

Inferência Bayesiana Objetiva e Frequentista  
Para a Probabilidade de Sucesso

RUBIANE MARIA PIRES

UFSCar - São Carlos/SP

Fevereiro/2009

Universidade Federal de São Carlos  
Centro de Ciências Exatas e de Tecnologia  
Departamento de Estatística

# Inferência Bayesiana Objetiva e Frequentista

## Para a Probabilidade de Sucesso

RUBIANE MARIA PIRES

PROF. DR. JOSEMAR RODRIGUES

Trabalho apresentado ao Departamento de Estatística da Universidade Federal de São Carlos - DEs/UFSCar como parte dos requisitos para obtenção do título de Mestre em Estatística.

UFSCar - São Carlos/SP

Fevereiro/2009

**Ficha catalográfica elaborada pelo DePT da  
Biblioteca Comunitária da UFSCar**

P667ib

Pires, Rubiane Maria.

Inferência bayesiana objetiva e freqüentista para a probabilidade de sucesso / Rubiane Maria Pires. -- São Carlos : UFSCar, 2009.

142 f.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2009.

1. Distribuição a priori menos favorável. 2. Distribuição binomial. 3. Distribuição binomial negativa. 4. Inferência bayesiana. I. Título.

CDD: 519.54 (20ª)



## DECLARAÇÃO

Declaramos, para os devidos fins, que Rubiane Maria Pires defendeu sua Dissertação de Mestrado no dia 10/02/2009, tendo sido **aprovada**. A aluna deverá apresentar a versão final da dissertação (com as correções e sugestões da Banca, e a ficha catalográfica anexada), e a Certidão Negativa da Biblioteca Comunitária, para formação do processo de homologação e emissão do Diploma do Título.

Igualmente, a aluna deverá apresentar a documentação da pesquisa (rotinas, arquivos em LaTeX, resultados complementares etc.) ao seu orientador, visando facilitar a confecção de relatórios técnicos que condensarão os resultados obtidos.

Toda a documentação citada deverá ser entregue em 30 (trinta) dias, a iniciar na data da Defesa.

Essa declaração é válida pelo período de 30 dias.

São Carlos, 10 de fevereiro de 2009.

Prof. Dr. Josemar Rodrigues  
Coordenador – PPG-Es / UFSCar

# Agradecimentos

Primeiramente agradeço a Deus, que me dá saúde e força para superar os obstáculos todos os dias.

A minha mãe e demais familiares, pela formação de caráter em boas maneiras de conduzir minha vida, principalmente aos meus irmãos Sueli e Rubens e meu cunhado Sebastião que me incentivaram na árdua tarefa de estudar.

Ao namorado e companheiro Eduardo, pelo amor, amizade e total apoio, compartilhados nas pacientes horas de trabalhos, incentivando-me a não desistir dos estudos.

A todos os meus amigos que sempre estiveram carinhosamente presentes contribuindo com críticas, sugestões e paciente tolerância.

Ao meu orientador, Josemar Rodrigues, pela orientação e incentivo na elaboração e condução do trabalho.

Aos professores Adriano Polpo e Mário de Castro, membros da banca do exame de qualificação, pelas sugestões feitas.

Aos professores do Departamento de Estatística da Universidade Federal de São Carlos que me abriram as portas e me ofereceram o ambiente acolhedor e sadio para que eu pudesse realizar o meu mestrado.

Aos diversos professores do Departamento de Estatística da Universidade Federal do Paraná, em especial os professores Anselmo Chaves Neto, Benito Olivares Aguilera, Fernando Lucambio Pérez e Sueli Giolo pela amizade, ensinamentos e pelo incentivo que me levaram a fazer a pós-graduação.

Finalmente, agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo auxílio concedido para este trabalho.

# Resumo

Neste estudo são abordadas duas distribuições discretas baseadas em ensaios de Bernoulli, a Binomial e a Binomial Negativa. São explorados intervalos de credibilidade e confiança para estimação da probabilidade de sucesso de ambas as distribuições. A principal finalidade é analisar nos contextos clássico e bayesiano o desempenho da probabilidade de cobertura e amplitude média gerada pelos intervalos de confiança e intervalos de credibilidade ao longo do espaço paramétrico. Considerou-se também a análise dos estimadores pontuais bayesianos e o estimador de máxima verossimilhança, cujo interesse é confirmar por meio de simulação a consistência e calcular o viés e o erro quadrático médio dos mesmos. A Inferência Bayesiana Objetiva é empregada neste estudo por meio das distribuições *a priori* não-informativas de Bayes-Laplace, de Haldane, de Jeffreys e menos favorável. Ao analisar as distribuições *a priori* no contexto de teoria de decisões minimax, a distribuição *a priori* menos favorável resgata as demais citadas ao empregar a função de perda quadrática e coincide com a distribuição *a priori* de Bayes-Laplace ao considerar a função de perda quadrática ponderada para o modelo Binomial, o que não foi encontrado até o momento na literatura. Para o modelo Binomial Negativa são consideradas as distribuições *a priori* não-informativas de Bayes-Laplace e de Jeffreys. Com os estudos desenvolvidos pôde-se observar que a Inferência Bayesiana Objetiva para a probabilidade de sucesso dos modelos Binomial e Binomial Negativa apresentou boas propriedades frequentistas, analisadas a partir da probabilidade de cobertura e amplitude média dos intervalos bayesianos e por meio das propriedades dos estimadores pontuais. A última etapa do trabalho consiste na análise da ocorrência de proporções correlacionadas em pares de eventos de Bernoulli (tabela  $2 \times 2$ ) com a finalidade de determinar um possível ganho de informação em relação as medidas consideradas para testar a ocorrência de proporções correlacionadas. Para tanto fez-se uso do modelo Trinomial e da função de

verossimilhança parcial tanto numa abordagem clássica quanto bayesiana. Nos conjuntos de dados analisados observou-se a medida de evidência bayesiana (FBST) como sensível à parametrização, já para os métodos clássicos essa comparação não foi possível, pois métodos distintos precisam ser aplicados para o modelo Trinomial e para a função de verossimilhança parcial.

**Palavras-chave:** Distribuição *a priori* menos favorável; Distribuição Binomial; Distribuição Binomial Negativa; Inferência Bayesiana Objetiva; Medida de evidência bayesiana (FBST).

# Abstract

This study considers two discrete distributions based on Bernoulli trials: the Binomial and the Negative Binomial. We explore credibility and confidence intervals to estimate the probability of success of each distribution. The main goal is to analyze their performance coverage probability and average range across the parametric space. We also consider point analysis of bayesian estimators and maximum likelihood estimators, whose interest is to confirm through simulation their consistency, bias and mean square error. In this paper the Objective Bayesian Inference is applied through the noninformative Bayes-Laplace prior, Haldane prior, reference prior and least favorable prior. By analyzing the prior distributions in the minimax decision theory context we verified that the least favorable prior distribution has every other considered prior distributions as particular cases when a quadratic loss function is applied, and matches the Bayes-Laplace prior in considering the quadratic weighed loss function for the Binomial model (which was never found in literature). We used the noninformative Bayes-Laplace prior and Jeffreys prior for the Negative Binomial model. Our findings show through coverage probability, average range of bayesian intervals and point estimation that the Objective Bayesian Inference has good frequentist properties for the probability of success of Binomial and Negative Binomial models. The last stage of this study discusses the presence of correlated proportions in matched-pairs ( $2 \times 2$  table) of Bernoulli with the goal of obtaining more information in relation of the considered measures for testing the occurrence of correlated proportions. In this sense the Trinomial model and the partial likelihood function were used from the frequentist and bayesian point of view. The Full Bayesian Significance Test (FBST) was used for real data sets and was shown sensitive to parameterization, however, this study was not possible for the frequentist method since distinct methods are needed to be applied to Trinomial model and the partial likelihood function.

**Keywords:** Least favorable prior distribution; Binomial distribution; Negative Binomial distribution; Objective Bayesian Inference; Full Bayesian Significance Test (FBST).

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Organização dos capítulos	3
<b>2</b>	<b>Inferência Bayesiana Objetiva</b>	<b>5</b>
2.1	Distribuições <i>a priori</i> não-informativa	6
2.1.1	Método de Bayes-Laplace	7
2.1.2	Método de Jeffreys	7
2.1.3	Distribuição <i>a priori</i> de referência	9
<b>3</b>	<b>Distribuição Binomial</b>	<b>10</b>
3.1	Distribuições <i>a priori</i> e distribuições <i>a posteriori</i>	10
3.1.1	Distribuição <i>a priori</i> menos favorável	11
3.1.2	Distribuição <i>a priori</i> de Bayes-Laplace	20
3.1.3	Distribuição <i>a priori</i> de Haldane	21
3.1.4	Distribuição <i>a priori</i> de Jeffreys	22
3.2	Estimação Intervalar	23
3.2.1	Intervalo de credibilidade	23
3.2.2	Intervalo de credibilidade a partir da distribuição $F$ de Fisher	24
3.2.3	Intervalo HPD (máxima densidade <i>a posteriori</i> )	26

---

3.2.4	Intervalo de confiança com aproximação pela distribuição Normal . . . . .	26
3.2.5	Intervalo de confiança exato de Clopper & Pearson . . . . .	27
3.2.6	Intervalo com probabilidade de cobertura garantida . . . . .	30
<b>4</b>	<b>Distribuição Binomial Negativa . . . . .</b>	<b>32</b>
4.1	Distribuições <i>a priori</i> e distribuições <i>a posteriori</i> . . . . .	33
4.1.1	Distribuição <i>a priori</i> menos favorável . . . . .	33
4.1.2	Distribuição <i>a priori</i> de Bayes-Laplace . . . . .	35
4.1.3	Distribuição <i>a priori</i> de Jeffreys . . . . .	35
4.2	Estimação Intervalar . . . . .	37
4.2.1	Intervalo de credibilidade . . . . .	37
4.2.2	Intervalo HPD (máxima densidade <i>a posteriori</i> ) . . . . .	38
4.2.3	Intervalo de confiança com aproximação pela distribuição Normal . . . . .	38
4.2.4	Intervalo de confiança exato de Clopper & Pearson . . . . .	39
<b>5</b>	<b>Estudo de Simulação . . . . .</b>	<b>42</b>
5.1	Distribuição Binomial . . . . .	43
5.1.1	Proporções de cobertura e amplitudes médias dos intervalos bayesianos . . . . .	43
5.1.2	Proporções de cobertura e amplitudes médias dos intervalos clássicos . . . . .	69
5.1.3	Consistência, viés e erro quadrático médio dos estimadores . . . . .	75
5.2	Distribuição Binomial Negativa . . . . .	85
5.2.1	Proporções de cobertura e amplitudes médias dos intervalos bayesianos . . . . .	85
5.2.2	Proporções de cobertura e amplitudes médias dos intervalos clássicos . . . . .	105
5.2.3	Consistência, viés e erro quadrático médio dos estimadores . . . . .	110

---

<b>6</b>	<b>Comparação de duas proporções correlacionadas</b>	<b>120</b>
6.1	Modelo Trinomial	121
6.2	Verossimilhança parcial da distribuição Trinomial	124
6.3	Teste de Hipóteses	125
6.3.1	Testes de Hipóteses Clássicos	125
6.3.2	Uma medida de evidência completamente bayesiana	127
6.4	Exemplos	127
6.4.1	Procedimentos clássicos	128
6.4.2	Medida de evidência bayesiana	128
<b>7</b>	<b>Considerações Finais</b>	<b>131</b>
	<b>Referências</b>	<b>133</b>
<b>A</b>	<b>Demonstrações</b>	<b>135</b>
A.1	Demonstração do Lema 3.1	135
A.2	Demonstração do Lema 3.2	136
A.3	Demonstração do Lema 4.1	137
A.4	Demonstração da igualdade 6.7, do Capítulo 6	138
<b>B</b>	<b>Esperança e Variância dos estimadores pontuais</b>	<b>139</b>

# Capítulo 1

## Introdução

Ao realizar um experimento o pesquisador pode ter interesse em respostas que assumam um número finito de valores, por exemplo a condição de uma peça (defeituosa e não defeituosa) e o desempenho de um candidato (bom, regular ou péssimo).

A análise de estudos em que a variável resposta apresenta esta natureza categórica é chamada de análise de dados discretos, pois as distribuições discretas de probabilidade encontram-se associadas às variáveis de interesse.

Este texto aborda distribuições discretas cujas distribuições de probabilidade estão baseadas em ensaios de Bernoulli, cujos resultados são binários. Para as distribuições Binomial e Binomial Negativa a finalidade principal é analisar, nos contextos clássico e bayesiano, o desempenho das probabilidades de cobertura e amplitudes médias dos intervalos de confiança e intervalos de credibilidade para a probabilidade de sucesso das duas distribuições ao longo do espaço paramétrico.

A motivação deste trabalho está baseada no princípio freqüentista (Neyman, 1977) que consiste em considerar réplicas para diferentes valores do parâmetro, não apenas réplicas em um único parâmetro fixo verdadeiro.

Adicionalmente este trabalho também deseja confirmar por meio de estudo de simulação que os estimadores pontuais bayesianos, média e moda da distribuição *a posteriori* da probabilidade de sucesso, e o estimador clássico de máxima verossimilhança são consistentes, e também é calculado o viés e o erro quadrático médio dos mesmos.

No caso dos intervalos de confiança para a probabilidade de sucesso, o intervalo com aproximação pela distribuição Normal é geralmente aplicado, conforme destacam Brown *et al.* (2001), isso também pode ser observado nos cursos introdutórios de Inferência Estatística. Brown *et al.* (2002) mostram exemplos em que as probabilidades de cobertura deste intervalo são irregulares no espaço paramétrico. Contudo, propostas alternativas à aproximação pela distribuição Normal são temas de estudos mais antigos, como o intervalo exato de Clopper & Pearson (1934), bem como estudos recentes, como o intervalo com probabilidade de cobertura garantida de Chen *et al.* (2008).

Com os estudos desenvolvidos neste texto também teve-se interesse em explorar algumas distribuições *a priori* não-informativas existentes para os modelos abordados (Agresti & Hitchcock, 2005; Lehmann & Casella, 1998, cap. 5). Berger (1985, p. 137) menciona que ao considerar distribuições *a priori* não-informativas, os resultados bayesianos podem, em vários casos, reproduzir os resultados clássicos numericamente, embora com interpretações diferentes.

Uma análise destas distribuições *a priori* no contexto de teoria de decisões minimax através da distribuição *a priori* menos favorável também foi considerado para o modelo Binomial. Como esta concepção depende da função de perda associada, uma proposta deste trabalho foi considerar uma função de perda na qual a distribuição *a priori* é mais realista e não depende do tamanho amostral conforme abordada por Lehmann & Casella (1998, p. 311).

A qualidade do procedimento bayesiano é analisada no contexto freqüentista, ou seja, amostras foram simuladas para a construção dos intervalos de credibilidade. Desses intervalos foi verificada a estabilidade da probabilidade de cobertura. Esse mesmo procedimento foi considerado para os intervalos de confiança.

A distribuição Binomial conta o número de sucessos em um número fixo de ensaios de Bernoulli e a distribuição Binomial Negativa conta o número de ensaios de Bernoulli necessários para se obter um número de sucessos fixo (Casella & Berger, 2002, p. 82).

Suponha agora que estamos interessados em eventos do conjunto  $\{(0, 0), (0, 1), (1, 0), (1, 1)\}$ . Um problema decorrente desses eventos, geralmente presentes quando se trata de mudança de comportamento ou mudança de opinião, é encontrar proporções correlacionadas, pois a dependência entre as medidas causa dificuldade na obtenção de

intervalos de confiança adequados, conforme citam Irony *et al.* (2000).

Geralmente para verificar a ocorrência de proporções correlacionadas é empregado o teste clássico de McNemar (1947). O interesse deste trabalho é analisar a percepção da ocorrência da mudança de opinião ao considerar esse teste clássico e o teste de significância genuinamente bayesiano (FBST - *Full Bayesian Significance Test*) (Pereira *et al.*, 2008) para duas proporções correlacionadas, por meio de conjunto de dados reais. Além disso, analisar a percepção do impacto da verossimilhança parcial (Altham, 1971).

## 1.1 Organização dos capítulos

O foco principal do Capítulo 2 é mostrar alguns métodos de obtenção de distribuições *a priori* não-informativas, mais consideradas na literatura, para os modelos Binomial e Binomial Negativa e proporcionar uma reflexão sobre a Análise Bayesiana Objetiva.

No Capítulo 3 serão exploradas construções de distribuições *a priori* não-informativas para a probabilidade de sucesso do modelo Binomial. O contexto de teoria de decisões minimax é abordado. Neste capítulo serão mostrados também alguns métodos para construção de intervalos clássicos e bayesianos para a probabilidade de sucesso da distribuição Binomial.

No Capítulo 4 serão abordadas distribuições *a priori* não-informativas, discutidas no Capítulo 2, para a distribuição Binomial Negativa, as dificuldades para construção da distribuição *a priori* menos favorável para esse modelo e algumas metodologias para construção de intervalos clássicos e bayesianos.

A verificação da probabilidade de cobertura e amplitude média dos intervalos será desenvolvida no Capítulo 5, no qual serão apresentados os resultados de simulações dos métodos clássicos e bayesianos mencionados nos Capítulos 3 e 4. Neste capítulo também são exploradas a consistência, o viés e o erro quadrático médio dos estimadores pontuais gerados por ambas as metodologias clássica e bayesiana.

No Capítulo 6 foi feita a última etapa do trabalho, na qual os eventos de interesse podem ser representados por uma tabela  $2 \times 2$  em que o tamanho amostral é fixo, ou seja, os pares de ensaios de Bernoulli são aleatoriamente distribuídos segundo um modelo

---

Trinomial. Além disso, são apresentados os testes clássicos de McNemar (1947) e Binomial e o teste de significância genuinamente bayesiano (FBST). Essas metodologias são aplicadas a conjuntos de dados para detectar a ocorrência de proporções correlacionadas. As distribuições *a priori* não-informativas são consideradas para os modelos utilizados (Agresti & Hitchcock, 2005).

## Capítulo 2

# Inferência Bayesiana Objetiva

Não há consenso quanto à definição de Análise Bayesiana Objetiva, nem mesmo quanto a sua meta (Berger, 2006). No entanto, um grupo de pesquisadores dedica-se ao estudo desta teoria, pois acreditam que a Inferência Bayesiana Objetiva seja uma alternativa reconciliadora com a inferência clássica, pois a distribuição *a priori* é não-informativa e a metodologia bayesiana possui boas qualidades freqüentistas para alguns modelos já estudados.

*Workshops* de âmbito internacional são realizados para discutir essa metodologia. O oitavo *workshop* foi realizado em Roma (Itália, 2007); em 2009 ocorrerá o 9<sup>th</sup> *International Workshop on Objective Bayes Methodology*, o qual será realizado na Pensilvânia (Filadélfia, PA, EUA). Seu principal objetivo será facilitar o intercâmbio de pesquisas recentes, desenvolvimentos da metodologia bayesiana objetiva, estabelecer novas parcerias e colaborações para canalizar os esforços em problemas pendentes e assim abrir novos rumos para um estudo mais aprofundado.

A análise bayesiana com distribuições *a priori* não-informativas tem sido cada vez mais reconhecida como um método para estatísticos clássicos obterem procedimentos com boas propriedades freqüentistas (Yang & Berger, 1997). Berger (2006) debate algumas posições filosóficas sobre o assunto:

1. A Análise Bayesiana Objetiva é o melhor método para sintetizar objetivamente e comunicar as incertezas que surgem em um cenário específico.
2. A Análise Bayesiana Objetiva é uma convenção que deverá ser adotada em situações

na qual uma análise subjetiva pode não ser razoável.

3. A Análise Bayesiana Objetiva é simplesmente uma coleção de metodologias de aprendizagem a partir de dados, com uma finalidade específica.

Conforme sugerido por Yang & Berger (1997) é possível utilizar a Inferência Bayesiana Objetiva para analisar a sensibilidade da informação disponível, ou seja, comparar as respostas de uma análise subjetiva com as respostas de uma análise com distribuição *a priori* não-informativa. Em muitas situações práticas é usual fazer uso das distribuições *a priori* não-informativas, principalmente quando não há *a priori* informação alguma sobre o problema e/ou há complexidade de elicitar distribuições *a priori* subjetivas.

Berger (2006) relata que diante das posições citadas, o item 2 necessita de um rigor através da definição de procedimentos objetivos bayesianos ótimos em função de um cenário bem definido. Uma abordagem nesse sentido é dada pela distribuição *a priori* de referência.

## 2.1 Distribuições *a priori* não-informativa

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

O conceito de *vago* (ignorância *a priori*) não é único e o problema de elicitar distribuições *a priori* com tais características pode se tornar bastante complexo. Por outro lado, é reconhecida a necessidade de alguma forma de análise que consiga captar esta noção através de uma distribuição *a priori* com efeito mínimo sobre os resultados. Esta análise pode ser pensada como um ponto de partida quando não se consegue uma descrição detalhada do *verdadeiro* conhecimento *a priori*.

A seguir são apresentados alguns métodos para obtenção de distribuições *a priori* não-informativas.

### 2.1.1 Método de Bayes-Laplace

A primeira idéia de ausência de informação *a priori* que se pode ter é pensar em uma distribuição *a priori* constante no espaço paramétrico. Essa idéia instigou Bayes e Laplace a sugerir  $h(\theta) \propto k$  para  $\theta$  variando em um subconjunto da reta.

Porém esta escolha pode trazer algumas dificuldades técnicas, como por exemplo, se o intervalo de variação de  $\theta$  for ilimitado, então a distribuição *a priori* é imprópria, ou melhor,  $\int h(\theta)d\theta = \infty$ .

### 2.1.2 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 transformações 1-1. Por conta disso ele propôs, em 1961, uma forma de obtenção de distribuições *a priori* que assegura invariância sob transformações injetivas (Paulino *et al.*, 2003, p. 99). Esse método baseia-se na informação de Fisher sobre  $\theta \in \mathfrak{R}$ .

**Definição 2.1** (Casella & Berger, 2002, p. 338; Paulino *et al.*, 2003, p. 99) Seja  $X$  uma única observação com função (densidade) de probabilidade  $h(x|\theta)$ . A medida de informação esperada de Fisher de  $\theta$  por meio de  $X$  é definida como

$$I(\theta) = \mathbb{E}_X \left[ \left( \frac{\partial \ln h(X|\theta)}{\partial \theta} \right)^2 \middle| \theta \right].$$

Alternativamente, o resultado do lema a seguir pode ser considerado para calcular a informação esperada de Fisher uniparamétrica.

**Lema 2.1** (Casella & Berger, 2002, p. 338) Se  $h(x|\theta)$  satisfaz

$$\frac{d}{d\theta} \mathbb{E}_X \left[ \left( \frac{\partial}{\partial \theta} \ln h(X|\theta) \right) \middle| \theta \right] = \int \frac{\partial}{\partial \theta} \left[ \left( \frac{\partial}{\partial \theta} \ln h(x|\theta) \right) h(x|\theta) \right] dx$$

(verdade para distribuições da família exponencial), então

$$\mathbb{E}_X \left[ \left( \frac{\partial}{\partial \theta} \ln h(X|\theta) \right)^2 \middle| \theta \right] = -\mathbb{E}_X \left( \frac{\partial^2}{\partial \theta^2} \ln h(X|\theta) \middle| \theta \right).$$

O conceito de informação é associado à curvatura média da função de verossimilhança; quanto maior a curvatura, mais precisa é a informação contida na verossimilhança, ou seja, maior o valor de  $I(\theta)$ . Observe que a esperança matemática é tomada em relação à distribuição da amostra  $h(X|\theta)$ .

**Definição 2.2** (Paulino *et al.*, 2003, p. 101) No caso em que  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_s)$  é um vetor paramétrico, define-se a matriz de informação esperada de Fisher de  $\boldsymbol{\theta}$  por meio de  $X$  por

$$I(\boldsymbol{\theta}) = \mathbb{E}_X \left[ \frac{\partial \ln h(X|\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \middle| \boldsymbol{\theta} \right].$$

**Definição 2.3** (Paulino *et al.*, 2003, p. 100; Berger, 1985, p. 87) Seja  $X$  uma variável aleatória com função (densidade) de probabilidade  $h(x|\theta)$ . A *a priori* de Jeffreys para o caso uniparamétrico é dada por  $h(\theta) \propto I(\theta)^{1/2}$ ; e no caso multiparamétrico por  $h(\boldsymbol{\theta}) \propto [\det I(\boldsymbol{\theta})]^{1/2}$ .

Vale ressaltar, contudo, 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  $\theta$  não precisa necessariamente pertencer a um espaço limitado.

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

1. 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*.
2. A existência da distribuição *a priori* de Jeffreys requer condições de regularidade fortes, como a normalidade assintótica da distribuição *a posteriori* de  $\theta$ , conforme lembra Bernardo & Smith (2000, p. 314).

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 de distribuição *a priori*. Bernardo na tentativa de superar as dificuldades encontradas por Jeffreys propôs, em 1979, a distribuição *a priori* de referência.

### 2.1.3 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, p. 304).

A especificação da distribuição *a priori* de referência é complexa computacionalmente, mas para as funções de verossimilhança abordadas neste trabalho a mesma coincide com o método de Jeffreys (ver em 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  $\theta$  possui normalidade assintótica.

# Capítulo 3

## Distribuição Binomial

É comum realizar aplicações em que é considerado um número fixo  $n$  de observações binárias. Sejam  $x_1, x_2, \dots, x_n$  respostas obtidas de  $n$  ensaios independentes e idênticos, em que  $\mathbb{P}(X_i = 1) = \theta$  e  $\mathbb{P}(X_i = 0) = 1 - \theta$ , comumente tratados de “sucesso” e “insucesso” para valores 1 e 0 (Agresti, 2007, p. 4). O número de sucessos total  $X|\theta$ , tem distribuição condicional de  $X$  dado  $\theta$ , adotado para unificar as notações quanto as abordagens clássica e bayesiana. Sua função de probabilidade para possíveis valores da variável aleatória  $X$  é dada por

$$h(x|\theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}, \quad (3.1)$$

em que  $0 < \theta < 1$  e  $x = 0, 1, 2, \dots, n$ .

Esse modelo será denotado por  $\text{Bin}(n, \theta)$ . Será convencionado neste estudo que o parâmetro  $n$  é conhecido e  $\theta$ , que descreve a probabilidade de sucesso, é o parâmetro a ser estimado, que tem como estatística suficiente  $\mathbf{X} = \sum_{i=1}^n X_i$ .

Para que a abordagem bayesiana possa ser construída algumas distribuições *a priori* para o parâmetro de interesse serão discutidas a seguir.

### 3.1 Distribuições *a priori* e distribuições *a posteriori*

A abordagem bayesiana para análise de dados discretos iniciou-se com a estimação da probabilidade de sucesso da distribuição Binomial a partir da proposta de Bayes-

Laplace. Mais tarde Haldane defendeu a distribuição *a priori* Beta(0,0) com aplicações em genética. Bernardo propôs a distribuição *a priori* de referência, que coincide com a distribuição *a priori* de Jeffreys proposta anteriormente, conforme comentam Agresti & Hitchcock (2005) e Gelman *et al.* (2004, p. 63).

### 3.1.1 Distribuição *a priori* menos favorável

No Capítulo 2 foram citados alguns métodos para obtenção de distribuições *a priori* não-informativas, consolidados na literatura, para estimação da probabilidade de sucesso no modelo Binomial. Contudo, estas distribuições *a priori* serão analisadas no contexto de teoria de decisões minimax (Lehmann & Casella, 1998, cap. 5).

A teoria de decisões minimax foi introduzida e desenvolvida extensivamente por Von Neumann e Morgenstern em 1947, conforme descreve DeGroot (1970, p. 136). A estimação de Bayes para um possível risco máximo foi considerada inicialmente por Hodges e Lehmann, em 1959, ao tratar de problemas computacionalmente difíceis, segundo Lehmann & Casella (1998, p. 429).

As informações *a priori* e dos dados são complementadas com um novo tipo de função (função de perda) que diz respeito às conseqüências da decisão tomada. A função de perda  $L(\theta, d)$  associa uma penalização, com valores positivos, para cada decisão  $d$  e cada possível valor do parâmetro  $\theta$ . Quando a perda é zero significa que o valor de  $\theta$  foi corretamente estimado.

A precisão de uma função de decisão  $\delta$  é mensurada pela função de risco  $R(\theta, \delta)$  (Lehmann & Casella, 1998, p. 5). A função de risco exprime a perda média sofrida pelo estatístico ao considerar o estimador  $\delta(X)$  em um estado da natureza  $\theta \in \Theta$  e é definida conforme segue.

**Definição 3.1** (Lehmann & Casella, 1998, p. 5) A função de risco da decisão  $\delta(\mathbf{X})$  é definida por

$$R(\theta, \delta) = \mathbb{E}_{\mathbf{X}|\theta} [L(\theta, \delta(\mathbf{X}))] = \int_{\chi} L(\theta, \delta(x))h(x|\theta)dx,$$

no qual  $\chi$  denota o espaço amostral.

Note que a função de risco é integrada em relação ao espaço amostral, o que não é bem visto por bayesianos. Uma alternativa é a solução minimax. A idéia do estimador minimax, para um determinado parâmetro, é minimizar o risco máximo.

**Definição 3.2** (Lehmann & Casella, 1998, p. 309) Um estimador  $\delta^M$  de  $\theta$ , que minimiza o risco máximo, quer dizer, o qual satisfaz

$$\inf_{\delta} \sup_{\theta} R(\theta, \delta) = \sup_{\theta} R(\theta, \delta^M),$$

é chamado de estimador minimax.

Encontrar este estimador não é simples; portanto, em muitos problemas não é possível obter sua expressão exata. Lehmann & Casella (1998, p. 310) questionam ao aplicar a idéia de estimadores minimax: Para qual distribuição *a priori*  $h(\theta)$  a solução de Bayes é minimax?

Para que se possa apresentar formalmente o conceito de distribuição *a priori* menos favorável é necessário utilizar o risco de Bayes (risco médio).

**Definição 3.3** (Berger, 1985, p. 11) O risco de Bayes da regra de decisão  $\delta$ , com distribuição *a priori* para  $\theta$  dada por  $h(\theta)$ , é definido por

$$r(h, \delta) = \mathbb{E}_{\theta}[R(\theta, \delta)] = \int_{\Theta} R(\theta, \delta)h(\theta)d\theta.$$

Esta definição pode ser reescrita em termos da Definição 3.1 e conjuntamente ao aplicar o Teorema de Bayes, tem-se que

$$r(h, \delta) = \int_{\mathcal{X}} \left[ \int_{\Theta} L(\theta, \delta(x))h(\theta|x)d\theta \right] m(x)dx,$$

em que  $m(x) = \int_{\Theta} h(x|\theta)h(\theta)d\theta$ .

Assim, encontrar uma regra que minimiza esta função (estimador de Bayes, denotado por  $\delta_h$ ) é o mesmo que obter a regra que minimiza

$$\int_{\Theta} L(\theta, \delta(x))h(\theta|x)d\theta. \quad (3.2)$$

Contudo, o risco da solução de Bayes é

$$r_h = r(h, \delta_h) = \int_{\Theta} R(\theta, \delta_h)h(\theta)d\theta.$$

A solução para a distribuição *a priori* menos favorável é admitir que a natureza é a mais hostil possível, ou seja, admitir que a natureza se coloca em um estado que maximiza a perda esperada deixando ao estatístico apenas a possibilidade de minimizar esse máximo. Em outras palavras, espera-se que esse estimador seja o estimador de Bayes para a pior distribuição possível.

**Definição 3.4** A distribuição *a priori*  $h(\theta)$  é menos favorável se  $r_h \geq r_{h'}$  para toda distribuição *a priori*  $h'(\theta)$ .

Pela Definição 3.4, uma distribuição *a priori*  $h(\theta)$  é menos favorável se qualquer outra distribuição *a priori* é mais informativa que ela. Essa distribuição *a priori* causa estatisticamente a maior perda média. Desta forma, intuitivamente é introduzida a idéia de distribuição *a priori* não-informativa.

O teorema a seguir fornece uma condição simples para que um estimador de Bayes seja minimax.

**Teorema 3.1** Suponha que  $h(\theta)$  é uma distribuição em  $\Theta$  tal que

$$r(h, \delta_h) = \sup_{\theta} R(\theta, \delta_h). \quad (3.3)$$

Então,

1.  $\delta_h$ , o estimador de Bayes, é minimax.
2. Se  $\delta_h$  é a única solução de Bayes para  $h(\theta)$ , este é o único procedimento minimax.
3.  $h(\theta)$  é menos favorável.

A demonstração do Teorema 3.1 está em Lehmann & Casella (1998, p. 310).

O Teorema 3.1 é válido quando  $h(\theta)$  assume probabilidade 1 no conjunto em que a função de risco tem o seu valor máximo. Também tem-se o seguinte corolário.

**Corolário 3.1** Se a solução de Bayes  $\delta_h$  tem risco constante em relação a  $\theta$ , então ela é minimax.

Demonstração do Corolário 3.1 ver Lehmann & Casella (1998, p. 311).

A partir do conhecimento sobre o parâmetro que descreve a probabilidade de sucesso no modelo Binomial, pode-se definir uma família paramétrica de densidades. 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 distribuições *a priori* são chamados de *hiperparâmetros* para distingui-los do parâmetro de interesse  $\theta$ .

Essa abordagem em geral facilita a análise e o caso mais importante é o de distribuições *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  $\theta$  envolve apenas uma mudança nos hiperparâmetros.

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

$$\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1 - \theta)^{\beta-1}, \quad 0 < \theta < 1.$$

Desta forma, ao considerar a função de verossimilhança  $L_x(\theta) = h(x|\theta)$ , a densidade *a posteriori*  $h(\theta|x)$  pode ser obtida pelo Teorema de Bayes:

$$\begin{aligned} h(\theta|x) &= \frac{L_x(\theta) h(\theta)}{\int_{\Theta} L_x(\theta) h(\theta) d\theta} \propto \frac{\theta^x (1 - \theta)^{n-x} \theta^{\alpha-1} (1 - \theta)^{\beta-1}}{\int_0^1 \theta^x (1 - \theta)^{n-x} \theta^{\alpha-1} (1 - \theta)^{\beta-1} d\theta} \\ &\propto \frac{\theta^{x+\alpha-1} (1 - \theta)^{n-x+\beta-1}}{\int_0^1 \theta^{x+\alpha-1} (1 - \theta)^{n-x+\beta-1} d\theta} \end{aligned} \quad (3.4)$$

em que  $x = 0, 1, \dots, n$  e  $0 < \theta < 1$ . Assim, a distribuição *a posteriori* de  $\theta$  tem distribuição Beta com hiperparâmetros  $\alpha^* = x + \alpha$  e  $\beta^* = n - x + \beta$ .

**Teorema 3.2** (Berger, 1985, p. 60) Ao considerar uma função de perda quadrática,

$$L(\theta, \delta(x)) = (\delta(x) - \theta)^2,$$

o estimador de Bayes  $\delta_h(\mathbf{X})$  é a média da distribuição *a posteriori*.

**Prova:** O estimador de Bayes é a regra que minimiza a função (3.2). Desta forma, basta calcular seu mínimo ao considerar a função de perda quadrática. Desta forma, basta calcular

$$\frac{\partial}{\partial \delta(x)} \left[ \int_{\Theta} (\delta(x) - \theta)^2 h(\theta|x) d\theta \right] = 0.$$

Logo tem-se que

$$\frac{\partial}{\partial \delta(x)} \left[ \int_{\Theta} (\delta(x)^2 - 2\delta(x)\theta + \theta^2) h(\theta|x) d\theta \right],$$

ou seja,

$$\frac{\partial}{\partial \delta(x)} \left[ \delta(x)^2 \int_{\Theta} h(\theta|x) - 2\delta(x) \int_{\Theta} \theta h(\theta|x) + \int_{\Theta} \theta^2 h(\theta|x) \right],$$

que leva a

$$\delta(x) = \int_{\Theta} \theta h(\theta|x) d\theta. \quad \blacksquare$$

Então a média da distribuição *a posteriori* de  $\theta|x$  é dada por

$$\delta_h(\mathbf{X}) = \frac{\mathbf{X} + \alpha}{\alpha + \beta + n}. \quad (3.5)$$

Observe no Apêndice B que a esperança desse estimador é diferente de  $\theta$ , ou seja, ele é viciado. Note também que com o aumento do valor de  $n$  o estimador tende a zero, ou seja, ele é consistente para estimar a probabilidade de sucesso do modelo Binomial.

Esta esperança pode ser reescrita como uma média ponderada da proporção amostral e da média da distribuição *a priori* (Agresti & Hitchcock, 2005), conforme segue:

$$\mathbb{E}(\theta|\mathbf{X}) = w^* \left( \frac{\mathbf{X}}{n} \right) + (1 - w^*) \left( \frac{\alpha}{\alpha + \beta} \right),$$

em que  $w^*$  é dado por  $w^* = \frac{n}{n + \alpha + \beta}$ .

