mais próxima da probabilidade de $2b$ na *priori* de Jeffreys do que na *priori* de Bayes - Laplace e, acontece o mesmo nos casos $5a$ e $5b$.
- Nota-se que os diferenciais da possível *a priori* $Beta(0, 0)$ entre a e b para os casos 3 e 4 são puramente devido a expressão (4.1). Como se esperava, este resultado causou uma diferença muito grande entre as $Beta(0, 0)$ nos cenários a e b . Esta diferença é menor no âmbito da *priori* de Bayes - Laplace, mas estes resultados são difíceis para serem justificados; por exemplo, no caso $3b$ a probabilidade de 0,017867; dado a evidência de apenas uma observação, é aproximadamente de $\frac{3}{2}$ vezes maior do que a probabilidade do caso $4b$ baseado em um fracasso em vez de um sucesso. Movendo-se do caso $3b$ para $3c$ há um acréscimo maior na probabilidade *a posteriori* preditiva em relação *a priori* de Bayes - Laplace do que na *priori* de Jeffreys, ao reduzir o número de fracassos na experiência futura de tamanho 20 para 18, na ausência de informação *a priori*, um pequeno acréscimo seria de ser esperar.

5 Considerações Finais

Neste trabalho foi proposto uma possível *a priori* $Beta(0,0)$ não informativa para a probabilidade de sucesso da distribuição Binomial Negativa. Para prôpor esta *a priori* não informativa foram observadas as seguintes características:

1. na *seção 2.8*, verificamos que a distribuição $Beta(a,b)$ é unimodal somente quando $a > 1$ e $b > 1$;
2. na *seção 3.1*, verificamos que a distribuição Binomial Negativa é unimodal;
3. verificamos que a moda *a posteriori*, ao considerar *a priori* de Bayes - Laplace, coincide com o estimador de máxima verossimilhança para o parâmetro θ da distribuição Binomial Negativa;
4. verificamos que a média *a posteriori*, ao considerar a possível *a priori* $Beta(0,0)$, coincide com o estimador de máxima verossimilhança para o parâmetro θ da distribuição Binomial Negativa;
5. na *seção 3.4*, verificamos que a distribuição Beta Binomial Negativa é unimodal quando $a > 1$ e $b > 1$. Quando a e b se aproxima de 1, o valor de r tem que crescer para garantir a afirmação anterior;
6. no segundo argumento verificamos que a média *a posteriori* é uma boa medida para distribuição $Beta(a,b)$, pois esta distribuição é limitada entre 0 e 1.
7. pelas afirmações anteriores podemos dizer que a distribuição $Beta(0,0)$ pode ser uma possível *a priori* não informativa para o parâmetro θ da distribuição Binomial Negativa;
8. quando aplicamos a distribuição $Beta(0,0)$ numa distribuição preditiva *a posteriori* para o modelo Binomial Negativa, verificamos um comportamento simétrico em relação aos números fracassos x e y . E de acordo com os resultados obtidos na *Tabela*

(3), podemos dizer que a distribuição $Beta(0, 0)$ pode ser uma possível *a priori* não informativa para o modelo Binomial Negativa.

A facilidade de utilização do Software Livre R propiciou grande facilidade para a execução deste trabalho e mostrou que cada vez mais existe uma tendência para a substituição do software proprietário pelo software livre. Esse software serviu para construção dos gráficos e tabelas deste trabalho.

Outro software que utilizamos neste trabalho para desenvolver expressões matemáticas foi o MAPLE, que é um programa de Computação Algébrico.