Um problema importante a ser discutido é encontrar os hiperparâmetros para que  $\delta_h(\mathbf{X})$  seja minimax. Lembre-se que pelo Corolário 3.1 uma forma de obter o estimador minimax é tornar a função de risco constante. Assim, pela Definição 3.1, ao considerar o estimador de Bayes com função de perda quadrática da equação (3.5), tem-se

$$R(\theta, \delta_h(\mathbf{X})) = \mathbb{E}_{\mathbf{X}|\theta} \left[ \frac{\mathbf{X} + \alpha}{\alpha + \beta + n} - \theta \right]^2 = \mathbb{E}_{\mathbf{X}|\theta} \left[ \frac{\mathbf{X} + \alpha - \theta(\alpha + \beta + n)}{\alpha + \beta + n} \right]^2$$

$$\begin{aligned}
&= \frac{1}{(\alpha + \beta + n)^2} [n\theta(1 - \theta) + \alpha^2 - 2\theta\alpha(\alpha + \beta) + \theta^2(\alpha + \beta)^2] \\
&= \frac{\alpha^2}{(\alpha + \beta + n)^2} + \frac{\theta^2((\alpha + \beta)^2 - n)}{(\alpha + \beta + n)^2} + \frac{\theta(-2\alpha(\alpha + \beta) + n)}{(\alpha + \beta + n)^2}.
\end{aligned}$$

Note, na última equação, que a função de risco não depende de  $\theta$  se duas condições ocorrem:

$$\begin{cases} (\alpha + \beta)^2 - n = 0 \\ 2\alpha(\alpha + \beta) - n = 0. \end{cases}$$

Desta forma, a distribuição *a priori* conjugada menos favorável para o parâmetro da probabilidade de sucesso do modelo Binomial, ao considerar a função de perda quadrática, é dada pela distribuição Beta  $\left(\frac{\sqrt{n}}{2}, \frac{\sqrt{n}}{2}\right)$ , pela resolução do sistema com as duas restrições. O estimador de Bayes é

$$\begin{aligned}
\delta_h(\mathbf{X}) &= \frac{\mathbf{X} + \frac{\sqrt{n}}{2}}{\sqrt{n} + n} = \left[ \mathbf{X} + \frac{\sqrt{n}}{2} \right] \frac{1}{\sqrt{n} + n} = \\
&= \frac{\mathbf{X}}{\sqrt{n} + n} + \frac{\sqrt{n}}{2} \frac{1}{\sqrt{n} + n} = \frac{\mathbf{X}}{n} \frac{\sqrt{n}}{1 + \sqrt{n}} + \frac{1}{2(1 + \sqrt{n})}.
\end{aligned}$$

O risco de Bayes é constante, portanto, minimax, conforme o Corolário 3.1. Observe que o estimador de Bayes é único; conseqüentemente, o estimador minimax também é único.

Mais adiante nesta seção seguem as discussões e desenvolvimentos das distribuições *a priori* de Haldane, Bayes-Laplace e de Jeffreys, para as quais são obtidas distribuições *a priori* Beta(0, 0), Beta(1, 1) e Beta  $\left(\frac{1}{2}, \frac{1}{2}\right)$ , respectivamente.

A construção da distribuição *a priori* menos favorável, com função de perda quadrática, permite que para alguns tamanhos amostrais específicos as demais distribuições *a priori* mencionadas possam ser recuperadas. Se não forem consideradas observações amostrais, ou seja,  $n = 0$ , tem-se a distribuição *a priori* de Haldane. Ao avaliar com tamanho amostral  $n = 4$ , ela coincide com a distribuição *a priori* de Bayes-Laplace. E ao atribuir apenas uma observação amostral, ela coincide com a distribuição *a priori* de Jeffreys. Contudo, pode-se considerar a distribuição *a priori* menos favorável, com função de perda quadrática, não muito realista, pois depende do tamanho amostral. E

não é muito interessante no sentido em que se aumentamos o valor de  $n$  essa distribuição *a priori* tende a se concentrar em  $\theta = 0,5$ .

Observe graficamente, na figura 3.1, a distribuição *a priori* menos favorável para alguns tamanhos amostrais.

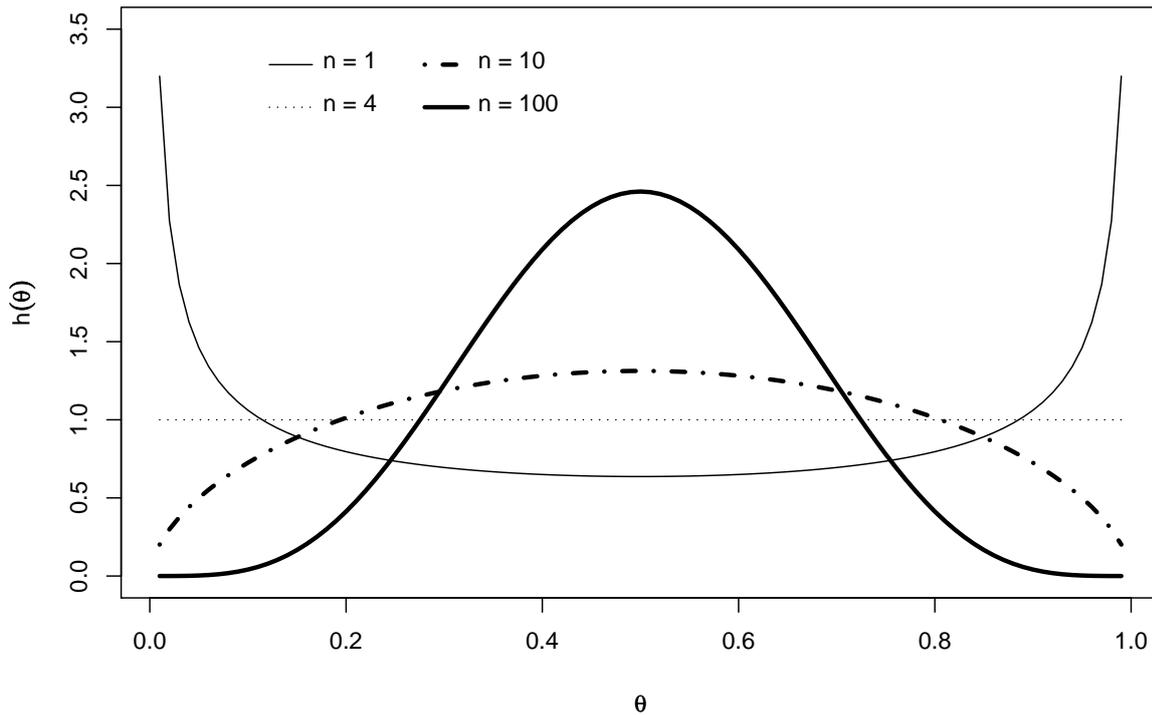


FIGURA 3.1: Distribuição *a priori* menos favorável, ao considerar a função de perda quadrática, para a probabilidade de sucesso do modelo Binomial.

Lembre-se que o estimador minimax depende da função de perda associada. Desta forma, uma proposta deste trabalho é atribuir uma função de perda mais realista, para a qual a distribuição *a priori* menos favorável independa do valor de  $n$ . Como a função de perda quadrática é criticada por penalizar demais o erro de estimação e a variância do modelo Binomial depende de sua média, a função de perda quadrática ponderada é mais razoável.

**Teorema 3.3** Seja a função de perda quadrática ponderada

$$L[\theta, \delta(x)] = w(\theta) [\delta(x) - \theta]^2,$$

em que  $w(\theta) > 0$ . O estimador de Bayes  $\delta_h(\mathbf{X})$  é dado por

$$\delta_h(\mathbf{X}) = \frac{\mathbb{E}_{\theta|\mathbf{X}}[\theta w(\theta)]}{\mathbb{E}_{\theta|\mathbf{X}}[w(\theta)]}.$$

**Prova:** O estimador de Bayes é a regra que minimiza a função (3.2). Desta forma, basta calcular seu mínimo ao considerar a função de perda quadrática ponderada. Então, é necessário calcular

$$\frac{\partial}{\partial \delta(x)} \left[ \int_{\Theta} w(\theta) [\theta - \delta(x)]^2 h(\theta|x) d\theta \right] = 0,$$

ou seja

$$\frac{\partial}{\partial \delta(x)} \left[ \int_{\Theta} \theta^2 w(\theta) h(\theta|x) d\theta - 2\delta(x) \int_{\Theta} \theta w(\theta) h(\theta|x) d\theta + \delta(x)^2 \int_{\Theta} w(\theta) h(\theta|x) d\theta \right] = 0.$$

Assim,  $\delta_h(\mathbf{X})$  é mínimo quando

$$-2 \int_{\Theta} \theta w(\theta) h(\theta|x) d\theta + 2\delta(x) \int_{\Theta} w(\theta) h(\theta|x) d\theta = 0,$$

implicando que

$$\delta_h(x) = \frac{\int_{\Theta} \theta w(\theta) h(\theta|x) d\theta}{\int_{\Theta} w(\theta) h(\theta|x) d\theta}. \quad \blacksquare$$

Ao considerar a distribuição *a posteriori* de  $\theta$  e  $w(\theta) = \frac{1}{\theta(1-\theta)}$ , pelo Teorema 3.3, tem-se que

$$\begin{aligned} \delta_h(x) &= \frac{\frac{\Gamma(\alpha + \beta + n)}{\Gamma(\alpha + x)\Gamma(\beta + n - x)} \int_0^1 \frac{\theta}{\theta(1-\theta)} \theta^{\alpha+x-1} (1-\theta)^{\beta+n-x-1} d\theta}{\frac{\Gamma(\alpha + \beta + n)}{\Gamma(\alpha + x)\Gamma(\beta + n - x)} \int_0^1 \frac{1}{\theta(1-\theta)} \theta^{\alpha+x-1} (1-\theta)^{\beta+n-x-1} d\theta} \\ &= \frac{\int_0^1 \theta^{\alpha+x-1} (1-\theta)^{\beta+n-x-2} d\theta}{\int_0^1 \theta^{\alpha+x-2} (1-\theta)^{\beta+n-x-2} d\theta} = \frac{\frac{\Gamma(\alpha + x)\Gamma(\beta + n - x - 1)}{\Gamma(\alpha + \beta + n - 1)}}{\frac{\Gamma(\alpha + x - 1)\Gamma(\beta + n - x - 1)}{\Gamma(\alpha + \beta + n - 2)}} \\ &= \frac{\alpha + x - 1}{\alpha + \beta + n - 2}. \end{aligned}$$

Portanto o estimador  $\delta_h(\mathbf{X})$  é dado pela moda da distribuição *a posteriori*

$$\delta_h(\mathbf{X}) = \frac{\alpha + \mathbf{X} - 1}{\alpha + \beta + n - 2}. \quad (3.6)$$

Observe no Apêndice B que a esperança desse estimador é diferente de  $\theta$ , ou seja, ele também é viciado, como a média da distribuição *a posteriori*. No entanto, ele é consistente para estimar a probabilidade de sucesso do modelo Binomial.

Voltando ao problema inicialmente mencionado, agora com função de perda quadrática ponderada, quais os valores dos hiperparâmetros  $\alpha$  e  $\beta$  para que o estimador (3.6) seja minimax? Pela Definição 3.1, ao considerar o estimador de Bayes da equação (3.6), tem-se

$$\begin{aligned}
R(\theta, \delta_h(\mathbf{X})) &= \frac{1}{\theta(1-\theta)} \mathbb{E}_{\mathbf{X}|\theta} \left[ \frac{\alpha + \mathbf{X} - 1}{\alpha + \beta + n - 2} - \theta \right]^2 \\
&= \frac{1}{\theta(1-\theta)} \mathbb{E}_{\mathbf{X}|\theta} \left[ \left( \frac{\alpha + \mathbf{X} - 1}{\alpha + \beta + n - 2} \right)^2 - 2\theta \left( \frac{\alpha + \mathbf{X} - 1}{\alpha + \beta + n - 2} \right) + \theta^2 \right] \\
&= \frac{1}{\theta(1-\theta)} \left[ \frac{(\alpha - 1)^2 + n\theta(1-\theta) + n^2\theta^2 + 2n\theta\alpha - 2n\theta}{(\alpha + \beta + n - 2)^2} \right] \\
&\quad - \frac{1}{\theta(1-\theta)} \left[ \frac{2\theta\alpha + 2n\theta^2 - 2\theta}{\alpha + \beta + n - 2} - \theta^2 \right]. \tag{3.7}
\end{aligned}$$

Na equação (3.7) observe que o termo

$$\frac{(\alpha - 1)^2}{\theta(1-\theta)(\alpha + \beta + n - 2)^2}$$

só será constante em relação a  $\theta$  se  $\alpha = 1$ , portanto é a única solução possível. Voltando à equação (3.7), obtém-se o valor de  $\beta$  que satisfaz a condição de risco constante, para que o estimador seja minimax. Tem-se assim

$$\begin{aligned}
R(\theta, \delta_h(\mathbf{X})) &= \frac{1}{\theta(1-\theta)} \left[ \frac{n\theta - n\theta^2 + n^2\theta^2}{(\beta + n - 1)^2} - \frac{2n\theta^2}{\beta + n - 1} + \theta^2 \right] \\
&= \frac{1}{\theta(1-\theta)(\beta + n - 1)^2} [n\theta + n\theta^2 - n^2\theta^2 - 2n\theta^2\beta + \theta^2(\beta + n - 1)^2] \\
&= \frac{1}{\theta(1-\theta)(\beta + n - 1)^2} [n\theta + \theta^2(n - n^2 - 2n\beta + (\beta + n - 1)^2)] \\
&= \frac{1}{\theta(1-\theta)(\beta + n - 1)^2} [n\theta + n\theta^2 + \theta^2(\beta - 1)^2] \\
&= \frac{1}{\theta(1-\theta)(\beta + n - 1)^2} [n(\theta(1-\theta)) + \theta^2(\beta - 1)^2]
\end{aligned}$$

$$= \frac{n}{(\beta + n - 1)^2} + \frac{\theta(\beta - 1)^2}{(1 - \theta)(\beta + n - 1)^2}.$$

Na última equação, a função de risco é constante, ou seja, não depende de  $\theta$  se  $(\beta - 1)^2 = 0$  que implica em  $\beta = 1$ . Logo, a distribuição *a priori* menos favorável para o parâmetro  $\theta$  do modelo Binomial, considerando a função de perda quadrática ponderada, é uma distribuição Beta(1, 1).

O estimador de Bayes é dado por

$$\delta_h(\mathbf{X}) = \frac{\alpha + \mathbf{X} - 1}{\alpha + \beta + n - 2} = \frac{\mathbf{X}}{n},$$

que coincide com o estimador de máxima verossimilhança e concilia a inferência bayesiana com a clássica. Diferentemente do estimador obtido com a função de perda quadrática, este não é viciado para o parâmetro  $\theta$  e também consistente.

O risco de Bayes para a função de perda quadrática ponderada também é constante e único. Portanto, tem-se que o estimador minimax existe e é único.

Por motivos referenciados anteriormente, a função de perda quadrática ponderada é mais interessante para o modelo Binomial do que a função de perda quadrática sem ponderação. Desta forma, seria interessante que o modelo ao utilizar a distribuição *a priori* menos favorável com perda quadrática ponderada mostre melhores propriedades freqüentistas que ao empregar as demais distribuições *a priori* abordadas para o modelo Binomial.

Após a obtenção da distribuição *a posteriori* para  $\theta$ , algumas vezes é necessário resumir esta informação. A seguir é apresentado o estimador pontual da variância, pois os estimadores da média e da moda da distribuição *a posteriori* já foram mencionadas.

O estimador da variância da distribuição *a posteriori* de  $\theta$  baseada na distribuição *a priori* conjugada é dado por

$$\text{Var}(\theta|\mathbf{X}) = \frac{(\mathbf{X} + \alpha)(n - \mathbf{X} + \beta)}{(\alpha + n + \beta)^2(\alpha + n + \beta + 1)}.$$

### 3.1.2 Distribuição *a priori* de Bayes-Laplace

Bayes e Laplace atribuíram densidade *a priori* para a probabilidade de sucesso do modelo Binomial a distribuição Uniforme, que também pode ser escrita em função da

distribuição Beta com  $(\alpha = \beta = 1)$ , conforme metodologia descrita no Capítulo 2.

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

$$\begin{aligned} h(\theta|x) &\propto L_x(\theta) h(\theta) \propto [\theta^x(1-\theta)^{n-x}][\theta^{1-1}(1-\theta)^{1-1}] \\ &\propto \theta^x(1-\theta)^{n-x}, \quad 0 < \theta < 1. \end{aligned} \quad (3.8)$$

A distribuição *a posteriori*  $h(\theta|x)$ , utilizando a distribuição *a priori* de Bayes-Laplace fornece os estimadores:

$$\mathbb{E}(\theta|\mathbf{X}) = \frac{\mathbf{X} + 1}{n + 2}, \quad \text{Moda}(\theta|\mathbf{X}) = \frac{\mathbf{X}}{n} \quad \text{e} \quad \text{Var}(\theta|\mathbf{X}) = \frac{(\mathbf{X} + 1)(n - \mathbf{X} + 1)}{(n + 2)^2(n + 3)}.$$

Note que a moda da distribuição *a posteriori* de  $\theta$  ao considerar a distribuição *a priori* Uniforme coincide com o estimador de máxima verossimilhança para  $\theta$ .

### 3.1.3 Distribuição *a priori* de Haldane

Haldane, em 1948, propôs uma distribuição *a priori* imprópria para  $\theta$  dada por

$$h(\theta) \propto \frac{1}{\theta(1-\theta)}.$$

Embora a distribuição Beta seja própria para valores de  $\alpha > 0$  e  $\beta > 0$ , Haldane sugeriu valores para  $\alpha = 0$  e  $\beta = 0$ , que apesar de gerarem uma distribuição imprópria, são razoáveis para aplicações genéticas, em que a esperança de  $\log(\theta)$  é aproximadamente uniforme para  $\theta$  próximo de zero (Agresti & Hitchcock, 2005). A distribuição *a posteriori*:

$$\begin{aligned} h(\theta|x) &\propto L_x(\theta) h(\theta) \propto [\theta^x(1-\theta)^{n-x}][\theta^{0-1}(1-\theta)^{0-1}] \\ &\propto \theta^{x-1}(1-\theta)^{n-x-1}, \quad 0 < \theta < 1. \end{aligned}$$

Essa distribuição *a posteriori* será imprópria quando  $x = 0$  ou  $x = n$ . Ao considerar a distribuição *a priori* proposta por Haldane, tem-se os estimadores:

$$\mathbb{E}(\theta|\mathbf{X}) = \frac{\mathbf{X}}{n}, \quad \text{Moda}(\theta|\mathbf{X}) = \frac{\mathbf{X} - 1}{n - 2}, \quad \text{para } \mathbf{X} > 0 \text{ e } n > 2 \quad \text{e} \quad \text{Var}(\theta|\mathbf{X}) = \frac{\mathbf{X}(n - \mathbf{X})}{n^2(n + 1)}.$$

### 3.1.4 Distribuição *a priori* de Jeffreys

Outra distribuição *a priori*, mais popular para inferência no modelo Binomial, é a distribuição *a priori* de Jeffreys, que em parte é popular por ser invariante para transformações injetivas para o parâmetro de escala.

A distribuição *a priori* de referência no caso da distribuição Binomial coincide com a distribuição *a priori* de Jeffreys, conforme mencionado no Capítulo 2.

Observe na figura 3.1, para  $n = 1$ , que a distribuição *a priori* obtida atribui maior peso para valores extremos que a probabilidade de sucesso pode assumir e é simétrica em torno de  $\theta = 0,5$ .

A distribuição *a priori* de Jeffreys é proporcional à raiz quadrada da informação de Fisher para o parâmetro de interesse, conforme dito no Capítulo 2. Neste caso, a distribuição *a priori* de Jeffreys é dada pela distribuição Beta  $\left(\frac{1}{2}, \frac{1}{2}\right)$ . Para se obter essa distribuição *a priori* é calculado o logaritmo da verossimilhança, que é

$$\ln h(x|\theta) = \ln \binom{n}{x} + x \ln(\theta) + (n-x) \ln(1-\theta).$$

Para obter a informação de Fisher é necessário conhecer a segunda derivada do logaritmo da verossimilhança:

$$\frac{\partial \ln h(x|\theta)}{\partial \theta} = \frac{x}{\theta} - \frac{n-x}{1-\theta} \quad \text{e} \quad \frac{\partial^2 \ln h(x|\theta)}{\partial \theta^2} = -\frac{x}{\theta^2} - \frac{n-x}{(1-\theta)^2}.$$

Sabendo que a distribuição Binomial pertence a família exponencial (Casella & Berger, 2002, p. 111) e ao utilizar a informação de Fisher do Lema 2.1, então é obtida

$$I(\theta) = -\mathbb{E}_{\mathbf{X}|\theta} \left[ -\frac{\mathbf{X}}{\theta^2} - \frac{n-\mathbf{X}}{(1-\theta)^2} \right] \propto \frac{1}{\theta(1-\theta)}.$$

Portanto, pela Definição 2.3, a distribuição *a priori* de Jeffreys é dada por

$$h(\theta) \propto \sqrt{\frac{1}{\theta(1-\theta)}}.$$

A distribuição *a posteriori* de  $\theta|x$ , utilizando a distribuição *a priori* de Jeffreys, pela equação (3.4) é dada por uma distribuição Beta  $\left(x + \frac{1}{2}, n-x + \frac{1}{2}\right)$ , assim:

$$\mathbb{E}(\theta|\mathbf{X}) = \frac{\mathbf{X} + \frac{1}{2}}{n + 1}, \quad \text{Moda}(\theta|\mathbf{X}) = \frac{\mathbf{X} - \frac{1}{2}}{n - 1}, \text{ para } \mathbf{X} > 0 \text{ e } n > 1$$

$$\text{e } \text{Var}(\theta|\mathbf{X}) = \frac{\left(\mathbf{X} + \frac{1}{2}\right) \left(n - \mathbf{X} + \frac{1}{2}\right)}{(n + 1)^2(n + 2)}.$$

## 3.2 Estimação Intervalar

Algumas metodologias têm sido desenvolvidas para se obter intervalos para a probabilidade de sucesso da distribuição Binomial. Neste texto foram explorados intervalos de confiança e de credibilidade obtidos pelas metodologias clássica e bayesiana.

No que se refere a intervalos bayesianos, foram considerados o intervalo de credibilidade, o intervalo HPD (máxima densidade *a posteriori*) e o intervalo de credibilidade a partir da distribuição  $F$  de Fisher. Quanto à metodologia clássica, foram abordados o intervalo obtido com aproximação pela distribuição Normal, o intervalo exato de Clopper & Pearson (1934) e o intervalo com probabilidade de cobertura garantida de Chen *et al.* (2008).

### 3.2.1 Intervalo de credibilidade

Um intervalo de credibilidade é dado pelos quantis  $(Q_{\epsilon/2}; Q_{1-\epsilon/2})$  da distribuição *a posteriori* de  $\theta|x$  fixando um nível de credibilidade  $\epsilon$ . É interpretado como uma probabilidade igual a  $100(1 - \epsilon)\%$  do verdadeiro parâmetro  $\theta$  pertencer ao intervalo.

Para as distribuições *a priori* mencionadas, as distribuições *a posteriori* são Beta, conforme visto; portanto, basta tomar os valores correspondentes aos quantis desejados para obtenção do intervalo de credibilidade.

### 3.2.2 Intervalo de credibilidade a partir da distribuição $F$ de Fisher

**Teorema 3.4** (Agresti & Hitchcock, 2005) A distribuição *a posteriori* de  $\theta$ , ou seja, a distribuição  $\text{Beta}(x+\alpha, n-x+\beta)$  pode ser reescrita como sendo uma função da distribuição  $F$  de Fisher, conforme segue

$$\frac{\left(\frac{\mathbf{X} + \alpha}{n - \mathbf{X} + \beta}\right) F^*}{1 + \left(\frac{\mathbf{X} + \alpha}{n - \mathbf{X} + \beta}\right) F^*}, \quad (3.9)$$

em que  $F^*$  descreve uma variável com distribuição  $F$  de Fisher com  $n_1 = 2(x + \alpha)$  e  $n_2 = 2(n - x + \beta)$  graus de liberdade.

A demonstração do Teorema 3.4 será omitida (ver demonstração do Lema 3.2 no Apêndice A).

Casella & Berger (2002, p. 456) mencionam como exercício a construção do intervalo de credibilidade com  $100(1 - \epsilon)\%$  baseado na função (3.9); em que o limite inferior  $\theta_I$  é obtido pela

$$\mathbb{P} \left[ \theta \geq \theta_I \mid \mathbf{X} = x \right] = 1 - \frac{\epsilon}{2}.$$

Substituindo  $\theta$  pela função (3.9), tem-se

$$\mathbb{P} \left[ \frac{\left(\frac{x + \alpha}{n - x + \beta}\right) F^*}{1 + \left(\frac{x + \alpha}{n - x + \beta}\right) F^*} \geq \theta_I \mid \mathbf{X} = x \right] = 1 - \frac{\epsilon}{2}.$$

Desta forma,

$$\mathbb{P} \left[ \left(\frac{x + \alpha}{n - x + \beta}\right) F^* \geq \theta_I + \theta_I \left(\frac{x + \alpha}{n - x + \beta}\right) F^* \mid \mathbf{X} = x \right] = 1 - \frac{\epsilon}{2},$$

então

$$\mathbb{P} \left[ \left(\frac{x + \alpha}{n - x + \beta}\right) F^* - \theta_I \left(\frac{x + \alpha}{n - x + \beta}\right) F^* \geq \theta_I \mid \mathbf{X} = x \right] = 1 - \frac{\epsilon}{2},$$

logo,

$$\mathbb{P} \left[ F^* \geq \frac{\theta_I}{\left(\frac{x + \alpha}{n - x + \beta}\right) (1 - \theta_I)} \mid \mathbf{X} = x \right] = 1 - \frac{\epsilon}{2}.$$

Como  $F^*$  representa uma variável com distribuição  $F$  de Fisher com  $n_1 = 2(x + \alpha)$  e  $n_2 = 2(n - x + \beta)$  graus de liberdade, esse quantil é igual a

$$F_{2(x+\alpha), 2(n-x+\beta), \epsilon/2} = \frac{\theta_I}{\left(\frac{x+\alpha}{n-x+\beta}\right) (1-\theta_I)}.$$

Então,

$$F_{2(x+\alpha), 2(n-x+\beta), \epsilon/2} \left(\frac{x+\alpha}{n-x+\beta}\right) - \theta_I F_{2(x+\alpha), 2(n-x+\beta), \epsilon/2} \left(\frac{x+\alpha}{n-x+\beta}\right) = \theta_I,$$

logo,

$$\theta_I + \theta_I F_{2(x+\alpha), 2(n-x+\beta), \epsilon/2} \left(\frac{x+\alpha}{n-x+\beta}\right) = F_{2(x+\alpha), 2(n-x+\beta), \epsilon/2} \left(\frac{x+\alpha}{n-x+\beta}\right).$$

Conclui-se que o limite inferior do intervalo de credibilidade é dado por

$$\theta_I = \frac{F_{2(x+\alpha), 2(n-x+\beta), \epsilon/2} \left(\frac{x+\alpha}{n-x+\beta}\right)}{1 + F_{2(x+\alpha), 2(n-x+\beta), \epsilon/2} \left(\frac{x+\alpha}{n-x+\beta}\right)}.$$

Analogamente, o limite superior  $\theta_S$  é obtido ao considerar

$$P \left[ \theta \leq \theta_S \mid \mathbf{X} = x \right] = \frac{\epsilon}{2}.$$

Portanto,

$$\theta_S = \frac{F_{2(x+\alpha), 2(n-x+\beta), 1-\epsilon/2} \left(\frac{x+\alpha}{n-x+\beta}\right)}{1 + F_{2(x+\alpha), 2(n-x+\beta), 1-\epsilon/2} \left(\frac{x+\alpha}{n-x+\beta}\right)}. \quad \blacksquare$$

Desta forma, os limites do intervalo com  $100(1 - \epsilon)\%$  de credibilidade são obtidos por

$$\left[ \frac{F_{2(\mathbf{X}+\alpha), 2(n-\mathbf{X}+\beta), \epsilon/2} \left(\frac{\mathbf{X}+\alpha}{n-\mathbf{X}+\beta}\right)}{1 + F_{2(\mathbf{X}+\alpha), 2(n-\mathbf{X}+\beta), \epsilon/2} \left(\frac{\mathbf{X}+\alpha}{n-\mathbf{X}+\beta}\right)}; \frac{F_{2(\mathbf{X}+\alpha), 2(n-\mathbf{X}+\beta), 1-\epsilon/2} \left(\frac{\mathbf{X}+\alpha}{n-\mathbf{X}+\beta}\right)}{1 + F_{2(\mathbf{X}+\alpha), 2(n-\mathbf{X}+\beta), 1-\epsilon/2} \left(\frac{\mathbf{X}+\alpha}{n-\mathbf{X}+\beta}\right)} \right].$$

Vale ressaltar que a escolha dos limites do intervalo baseia-se em iguais probabilidades nas caudas, no entanto outras possibilidades poderiam ser consideradas. Esse intervalo é equivalente ao intervalo obtido diretamente pela distribuição Beta. Porém, uma “vantagem” dele é a facilidade de encontrar tabelas com os quantis da distribuição  $F$  de Fisher em livros de estatística, no entanto com o avanço computacional essa vantagem pode ser diminuída.

### 3.2.3 Intervalo HPD (máxima densidade *a posteriori*)

Os intervalos HPD (Bernardo & Smith, 2000, p. 260) têm duas propriedades:

1. A densidade para qualquer ponto pertencente ao intervalo é maior que a densidade para qualquer não pertencente ao intervalo.
2. Para uma dada probabilidade de cobertura  $100(1 - \epsilon)\%$  o intervalo é o de menor comprimento.

Quando a distribuição *a posteriori* é unimodal e simétrica, o intervalo HPD coincide com o intervalo de credibilidade e tem a mesma interpretação que o intervalo de credibilidade citado na Seção 3.2.1. Caso a distribuição *a posteriori* não seja unimodal, pode resultar na união de intervalos.

Um método para obtenção dos intervalos HPD é a aproximação de Monte Carlo (Chen & Shao, 1998), aplicado quando a distribuição *a posteriori* é unimodal e contínua, conforme o seguinte algoritmo:

1. Obter uma amostra aleatória de tamanho  $n$ , para  $\theta_i, i = 1, 2, \dots, n$  para a distribuição *a posteriori* de  $\theta|x$ .
2. Ordenar os valores  $\theta_i, i = 1, 2, \dots, n$ , resultando em  $\theta_{(1)} \leq \dots \leq \theta_{(n)}$ .
3. Calcular os intervalos com  $100(1 - \epsilon)\%$  de cobertura

$$C_{j(n)} = (\theta_{(j)}, \theta_{(j)+[(1-\epsilon)n]})$$

para  $j = 1, 2, \dots, n - [(1 - \epsilon)n]$ , em que  $(1 - \epsilon)n$  é um número inteiro.

4. O intervalo HPD  $100(1 - \epsilon)\%$  de cobertura é o intervalo de menor amplitude dentre os obtidos.

### 3.2.4 Intervalo de confiança com aproximação pela distribuição Normal

Um método clássico para obtenção de um intervalo de confiança para a probabilidade de sucesso da distribuição Binomial é dado pela aproximação com a distribuição

Normal, cuja base é o Teorema Central do Limite, conforme mostram Chen *et al.* (2008), entre outros.

Esse intervalo é conhecido na literatura como intervalo de confiança de Wald, historicamente foi um dos primeiros intervalos de confiança propostos para um parâmetro (Agresti & Coull, 1998). Portanto, pelo Teorema Central do Limite tem-se que

$$\lim_{n \rightarrow \infty} \mathbb{P} \left\{ \frac{|\hat{\theta} - \theta|}{\sqrt{\frac{\theta(1-\theta)}{n}}} < z \right\} = 2\Phi(z) - 1,$$

em que  $\hat{\theta}$  é o estimador de máxima verossimilhança de  $\theta$ , dado por  $\frac{\mathbf{X}}{n}$ ;  $z > 0$  e  $\Phi(\cdot)$  é a função distribuição acumulada da distribuição Normal padrão.

Seja  $z_{\epsilon/2}$  o valor que satisfaz  $\Phi(z_{\epsilon/2}) = 1 - \frac{\epsilon}{2}$ . Então, segue que

$$\lim_{n \rightarrow \infty} \mathbb{P} \left\{ \hat{\theta} - z_{\epsilon/2} \sqrt{\frac{\theta(1-\theta)}{n}} < \theta < \hat{\theta} + z_{\epsilon/2} \sqrt{\frac{\theta(1-\theta)}{n}} \right\} = 1 - \epsilon.$$

Esses limites podem ser obtidos por

$$\left[ \hat{\theta} - z_{\epsilon/2} \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} \quad ; \quad \hat{\theta} + z_{\epsilon/2} \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} \right].$$

A qualidade da aproximação para amostras pequenas pode ser baixa devido ao resultado assintótico. Um problema em aplicações com eventos raros é que não se sabe quão suficientemente grande é o tamanho da amostra para o erro da aproximação ser desprezível.

Um intervalo de confiança é interpretado da seguinte maneira: se replicarmos a amostra  $n$  vezes, o intervalo de confiança, que é aleatório, conterá o verdadeiro parâmetro  $\theta$  fixo e desconhecido em  $100(1 - \epsilon)\%$  das vezes; o intervalo de confiança é, portanto, baseado em um contexto freqüentista.

### 3.2.5 Intervalo de confiança exato de Clopper & Pearson

Clopper & Pearson (1934) propuseram um intervalo de confiança exato, alternativo ao intervalo com aproximação pela distribuição Normal, para a probabilidade de sucesso da distribuição Binomial, conforme o teorema a seguir.

**Teorema 3.5** (Casella & Berger, 2002, p. 454) Seja  $X \sim \text{Bin}(n, \theta)$ . Aplicando a relação entre a distribuição Binomial e a distribuição Beta, tem-se o intervalo aleatório exato

$$\left[ \frac{1}{1 + \frac{n - \mathbf{X} + 1}{\mathbf{X}} F_{2(n - \mathbf{X} + 1), 2\mathbf{X}, \epsilon/2}} ; \frac{\frac{\mathbf{X} + 1}{n - \mathbf{X}} F_{2(\mathbf{X} + 1), 2(n - \mathbf{X}), \epsilon/2}}{1 + \frac{\mathbf{X} + 1}{n - \mathbf{X}} F_{2(\mathbf{X} + 1), 2(n - \mathbf{X}), \epsilon/2}} \right],$$

em que  $\mathbf{X} = 1, 2, \dots, n - 1$ . Para  $\mathbf{X} = 0$  o limite inferior é 0 e para  $\mathbf{X} = n$ , o limite superior é 1.

Para obter esse intervalo exato são necessários os seguintes lemas:

**Lema 3.1** (Casella & Berger, 2002, p. 454) Se  $X \sim \text{Bin}(n, \theta)$ , então  $\mathbb{P}_\theta(X \geq x) = \mathbb{P}(X_1 \leq \theta)$ , em que  $X_1 \sim \text{Beta}(x, n - x + 1)$ .

**Lema 3.2** (Casella & Berger, 2002, p. 225)

$$\text{Se } X_2 \sim F_{(p,q)}, \text{ então } Y_1 = \frac{\frac{p}{q} X_2}{1 + \frac{p}{q} X_2} \sim \text{Beta}\left(\frac{p}{2}, \frac{q}{2}\right).$$

**Lema 3.3** (Casella & Berger, 2002, p. 225) Se  $U \sim F_{(p,q)}$ , então  $\frac{1}{U} \sim F_{(q,p)}$ ; ou seja, a recíproca de uma variável aleatória  $F$  ainda é uma variável aleatória  $F$ .

As demonstrações dos Lemas 3.1 e 3.2 encontram-se no Apêndice A e a demonstração do Lema 3.3 será omitida.

Para construir um intervalo de confiança para  $\theta$  da forma

$$\{\theta : \theta_I \leq \theta \leq \theta_S | \mathbf{X} = x\} \text{ com } \mathbb{P}[\theta_I \leq \theta \leq \theta_S | \mathbf{X} = x] = 1 - \epsilon,$$

ao considerar o método do pivô baseado na função distribuição acumulada (Cox, 2006, p. 25; Casella & Berger, 2002, p. 434), é necessário obter  $\theta_I$  e  $\theta_S$  nas equações

$$\sum_{k=0}^x \binom{n}{k} \theta_S^k (1 - \theta_S)^{n-k} = \frac{\epsilon}{2} \quad (3.10)$$

e

$$\sum_{k=x}^n \binom{n}{k} \theta_I^k (1 - \theta_I)^{n-k} = \frac{\epsilon}{2}. \quad (3.11)$$

Pela equação (3.10) tem-se que

$$\frac{\epsilon}{2} = \mathbb{P}(K \leq x) = 1 - \mathbb{P}(K > x) = 1 - \mathbb{P}(K \geq x + 1).$$

Pela aplicação do Lema 3.1

$$1 - \mathbb{P}(K \geq x + 1) = 1 - \mathbb{P}(X_1 \leq \theta_S),$$

em que  $X_1 \sim \text{Beta}(x + 1, n - x)$ ,  $K \sim \text{Bin}(n, \theta_S)$  e  $\theta_S$  corresponde ao limite superior do intervalo. A última igualdade pode ser escrita como  $1 - \mathbb{P}(X_1 \leq \theta_S) = \mathbb{P}(X_1 > \theta_S)$ .

Sejam  $U \sim F_{2(x+1), 2(n-x)}$  e  $c = \left(\frac{x+1}{n-x}\right)$ . Pelo Lema 3.2 tem-se  $\frac{cU}{1+cU} \sim \text{Beta}(x+1, n-x)$ . Então, ao substituir  $X_1$  por  $\frac{cU}{1+cU}$  tem-se

$$\frac{\epsilon}{2} = \mathbb{P}\left[\frac{cU}{1+cU} > \theta_S\right] = \mathbb{P}\left[U > \frac{\theta_S}{c(1-\theta_S)}\right].$$

Desta forma, conclui-se que

$$\frac{\theta_S}{c(1-\theta_S)} = F_{2(x+1), 2(n-x), \epsilon/2}.$$

Resolvendo esta equação em  $\theta_S$  obtém-se

$$\theta_S = \frac{\frac{x+1}{n-x} F_{2(x+1), 2(n-x), \epsilon/2}}{1 + \frac{x+1}{n-x} F_{2(x+1), 2(n-x), \epsilon/2}}.$$

Portanto  $\theta_S$  é dado por 
$$\frac{\frac{\mathbf{X}+1}{n-\mathbf{X}} F_{2(\mathbf{X}+1), 2(n-\mathbf{X}), \epsilon/2}}{1 + \frac{\mathbf{X}+1}{n-\mathbf{X}} F_{2(\mathbf{X}+1), 2(n-\mathbf{X}), \epsilon/2}}.$$

Para calcular o limite inferior, pela equação (3.11) é obtida a relação  $\frac{\epsilon}{2} = \mathbb{P}(K \geq x)$ , que pela aplicação do Lema 3.1 conduz a

$$\mathbb{P}(K \geq x) = \mathbb{P}(X_1 \leq \theta_I),$$

em que  $X_1 \sim \text{Beta}(x, n - x + 1)$  e  $K \sim \text{Bin}(n, \theta_I)$ . Seja  $U \sim F_{2x, 2(n-x+1)}$  e  $c = \left(\frac{x}{n-x+1}\right)$ , pelo Lema 3.2 tem-se  $\frac{cU}{1+cU} \sim \text{Beta}(x, n-x+1)$ , então

$$\frac{\epsilon}{2} = \mathbb{P}\left[\frac{cU}{1+cU} \leq \theta_I\right] = \mathbb{P}\left[U \leq \frac{\theta_I}{c(1-\theta_I)}\right].$$

Desta forma, conclui-se que

$$\frac{\theta_I}{c(1 - \theta_I)} = F_{2x, 2(n-x+1), \epsilon/2}.$$

Resolvendo esta equação em  $\theta_I$  obtém-se

$$\theta_I = \frac{\frac{x}{n-x+1} F_{2x, 2(n-x+1), \epsilon/2}}{1 + \frac{x}{n-x+1} F_{2x, 2(n-x+1), \epsilon/2}}. \quad (3.12)$$

Ao dividir o numerador e denominador do lado direito da equação (3.12) por  $1 + \frac{x}{n-x+1} F_{2x, 2(n-x+1), \epsilon/2}$  e considerar o resultado do Lema 3.3 é obtido

$$\theta_I = \frac{1}{1 + \frac{n-x+1}{x} F_{2(n-x+1), 2x, \epsilon/2}}.$$

Assim,  $\theta_I$  é dado por 
$$\frac{1}{1 + \frac{n - \mathbf{X} + 1}{\mathbf{X}} F_{2(n - \mathbf{X} + 1), 2\mathbf{X}, \epsilon/2}}.$$

### 3.2.6 Intervalo com probabilidade de cobertura garantida

Intervalos de confiança com probabilidade de cobertura garantida foram propostos por Chen *et al.* (2008). Esses autores sugeriram uma fórmula explícita para a construção do intervalo de confiança para a probabilidade de sucesso da distribuição Binomial. Eles afirmam que essa fórmula supera a limitação encontrada na aproximação assintótica pela distribuição Normal, apresenta-se menos conservadora em comparação com a amplitude do intervalo exato clássico de Clopper & Pearson (1934) e ainda possibilita a obtenção de um excelente desempenho da probabilidade de cobertura.

**Teorema 3.6** (Chen *et al.*, 2008) Os limites do intervalo para  $\theta$  (probabilidade de sucesso da distribuição Binomial), com  $\mathbb{P}\{\theta_I < \theta < \theta_S | \mathbf{X} = x\} > 1 - \epsilon$ , são definidos por

$$\theta_I = \frac{\mathbf{X}}{n} + \frac{3}{4} \frac{1 - \frac{2\mathbf{X}}{n} + \sqrt{1 + 4\eta\mathbf{X} \left(1 - \frac{\mathbf{X}}{n}\right)}}{1 + \eta n}$$

e

$$\theta_S = \frac{\mathbf{X}}{n} + \frac{3}{4} \frac{1 - \frac{2\mathbf{X}}{n} - \sqrt{1 + 4\eta\mathbf{X} \left(1 - \frac{\mathbf{X}}{n}\right)}}{1 + \eta n},$$

em que  $\mathbf{X} = 0, 1, \dots, n$ , com  $\eta = \frac{9}{8 \ln \frac{2}{\epsilon}}$ .

A demonstração do intervalo está apresentada em Chen *et al.* (2008).

O intervalo com 95% de confiança é aproximado por

$$\left[ \frac{\mathbf{X}}{n} + \frac{3}{4} \frac{1 - \frac{2\mathbf{X}}{n} - \sqrt{1 + 2\mathbf{X} \left(1 - \frac{\mathbf{X}}{n}\right)}}{1 + \frac{n}{2}} ; \frac{\mathbf{X}}{n} + \frac{3}{4} \frac{1 - \frac{2\mathbf{X}}{n} + \sqrt{1 + 2\mathbf{X} \left(1 - \frac{\mathbf{X}}{n}\right)}}{1 + \frac{n}{2}} \right].$$

# Capítulo 4

## Distribuição Binomial Negativa

A distribuição Binomial conta o número de sucessos em um número fixo de ensaios de Bernoulli independentes. Suponha agora que o interesse seja contar o número de ensaios de Bernoulli necessários para se obter um número de sucessos fixo. Essa formulação conduz à distribuição Binomial Negativa (Casella & Berger, 2002, p. 82).

Em uma seqüência de ensaios de Bernoulli independentes, a variável aleatória  $X$  denota a quantidade de ensaios até a ocorrência do  $r$ -ésimo sucesso, em que  $r$  é um número inteiro fixo (Casella & Berger, 2002, p. 82). Assim,  $X|\theta$  tem distribuição condicional Binomial Negativa, com parâmetros  $r$  e  $\theta$ ,

$$h(x|\theta) = \binom{x-1}{r-1} \theta^r (1-\theta)^{x-r}, \quad (4.1)$$

no qual  $0 < \theta < 1$ ,  $r$  é inteiro,  $x = r, r+1, r+2, \dots$  e a estatística suficiente para  $\theta$  é dada por  $\mathbf{X} = \sum_{i=1}^n X_i$ .

Ao considerar o número de insucessos igual a 1 ( $r = 1$ ) a distribuição Geométrica é obtida, sendo portanto um caso particular da distribuição Binomial Negativa.

O parâmetro  $\theta$ , presente nas distribuições Binomial e Binomial Negativa, tem a mesma interpretação. Ele retrata a probabilidade de sucesso, a diferença está associada ao planejamento amostral.

## 4.1 Distribuições *a priori* e distribuições *a posteriori*

A distribuição *a priori* conjugada para a probabilidade de sucesso do modelo Binomial Negativa é a distribuição Beta, pois o núcleo da distribuição Binomial Negativa é proporcional ao núcleo da distribuição Beta. Assim, a distribuição *a posteriori* também será uma distribuição Beta com parâmetros  $\alpha$  e  $\beta$ .

Ao considerar a função de verossimilhança  $L_x(\theta) = h(x|\theta)$ , a densidade *a posteriori*  $h(\theta|x)$  pode ser obtida pelo Teorema de Bayes, como segue:

$$\begin{aligned} h(\theta|x) &= \frac{L_x(\theta) h(\theta)}{\int_{\Theta} L_x(\theta) h(\theta) d\theta} \propto L_x(\theta) h(\theta) \\ &\propto [\theta^r (1-\theta)^{x-r}] [\theta^{\alpha-1} (1-\theta)^{\beta-1}] \propto \theta^{r+\alpha-1} (1-\theta)^{x-r+\beta-1}, \end{aligned}$$

em que  $x = r, r+1, \dots$  e  $0 < \theta < 1$ . Assim, a distribuição *a posteriori* satisfaz  $\theta|x \sim \text{Beta}(r+\alpha, x-r+\beta)$ . Após a obtenção da distribuição *a posteriori* para  $\theta$ , alguns estimadores pontuais para a probabilidade de sucesso da distribuição Binomial Negativa são dados por

$$\begin{aligned} \mathbb{E}(\theta|\mathbf{X}) &= \frac{r+\alpha}{\alpha+\beta+\mathbf{X}}, & \text{Moda}(\theta|\mathbf{X}) &= \frac{r+\alpha-1}{\alpha+\beta+\mathbf{X}-2} \quad \text{e} \\ \text{Var}(\theta|\mathbf{X}) &= \frac{(r+\alpha)(\mathbf{X}-r+\beta)}{(\alpha+\beta+\mathbf{X})^2(\alpha+\beta+\mathbf{X}+1)}. \end{aligned}$$

Observe no Apêndice B que a esperança dos estimadores média e moda da distribuição *a posteriori* de  $\theta$  são diferente de  $\theta$ , ou seja, eles são viciados.

### 4.1.1 Distribuição *a priori* menos favorável

No Capítulo 3 foram feitas referências à distribuição *a priori* menos favorável baseada no contexto de teoria de decisões minimax, no qual foi possível obter uma distribuição *a priori* não-informativa que permite resgatar as demais mencionadas no Capítulo 2 ao se considerar a função de perda quadrática para o modelo Binomial. Agora, o interesse é desenvolver um estudo semelhante com o modelo Binomial Negativa.

Para obter a distribuição *a priori* menos favorável, pelo Corolário 3.1, é necessário encontrar valores para os hiperparâmetros  $\alpha$  e  $\beta$  da distribuição *a priori* para que a função

de risco seja constante.

Ao considerar a função de perda quadrática, pelo Teorema 3.2, tem-se que  $\delta_h(\mathbf{X})$  é dado pela média da distribuição *a posteriori*. Desta forma, pela Definição 3.1, ao considerar esse estimador de Bayes, tem-se que

$$R(\theta, \delta_h(\mathbf{X})) = \mathbb{E}_{\mathbf{X}|\theta} \left[ \frac{r + \alpha}{\alpha + \beta + \mathbf{X}} - \theta \right]^2.$$

Tanto no modelo Binomial como no modelo Binomial Negativo a variância depende de sua média, sugerindo a investigação da função de perda quadrática ponderada. O estimador de Bayes ao considerar a função de perda quadrática ponderada com fator de ponderação  $w(\theta) = \frac{(1 - \theta)}{\theta^2}$  é obtido, pelo Teorema 3.3, conforme segue:

$$\begin{aligned} \delta_h(x) &= \frac{\frac{\Gamma(\alpha + \beta + x)}{\Gamma(r + \alpha)\Gamma(x - r + \beta)} \int_0^1 \frac{\theta(1 - \theta)}{\theta^2} \theta^{r+\alpha-1} (1 - \theta)^{x-r+\beta-1} d\theta}{\frac{\Gamma(\alpha + \beta + x)}{\Gamma(r + \alpha)\Gamma(x - r + \beta)} \int_0^1 \frac{(1 - \theta)}{\theta^2} \theta^{r+\alpha-1} (1 - \theta)^{x-r+\beta-1} d\theta} \\ &= \frac{\int_0^1 \theta^{r+\alpha-2} (1 - \theta)^{x-r+\beta} d\theta}{\int_0^1 \theta^{r+\alpha-3} (1 - \theta)^{x-r+\beta} d\theta} = \frac{\frac{\Gamma(r + \alpha - 1)\Gamma(x - r + \beta + 1)}{\Gamma(\alpha + \beta + x)}}{\frac{\Gamma(r + \alpha - 2)\Gamma(x - r + \beta + 1)}{\Gamma(\alpha + \beta + x - 1)}} \\ &= \frac{r + \alpha - 2}{\alpha + \beta + x - 1}. \end{aligned}$$

Portanto o estimador é dado por

$$\delta_h(\mathbf{X}) = \frac{r + \alpha - 2}{\alpha + \beta + \mathbf{X} - 1}. \quad (4.2)$$

Novamente estamos interessados nos valores de  $\alpha$  e  $\beta$  para que este estimador seja minimax. Pela Definição 3.1, ao considerar o estimador de Bayes da equação (4.2), tem-se que

$$R(\theta, \delta_h(\mathbf{X})) = \frac{(1 - \theta)}{\theta^2} \mathbb{E}_{\mathbf{X}|\theta} \left[ \frac{r + \alpha - 2}{\alpha + \beta + \mathbf{X} - 1} - \theta \right]^2.$$

Foi mencionada no Capítulo 3 a dificuldade em obter a expressão exata dos estimadores minimax. Diante das dificuldades observadas nos cálculos considerados para encontrar os valores de  $\alpha$  e  $\beta$  que tornassem as funções de risco constante, na função de perda quadrática e função de perda quadrática ponderada, não foi possível identificar os estimadores minimax. Conseqüentemente deixamos como uma proposta futura apresentar a distribuição *a priori* menos favorável para o modelo Binomial Negativa.

### 4.1.2 Distribuição *a priori* de Bayes-Laplace

A distribuição *a priori* de Bayes-Laplace foi inicialmente proposta como 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 ( $\alpha = \beta = 1$ ) e é constante em relação ao espaço paramétrico, conforme pode ser observado na figura 3.1, para  $n = 4$ .

Com a distribuição *a priori* Uniforme, a distribuição *a posteriori* é proporcional à função de verossimilhança. A demonstração é análoga ao caso Binomial, equação (3.8).

A distribuição *a posteriori*  $\theta|x \sim \text{Beta}(r + 1, x - r + 1)$  fornece

$$\mathbb{E}(\theta|\mathbf{X}) = \frac{r + 1}{\mathbf{X} + 2}, \quad \text{Moda}(\theta|\mathbf{X}) = \frac{r}{\mathbf{X}} \quad \text{e} \quad \text{Var}(\theta|\mathbf{X}) = \frac{(r + 1)(\mathbf{X} - r + 1)}{(\mathbf{X} + 2)^2(\mathbf{X} + 3)}.$$

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

### 4.1.3 Distribuição *a priori* de Jeffreys

A distribuição *a priori* de Jeffreys é 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 o Capítulo 2. Neste caso, a distribuição *a priori* de Jeffreys fica como demonstrado abaixo:

$$h(\theta) \propto \frac{1}{\theta(1 - \theta)^{1/2}}.$$

Calcula-se o logaritmo da função de verossimilhança

$$\ln h(x|\theta) = \ln \binom{x-1}{r-1} + r \ln(\theta) + (x-r) \ln(1-\theta).$$

Como a distribuição Binomial Negativa pertence à família exponencial e, devido ao Lema 2.1 e aos resultados das derivadas abaixo:

$$\frac{\partial \ln h(x|\theta)}{\partial \theta} = \frac{r}{\theta} - \frac{(x-r)}{(1-\theta)}$$

e

$$\frac{\partial^2 \ln h(x|\theta)}{\partial \theta^2} = -\frac{r}{\theta^2} - \frac{(x-r)}{(1-\theta)^2},$$

obteve-se

$$I(\theta) = -\mathbb{E}_{\mathbf{X}|\theta} \left[ -\frac{r}{\theta^2} - \frac{(\mathbf{X}-r)}{(1-\theta)^2} \right] = \frac{r}{\theta^2(1-\theta)}.$$

A distribuição *a priori* de Jeffreys é dada por  $h(\theta) = I(\theta)^{1/2}$ , conforme a Definição 2.3, ou seja,

$$h(\theta) \propto \sqrt{\frac{1}{\theta^2(1-\theta)}}.$$

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

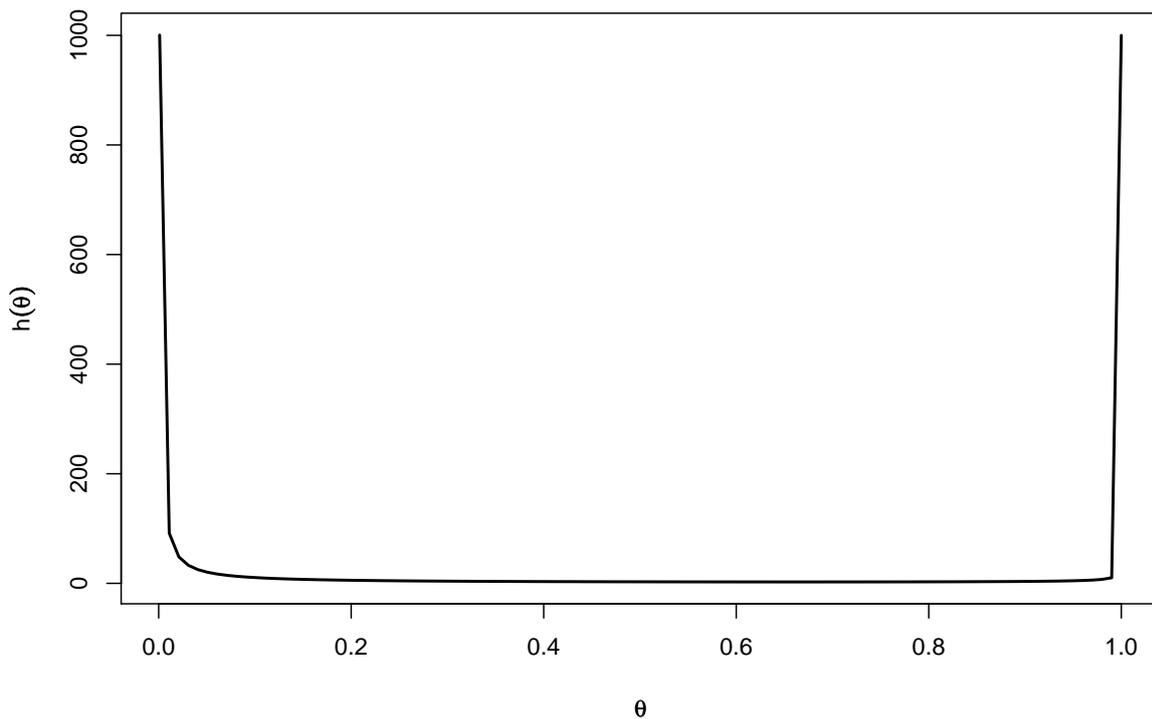


FIGURA 4.1: Distribuição *a priori* de Jeffreys para a probabilidade de sucesso do modelo Binomial Negativa.

Na figura 4.1 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 de a 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 em (4.1), tem-se a distribuição *a posteriori* para  $\theta|x$ :

$$h(\theta|x) \propto L_x(\theta) h(\theta) \propto [\theta^r(1-\theta)^{x-r}][\theta^{-1}(1-\theta)^{-1/2}] \propto \theta^{r-1}(1-\theta)^{x-r-1/2},$$

isto é,  $\theta|x \sim \text{Beta}\left(r, x-r+\frac{1}{2}\right)$ . Assim,

$$\mathbb{E}(\theta|\mathbf{X}) = \frac{r}{\mathbf{X} + \frac{1}{2}}, \quad \text{Moda}(\theta|\mathbf{X}) = \frac{r-1}{\mathbf{X} - \frac{3}{2}}, \quad \text{para } \mathbf{X} > 1 \quad \text{e}$$

$$\text{Var}(\theta|\mathbf{X}) = \frac{r\left(\mathbf{X} - r + \frac{1}{2}\right)}{\left(\mathbf{X} + \frac{1}{2}\right)^2 \left(\mathbf{X} + \frac{3}{2}\right)}.$$

## 4.2 Estimação Intervalar

Para a estimação intervalar da probabilidade de sucesso de uma distribuição Binomial Negativa foram utilizadas as metodologias clássica e bayesiana. Na abordagem bayesiana foram considerados o intervalo de credibilidade e o intervalo HPD (máxima densidade *a posteriori*). Já para obtenção de intervalos clássicos, adotaram-se intervalos com aproximação pela distribuição Normal e o intervalo exato de Clopper & Pearson.

### 4.2.1 Intervalo de credibilidade

Para obtenção dos intervalos de credibilidade a partir das distribuições *a priori* Uniforme e de Jeffreys, as distribuições *a posteriori* são as distribuições  $\text{Beta}(r+1, x-r+1)$  e  $\text{Beta}\left(r, x-r+\frac{1}{2}\right)$ , respectivamente. Para isso, basta tomar os valores correspondentes aos quantis desejados para a obtenção do intervalo de credibilidade, conforme exposto na Seção 3.2.1.

### 4.2.2 Intervalo HPD (máxima densidade *a posteriori*)

Para a construção dos intervalos HPD foi utilizada a aproximação de Monte Carlo (Chen & Shao, 1998) apresentada na Seção 3.2.3, pois as distribuições *a posteriori* são contínuas e unimodais.

### 4.2.3 Intervalo de confiança com aproximação pela distribuição Normal

Um método clássico para obtenção de um intervalo de confiança para a probabilidade de sucesso do modelo Binomial Negativa é dado pela aproximação com a distribuição Normal, cuja base é o Teorema Central do Limite, que fornece

$$\lim_{n \rightarrow \infty} \mathbb{P} \left\{ \frac{|\hat{\theta} - \theta|}{\sqrt{\frac{\theta^2(1-\theta)}{r}}} < z \right\} = 2\Phi(z) - 1,$$

em que  $z > 0$  e  $\Phi(\cdot)$  é a função distribuição acumulada da distribuição Normal padrão.

Seja  $z_{\epsilon/2}$  o valor que satisfaz  $\Phi(z_{\epsilon/2}) = 1 - \frac{\epsilon}{2}$ . Então, segue que

$$\lim_{n \rightarrow \infty} \mathbb{P} \left\{ \hat{\theta} - z_{\epsilon/2} \sqrt{\frac{\theta^2(1-\theta)}{r}} < \theta < \hat{\theta} + z_{\epsilon/2} \sqrt{\frac{\theta^2(1-\theta)}{r}} \right\} = 1 - \epsilon,$$

no qual  $\hat{\theta}$  é o estimador de máxima verossimilhança de  $\theta$  e a variância de  $\hat{\theta}$  é dada pelo inverso da informação de Fisher. Então,

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

Esses limites podem ser obtidos por

$$\left[ \hat{\theta} - z_{\epsilon/2} \sqrt{\frac{\hat{\theta}^2(1-\hat{\theta})}{r}} \quad ; \quad \hat{\theta} + z_{\epsilon/2} \sqrt{\frac{\hat{\theta}^2(1-\hat{\theta})}{r}} \right].$$

A qualidade da aproximação para amostras pequenas pode ser baixa devido ao resultado assintótico.

#### 4.2.4 Intervalo de confiança exato de Clopper & Pearson

Casella & Berger (2002, p. 454) mencionam um intervalo de confiança exato, alternativo ao intervalo com aproximação pela distribuição Normal, para a probabilidade de sucesso da distribuição Binomial Negativa, semelhante ao intervalo proposto por Clopper & Pearson (1934), conforme o teorema a seguir.

**Teorema 4.1** (Casella & Berger, 2002, p. 454) Se  $X \sim \text{BinNeg}(r, \theta)$  com  $r$  inteiro fixo. Aplicando a relação entre as distribuições Binomial e Beta, tem-se o intervalo aleatório exato para  $x \neq 0$ :

$$\left[ \frac{1}{1 + \frac{\mathbf{X} + 1}{r} F_{2(\mathbf{X}+1), 2r, \epsilon/2}} ; \frac{\frac{r}{\mathbf{X}} F_{2r, 2\mathbf{X}, \epsilon/2}}{1 + \frac{r}{\mathbf{X}} F_{2r, 2\mathbf{X}, \epsilon/2}} \right].$$

A obtenção deste intervalo é muito similar ao realizado para a distribuição Binomial, embora seja necessário um outro lema, conforme segue.

**Lema 4.1** (Casella & Berger, 2002, p. 130) Seja  $X \sim \text{Bin}(n, \theta)$  e seja  $Y \sim \text{BinNeg}(r, \theta)$ . Então,  $\mathbb{P}(X \leq r - 1) = 1 - \mathbb{P}(Y \leq n - r)$ .

A demonstração do Lema 4.1 encontra-se no Apêndice A.

Para construir um intervalo de confiança para  $\theta$  da forma

$$\{\theta : \theta_I \leq \theta \leq \theta_S | Y = y\} \quad \text{com} \quad \mathbb{P}[\theta_I \leq \theta \leq \theta_S | Y = y] = 1 - \epsilon,$$

ao considerar o método do pivô baseado na função distribuição acumulada (Cox, 2006, p. 25; Casella & Berger, 2002, p. 434); é necessário obter  $\theta_I$  e  $\theta_S$  das equações

$$\sum_{k=r}^y \binom{k-1}{r-1} \theta_I^r (1 - \theta_I)^{k-r} = \frac{\epsilon}{2} \quad (4.3)$$

e

$$\sum_{k=y}^{\infty} \binom{k-1}{r-1} \theta_S^r (1 - \theta_S)^{k-r} = \frac{\epsilon}{2}. \quad (4.4)$$

Pela equação (4.3) tem-se que  $\frac{\epsilon}{2} = \mathbb{P}(K \leq y)$ , em que  $K \sim \text{BinNeg}(r, \theta_I)$  e  $\theta_I$  corresponde ao limite inferior do intervalo. Ao aplicar o Lema 4.1 é obtida

$$\mathbb{P}(K \leq y) = 1 - \mathbb{P}(X \leq r - 1),$$

no qual  $X \sim \text{Bin}(y + r, \theta_I)$ . Ao reescrever a última igualdade tem-se que

$$1 - \mathbb{P}(X \leq r - 1) = \mathbb{P}(X > r - 1) = \mathbb{P}(X \geq r).$$

No entanto, pelo Lema 3.1,  $\mathbb{P}(X \geq r) = \mathbb{P}(X_1 \leq \theta_I)$ , com  $X_1 \sim \text{Beta}(r, y + 1)$ .

Sejam  $U \sim F_{2r, 2(y+1)}$  e  $c = \frac{r}{y+1}$ . Ao aplicar o Lema 3.2 tem-se  $\frac{cU}{1+cU} \sim \text{Beta}(r, y+1)$ . Então, ao substituir  $X_1$  por  $\frac{cU}{1+cU}$  tem-se

$$\frac{\epsilon}{2} = \mathbb{P} \left[ \frac{cU}{1+cU} \leq \theta_I \right] = \mathbb{P} \left[ U \leq \frac{\theta_I}{c(1-\theta_I)} \right].$$

Desta forma,

$$\frac{\theta_I}{c(1-\theta_I)} = F_{2r, 2(y+1), \epsilon/2}.$$

Resolvendo a equação em  $\theta_I$  obtem-se

$$\theta_I = \frac{\frac{r}{y+1} F_{2r, 2(y+1), \epsilon/2}}{1 + \frac{r}{y+1} F_{2r, 2(y+1), \epsilon/2}}. \quad (4.5)$$

Ao dividir o numerador e denominador do lado direito da equação (4.5) por  $1 + \frac{r}{y+1} F_{2r, 2(y+1), \epsilon/2}$  e considerar o resultado do Lema 3.3 é obtido

$$\theta_I = \frac{1}{1 + \frac{y+1}{r} F_{2(y+1), 2r, \epsilon/2}},$$

o limite inferior do intervalo.

Portanto  $\theta_I$  é dado por

$$\theta_I = \frac{1}{1 + \frac{\mathbf{X} + 1}{r} F_{2(\mathbf{X}+1), 2r, \epsilon/2}}.$$

Para calcular o limite superior, pela equação (4.4) é obtida a relação

$$\frac{\epsilon}{2} = \mathbb{P}(K \geq y) = 1 - \mathbb{P}(K < y) = 1 - \mathbb{P}(K \leq y - 1),$$

pela aplicação do Lema 4.1 verifica-se que

$$1 - \mathbb{P}(K \leq y - 1) = \mathbb{P}(X \leq r - 1),$$

no qual  $X \sim \text{Bin}(y + r - 1, \theta_S)$  e  $\theta_S$  corresponde ao limite superior do intervalo. Manipulando a igualdade tem-se que

$$\mathbb{P}(X \leq r - 1) = 1 - \mathbb{P}(X > r - 1) = 1 - \mathbb{P}(X \geq r).$$

Ao aplicar o Lema 3.1 obtemos

$$1 - \mathbb{P}(X \geq r) = 1 - \mathbb{P}(X_1 \leq \theta_S) = \mathbb{P}(X_1 > \theta_S),$$

em que  $X_1 \sim \text{Beta}(r, y)$ .

Sejam  $U \sim F_{2r, 2y}$  e  $c = \left(\frac{x+1}{n-x}\right)$ . Pelo Lema 3.2 tem-se  $\frac{cU}{1+cU} \sim \text{Beta}(r, y)$ .

Então, ao substituir  $X_1$  por  $\frac{cU}{1+cU}$  tem-se

$$\frac{\epsilon}{2} = \mathbb{P} \left[ \frac{cU}{1+cU} > \theta_S \right] = \mathbb{P} \left[ U > \frac{\theta_S}{c(1-\theta_S)} \right].$$

Desta forma, conclui-se que

$$\frac{\theta_S}{c(1-\theta_S)} = F_{2r, 2y, \epsilon/2}.$$

Resolvendo a equação em  $\theta_S$  obtém-se

$$\theta_S = \frac{\frac{r}{y} F_{2r, 2y, \epsilon/2}}{1 + \frac{r}{y} F_{2r, 2y, \epsilon/2}}. \quad (4.6)$$

Portanto  $\theta_S$  é dado por

$$\theta_S = \frac{\frac{r}{\mathbf{X}} F_{2r, 2\mathbf{X}, \epsilon/2}}{1 + \frac{r}{\mathbf{X}} F_{2r, 2\mathbf{X}, \epsilon/2}}.$$

# Capítulo 5

## Estudo de Simulação

Este trabalho traz, da literatura, algumas formas de obter intervalos e estimadores pontuais para a probabilidade de sucesso das distribuições Binomial e Binomial Negativa. O interesse deste capítulo é analisar a probabilidade de cobertura gerada por esses intervalos ao longo do espaço paramétrico e confirmar por meio de simulação a consistência e calcular o viés e o erro quadrático médio dos estimadores pontuais.

A proporção de cobertura descreve a proporção de intervalos gerados que contêm o verdadeiro valor do parâmetro, o qual é fixado ao realizar o estudo de simulação. O ideal é que esta taxa de cobertura dos intervalos de credibilidade e de confiança estejam muito próximas do nível estipulado,  $100(1 - \epsilon)\%$ , independente do valor do parâmetro.

Bayarri & Berger (2004) destacam que o cenário mais comum de ligação entre freqüentistas e bayesianos é quando não há informações externas além dos dados e do modelo, introduzindo na análise a distribuição *a priori* objetiva.

Para se avaliar as probabilidades de cobertura dos intervalos clássicos e bayesianos foram realizadas simulações de amostras das distribuições Binomial e Binomial Negativa. Para isso, foram utilizados os geradores de amostras pseudo-aleatórias cujas funções são *rbinom()* e *rnbinom()*, respectivamente, do pacote *stats* em R (R, 2008).

Foram geradas 10.000 réplicas para cada um de 99 valores selecionados do parâmetro  $\theta$ , pertencente ao intervalo  $(0, 1)$ , da distribuição Binomial e, analogamente, para a distribuição Binomial Negativa. Para os intervalos HPD, foram consideradas amostras de 1.000 valores gerados das distribuições *a posteriori* de  $\theta$ .

No caso da distribuição Binomial, a simulação foi repetida para tamanhos amostrais  $n = 10, 20, 50$  e  $100$  com distribuições *a priori* de Bayes-Laplace, de Jeffreys e menos favorável. Para a distribuição Binomial Negativa, foram adotadas nas simulações amostras de tamanho 1 com  $r = 1, 2, 3$  e  $4$  para as distribuições *a priori* de Bayes-Laplace e de Jeffreys.

Será convencionado neste capítulo que, ao mencionar distribuição *a priori* menos favorável, trata-se da distribuição *a priori* menos favorável para o modelo Binomial com função de perda quadrática (o valor de  $n$  corresponde ao tamanho amostral considerado na simulação, por exemplo, para o gráfico (d) da figura 5.2 foi utilizada a distribuição *a priori* Beta(5,5)). Ao tratar-se da distribuição *a priori* de Bayes-Laplace, pode se referir também à distribuição *a priori* menos favorável com função de perda quadrática ponderada, pois ambas são coincidentes no modelo Binomial.

O nível de confiança e credibilidade fixados para construção dos intervalos foi de 95%, ou seja, essa é a cobertura nominal ou esperada. A partir daí, observou-se a proporção de intervalos que contêm o verdadeiro valor do parâmetro em cada situação simulada.

Os resultados retratam que as probabilidades de cobertura obtidas nas simulações são pontuais, ou seja, para cada um dos 99 valores assumidos para o estudo de simulação, foi obtida uma proporção de cobertura.

## 5.1 Distribuição Binomial

### 5.1.1 Proporções de cobertura e amplitudes médias dos intervalos bayesianos

No Capítulo 3 foram considerados os intervalos bayesianos de credibilidade, de credibilidade a partir da distribuição  $F$  de Fisher e intervalos HPD, para os quais foram aplicadas as três distribuições *a priori* mencionadas. Nesse capítulo não será utilizada a distribuição *a priori* de Haldane, uma vez que se trata de um caso específico e leva a uma distribuição *a posteriori* imprópria em dois casos particulares, conforme descrito na Seção 3.1.3.

## Intervalos de credibilidade

Será omitido graficamente o desempenho das proporções de cobertura e amplitudes médias do intervalo de credibilidade a partir da distribuição  $F$  de Fisher, pois o mesmo é equivalente ao intervalo obtido diretamente pela distribuição Beta (Seção 4.2.1).

Pelas figuras 5.1 e 5.2 constata-se que as proporções de cobertura dos intervalos de credibilidade com distribuição *a priori* menos favorável para  $n = 10$  estão mais próximas da nominal se comparadas com os demais tamanhos amostrais considerados. Para  $n = 50$  e 100 a proporção de cobertura para valores extremos da probabilidade de sucesso, ou seja, próximos de 0 e 1, diminui sensivelmente. Explicado pelo fato desta distribuição *a priori* tender a se concentrar em  $\theta = 0,5$  com o aumento de  $n$ .

Ao observar as amplitudes médias obtidas pelos intervalos de credibilidade com a distribuição *a priori* menos favorável em relação às demais distribuições *a priori* estudadas, pode-se ver, nas figuras 5.3 e 5.4, que apresentam as maiores amplitudes médias para valores de  $\theta$  próximos de 0 e 1 independente do tamanho amostral.

Assim, pode-se dizer, que com o estudo de simulação desenvolvido, os intervalos de credibilidade com distribuição *a priori* menos favorável, com função de perda quadrática, não possuem boas propriedades freqüentistas, pois não há estabilidade da proporção de cobertura ao longo do espaço paramétrico, com o aumento do tamanho amostral.