Referências

- Albert, J. (2009). *Bayesian Computation with R*. Springer, New York, second edition.
- Berger, J. and Bernardo, J. *Estimating a product of means: bayesian analysis with reference priors*. Journal of the American Statistical Association, v. 84, p. 200 - 207, 1989.
- Bernardo, J.M. and Smith, A.F.M. (1994), *Bayesian Theory*. New York: John Wiley.
- Bernardo, J. M. & Smith, A. F. M. (2000). *Bayesian Theory*. Wiley, Chichester.
- Bernardo, J. M. (1997). *Noninformative priors do not exist: a dialogue with Jose M. Bernardo*, Journal Statistics Planning and Inference 65, 159-189 (with discussion).
- Bernardo, J. M. (2005). *Reference Analysis*. Handbook of Statistics, pages 17 - 90.
- Bernardo, J. M. *Reference posterior distributions for bayesian inference*. Journal of the Royal Statistjcai Society, Serie B, v. 41, p. 113 - 147, 1979.
- Bernardo, J.M. (1989), [*The Geometry of Asymptotic Inference*]: *Comment: On Multivariate Jeffreys' Priors*. Statistical Science, Vol. 4, N^o 3, pages 227 - 229.
- Box, G. E. P. & Tiao, G. C. (1973), *Bayesian Inference in Statistical Analysis*, Addison-Wesley Publishing Co., Reading, Mass. London-Don Mills, Ont.. Addison-Wesley Series in Behavioral Science: Quantitative Methods.
- Casella, G., and Berger, R. L. (1994). *Estimation with selected binomial information or do you really believe that Winfield*. Dave is batting. 471. J. Am. Stat. Assoc. 89, 1080 - 1090.
- Casella, G. & Berger, R. L. (2002). *Statistical Inference*. Duxbury, Pacific Grove, second edition.
- DeGroot, M. H. (1970). *Optimal Statistical Decisions*. McGraw - Hill, New York.
- Ehlers, R. S. (2007). *Introdução à Inferência Bayesiana*. 5 ed., Texto publicado pela Universidade Federal do Paraná. Departamento de Estatística, UFPR.
- Feller, W. (1968). *In An Introduction to Probability Theory and Its Applications vol. I*. pp. 123-124, John Wiley & Sons, Inc., New York.
- Gamerman, D. and Lopes. H.F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. 2.ed. New York: Chapman and Hall/CRC.

Haldane, J. B. S. *On a Method of Estimating Frequencies*. Biometrika, Vol. 33, No. 3 (Nov., 1945), pp. 222 - 225.

Hassan, Anwar and Bilal, Sheikh *On Estimation of Negative Polya-Eggenberger Distribution and Its Applications*. J. Ksiam, vol. 12, N^o 2, 81 - 95, 2008.

Johnson, N. L., Kotz, S. and Adrienne W. K. *Univariate Discrete Distributions*. Third edition, Wiley, New York, 2005.

Kinas, Paul G. & Andrade, Humber A., *Intrdução à análise bayesiana (com R)*, ed. maisQnada, Porto Alegre - 2010.

Lindley, D. V. (1990). *The Present Position in Bayesian Statistics*. Statistical science 5, N^o 1, 44 - 89.

Madeira, A. P. Coelho. (2009). *A Distribuição beta binomial negativa*. 81 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, MG.

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

Ross, G. J. S. and Preence, D. A., *The Negative Binomial Distribution*, The Statistician, vol. 34, N^o 3 (1985), p. 323 - 335.

Teerapabolarn, K. (2008). *Poisson approximation to the beta-negative binomial distribution*. International Journal of Contemporary Mathematical Sciences, vol. 3, N^o 3 457-461.

Thatcher, A. R. (1964). *Relationships between Bayesian and confidence limits for prediction*. Journal of the Royal Statistical Society, Series B, 26: 126 - 210.

Tuyl F., Gerlach R. and Mengersen K. (2009) *Posterior predictive arguments in favor of the Bayes-Laplace prior as the consensus prior for binomial and multinomial parameters*. Bayesian Analysis, vol.4:1, pp. 151-58.

Yang, R. & Berger, J. O. (1997). *A Catalog of Noninformative Priors*. Tech. Rep., pages ISDS 97 - 42.

Zellner, A. (1996), *An Introduction to Bayesian Inference in Econometrics*, New York; Chichester: John Wiley.

Zhu, Mu and Lu, Arthur Y. (2004). *The counter-intuitive non-informative prior for the Bernoulli family*. Journal of Statistics Education, Volume 12, Number 2.