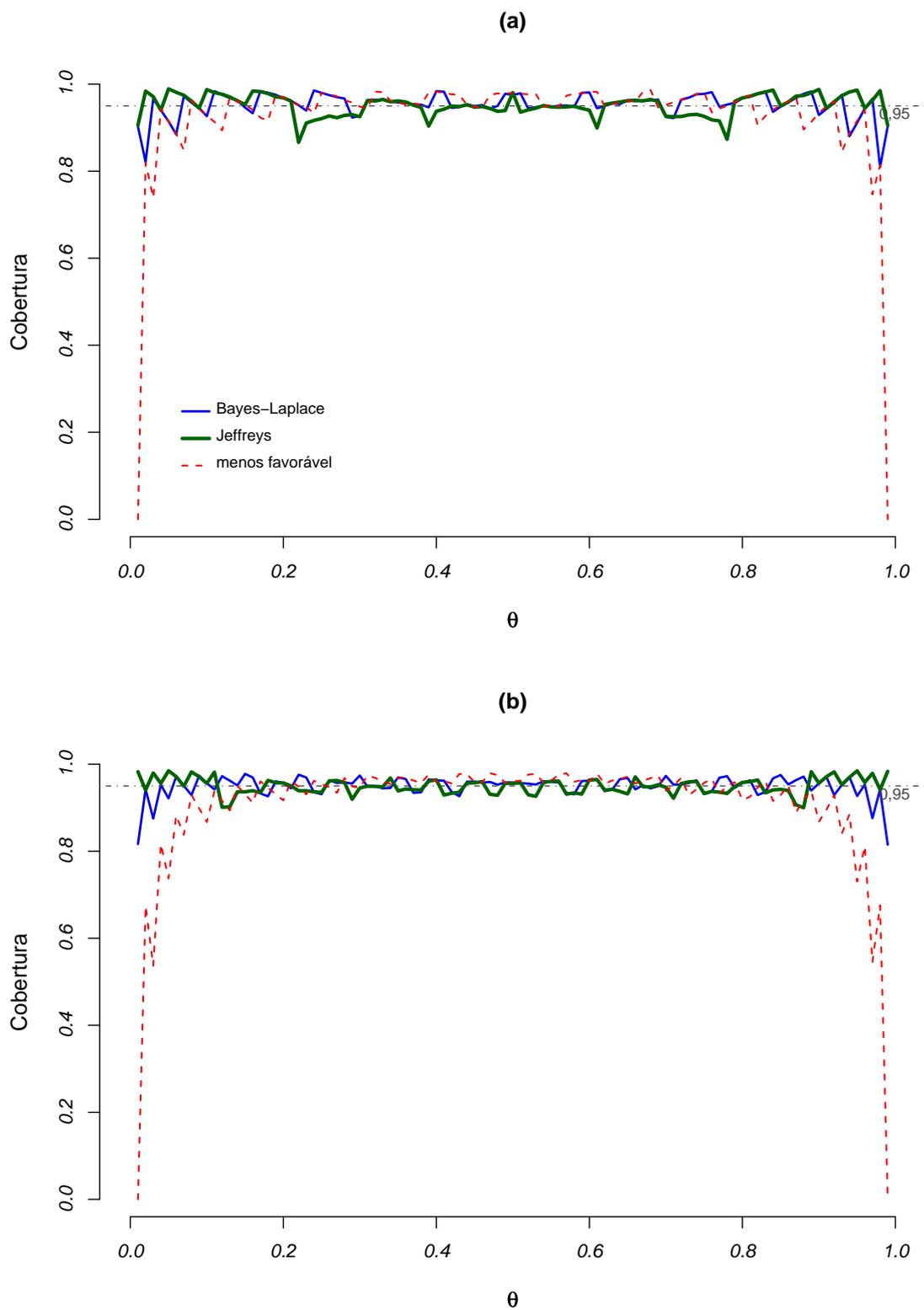


FIGURA 5.1: Proporções de cobertura dos intervalos de credibilidade para a probabilidade de sucesso do modelo Binomial: (a)  $n = 10$  e (b)  $n = 20$ .

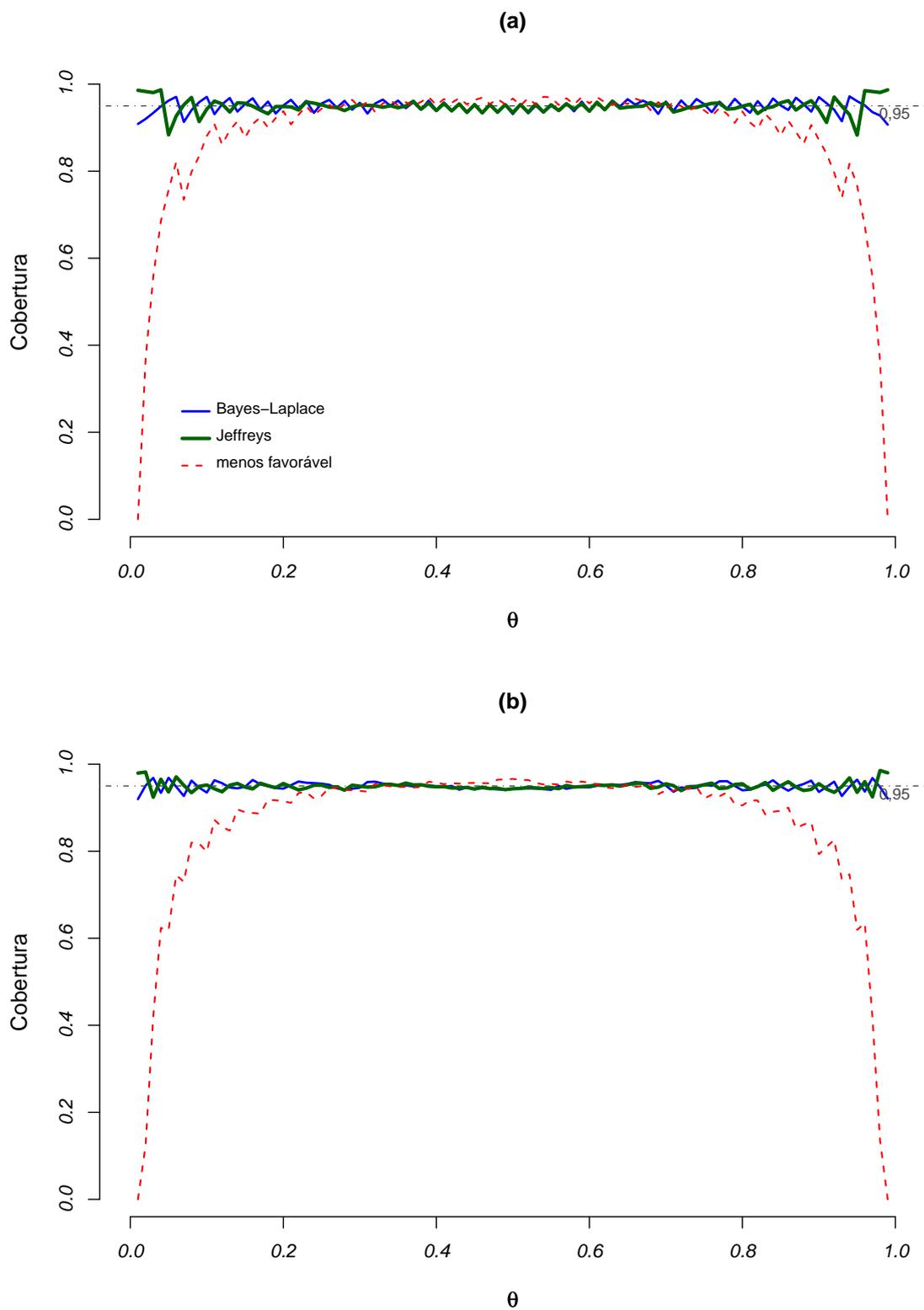


FIGURA 5.2: Proporções de cobertura dos intervalos de credibilidade para a probabilidade de sucesso do modelo Binomial: (a)  $n = 50$  e (b)  $n = 100$ .

As proporções de cobertura dos intervalos de credibilidade para  $n = 10$  (figura 5.1), que consideram a distribuição *a priori* de Jeffreys, apresentam algumas tendências de subestimação da cobertura nominal para valores de  $\theta$  entre  $(0, 2; 0, 8)$ , ou seja, mostram coberturas interiores a 95%, o que não é visível para os que empregam distribuição *a priori* de Bayes-Laplace. No entanto, as proporções de cobertura dos intervalos de credibilidade com distribuição *a priori* de Jeffreys evidenciam coberturas mais próximas da nominal para valores extremos inferiores e superiores em relação às proporções de cobertura dos intervalos de credibilidade com distribuição *a priori* de Bayes-Laplace.

A proporção de cobertura média observada para a distribuição *a priori* de Jeffreys foi de 95,13% para  $n = 10$ . Para os demais tamanhos amostrais, note que a proporção de cobertura está muito próxima da cobertura nominal. O fato da cobertura estar próxima de 95% para valores extremos pode estar relacionado com a forma desta distribuição *a priori*, que atribui mais peso para valores da probabilidade de sucesso próximos de 0 e 1.

Para os intervalos de credibilidade com distribuição *a priori* de Bayes-Laplace (figuras 5.1 e 5.2), apesar de a proporção de cobertura média ser de 95,28%, superior à verificada com a distribuição *a priori* de Jeffreys, ela se mostra mais sutilmente estável ao longo do espaço paramétrico. Para  $n = 50$  e 100, as proporções de cobertura dos intervalos de credibilidade ao considerar as distribuições *a priori* de Bayes-Laplace e de Jeffreys estão muito próximas entre si e também da cobertura nominal. É importante lembrar que estas distribuições *a priori* são não-informativas para a probabilidade de sucesso do modelo Binomial.

Nas figuras 5.3 e 5.4 podem ser observadas amplitudes médias dos intervalos de credibilidade, dos quais foram obtidas as proporções de cobertura apresentadas nas figuras 5.1 e 5.2. É interessante ressaltar que para  $n = 10$  e 20 as amplitudes médias dos intervalos de credibilidade com distribuição *a priori* de Bayes-Laplace são inferiores às amplitudes médias observadas para os intervalos de credibilidade com distribuição *a priori* de Jeffreys quando  $\theta \in (0, 25; 0, 75)$ . Nos intervalos restantes do espaço paramétrico ocorre o oposto. Lembramos que a distribuição *a priori* de Jeffreys agrega maior densidade às caudas, o que as torna mais informativas para construção dos intervalos de credibilidade em relação aos intervalos de credibilidade com distribuição *a priori* de Bayes-Laplace.

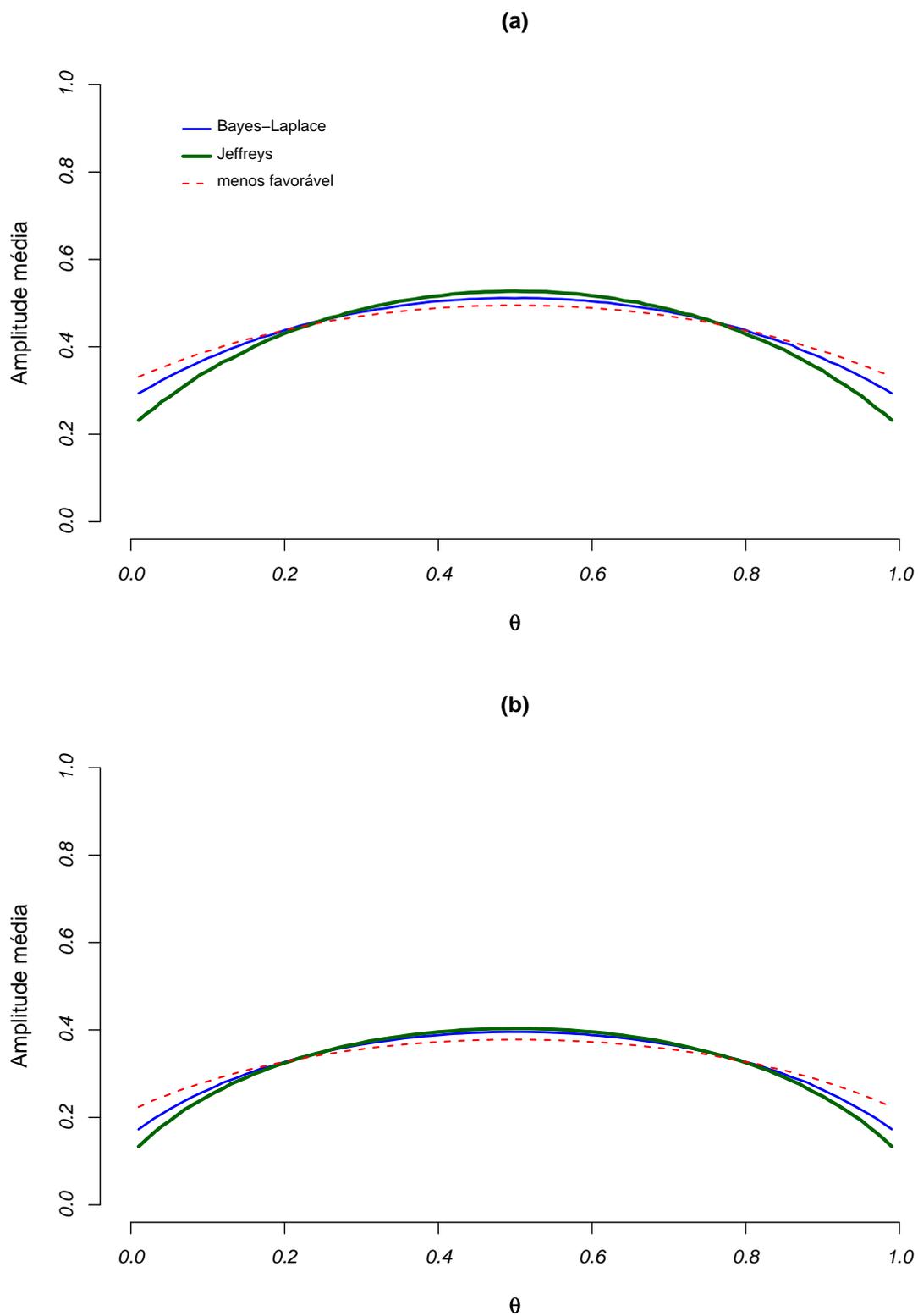


FIGURA 5.3: Amplitudes médias dos intervalos de credibilidade para a probabilidade de sucesso do modelo Binomial: (a)  $n = 10$  e (b)  $n = 20$ .

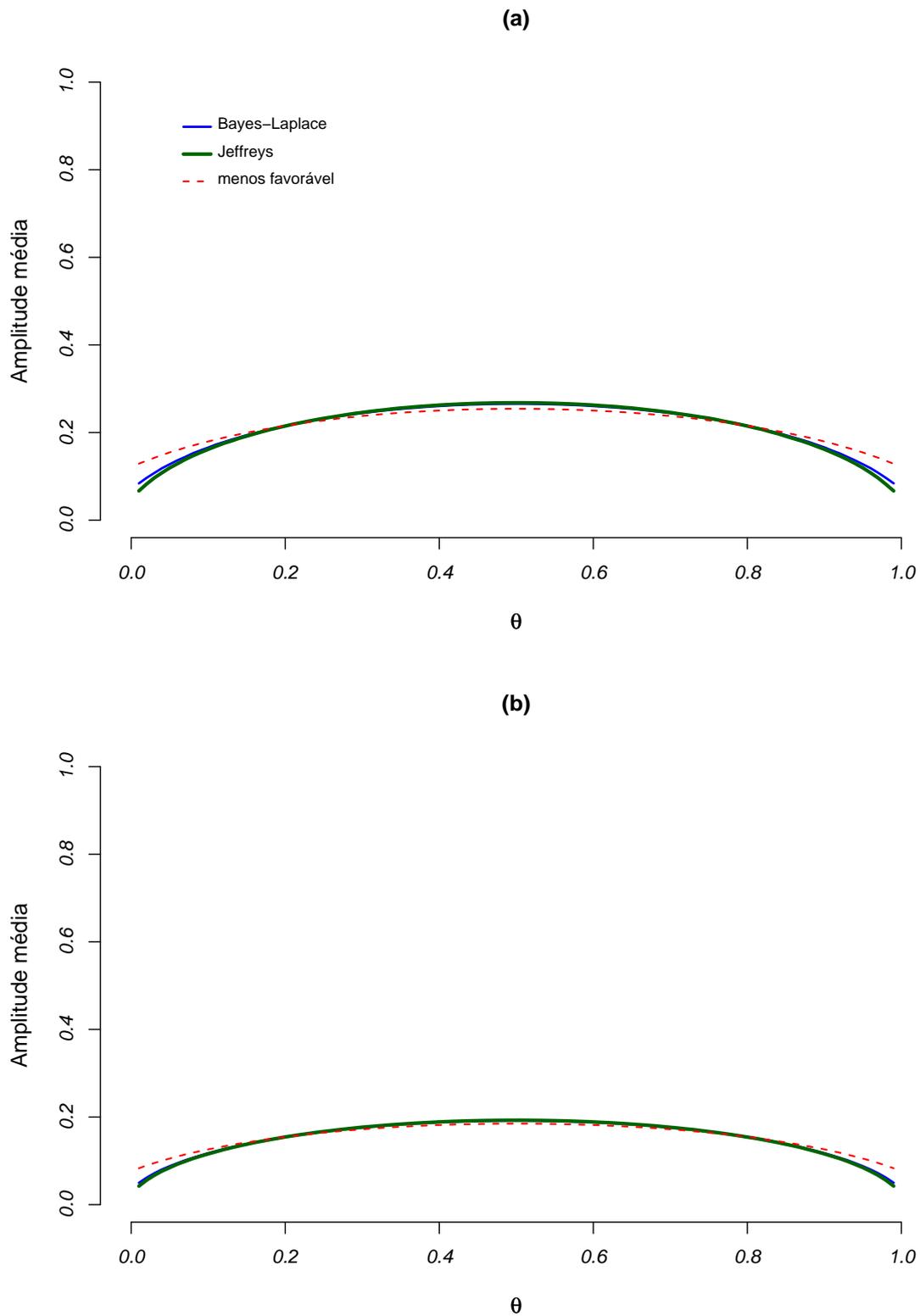


FIGURA 5.4: Amplitudes médias dos intervalos de credibilidade para a probabilidade de sucesso do modelo Binomial: (a)  $n = 50$  e (b)  $n = 100$ .

Assim, pode-se verificar um desempenho freqüentista sutilmente melhor da proporção de cobertura nos intervalos de credibilidade ao considerar a distribuição *a priori* de Bayes-Laplace quando a probabilidade de sucesso não assume valores extremos.

### Intervalos HPD

Nas figuras 5.5 e 5.6 são expostas as proporções de cobertura dos intervalos HPD. Ao considerar a distribuição *a priori* menos favorável observa-se que as proporções de cobertura para  $n = 10$  e  $20$  estão próximas da cobertura nominal para  $\theta \in (0, 25; 0, 75)$  e muito abaixo de 95% fora desse intervalo, independente do tamanho amostral considerado, pois esta distribuição *a priori* tender a se concentrar em  $\theta = 0, 5$  com o aumento de  $n$ .

Nas proporções de cobertura ao considerar os intervalos HPD com distribuição *a priori* de Bayes-Laplace é possível observar pontos em que as proporções de cobertura estão muito próximas ou sobrestimam a cobertura nominal para  $n = 10$  (figura 5.5), porém na maior parte há subestimação.

Nas figuras 5.5 e 5.6 também observa-se que as proporções de cobertura para os intervalos HPD com distribuição *a priori* de Jeffreys subestimam 95% quando  $\theta \in (0, 15; 0, 83)$  e  $n = 10$ , para  $n = 20$  esse intervalo é maior  $(0, 08; 0, 92)$ . Nos outros tamanhos amostrais considerados as proporções de cobertura estão mais próximas da nominal, mas ainda com leve subestimação. Em contrapartida, existem proporções de cobertura para valores de  $\theta$  próximos dos limites do intervalo paramétrico com sobrestimação das proporções de cobertura para os intervalo HPD com distribuição *a priori* de Jeffreys para  $n = 10$  e  $20$ , diminuindo consideravelmente com o aumento do tamanho amostral.

As proporções de cobertura dos intervalos HPD com distribuição *a priori* de Bayes-Laplace para  $n = 10$ , mostrados nas figuras 5.5, se comparados com os intervalos HPD com distribuição *a priori* de Jeffreys, evidenciam proporções de cobertura para valores extremos inferiores e superiores do espaço paramétrico mais próximas da nominal. Já o inverso ocorre para os intervalos de credibilidade com essas distribuições *a priori* (figura 5.3).

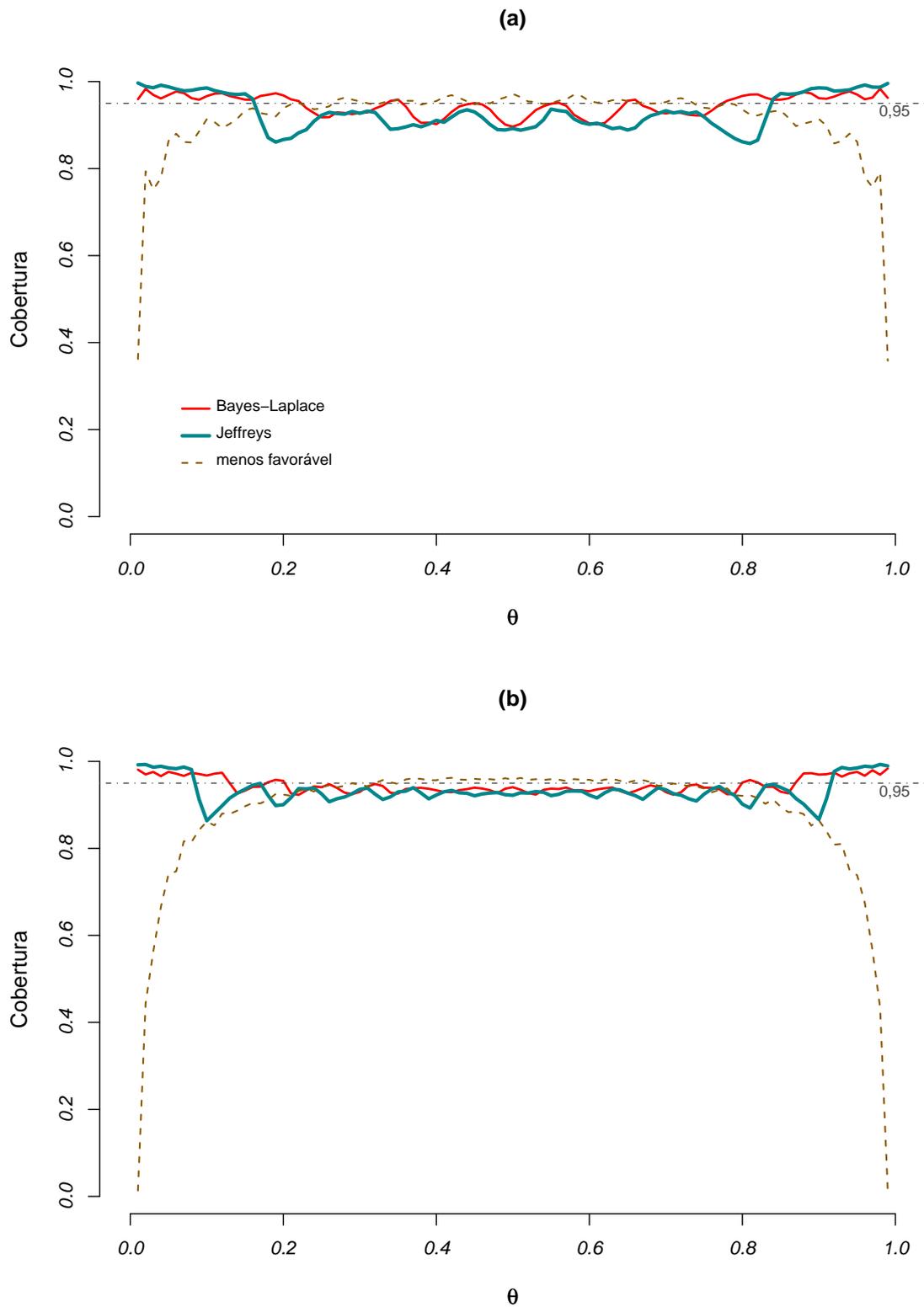


FIGURA 5.5: Proporções de cobertura dos intervalos HPD (máxima densidade *a posteriori*) para a probabilidade de sucesso do modelo Binomial: (a)  $n = 10$  e (b)  $n = 20$ .

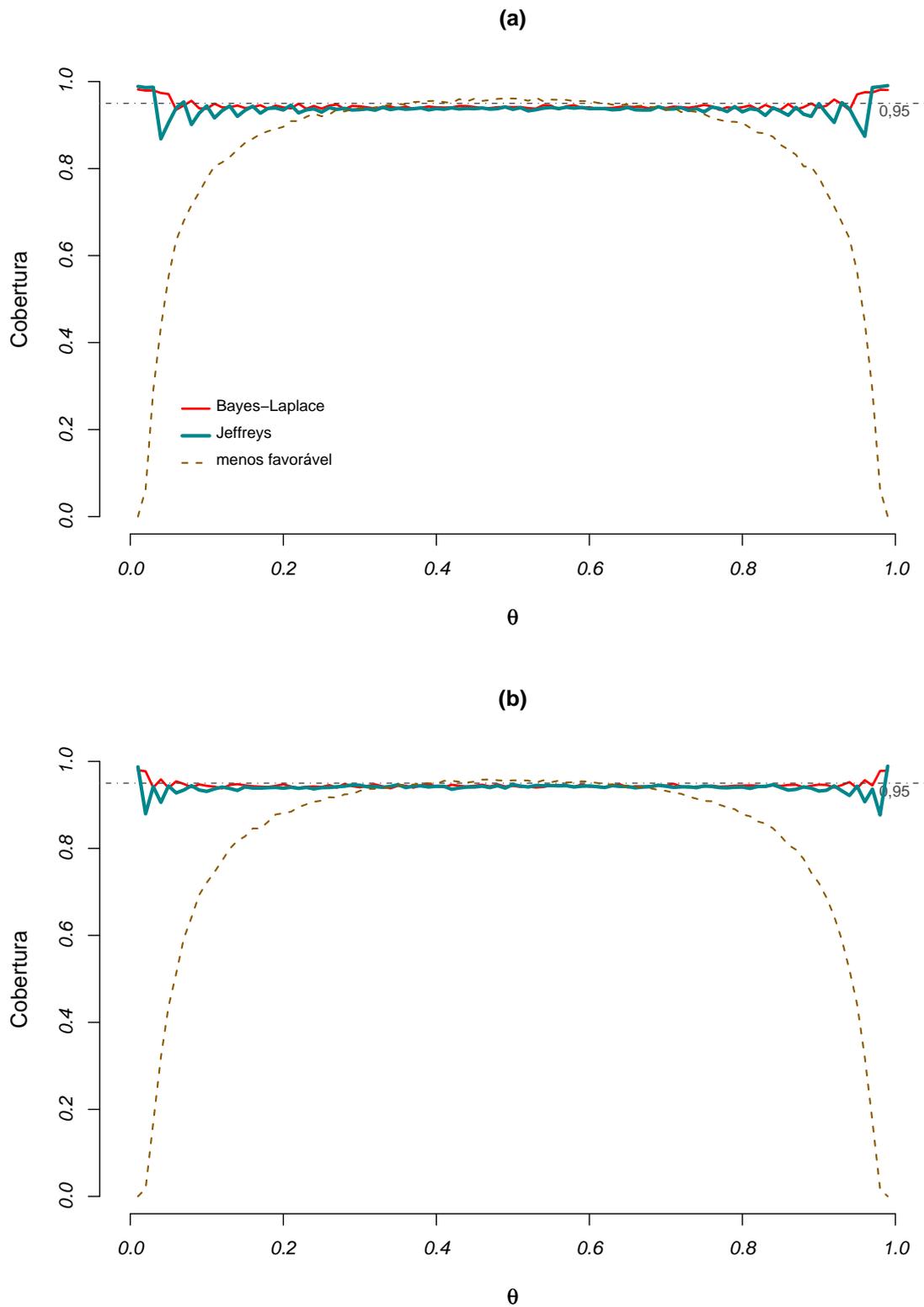


FIGURA 5.6: Proporções de cobertura dos intervalos HPD (máxima densidade *a posteriori*) para a probabilidade de sucesso do modelo Binomial: (a)  $n = 50$  e (b)  $n = 100$ .

Brown *et al.* (2001) citam que os intervalos HPD são mais difíceis de calcular e, ao empregar a distribuição *a priori* de Jeffreys, não são tão bons em termos da probabilidade de cobertura. Pôde ser verificado que, computacionalmente, a obtenção dos intervalos HPD demanda mais tempo do que para se obter intervalos de credibilidade e que as proporções de cobertura dos intervalos de credibilidade são mais próximas da nominal em relação às proporções de cobertura obtidas pelos intervalos HPD ao considerar a distribuição *a priori* de Jeffreys. É interessante ressaltar que Brown *et al.* (2001) não fazem uso da distribuição *a priori* de Bayes-Laplace, a qual mostrou neste estudo ter proporções de cobertura mais próximas da nominal para os intervalos HPD.

Para  $n = 50$  e  $100$  as amplitudes médias dos intervalos HPD apresentaram-se muito próximas ao considerar a distribuição *a priori* de Bayes-Laplace e de Jeffreys, conforme figura 5.8. Mas para  $n = 10$  e  $20$  observa-se que as amplitudes médias dos intervalos HPD para a distribuição *a priori* de Jeffreys são menores para  $\theta > 0,8$  e  $\theta < 0,2$ , ou seja, quando apresentou proporções de cobertura sobrestimada. No intervalo central ambas as amplitudes médias dos intervalos HPD apresentam-se muito próximas, com amplitudes médias levemente inferiores para intervalos HPD com distribuição *a priori* de Bayes-Laplace para  $\theta$  próximo de  $0,5$ .

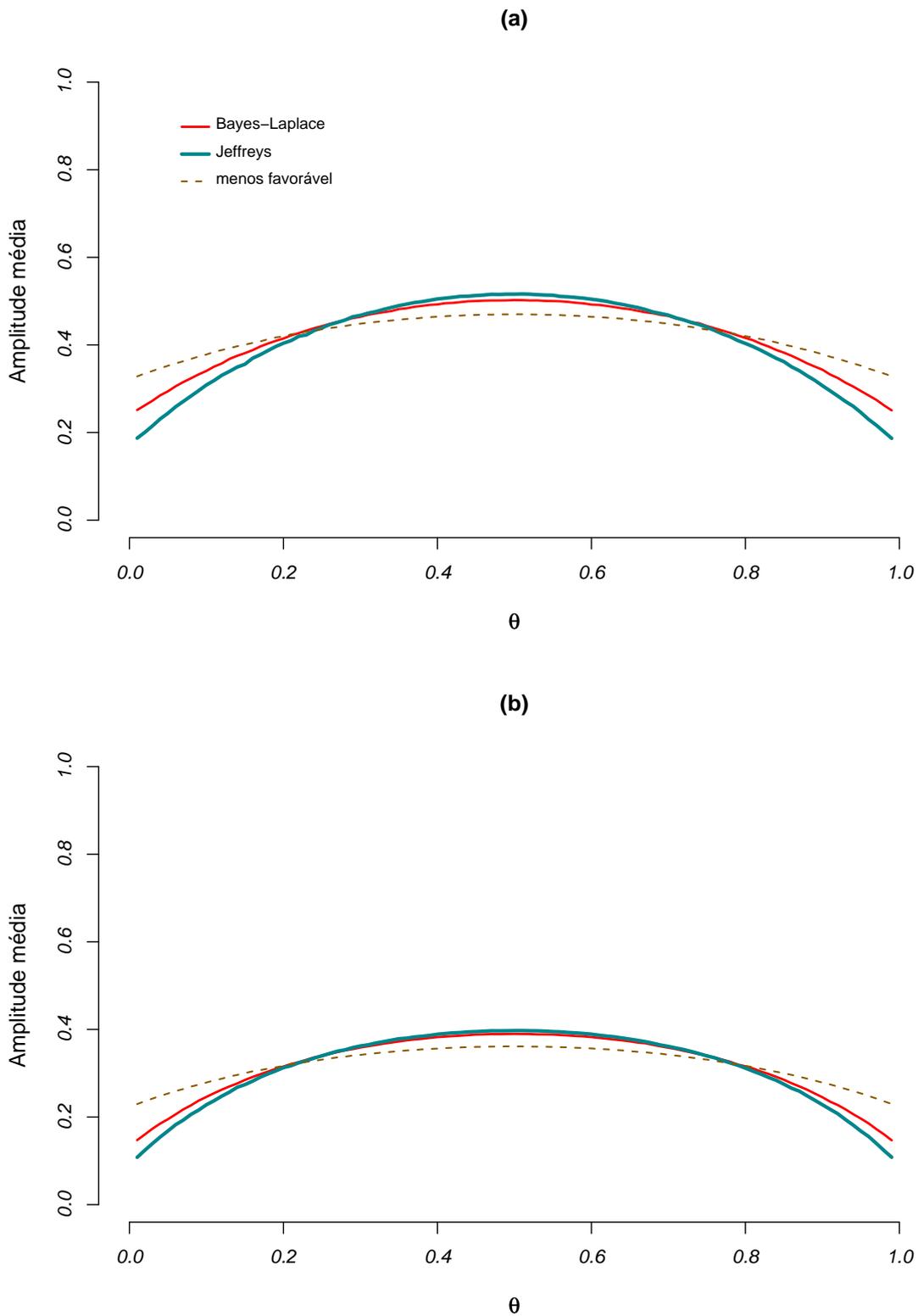


FIGURA 5.7: Amplitudes médias dos intervalos HPD (máxima densidade *a posteriori*) para a probabilidade de sucesso do modelo Binomial: (a)  $n = 10$  e (b)  $n = 20$ .

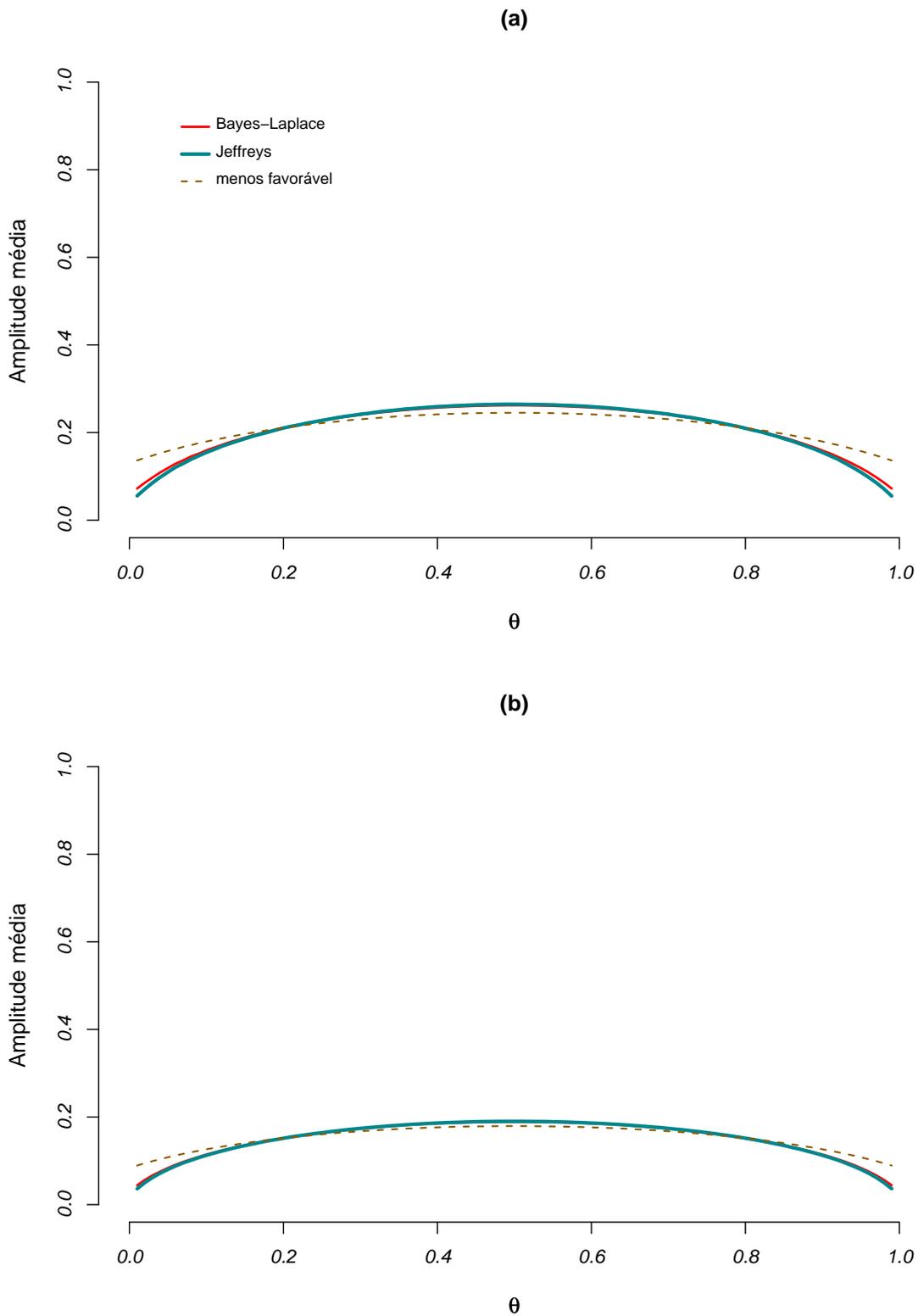


FIGURA 5.8: Amplitudes médias dos intervalos HPD (máxima densidade *a posteriori*) para a probabilidade de sucesso do modelo Binomial: (a)  $n = 50$  e (b)  $n = 100$ .

## Intervalos de credibilidade e intervalos HPD

É esperado que amplitudes médias dos intervalos HPD não sejam superiores às amplitudes médias obtidas pelos intervalos de credibilidade. Isso foi verificado para todas as situações em que  $n = 10$  e  $20$  (figuras 5.11, 5.15 e 5.19). Já para  $n = 50$  e  $100$  as amplitudes médias são muito próximas, conforme figuras 5.12, 5.16 e 5.20.

Não é observada grande diferença entre as proporções de cobertura dos intervalos HPD e dos intervalos de credibilidade para  $n = 50$  e  $100$  para a distribuição *a priori* de Bayes-Laplace e de Jeffreys, mostradas nas figuras 5.10 e 5.14. Entretanto, os intervalos HPD apresentam-se com amplitudes médias visivelmente inferiores para esses tamanhos amostrais ao considerar a distribuição *a priori* menos favorável.

Note nas figuras 5.9 e 5.10 que as proporções de cobertura dos intervalos com distribuição *a priori* de Bayes-Laplace têm proporções de cobertura mais próximas da nominal que as proporções de cobertura dos intervalos que consideram a distribuição *a priori* de Jeffreys (figuras 5.13 e 5.14).

As proporções de cobertura, observadas nas figuras 5.9 e 5.10, sobrestimam a proporção de cobertura nominal no caso dos intervalos HPD quando a probabilidade de sucesso assume valores muito próximos de 0 e 1 e o oposto ocorre com os intervalos de credibilidade, para a distribuição *a priori* de Bayes-Laplace. Por outro lado, para esses valores de  $\theta$ , existe sobrestimação da proporção de cobertura nominal para ambos os métodos bayesianos ao considerar a distribuição *a priori* de Jeffreys (figuras 5.13 e 5.14).

Embora os procedimentos bayesianos não tenham exibido muito boas qualidades freqüentistas para valores extremos da probabilidade de sucesso, no caso da distribuição *a priori* menos favorável principalmente, para os tamanhos amostrais analisados, os demais intervalos que consideram a distribuição *a priori* de Bayes-Laplace e Jeffreys apresentaram-se com proporções de cobertura satisfatórias. Vale ressaltar que em teoria, conforme comentado no Capítulo 3, era esperado a qualidade superior dos intervalos bayesianos empregando a distribuição *a priori* de Bayes-Laplace.

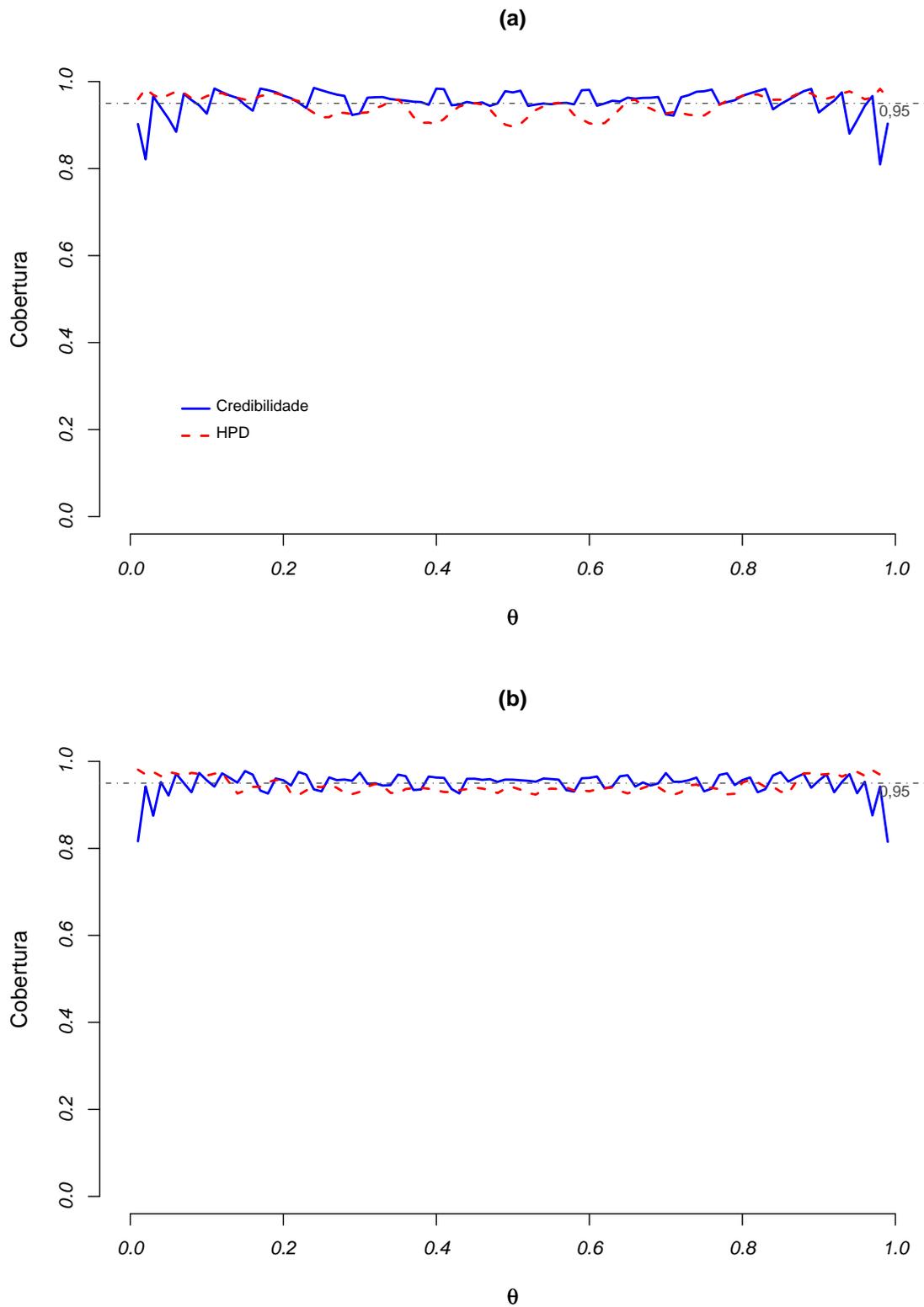


FIGURA 5.9: Proporções de cobertura dos intervalos de credibilidade e HPD com distribuição *a priori* de Bayes-Laplace para a probabilidade de sucesso do modelo Binomial: (a)  $n = 10$  e (b)  $n = 20$ .

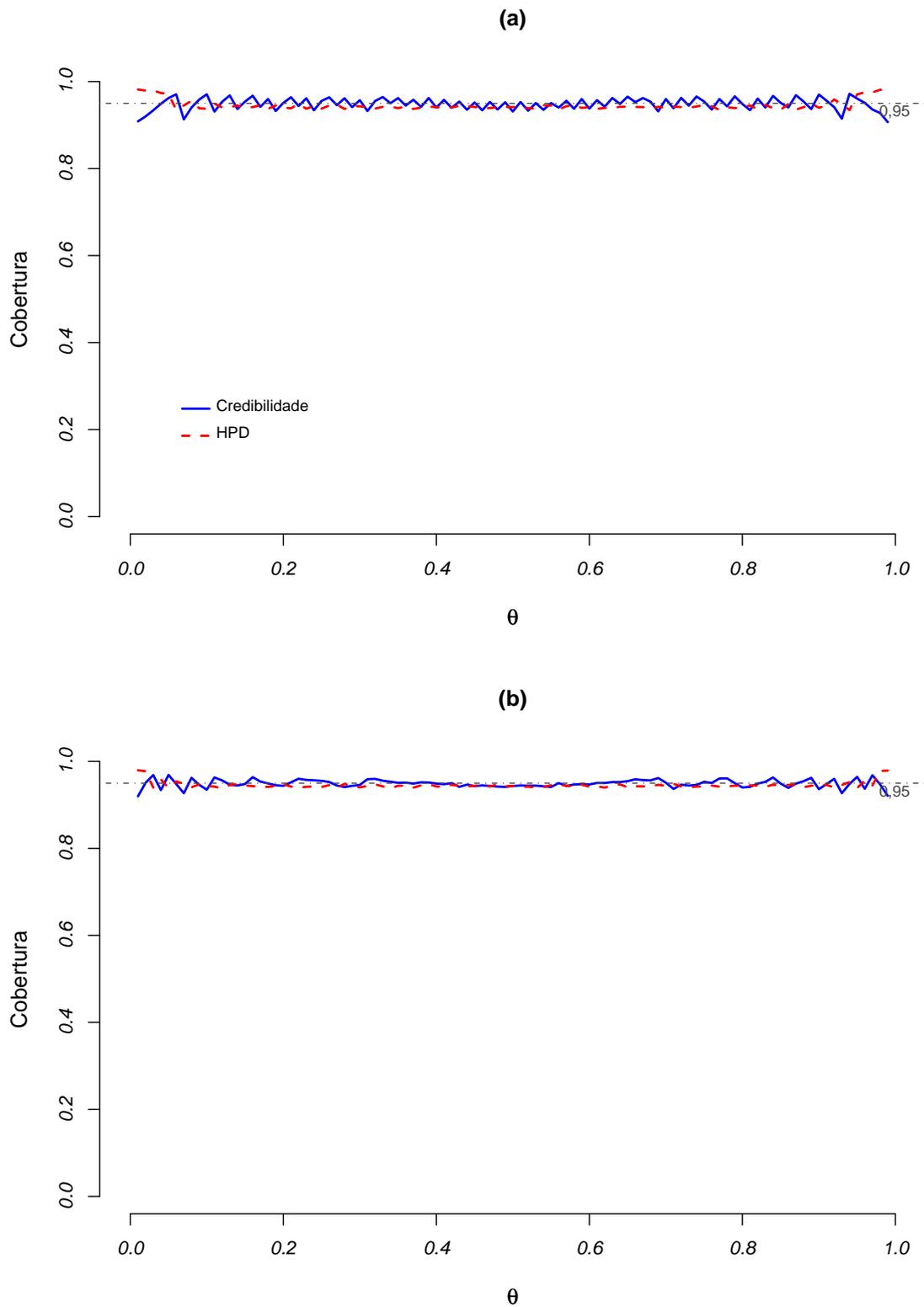


FIGURA 5.10: Proporções de cobertura dos intervalos de credibilidade e HPD com distribuição *a priori* de Bayes-Laplace para a probabilidade de sucesso do modelo Binomial: (a)  $n = 50$  e (b)  $n = 100$ .

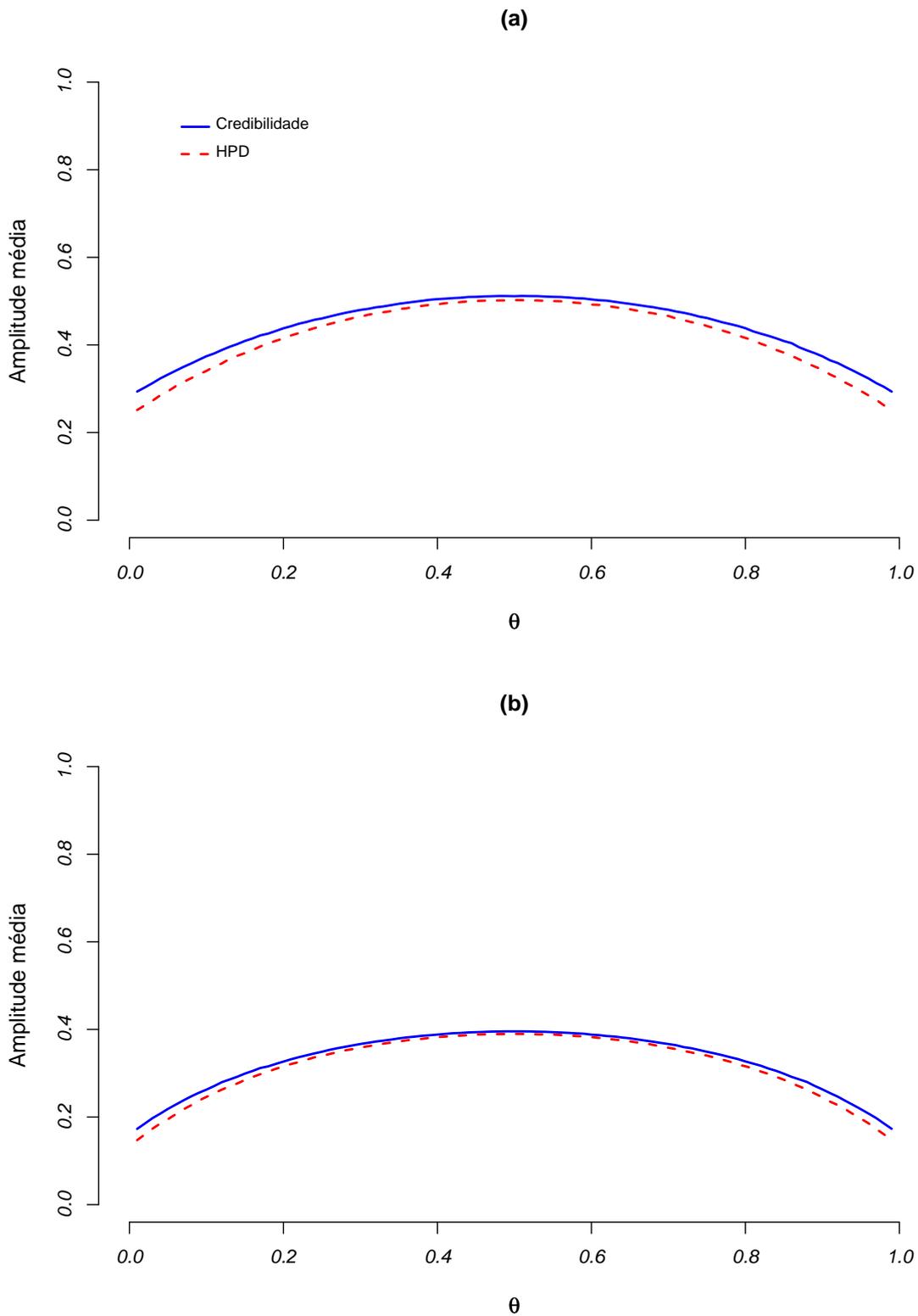


FIGURA 5.11: Amplitudes médias dos intervalos de credibilidade e HPD com distribuição *a priori* de Bayes-Laplace para a probabilidade de sucesso do modelo Binomial: (a)  $n = 10$  e (b)  $n = 20$ .

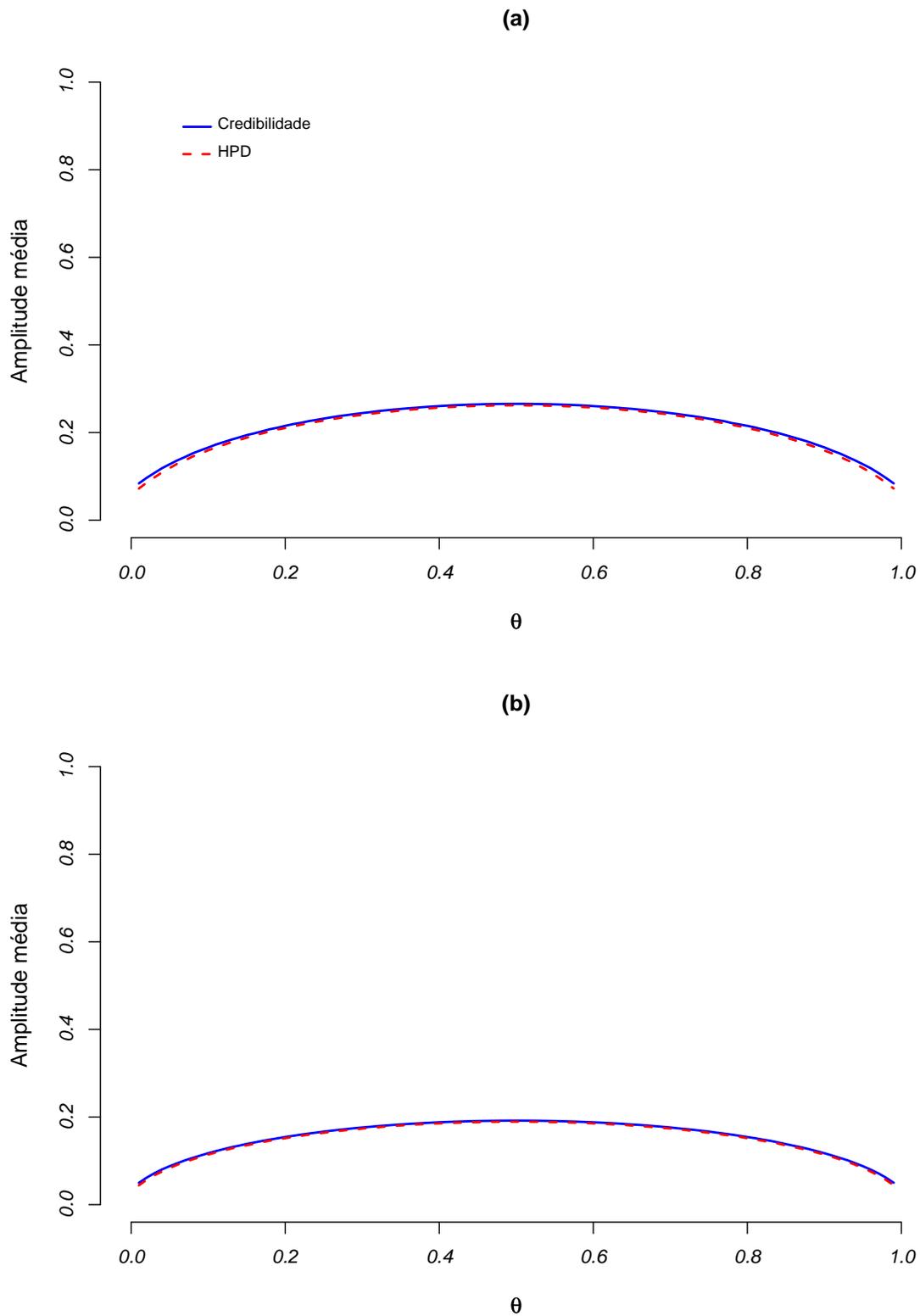


FIGURA 5.12: Amplitudes médias dos intervalos de credibilidade e HPD com distribuição *a priori* de Bayes-Laplace para a probabilidade de sucesso do modelo Binomial: (a)  $n = 50$  e (b)  $n = 100$ .

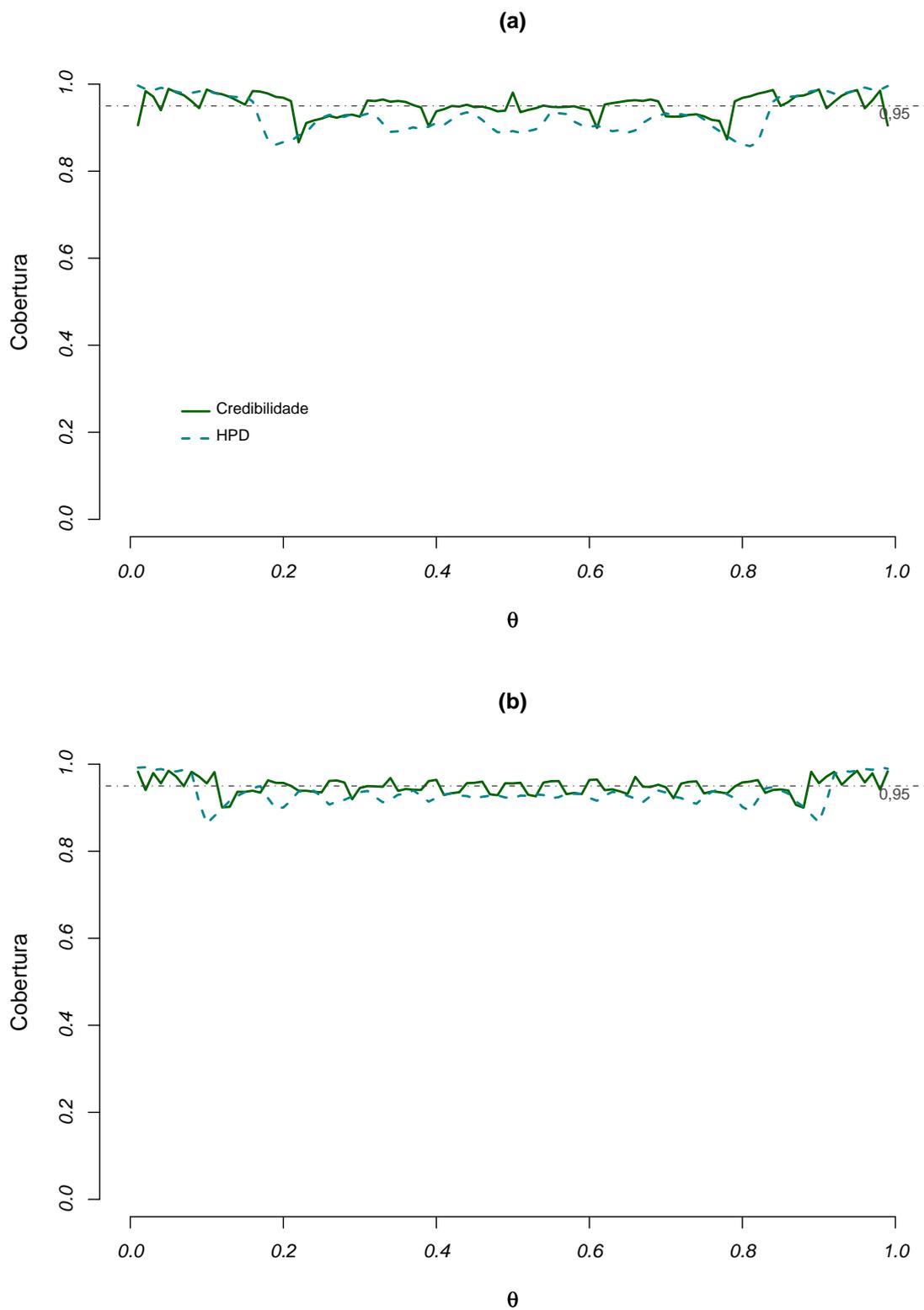


FIGURA 5.13: Proporções de cobertura dos intervalos de credibilidade e HPD com distribuição *a priori* de Jeffreys para a probabilidade de sucesso do modelo Binomial: (a)  $n = 10$  e (b)  $n = 20$ .

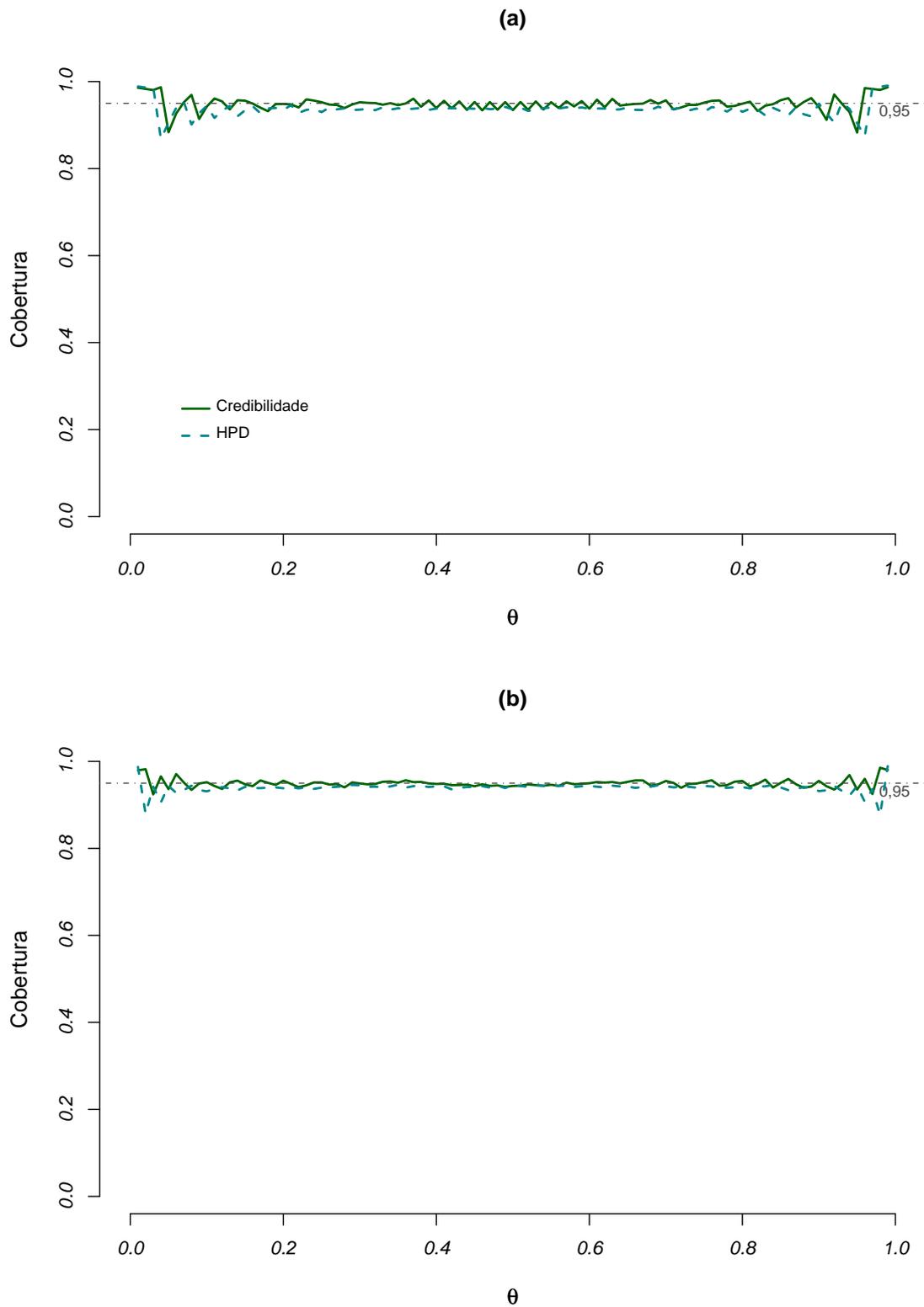


FIGURA 5.14: Proporções de cobertura dos intervalos de credibilidade e HPD com distribuição *a priori* de Jeffreys para a probabilidade de sucesso do modelo Binomial: (a)  $n = 50$  e (b)  $n = 100$ .

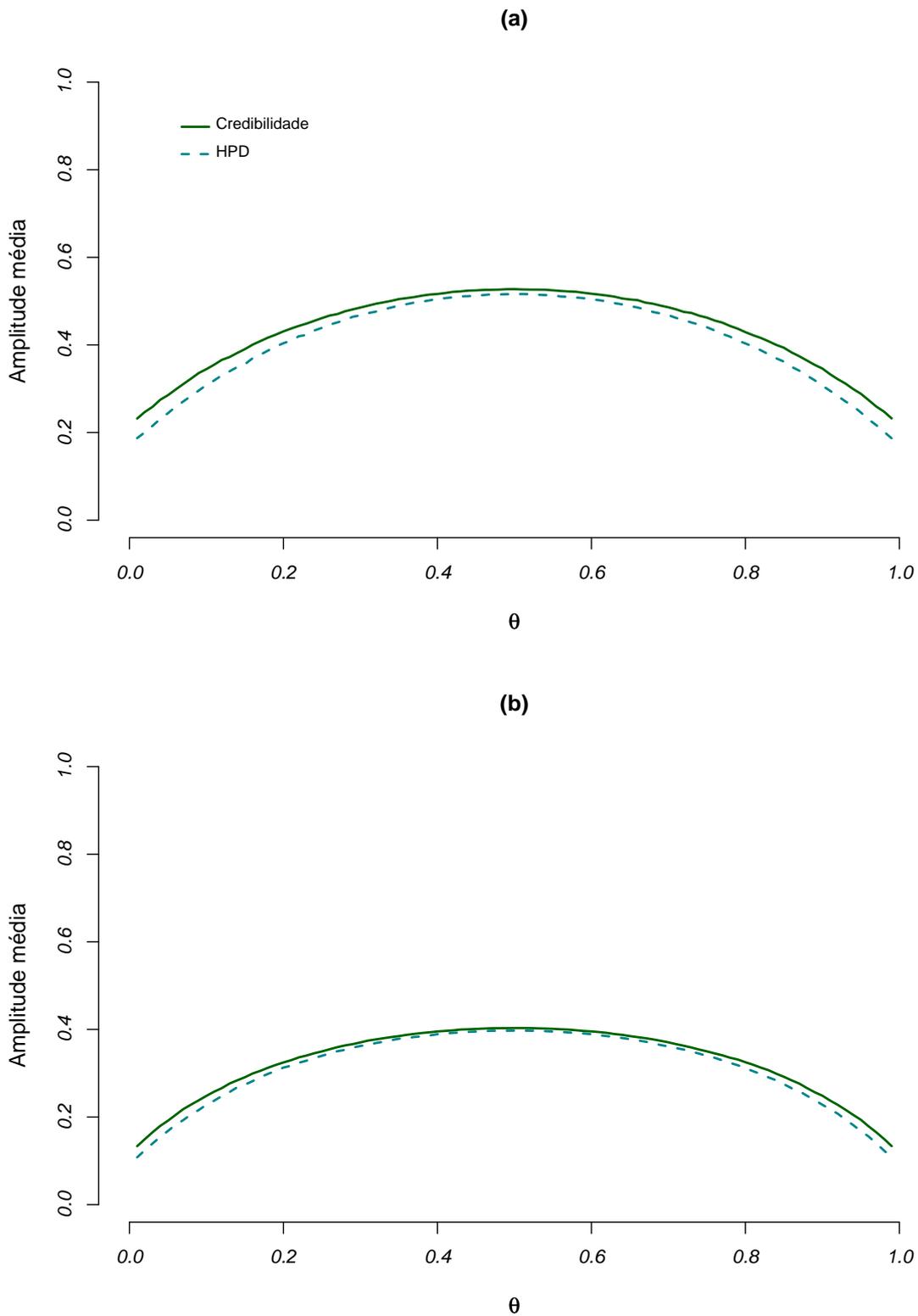


FIGURA 5.15: Amplitudes médias dos intervalos de credibilidade e HPD com distribuição *a priori* de Jeffreys para a probabilidade de sucesso do modelo Binomial: (a)  $n = 10$  e (b)  $n = 20$ .

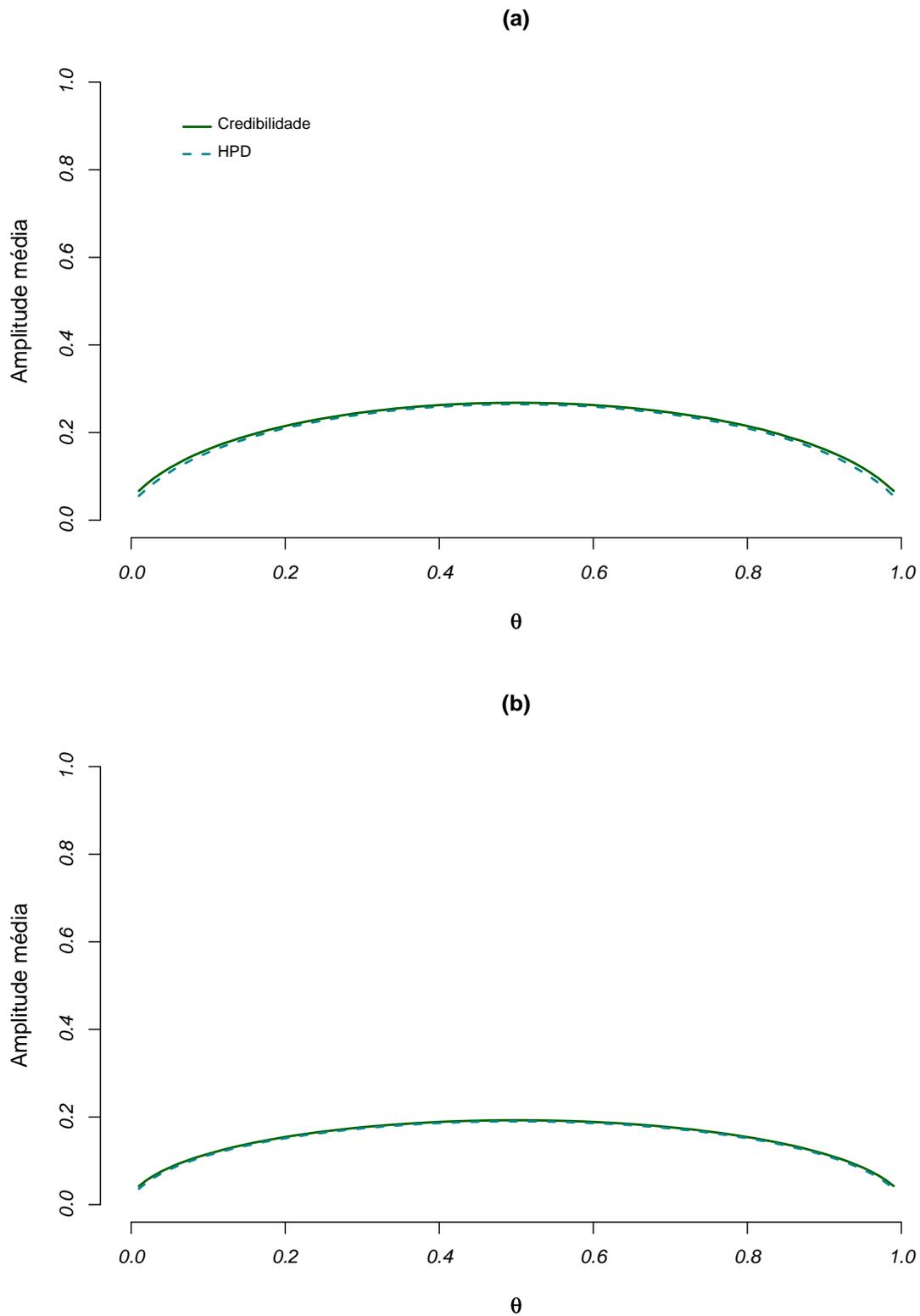


FIGURA 5.16: Amplitudes médias dos intervalos de credibilidade e HPD com distribuição *a priori* de Jeffreys para a probabilidade de sucesso do modelo Binomial: (a)  $n = 50$  e (b)  $n = 100$ .

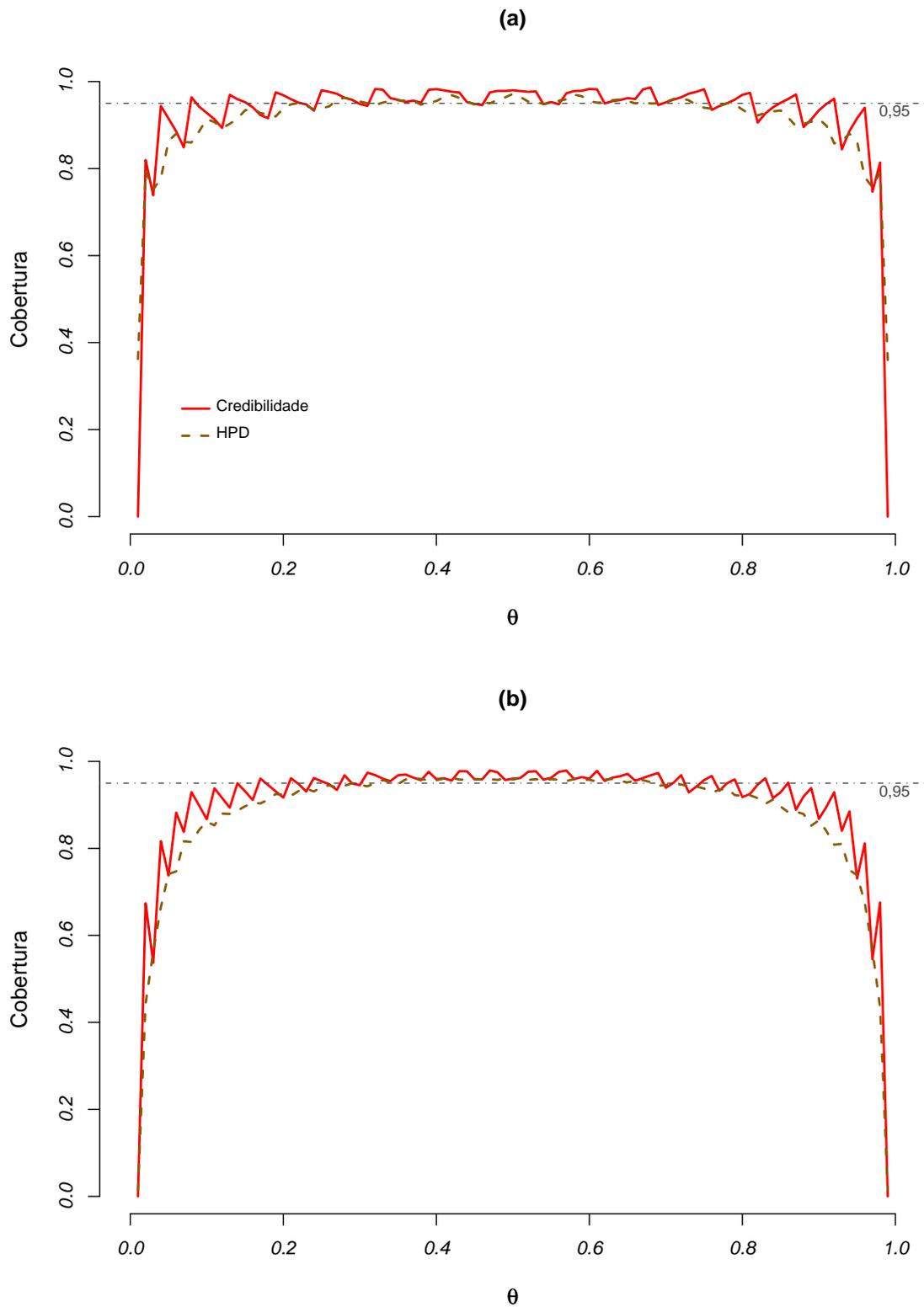


FIGURA 5.17: Proporções de cobertura dos intervalos de credibilidade e HPD com distribuição *a priori* menos favorável para a probabilidade de sucesso do modelo Binomial: (a)  $n = 10$  e (b)  $n = 20$ .