Anexo

```
##### Script do R #####
```

```
##### Gráfico da distribuição  $Beta(0.001, 0.001)$  #####
```

```
x <- seq(0, 1, length.out = 100)
plot(x, dbeta(x, shape1 = 0.001, shape2 = 0.001), xlab = "theta", ylab =
"Densidade da distribuição", main = "a < 1 e b < 1", col = "blue", type =
"l", lwd = 2)
abline(h = 0, col = "gray")
remove(x)
```

```
##### Gráfico da distribuição  $Beta(0.001, 0.5)$  #####
```

```
x <- seq(0, 1, length.out = 100)
plot(x, dbeta(x, shape1 = 0.001, shape2 = 0.5), xlab = "theta", ylab =
"Densidade da distribuição", main = "a < 1 e b < 1", col = "blue", type =
"l", lwd = 2)
abline(h = 0, col = "gray")
remove(x)
```

```
##### Gráfico da distribuição  $Beta(1, 1)$  #####
```

```
x <- seq(0, 1, length.out = 100)
plot(x, dbeta(x, shape1 = 1, shape2 = 1), xlab = "theta", ylab =
"Densidade da distribuição", main = "a = 1 e b = 1", col = "blue", type =
"l", lwd = 2)
abline(h = 0, col = "gray")
remove(x)
```

```
##### Gráfico da distribuição  $Beta(0.3, 1.3)$  #####
```

```
x <- seq(0, 1, length.out = 100)
plot(x, dbeta(x, shape1 = 0.3, shape2 = 1.3), xlab = "theta", ylab =
"Densidade da distribuição", main = "a < 1 e b > 1", col = "blue", type =
"l", lwd = 2)
```



```
abline(h = 0, col = "gray")
remove(x)
```

```
##### Gráfico da distribuição  $Beta(2,0.6)$  #####
```

```
x <- seq(0, 1, length.out = 100)
plot(x, dbeta(x, shape1 = 2, shape2 = 0.6), xlab = "theta", ylab =
"Densidade da distribuição", main = "a > 1 e b < 1", col = "blue", type =
"l", lwd = 2)
abline(h = 0, col = "gray")
remove(x)
```

```
##### Gráfico da distribuição  $Beta(3,2)$  #####
```

```
x <- seq(0, 1, length.out = 100)
plot(x, dbeta(x, shape1 = 3, shape2 = 2), xlab = "theta", ylab =
"Densidade da distribuição", main = "a > 1 e b > 1", col = "blue", type =
"l", lwd = 2)
abline(h = 0, col = "gray")
remove(x)
```

```
##### Gráfico da distribuição a priori de Bayes - Laplace  $Beta(1,1)$  #####
```

```
par(cex = 1.1, cex.lab = 1.55)
x <- seq(0, 1, length.out = 100)
plot(x, dbeta(x, shape1 = 1, shape2 = 1), xlab = expression(theta), ylab
= expression(p(theta)), col = "blue", type = "l", lwd = 2)
abline(h = 0, col = "gray")
remove(x)
```

```
##### Gráfico da distribuição a priori de Jeffreys  $Beta(0.0001,0.5)$  #####
```

```
par(cex = 1.1, cex.lab = 1.6)
x <- seq(0, 1, length.out = 100)
plot(x, dbeta(x, shape1 = 0.0001, shape2 = 0.5), xlab =
expression(theta), ylab = expression(p(theta)), col = "blue", type = "l",
lwd = 2)
abline(h = 0, col = "gray")
remove(x)
```

```
##### Gráfico da possível distribuição a priori  $Beta(0.0001,0.0001)$  #####
```

```

par(cex = 1.1, cex.lab = 1.6)
x <- seq(0, 1, length.out = 100)
plot(x, dbeta(x, shape1 = 0.0001, shape2 = 0.0001), xlab =
expression(theta), ylab = expression(p(theta)), col = "blue", type = "l",
lwd = 2)
abline(h = 0, col = "gray")
remove(x)

##### Distribuições a posteriori de Bayes - Laplace, de Jeffreys, #####
##### uma possível a priori  $Beta(0,0)$ , para  $n = 10$  e  $r = 3$ . #####

par(cex = 1.1, cex.lab = 1.2)
curve(dbeta(x, 4, 8), 0, 1, xlab = expression(theta), main = "Para  $n = 10$ 
e  $r = 3$ ", ylab = expression(paste("p(",theta,"|x)")), col = 2)
curve(dbeta(x, 3, 7.5), 0, 1, add = T, col = 3)
curve(dbeta(x, 3, 7), 0, 1, add = T, col = 4)
legend(0.45, 3, c("Bayes - Laplace", "Jeffreys", "Uma possível a
posteriori"), fill = 2:4)

##### Distribuições a posteriori de Bayes - Laplace, de Jeffreys, #####
##### uma possível a posteriori, para  $n = 100$  e  $r = 30$ . #####

par(cex = 1.1, cex.lab = 1.2)
curve(dbeta(x, 31, 71), 0, 1, xlab = expression(theta), main = "Para  $n =$ 
 $100$  e  $r = 30$ ", ylab = expression(paste("p(",theta,"|x)")), col = 2)
curve(dbeta(x, 30, 70.5), 0, 1, add = T, col = 3)
curve(dbeta(x, 30, 70), 0, 1, add = T, col = 4)
legend(0.45, 8, c("Bayes - Laplace", "Jeffreys", "Uma possível a
posteriori"), fill = 2:4)

##### Densidade da Binomial Negativa com  $r = 8$  e  $\theta = 0.15$  #####

.x <- 4:130
plot(.x, dnbinom(.x, size = 8, prob = 0.15), xlab = "x", ylab =
"Densidade da distribuição", type = "h", col = "blue")
points(.x, dnbinom(.x, size = 8, prob = 0.15), pch = 20, col = "blue")
abline(h = 0, col = "gray")
remove(.x)