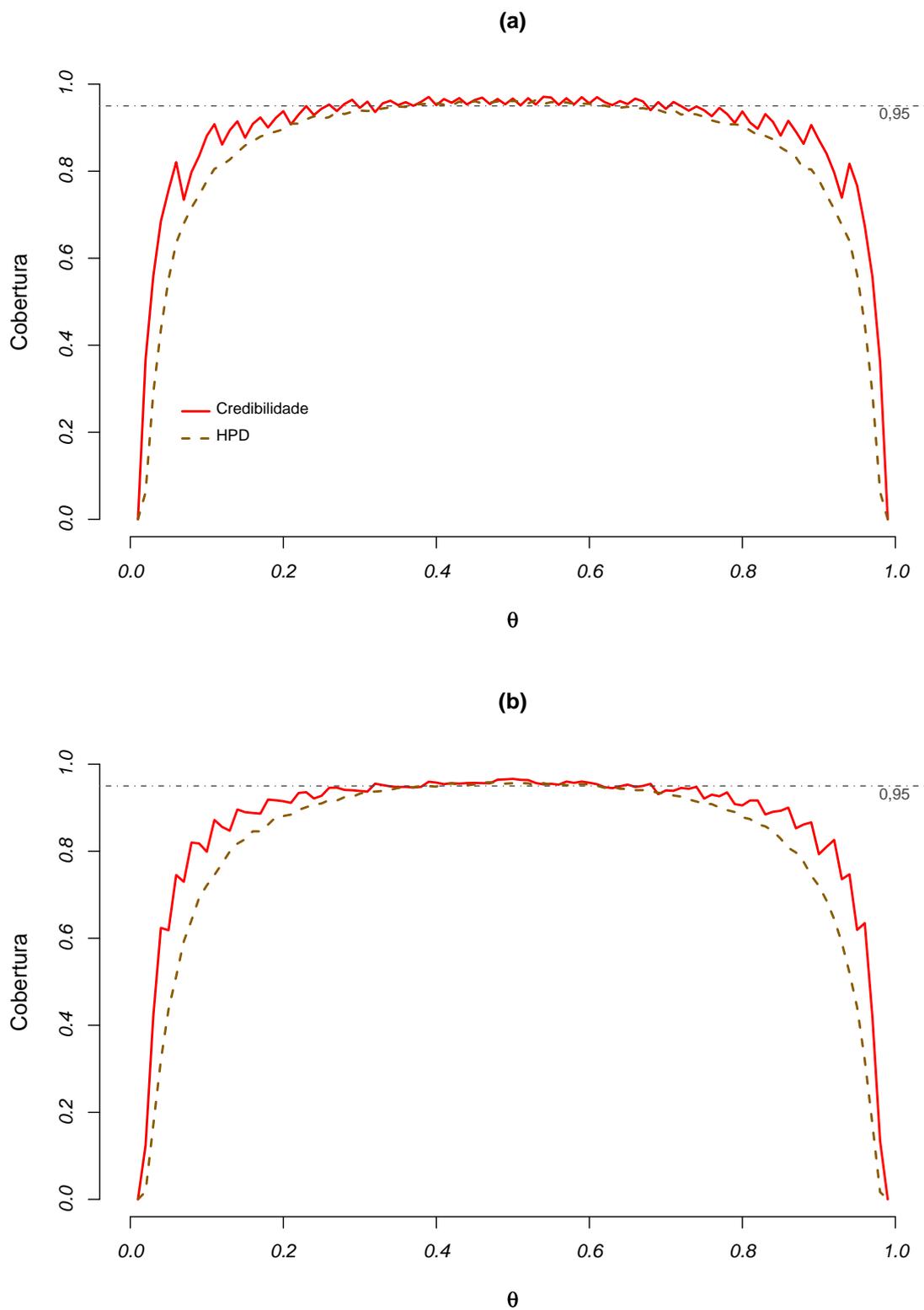


FIGURA 5.18: Proporções de cobertura dos intervalos de credibilidade e HPD com distribuição *a priori* menos favorável para a probabilidade de sucesso do modelo Binomial: (a)  $n = 50$  e (b)  $n = 100$ .

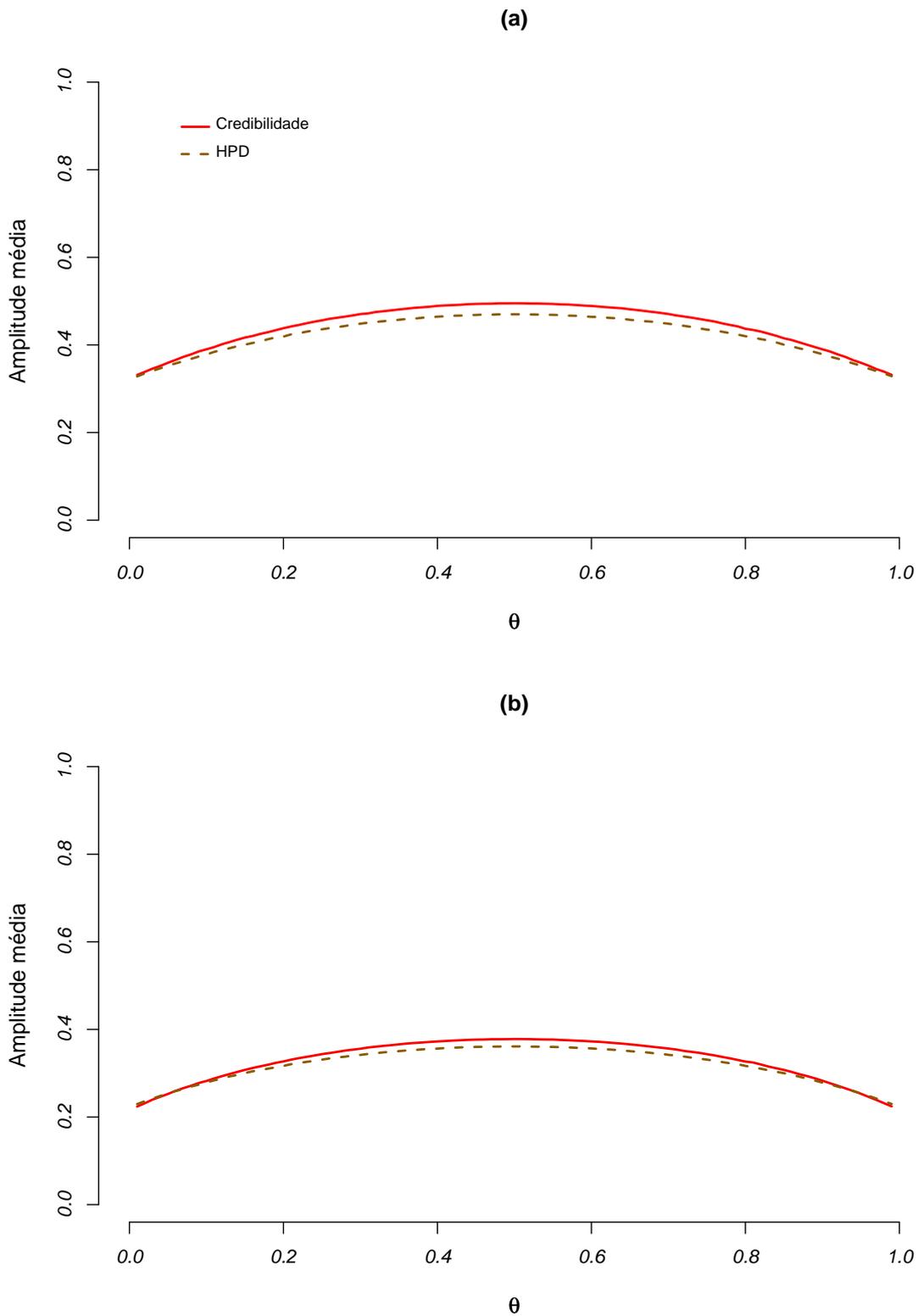


FIGURA 5.19: Amplitudes médias dos intervalos de credibilidade e HPD com distribuição *a priori* menos favorável para a probabilidade de sucesso do modelo Binomial: (a)  $n = 10$  e (b)  $n = 20$ .

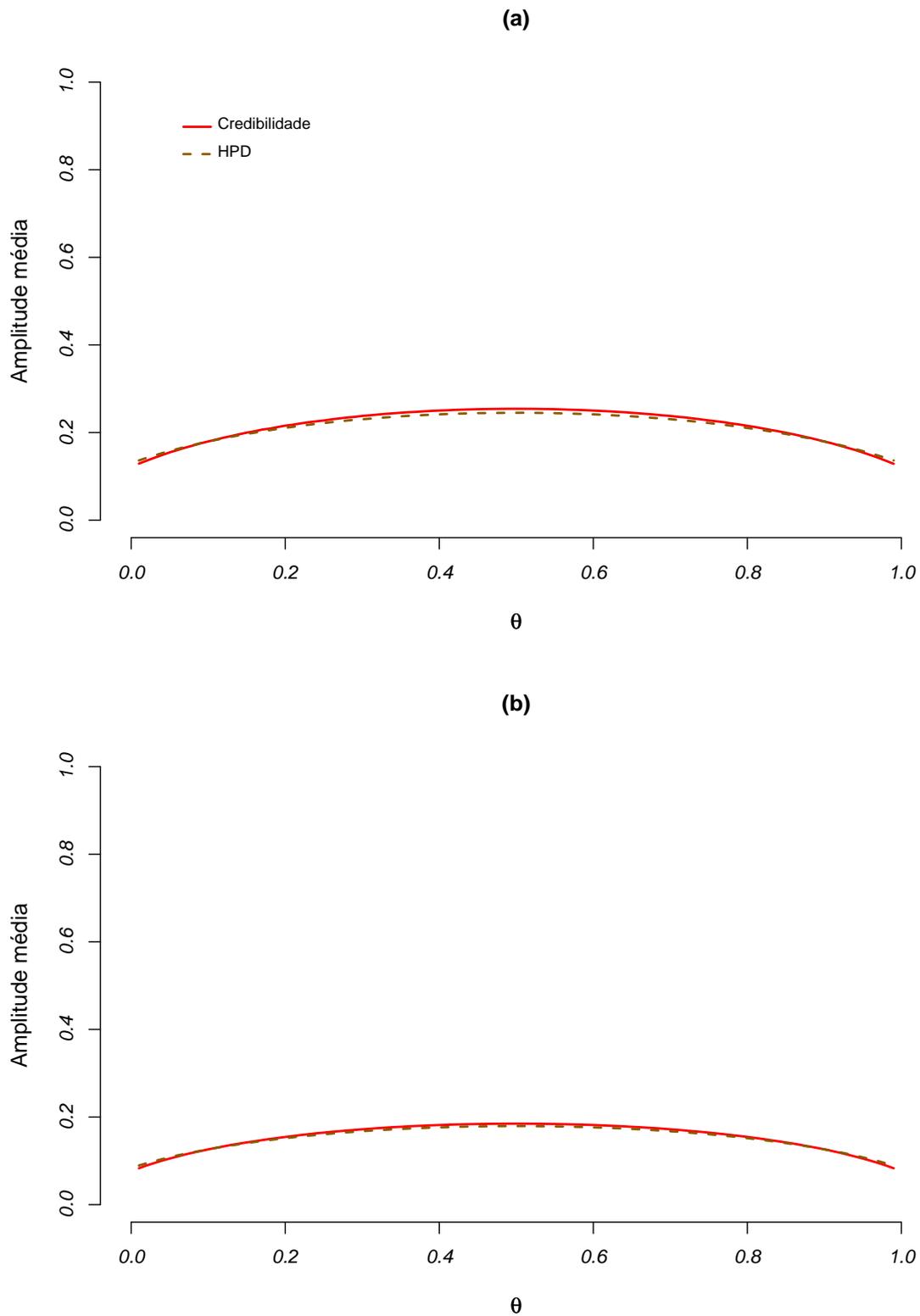


FIGURA 5.20: Amplitudes médias dos intervalos de credibilidade e HPD com distribuição *a priori* menos favorável para a probabilidade de sucesso do modelo Binomial: (a)  $n = 50$  e (b)  $n = 100$ .

### 5.1.2 Proporções de cobertura e amplitudes médias dos intervalos clássicos

Os intervalos considerados na metodologia clássica foram os intervalos com aproximação pela distribuição Normal, exato de Clopper & Pearson (1934) e de probabilidade de cobertura garantida de Chen *et al.* (2008), conforme mencionados no Capítulo 3.

Lembre-se que nos cursos introdutórios de Inferência Estatística é muito comum a aplicação da aproximação pela distribuição Normal para a estimação intervalar da probabilidade de sucesso do modelo Binomial, mas também é ressaltado que a qualidade da aproximação para amostras pequenas pode ser baixa devido ao resultado assintótico.

Diante dos resultados das simulações observados na figura 5.21, nota-se que as proporções de cobertura dos intervalos de confiança com aproximação pela distribuição Normal não são nem um pouco satisfatórias para amostras de tamanho 10, o que confirma a teoria. É verificado também que, para os intervalos de confiança com a aproximação pela distribuição Normal, as proporções de cobertura estão mais afastadas da nominal do que as demais metodologias analisadas neste estudo para a distribuição Binomial, mesmo com sensível melhora para  $n$  maior.

Observe as amplitudes médias destes intervalos na figura 5.23. Para  $n = 10$  os intervalos de confiança com aproximação pela distribuição Normal apresentam amplitudes dos intervalos sempre menores ou iguais às amplitudes médias dos demais intervalos.

Bayarri & Berger (2004), ao tratarem da oscilação da probabilidade de cobertura do intervalo de confiança para a probabilidade de sucesso da distribuição Binomial, citam o fato de a distribuição Binomial ser discreta, mais precisamente mencionam sobre a estrutura da distribuição.

Nas figuras 5.21 e 5.22 observe que as proporções de cobertura nos intervalos de confiança com aproximação pela distribuição Normal para valores de  $\theta$  próximos a 0,5 são mais próximas do desejado. Nestes casos a distribuição Binomial é simétrica, assim como a distribuição Normal pela qual é aproximada.

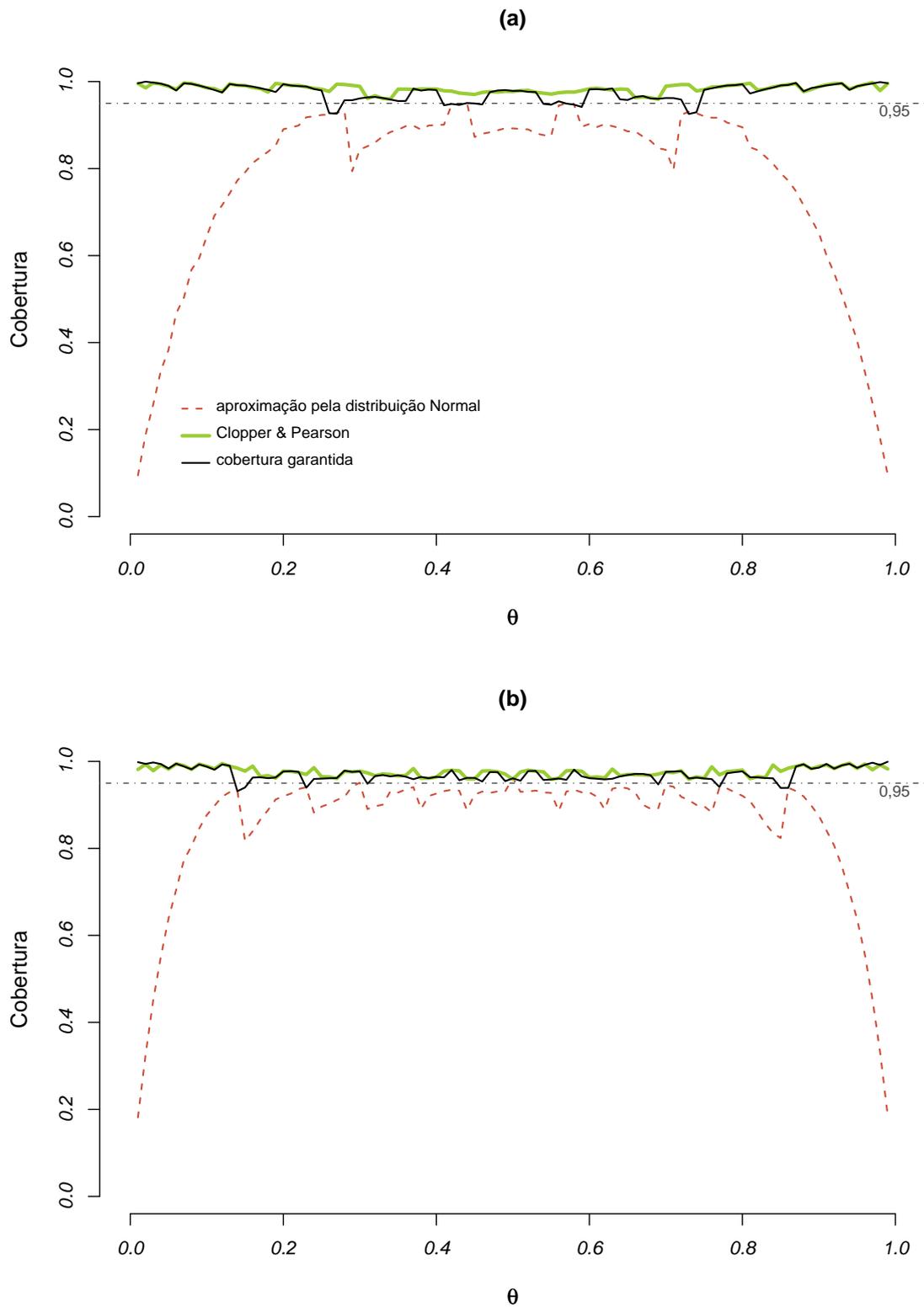


FIGURA 5.21: Proporções de cobertura dos intervalos clássicos para a probabilidade de sucesso do modelo Binomial: (a)  $n = 10$  e (b)  $n = 20$ .

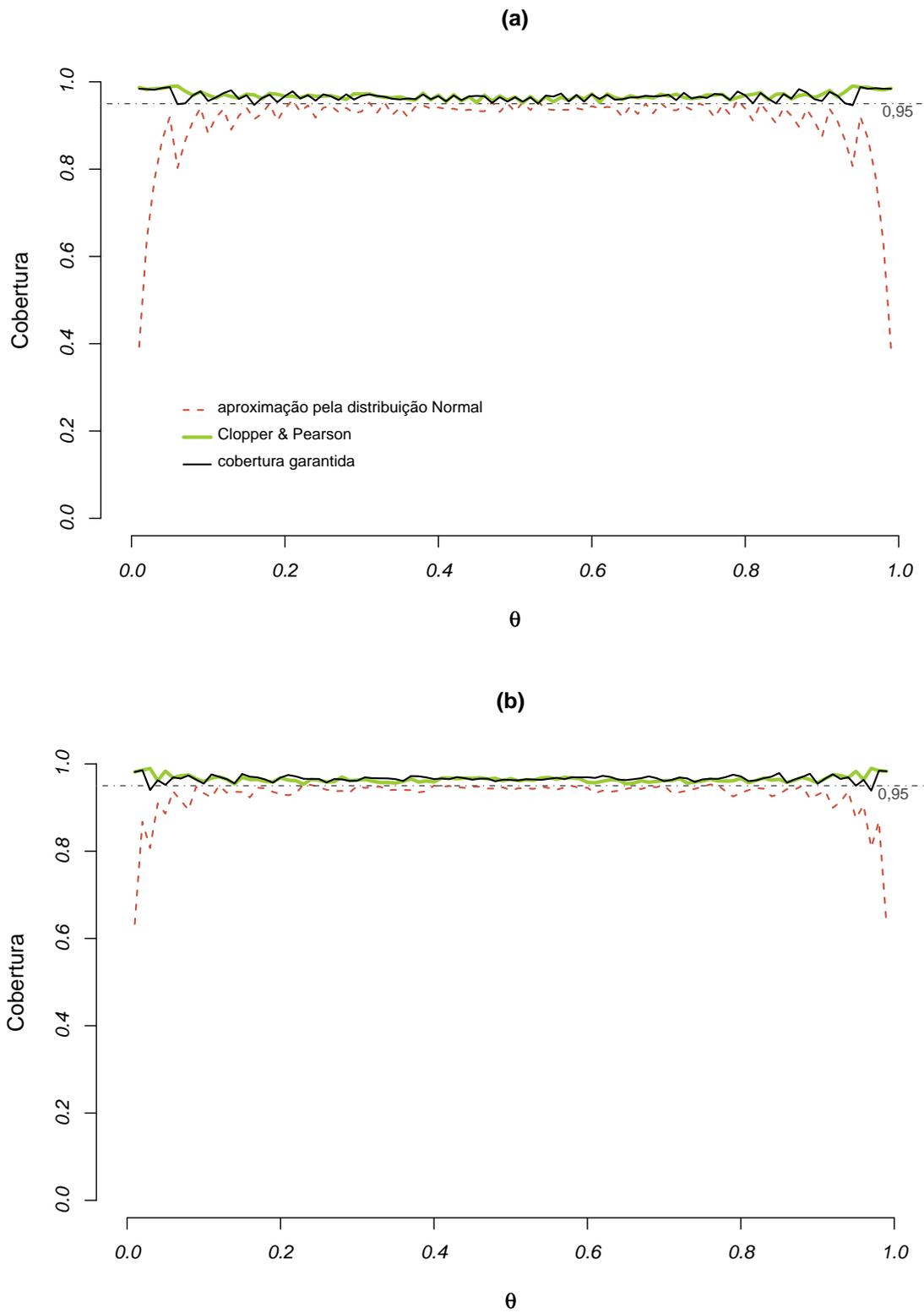


FIGURA 5.22: Proporções de cobertura dos intervalos clássicos para a probabilidade de sucesso do modelo Binomial: (a)  $n = 50$  e (b)  $n = 100$ .

Para o intervalo de confiança com aproximação pela distribuição Normal com valores de  $\theta$  próximos de 0 ou de 1, a amplitude média dos intervalos é menor que as dos intervalos bayesianos para amostras de tamanho  $n = 10$  e  $20$ , mas como o verdadeiro valor do parâmetro é desconhecido e como as proporções de cobertura dos intervalos de credibilidade e HPD mostraram-se mais estáveis, estes são mais aconselháveis, diante das informações obtidas pelo estudo de simulação. Para os demais tamanhos amostrais, as amplitudes são muito próximas.

O intervalo exato de Clopper & Pearson (1934) geralmente é tratado como *padrão ouro*. No entanto, este procedimento é necessariamente conservador, conforme Brown *et al.* (2002) e Chen *et al.* (2008). Para qualquer valor fixo do parâmetro, a probabilidade de cobertura pode ser muito maior que a nominal, a não ser que  $n$  seja bastante grande. Acredita-se que é inadequada esta abordagem estatística como ideal para a prática, conforme destacam Agresti & Coull (1998).

O que se pode observar na figura 5.21 é que as proporções de cobertura em ambos os intervalos exatos é sobrestimada para  $n = 10$ , embora o intervalo exato de Clopper & Pearson (1934) se mostre mais conservador. Um fator positivo a favor do intervalo de confiança com cobertura garantida é que suas amplitudes médias são sempre inferiores às apresentadas pelo intervalo exato de Clopper & Pearson (1934), para esse tamanho amostral.

Ao observar as proporções de cobertura obtidas para o intervalo com probabilidade de cobertura garantida de Chen *et al.* (2008) pode-se afirmar as mesmas são mais próximas da nominal em relação ao exato de Clopper & Pearson (1934), exceto para  $n = 100$ .

Note ainda que os métodos gerados pela teoria clássica para se obter intervalos para a probabilidade de sucesso do modelo Binomial ou são muito conservadores (exato de Clopper & Pearson (1934) e probabilidade de cobertura garantida Chen *et al.* (2008)), pois sobrestimam as proporções de cobertura em quase todo espaço paramétrico, ou subestimam em todo espaço paramétrico no caso da aproximação pela distribuição Normal.

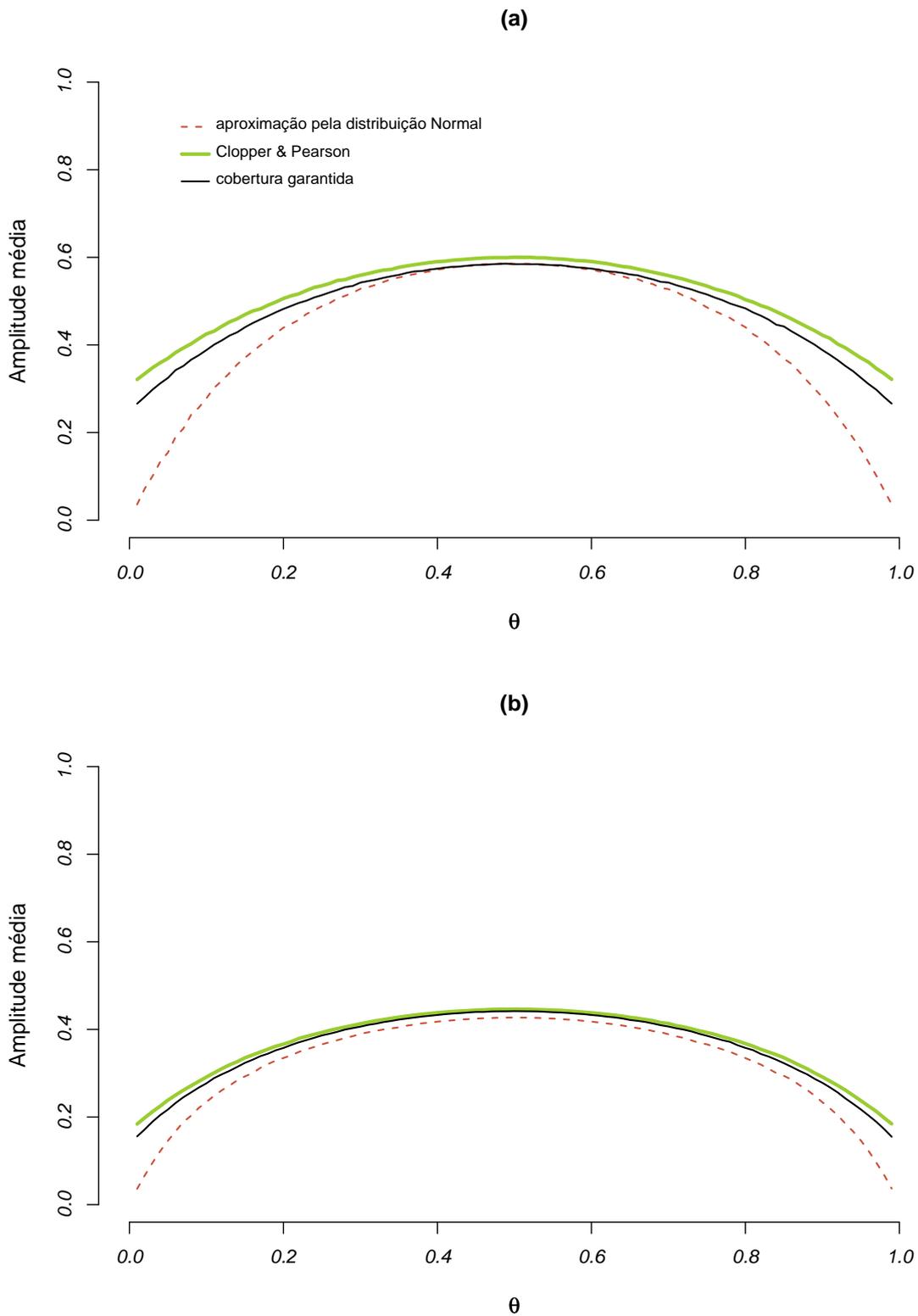


FIGURA 5.23: Amplitudes médias dos intervalos clássicos para a probabilidade de sucesso do modelo Binomial: (a)  $n = 10$  e (b)  $n = 20$ .

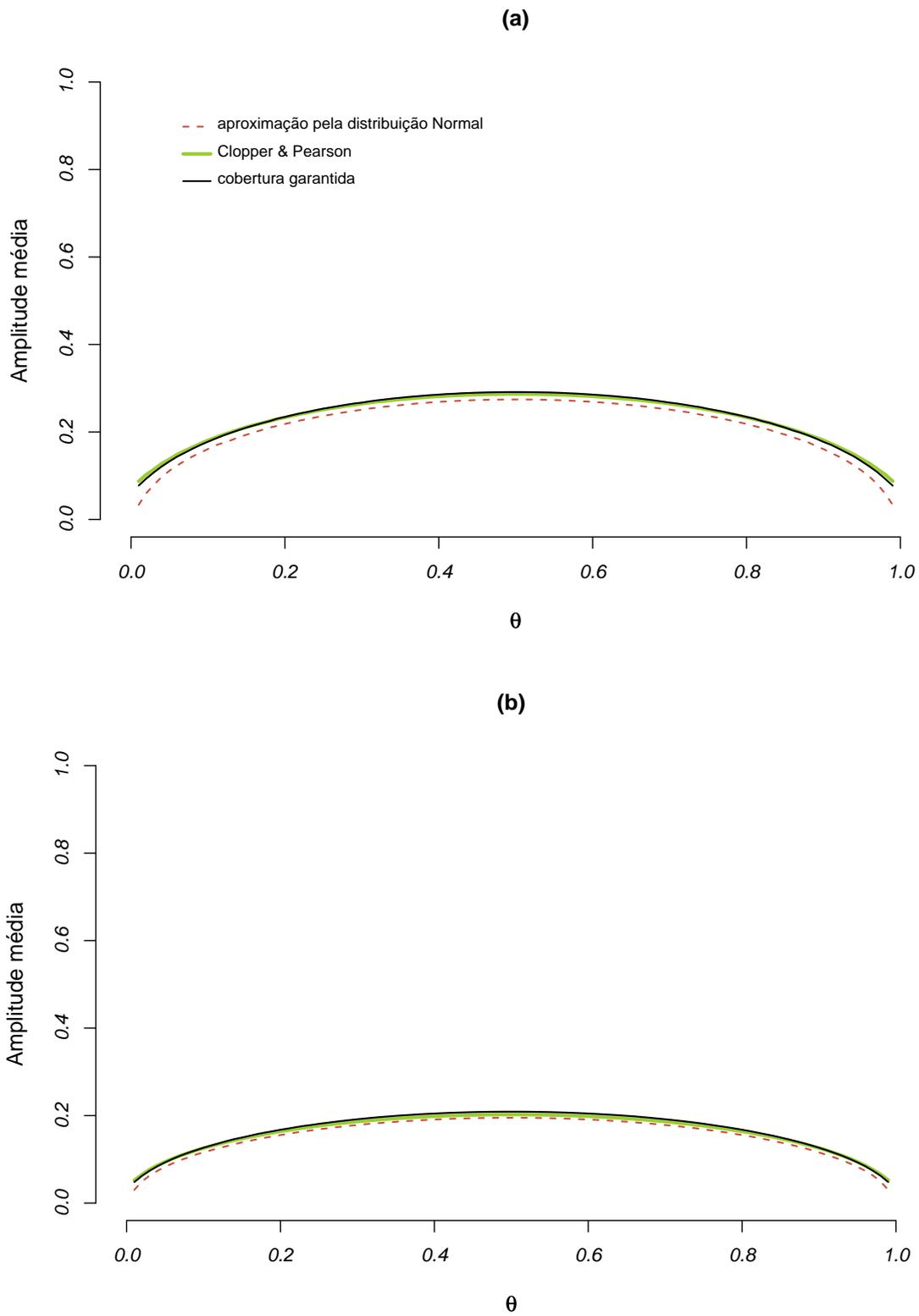


FIGURA 5.24: Amplitudes médias dos intervalos clássicos para a probabilidade de sucesso do modelo Binomial: (a)  $n = 50$  e (b)  $n = 100$ .

### 5.1.3 Consistência, viés e erro quadrático médio dos estimadores

As funções de perda utilizadas para obtenção da distribuição *a priori* menos favorável, expostas no Capítulo 3, evidenciaram dois estimadores bayesianos, a saber: a média da distribuição *a posteriori* de  $\theta$  ao empregar a função de perda quadrática e a moda da distribuição *a posteriori* de  $\theta$  no caso da função de perda quadrática ponderada.

Portanto, é de interesse confirmar por meio de simulação a consistência e calcular o viés e o erro quadrático médio desses estimadores bayesianos, que são viciados, para estimar pontualmente a probabilidade de sucesso do modelo Binomial (ver Apêndice B).

O erro quadrático médio incorpora duas componentes: a variabilidade do estimador e a medida de vício, considera-se o estimador com melhores propriedades o qual apresentar o menor erro quadrático médio (Casella & Berger, 2002, p. 311). Outra propriedade desejável para um estimador pontual é a consistência, pois implica que o estimador convirja para o “correto” valor quando o tamanho da amostra tende ao infinito (Casella & Berger, 2002, p. 330).

No Apêndice B encontram-se calculadas as esperanças e variâncias dos estimadores pontuais bayesianos e de máxima verossimilhança. No estudo de simulação realizado (figuras 5.25 e 5.26) pode-se observar vícios muito próximos de zero para o estimador de máxima verossimilhança, razão evidenciada pelo cálculo da sua esperança (ele é não viciado para  $\theta$ ). Sua consistência também foi confirmada no estudo de simulação.

Nas figuras 5.27, 5.28, 5.31 e 5.32 ao analisar os erros quadráticos médios do estimador de máxima verossimilhança diante dos estimadores bayesianos observa-se que para  $n = 10$  e  $20$  é perceptível a diferença entre eles, mas para  $n = 50$  e  $100$  são muito similares. O estimador clássico apresentou erros quadráticos médios superiores ao estimador média da distribuição *a posteriori* de  $\theta$  independente da distribuição *a priori* utilizada, mas inferiores ao estimador bayesiano moda da distribuição *a posteriori* ao considerar a distribuição *a priori* de Jeffreys.

As figuras 5.25 e 5.26 apresentam os vícios e as figuras 5.27 e 5.28 os erros quadráticos médios dos estimadores de máxima verossimilhança e da média da distribuição *a posteriori* para a probabilidade de sucesso do modelo Binomial com distribuição *a priori* de Bayes-Laplace, de Jeffreys e menos favorável. Nas figuras 5.29 e 5.30 seguem os

vícios e nas figuras 5.31 e 5.32 os erros quadráticos médios dos estimadores de máxima verossimilhança e da moda da distribuição *a posteriori* de  $\theta$  para as mesmas distribuições *a priori*.

Lembre-se que a moda da distribuição *a posteriori* de  $\theta$  ao considerar a distribuição *a priori* menos favorável com função de perda quadrática ponderada, coincidente com a distribuição *a priori* de Bayes-Laplace (Seção 3.1.2), tem estimador comum ao de máxima verossimilhança. Por esse motivo os valores obtidos para os vícios (figuras 5.29 e 5.30) e erros quadráticos médios (figuras 5.31 e 5.32) se sobrepõem.

A média da distribuição *a posteriori* de  $\theta$  é um estimador viciado e consistente ao considerar a distribuição *a priori* de Bayes-Laplace e de Jeffreys, confirmado no estudo de simulação. No entanto, não parece razoável, neste estudo de simulação, a consistência desse estimador ao empregar a distribuição *a priori* menos favorável, pois com o aumento de  $n$  o vício tende mais lentamente a zero.

Ao comparar os vícios da média da distribuição *a posteriori* de  $\theta$  nota-se, pelas figuras 5.25 e 5.26, que ao utilizar a distribuição *a priori* de Jeffreys ocorrem os menores vícios. Vícios muito próximos de zero quando  $\theta = 0,5$  (independente da distribuição *a priori* empregada), com o afastamento desse valor, o vício apresenta-se de forma simétrica sempre sobrestimando o valor “correto” do parâmetro quando  $\theta < 0,5$  e subestimando quando  $\theta > 0,5$ .

Os erros quadráticos médios para a média da distribuição *a posteriori* de  $\theta$  (figuras 5.27 e 5.28) são constantes em relação ao espaço paramétrico ao utilizar a distribuição *a priori* menos favorável, e para as demais distribuições *a priori* tem comportamento simétrico em torno de  $\theta = 0,5$ , em que assume valores máximos, diminuindo gradativamente para valores extremos de  $\theta$ . Também pode-se observar que com o aumento de  $n$  os erros quadráticos médios diminuem sensivelmente. Comportamento similar são observados para a moda da distribuição *a posteriori* de  $\theta$  (figuras 5.31 e 5.32).

Ao considerar a moda da distribuição *a posteriori* de  $\theta$  (figuras 5.29 e 5.30) é observado ao empregar a distribuição *a priori* menos favorável um comportamento muito similar ao observado por ela no caso da média. No entanto, ao considerar a distribuição *a priori* de Jeffreys ocorre o oposto, mas também apresenta-se consistente para estimar a probabilidade de sucesso do modelo Binomial.

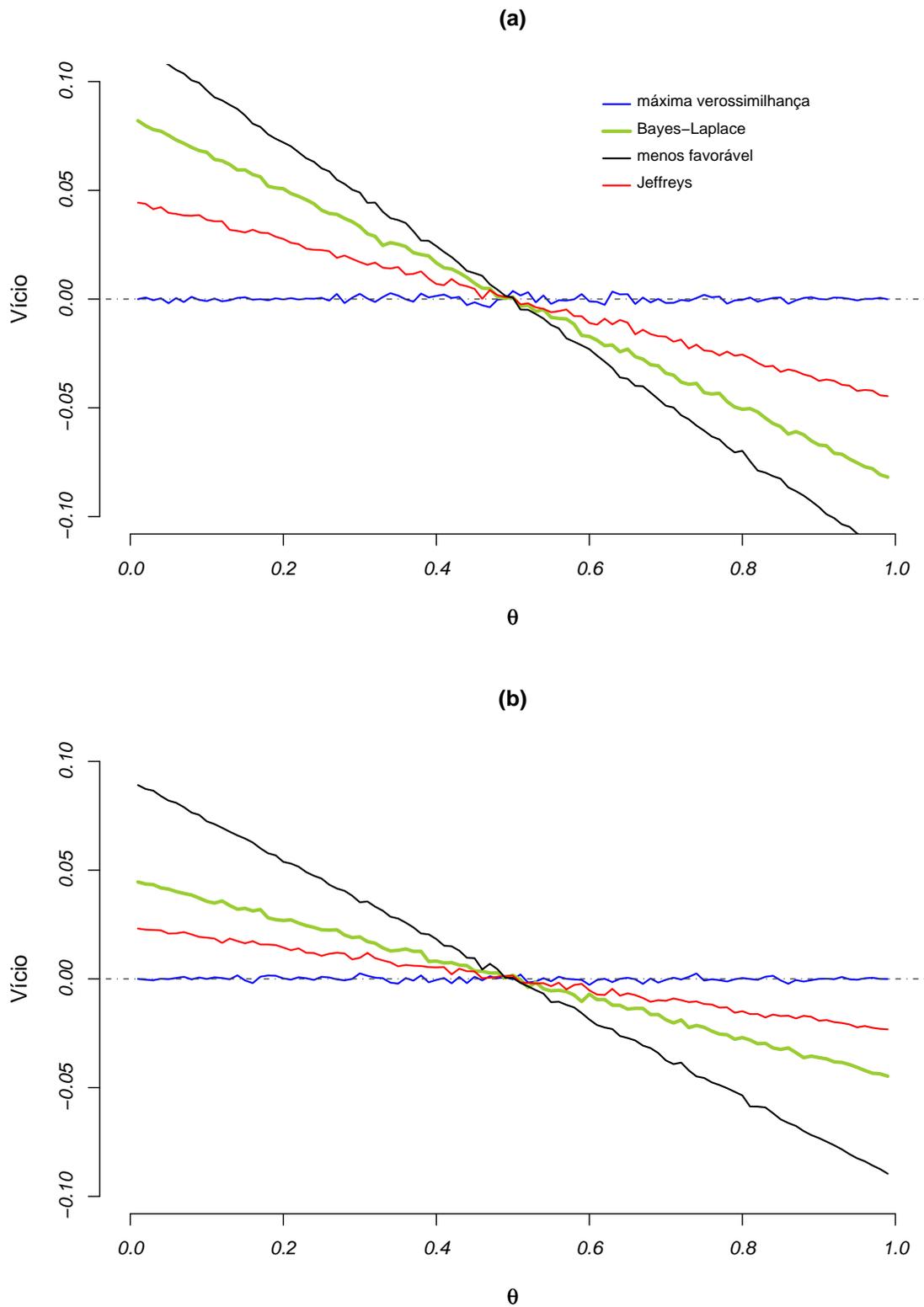


FIGURA 5.25: Vícios dos estimadores de máxima verossimilhança e da média da distribuição *a posteriori* para a probabilidade de sucesso do modelo Binomial: (a)  $n = 10$  e (b)  $n = 20$ .

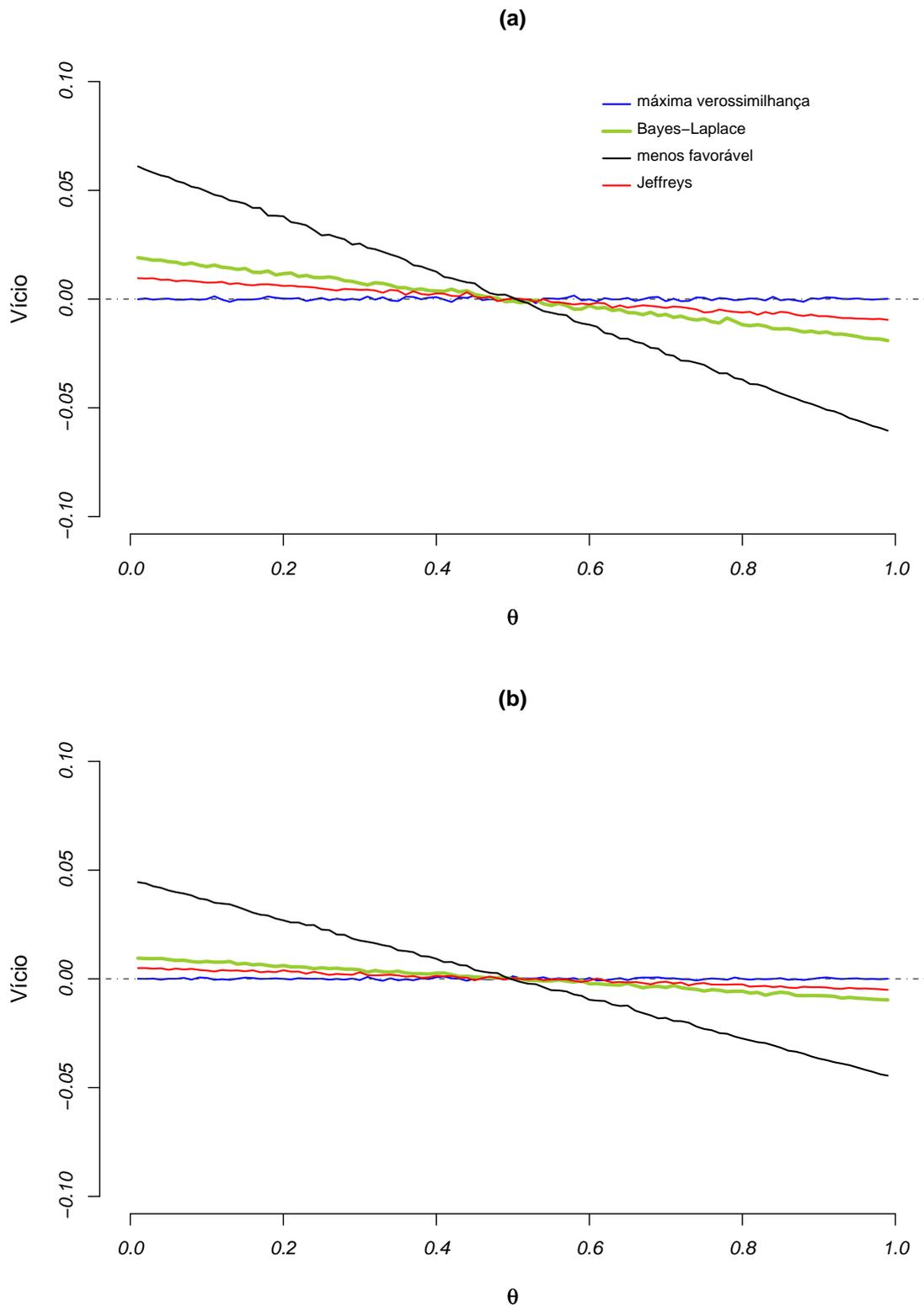


FIGURA 5.26: Vícios dos estimadores de máxima verossimilhança e da média da distribuição *a posteriori* para a probabilidade de sucesso do modelo Binomial: (a)  $n = 50$  e (b)  $n = 100$ .

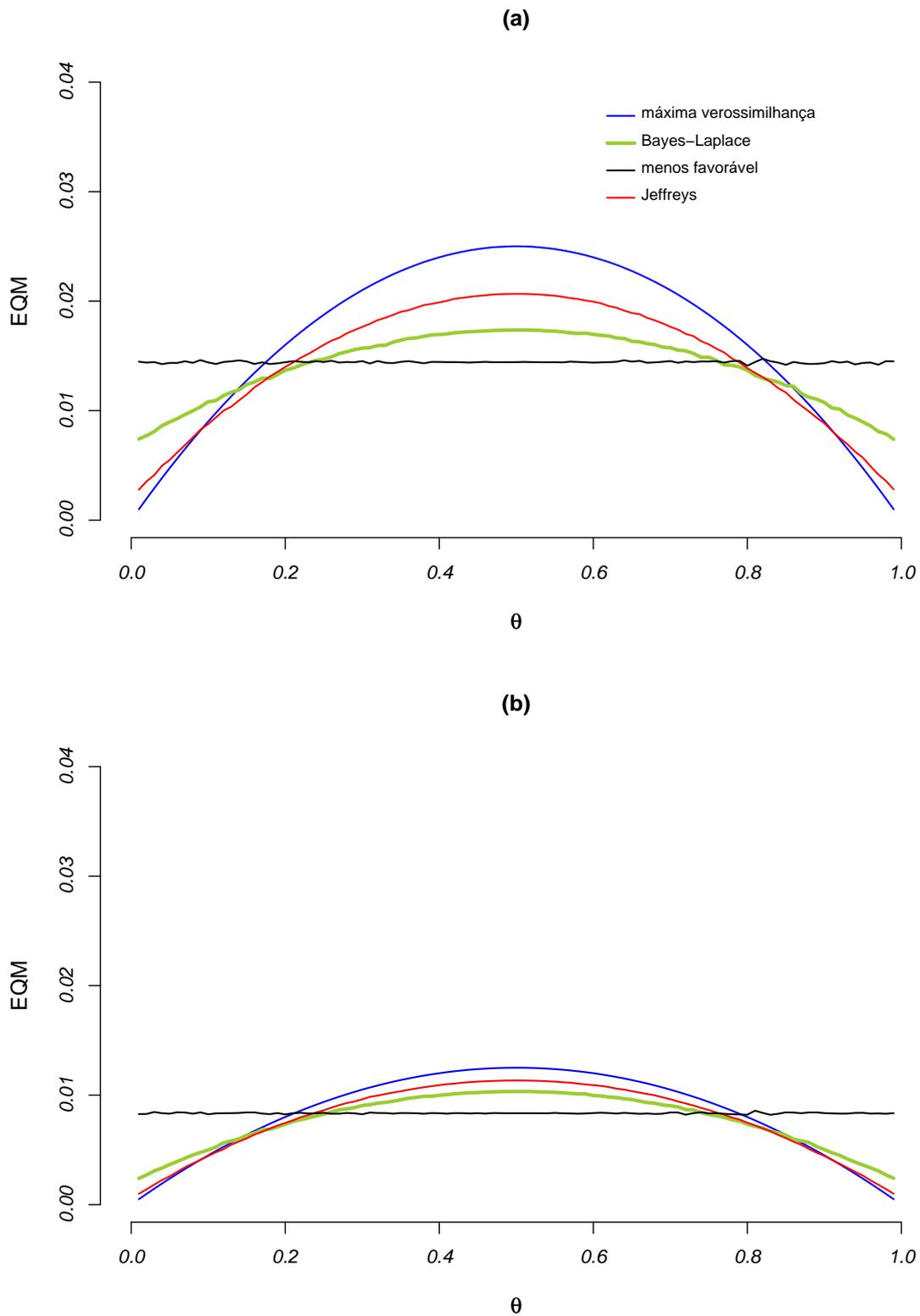


FIGURA 5.27: Erros quadráticos médios dos estimadores de máxima verossimilhança e da média da distribuição *a posteriori* para a probabilidade de sucesso do modelo Binomial: (a)  $n = 10$  e (b)  $n = 20$ .

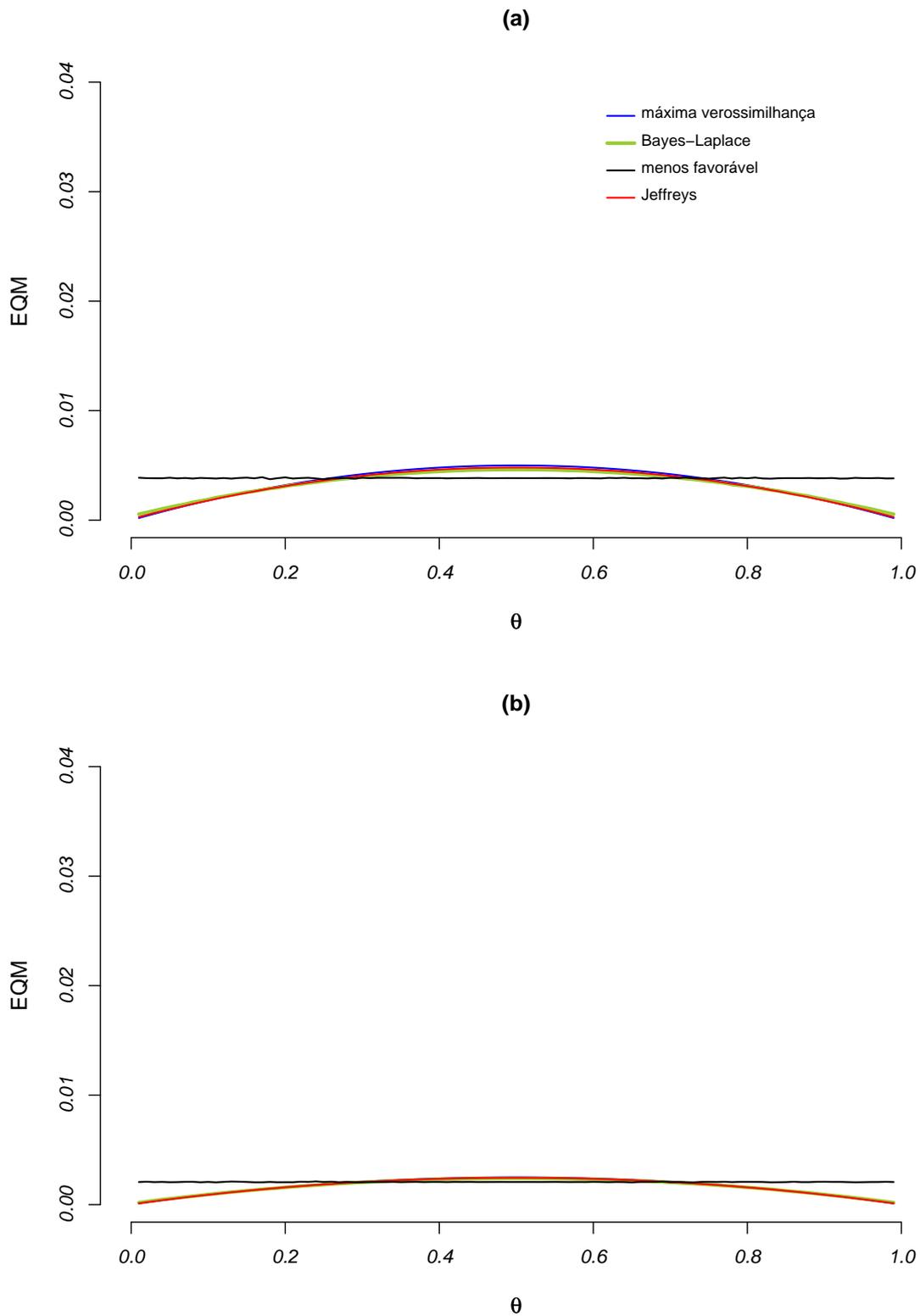


FIGURA 5.28: Erros quadráticos médios dos estimadores de máxima verossimilhança e da média da distribuição *a posteriori* para a probabilidade de sucesso do modelo Binomial: (a)  $n = 50$  e (b)  $n = 100$ .

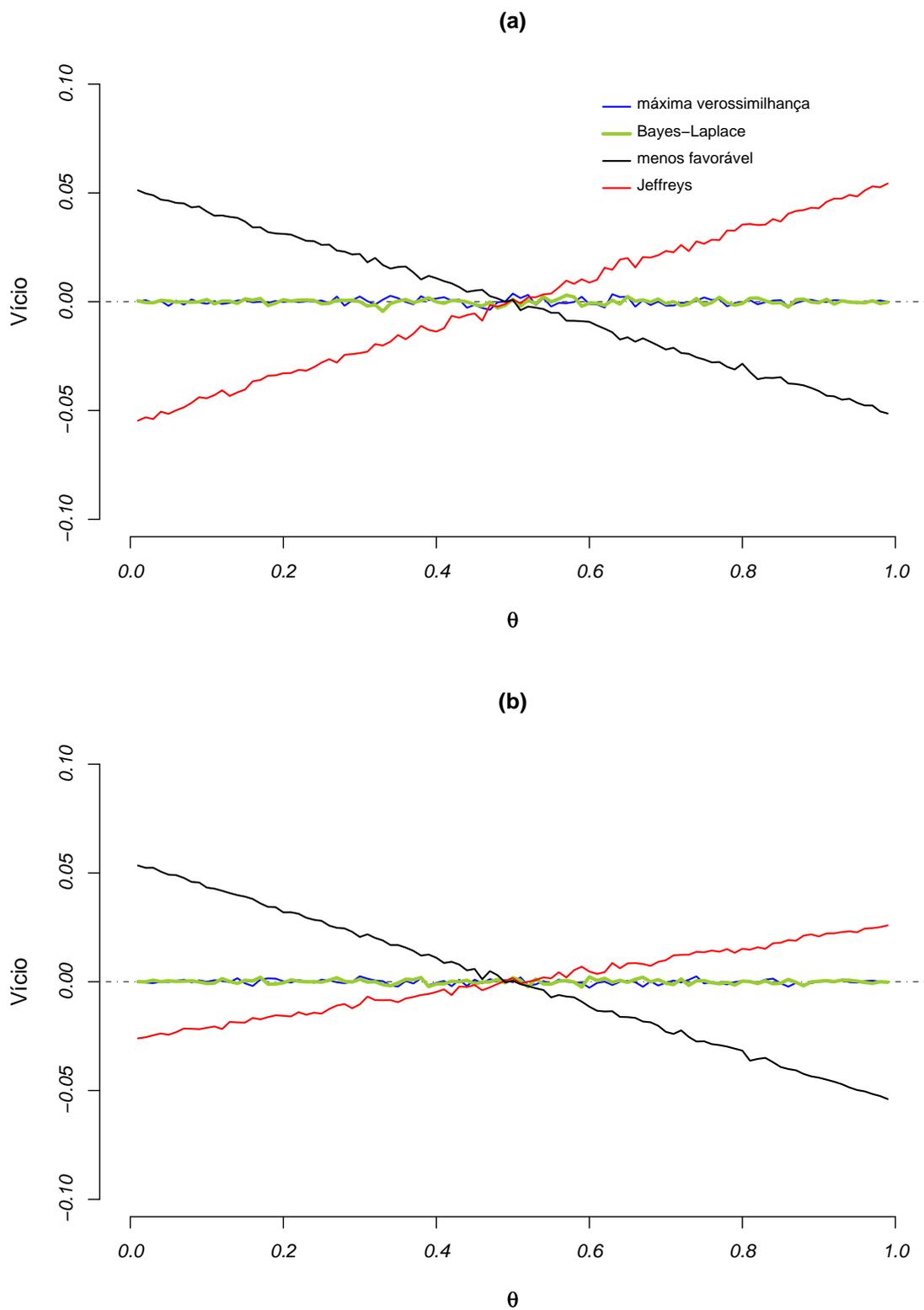


FIGURA 5.29: Vícios dos estimadores de máxima verossimilhança e da moda da distribuição *a posteriori* para a probabilidade de sucesso do modelo Binomial: (a)  $n = 10$  e (b)  $n = 20$ .

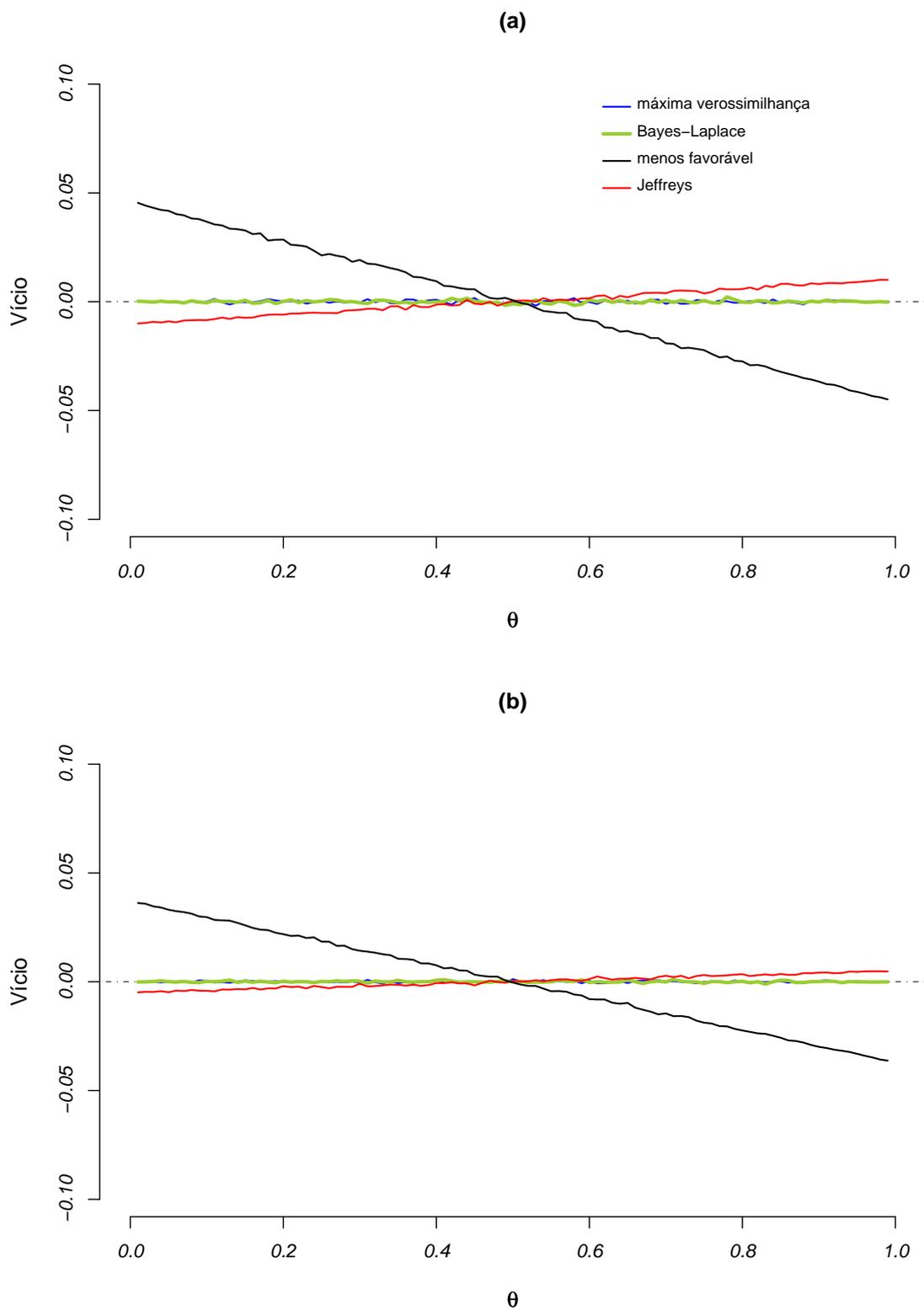


FIGURA 5.30: Vícios dos estimadores de máxima verossimilhança e da moda da distribuição *a posteriori* para a probabilidade de sucesso do modelo Binomial: (a)  $n = 50$  e (b)  $n = 100$ .

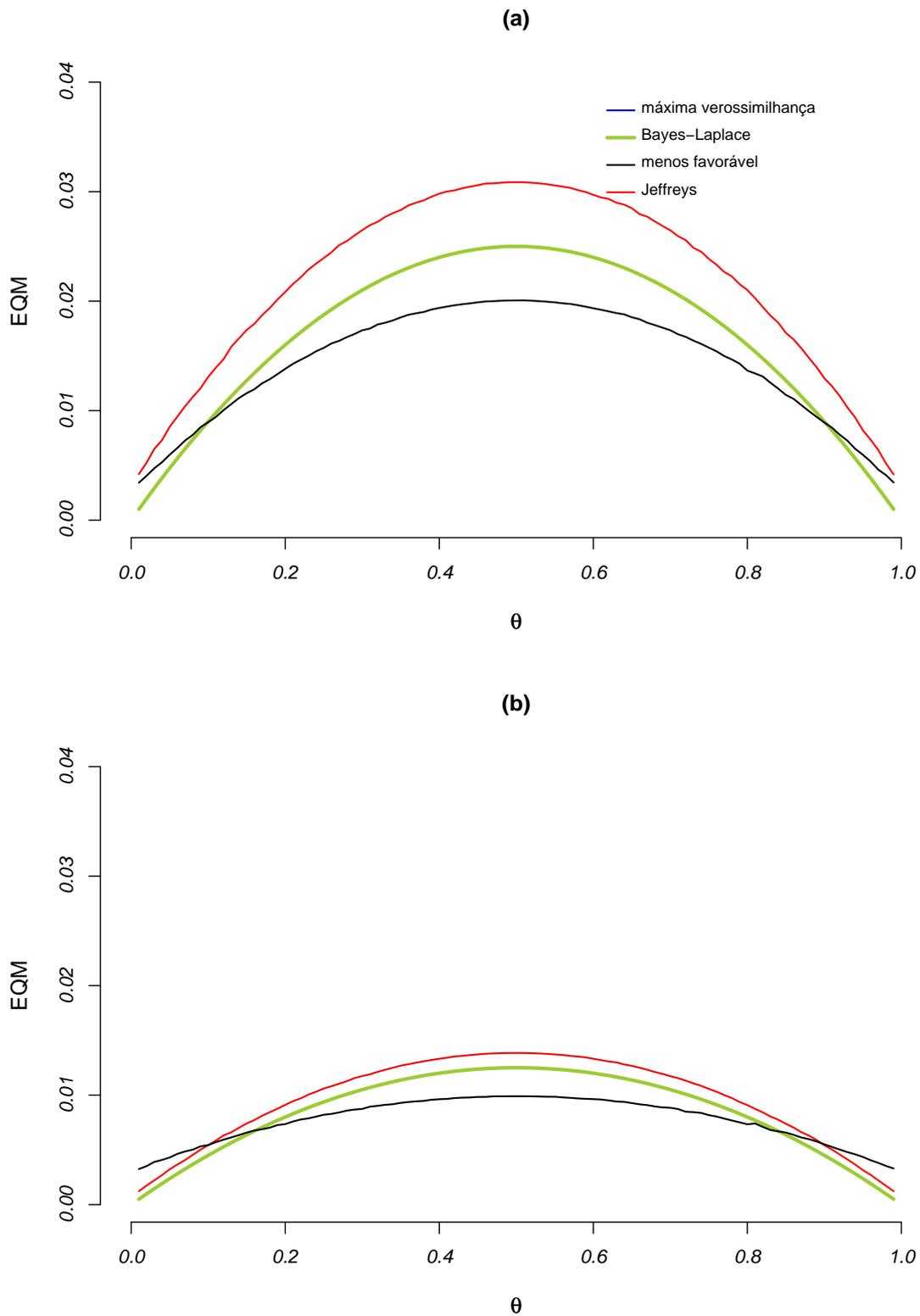


FIGURA 5.31: Erros quadráticos médios dos estimadores de máxima verossimilhança e da moda da distribuição *a posteriori* para a probabilidade de sucesso do modelo Binomial: (a)  $n = 10$  e (b)  $n = 20$ .

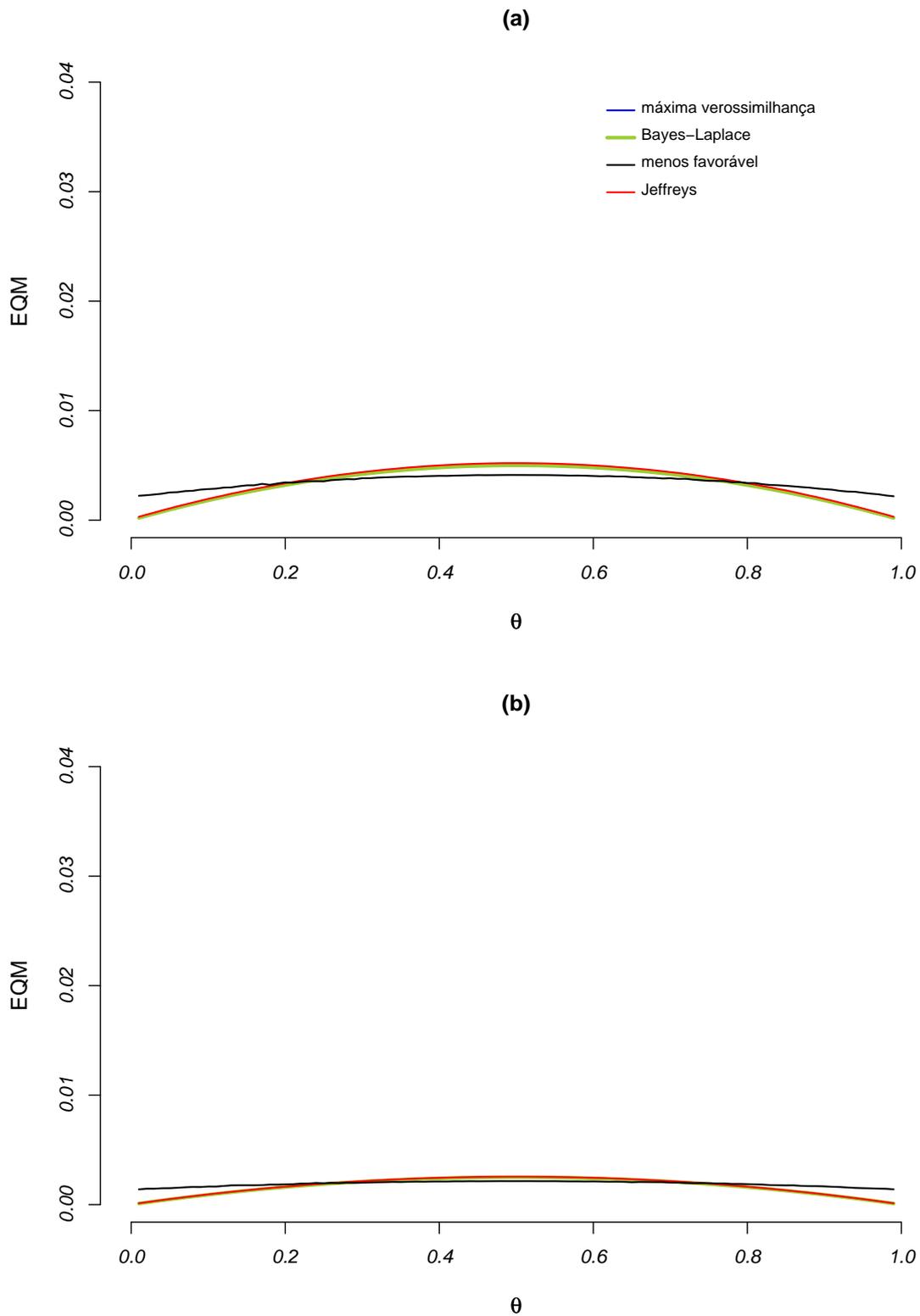


FIGURA 5.32: Erros quadráticos médios dos estimadores de máxima verossimilhança e da moda da distribuição *a posteriori* para a probabilidade de sucesso do modelo Binomial: (a)  $n = 50$  e (b)  $n = 100$ .

## 5.2 Distribuição Binomial Negativa

### 5.2.1 Proporções de cobertura e amplitudes médias dos intervalos bayesianos

Para a estimação intervalar da probabilidade de sucesso do modelo Binomial Negativa foram considerados os intervalos bayesianos de credibilidade e HPD, para os quais foram aplicadas a distribuição *a priori* de Bayes-Laplace e de Jeffreys, conforme Capítulo 4.

#### Intervalos de credibilidade

As figuras 5.33 e 5.34 apresentam as proporções de cobertura obtidas pelos intervalos de credibilidade para a distribuição *a priori* de Bayes-Laplace e de Jeffreys para os diferentes valores do parâmetro. Espera-se que a cobertura esteja próxima de 95%.

É importante salientar que nos casos em que  $r = 3$  e  $r = 4$  é necessário um número maior de ensaios para que se possa alcançar o número fixado de sucessos, portanto há mais informação para construção dos intervalos.

Observe para  $r = 1$ , na figura 5.33, que as proporções de cobertura dos intervalos de credibilidade com distribuição *a priori* de Bayes-Laplace sobrestimam a cobertura nominal no intervalo  $\theta \in (0, 16; 0, 90)$  e para valores de  $\theta$  não pertencentes a esse intervalo as proporções de cobertura são instáveis, principalmente para extremos inferiores ( $\theta < 0, 2$ ) (com evidente subestimação). Para  $r = 2$  a sobrestimação ocorre em  $\theta \in (0, 30; 0, 93)$ , porém é visível que a proporção de cobertura se aproxima de 95%, em relação ao observado para  $r = 1$  com  $\theta < 0, 2$ .

Para os intervalos de credibilidade com distribuição *a priori* de Jeffreys a proporção de cobertura também apresenta-se sobestimando quase todo espaço paramétrico, embora com menos instabilidade, conforme figura 5.33 para  $r = 1$ . Ao observar as amplitudes médias geradas por estes intervalos nota-se que eles apresentam-se mais conservadores em relação aos intervalos de credibilidade que utilizam a distribuição *a priori* de Bayes-Laplace (figuras 5.35 e 5.36).