```

```
##### Densidade da distribuição Beta Binomial Negativa p(x|1,1,40) #####
```

```
x <- 0:80
r <- 40
a <- 1
b <- 1
plot( (r/(x+r)) * ((choose(x+b-1, x) * choose(r+a-1, r)) /
(choose(x+r+a+b-1, x+r))), xlab = expression(x), ylab =
expression(p("x|1",1,40)), col = "blue", type = "h", lwd = 0, main="BBN")
points(((r/(x+r)) * (choose(x+b-1, x)*choose(r+a-1, r))) /
(choose(x+r+a+b-1, x+r)), type = "l", lwd = 2, pch = 1000, col = "blue")
abline(h = 0, col = "gray")
remove(x)
```

```
#### Densidade da distribuição Beta Binomial Negativa p(x|1.1,1.1,40) ####
```

```
x <- 0:80
r <- 40
a <- 1.1
b <- 1.1
plot((r/(x+r)) * ((choose(x+b-1, x) * choose(r+a-1, r)) /
(choose(x+r+a+b-1, x+r))), xlab = expression(x), ylab =
expression(p("x|1.1",1.1,40)), col="blue", type="h", lwd=0, main="BBN")
points(((r/(x+r)) * (choose(x+b-1, x)*choose(r+a-1, r))) /
(choose(x+r+a+b-1, x+r)), type = "l", lwd = 2, pch = 1000, col = "blue")
abline(h=0, col = "gray")
remove(x)
```

```
##### Densidade da distribuição Beta Binomial Negativa p(x|3,5,10) #####
```

```
x <- 0:70
r <- 10
a <- 3
b <- 5
plot((r/(x+r)) * (( choose(x+b-1, x) * choose(r+a-1, r)) /
(choose(x+r+a+b-1, x+r))), xlab = expression(x), ylab =
expression(p("x|3",5,10) ), ylim = c(0, 0.04), col = "blue", type = "h",
lwd = 0, main = "BBN")
points(((r/(x+r)) * (choose(x+b-1, x)*choose(r+a-1, r))) /
```

```
(choose(x+r+a+b-1, x+r)), type = 'l', lwd = 2, pch = 1000, col = "blue")
abline(h = 0, col = "gray")
remove(x)
```

```
##### Densidade da distribuição Beta Binomial Negativa p(x|3,5,40) #####
```

```
x <- 0:143
r <- 40
a <- 3
b <- 5
plot((r/(x+r)) * ((choose(x+b-1, x) * choose(r+a-1, r)) /
(choose(x+r+a+b-1, x+r))), xlab = expression(x), ylab =
expression(p("x|3",5,40)), col = "blue", type="h", lwd=0, main="BBN")
points(((r/(x+r)) * (choose(x+b-1, x) * choose(r+a-1, r))) /
(choose(x+r+a+b-1, x+r)), type = 'l', lwd = 2, pch = 1000, col = "blue")
abline(h = 0, col = "gray")
remove(x)
```