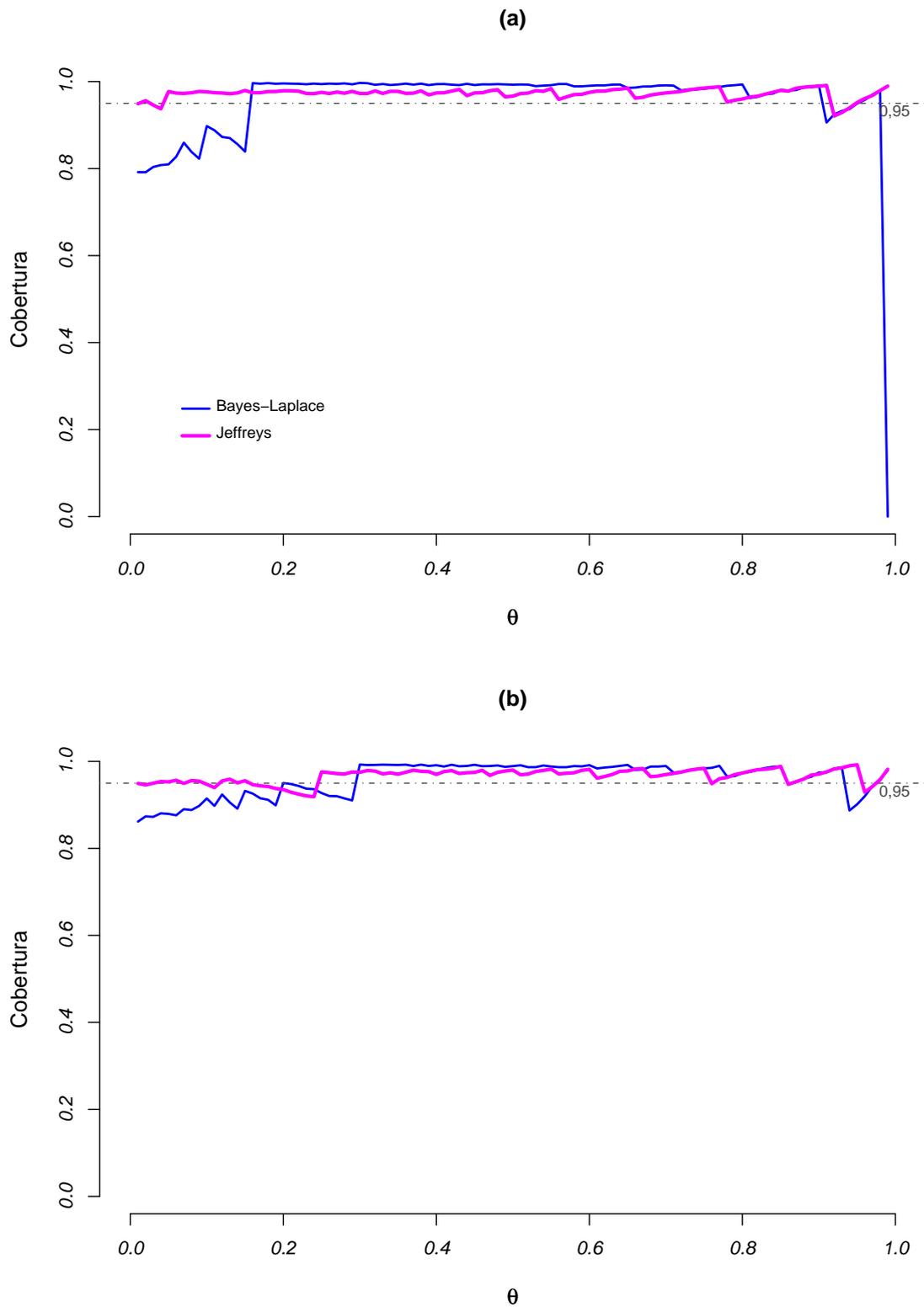


FIGURA 5.33: Proporções de cobertura dos intervalos de credibilidade para a probabilidade de sucesso do modelo Binomial Negativa: (a)  $r = 1$  e (b)  $r = 2$ .

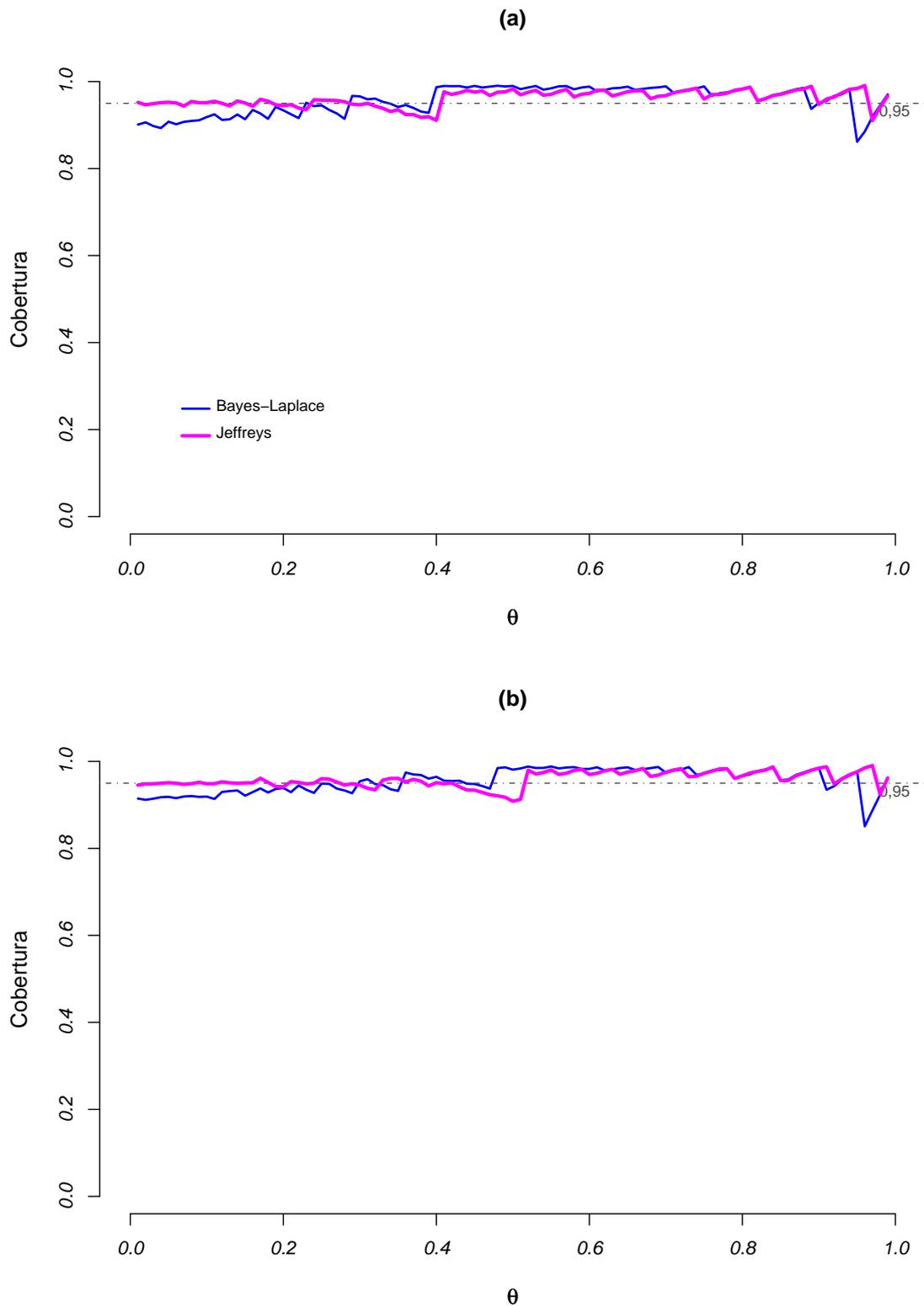


FIGURA 5.34: Proporções de cobertura dos intervalos de credibilidade para a probabilidade de sucesso do modelo Binomial Negativa: (a)  $r = 3$  e (b)  $r = 4$ .

Observe que as proporções de cobertura, expostas nas figuras 5.33 e 5.34, não mudam sensivelmente com o aumento de  $r$ , para  $\theta > 0,5$ , no estudo de simulação analisado. Isso pode estar relacionado com o tamanho amostral para valores de  $\theta$  no modelo Binomial Negativo próximos de 1, pois menos amostras precisam ser retiradas para se obter o número de sucessos desejados.

As proporções de cobertura dos intervalos de credibilidade, expostas na figura 5.34, mostram que ambas as distribuições *a priori* têm comportamentos muito similares, embora as proporções de cobertura dos intervalos de credibilidade com distribuição *a priori* de Jeffreys apresentem-se muito mais próximas da nominal para  $\theta < 0,30$  e  $\theta < 0,40$ , para  $r = 3$  e  $r = 4$ , respectivamente. Um dos fatores pelos quais as proporções de cobertura dos intervalos de credibilidade com essa distribuição *a priori* são mais próximos de 95% pode ser a forma da distribuição *a priori* de Jeffreys que atribui maior peso para valores da probabilidade de sucesso menores que 0,2 (figura 4.1).

Em geral as proporções de cobertura dos intervalos de credibilidade com distribuição *a priori* de Jeffreys apresentaram proporções de cobertura mais próximas da nominal ao longo do espaço paramétrico no estudo de simulação analisado. Mas ao observar as figuras 5.35 e 5.36 pode-se ver que os intervalos de credibilidade com distribuição *a priori* de Jeffreys têm amplitudes médias maiores ou muito próximas das amplitudes médias observadas para os intervalos de credibilidade com distribuição *a priori* de Bayes-Laplace, para  $r = 1$  e  $r = 2$ .

Com o aumento do número de sucessos no modelo Binomial Negativa, ou seja,  $r = 3$  e  $r = 4$ , pode-se perceber, pela figura 5.36, que as amplitudes médias dos intervalos de credibilidade com distribuição *a priori* de Jeffreys tendem a se igualar às amplitudes médias obtidas pelos intervalos de credibilidade empregando a distribuição *a priori* de Bayes-Laplace.

Nas figuras 5.35 e 5.36 também fica evidente que valores da probabilidade de sucesso do modelo Binomial Negativa maiores que 0,2 para os intervalos de credibilidade, independente da distribuição *a priori* estudada, têm amplitudes médias bastante superiores as amplitudes obtidas nos intervalos de credibilidade para essa distribuições *a priori* no modelo Binomial (figuras 5.3 e 5.4).

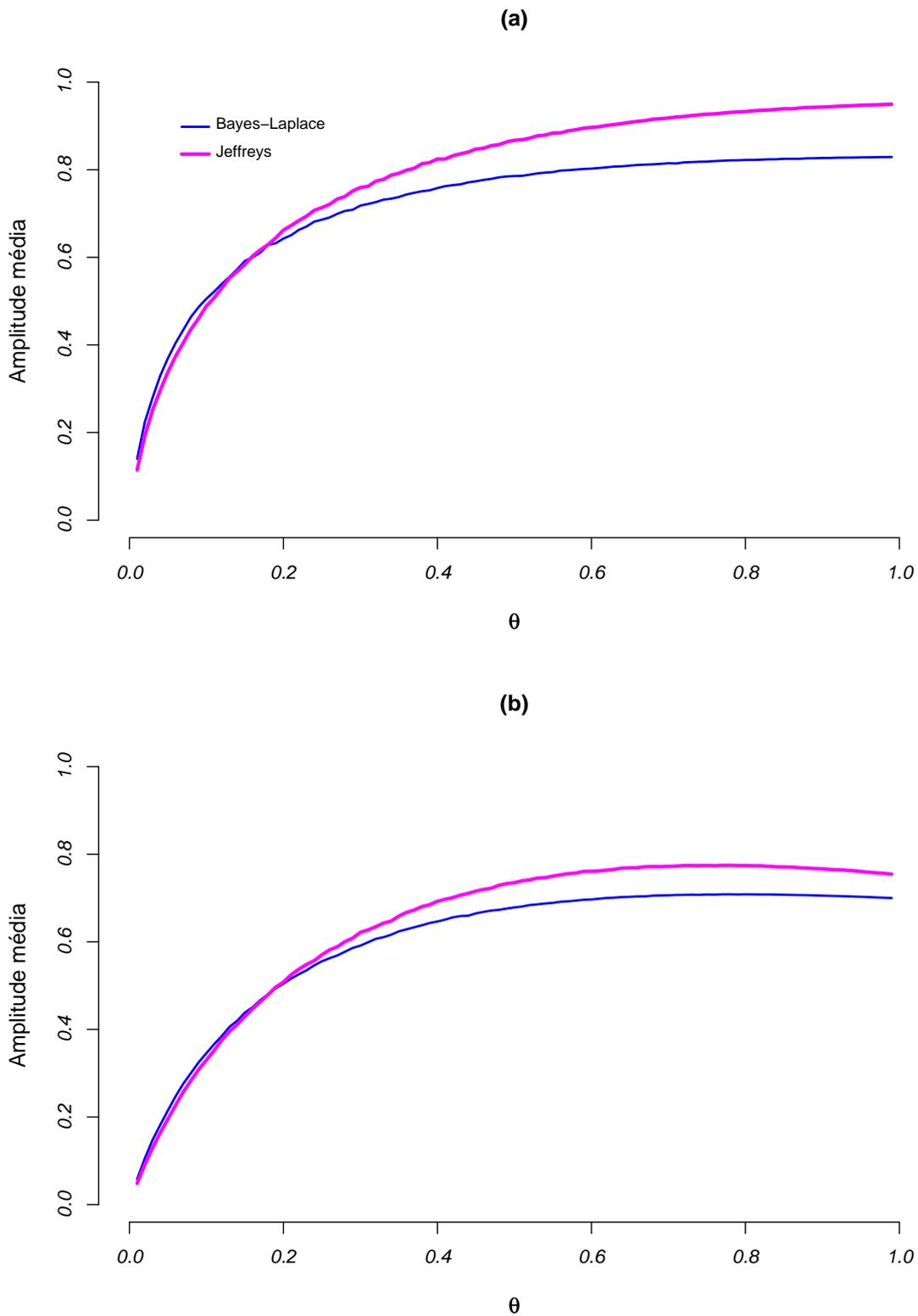


FIGURA 5.35: Amplitudes médias dos intervalos de credibilidade para a probabilidade de sucesso do modelo Binomial Negativa: (a)  $r = 1$  e (b)  $r = 2$ .

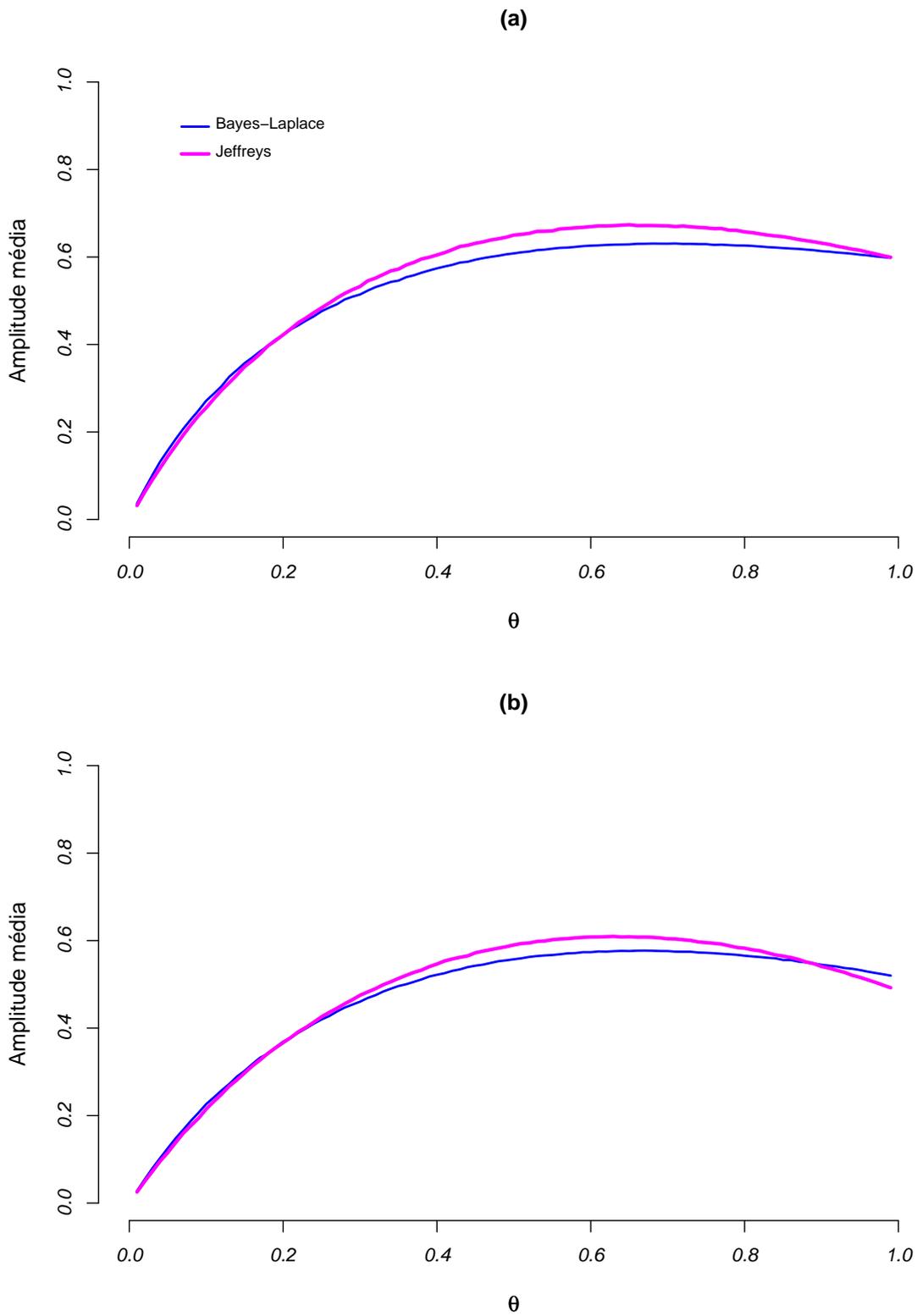


FIGURA 5.36: Amplitudes médias dos intervalos de credibilidade para a probabilidade de sucesso do modelo Binomial Negativa: (a)  $r = 3$  e (b)  $r = 4$ .

## Intervalos HPD

As proporções de cobertura dos intervalos HPD para a distribuição *a priori* de Bayes-Laplace e de Jeffreys estão expostas nas figuras 5.37 e 5.38 e as amplitudes médias desses intervalos são mostradas nas figuras 5.39 e 5.40.

As proporções de cobertura dos intervalos HPD com distribuição *a priori* de Jeffreys e  $r = 1$  (figura 5.37) apresentam-se muito próximos de 95% quando  $\theta \in (0, 11; 0, 81)$  (em relação aos intervalos HPD com distribuição *a priori* de Jeffreys), fora desse intervalo em geral ocorre subestimação da cobertura nominal. Para os intervalos HPD com distribuição *a priori* de Bayes-Laplace as proporções de cobertura estão demasiadamente subestimadas, para  $r = 1$  e  $\theta < 0,23$  e em valores de  $\theta > 0,23$  ocorre sobrestimação, exceto para  $\theta \approx 1$ .

Nas figuras 5.37 e 5.38, para  $r = 2, 3$  e  $4$  os comportamentos das proporções de cobertura dos intervalos HPD para ambas as distribuições *a priori* têm tendências parecidas. Entretanto, para a probabilidade de sucesso próxima de 0,5 e para  $r = 1$  e  $r = 2$  existe sobrestimação das proporções de cobertura por parte dos intervalos HPD com distribuição *a priori* de Bayes-Laplace.

As amplitudes médias dos intervalos HPD (figuras 5.39 e 5.40) apresentam-se com comportamento não diferente do que foi observado para os intervalos de credibilidade. Também verifica-se que as amplitudes médias para os intervalos HPD com distribuição *a priori* de Jeffreys são maiores ou muito próximos das amplitudes médias dos intervalos HPD com distribuição *a priori* de Bayes-Laplace, para  $r = 1$  e  $r = 2$ .

Pode-se verificar que, no geral, para ambas as distribuições *a priori*, a proporção de cobertura dos intervalos HPD são próximas de 95%, embora os intervalos HPD com distribuição *a priori* de Jeffreys tenham melhor desempenho e qualidades freqüentistas, ou seja, mais estabilidade quanto à proporção de cobertura do que os intervalos HPD com a distribuição *a priori* de Bayes-Laplace, para este estudo de simulação. No entanto, deve-se destacar que para  $r = 1$  (figura 5.39) que os intervalos HPD com distribuição *a priori* de Jeffreys possuem amplitudes médias dos intervalos muito grande, ou seja, amplitudes muito próximas de 1, o que os torna intervalos pouco informativos para estimação da probabilidade de sucesso do modelo Binomial Negativa.

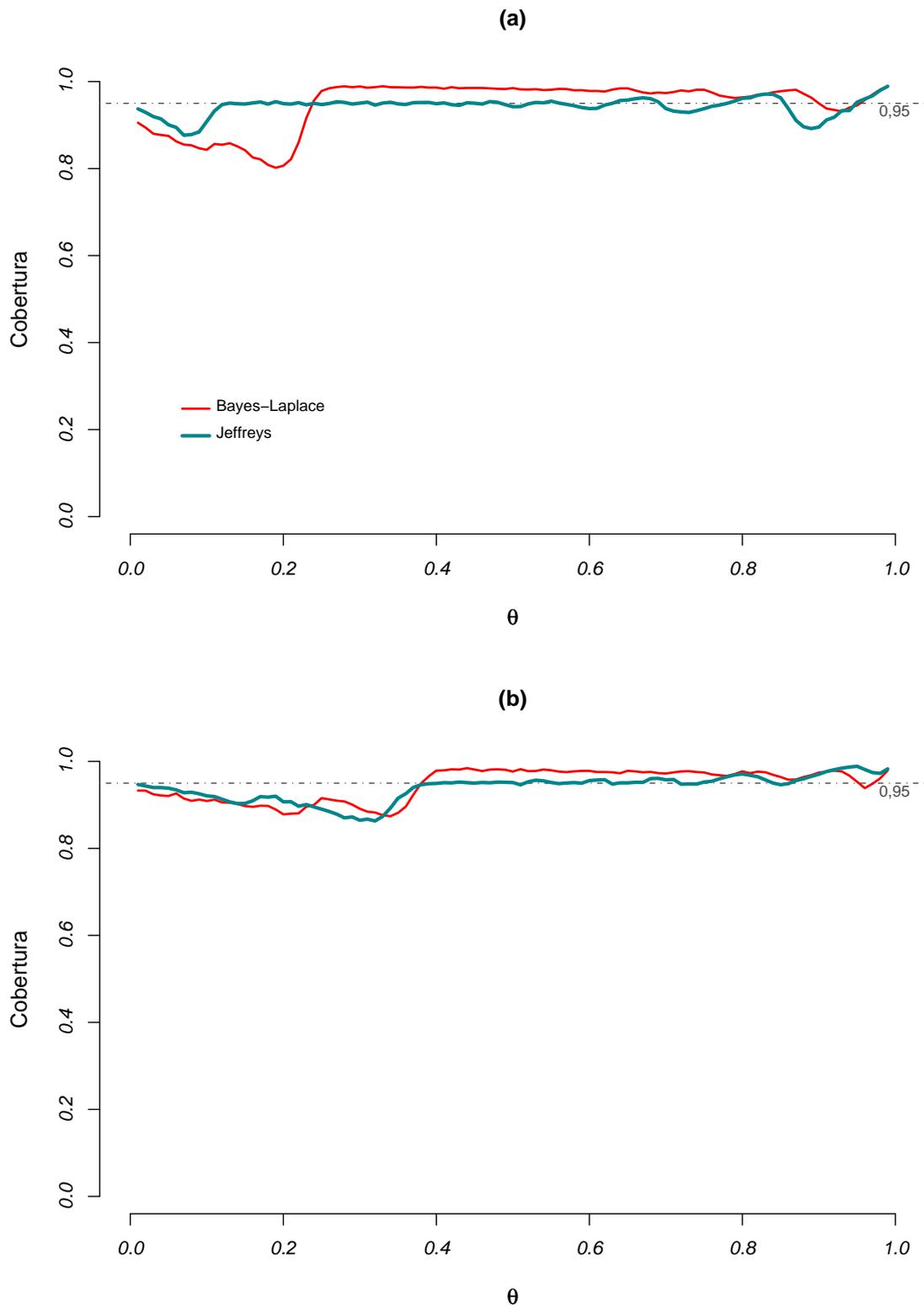


FIGURA 5.37: Proporções de cobertura dos intervalos HPD (máxima densidade *a posteriori*) para a probabilidade de sucesso do modelo Binomial Negativa: (a)  $r = 1$  e (b)  $r = 2$ .

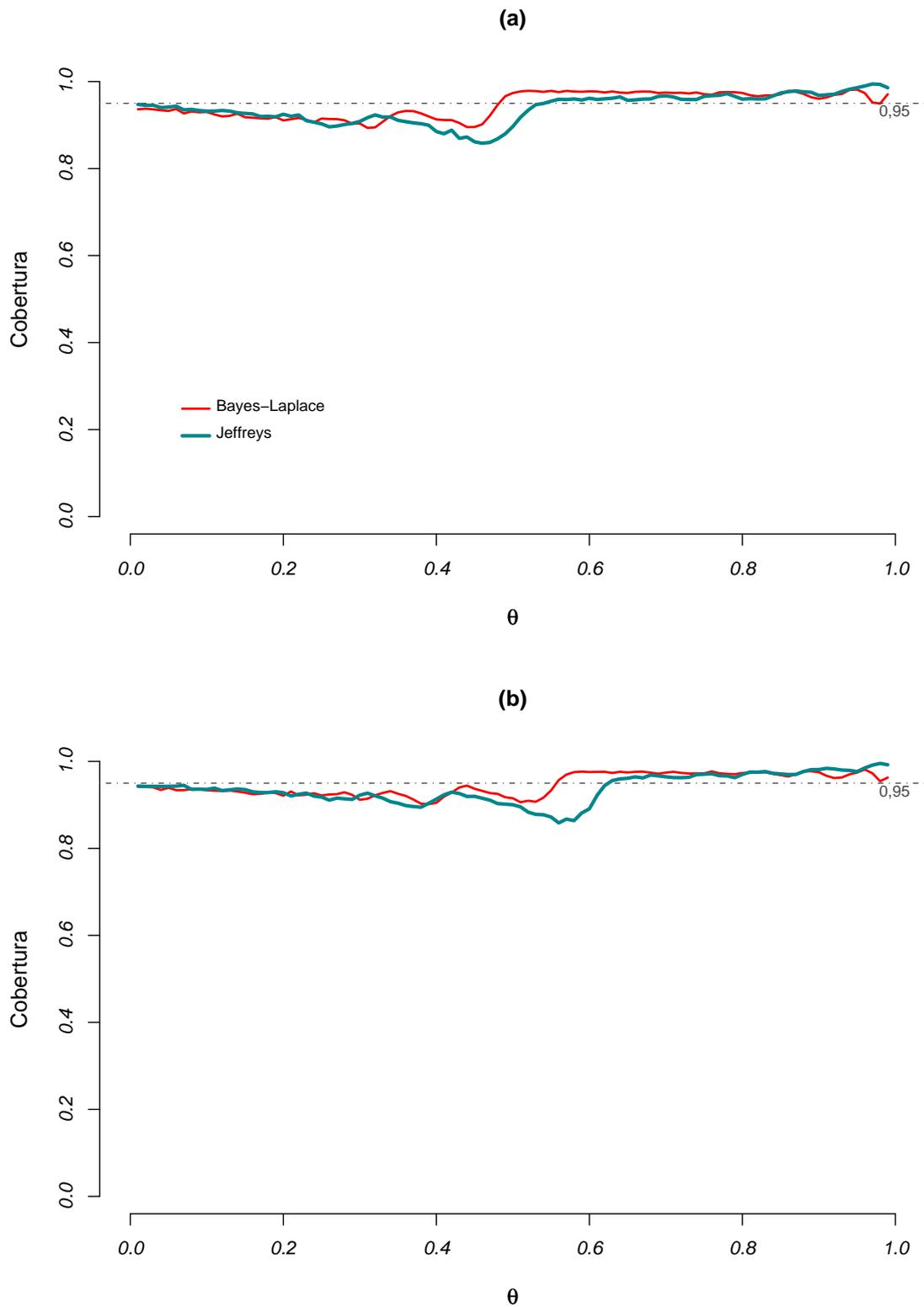


FIGURA 5.38: Proporções de cobertura dos intervalos HPD (máxima densidade *a posteriori*) para a probabilidade de sucesso do modelo Binomial Negativa: (a)  $r = 3$  e (b)  $r = 4$ .

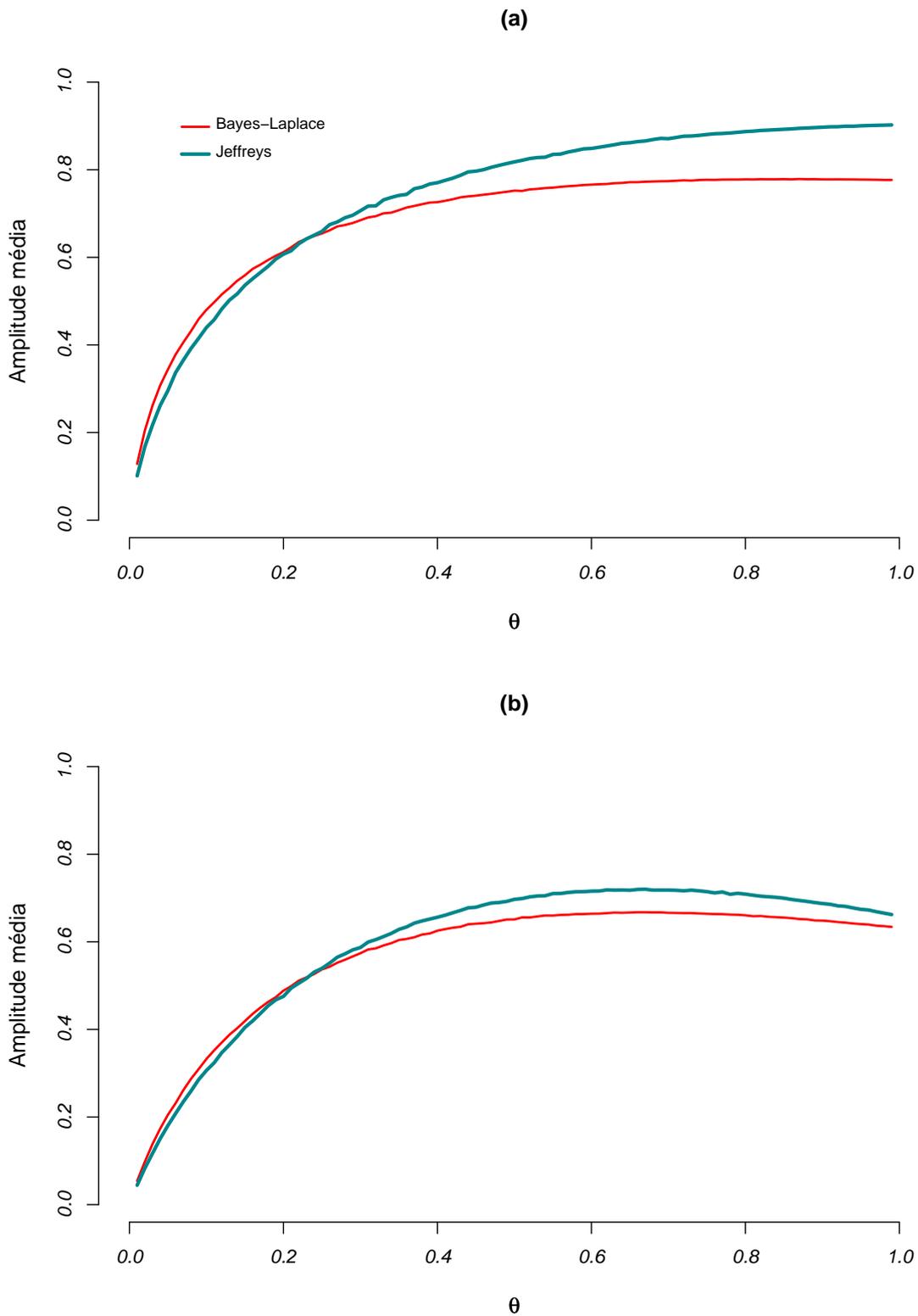


FIGURA 5.39: Amplitudes médias dos intervalos HPD (máxima densidade *a posteriori*) para a probabilidade de sucesso do modelo Binomial Negativa: (a)  $r = 1$  e (b)  $r = 2$ .

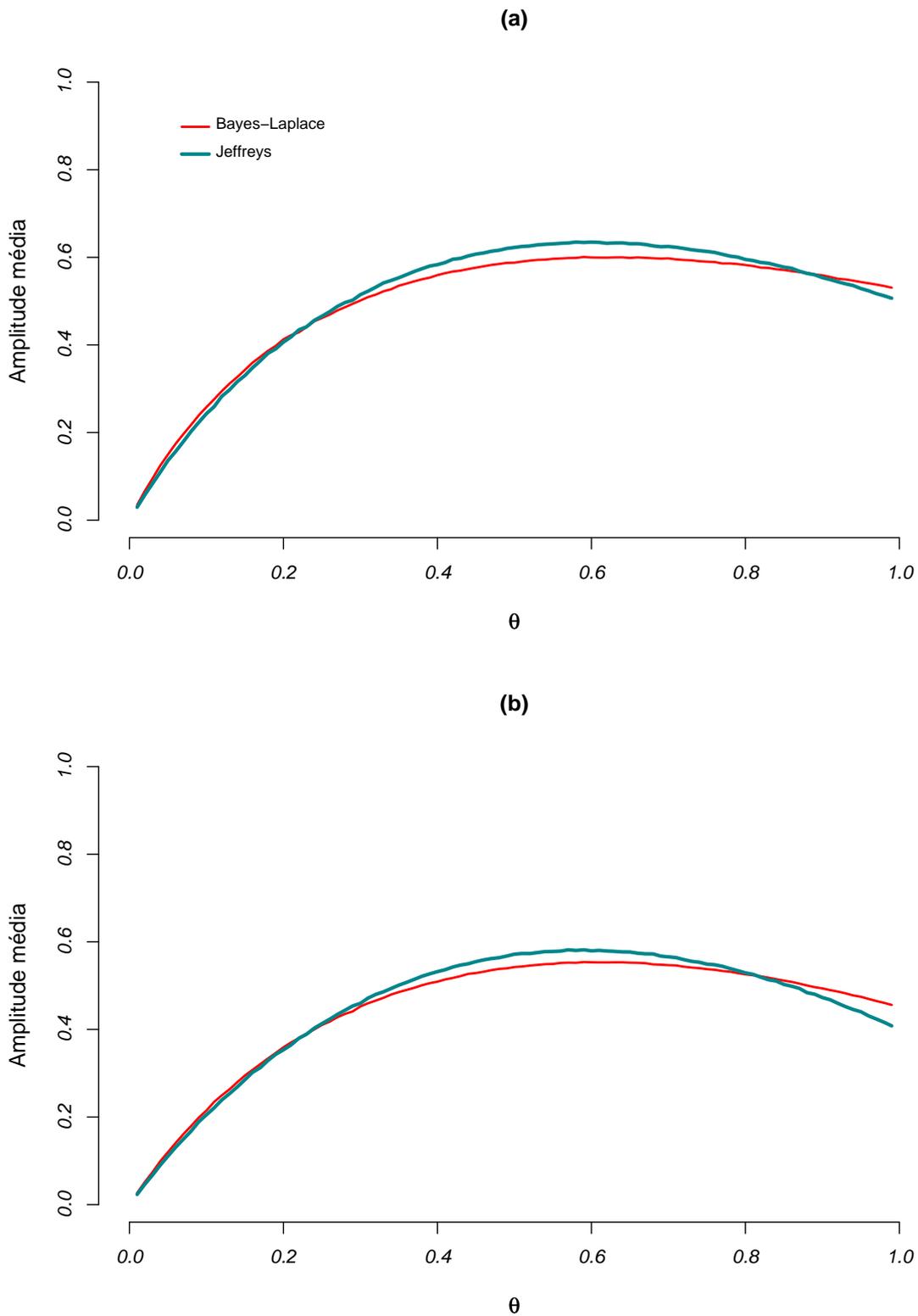


FIGURA 5.40: Amplitudes médias dos intervalos HPD (máxima densidade *a posteriori*) para a probabilidade de sucesso do modelo Binomial Negativa: (a)  $r = 3$  e (b)  $r = 4$ .

## Intervalos de credibilidade e intervalos HPD

Verifica-se, como esperado, pelas figuras 5.43, 5.44, 5.47 e 5.48 que as amplitudes médias dos intervalos construídos pela abordagem bayesiana sempre não superiores para os intervalos HPD em relação aos intervalos de credibilidade, independente do valor de  $r$  e de  $\theta$ . Isso porque os intervalos HPD são intervalos com menores amplitudes possíveis com credibilidade igual ou superior a  $100(1 - \epsilon)\%$ . Observa-se também que, para valores de  $\theta > 0,5$ , nos quais a probabilidade de sucesso é maior do que a probabilidade de insucesso e as amostras são menores, a amplitude é muito maior que para valores em que a probabilidade de sucesso é menor.

Nas figuras 5.41 e 5.42 observa-se que em quase todo espaço paramétrico ocorre sobestimação da cobertura por parte de ambos os intervalos bayesianos, para  $r = 1$ . No entanto, para  $\theta < 0,2$  ocorre o oposto com as proporções de cobertura tanto para o intervalo HPD quanto para o intervalo de credibilidade. Uma sutil melhora, ou seja, as proporções de cobertura dos métodos bayesianos estão mais próximos da cobertura nominal para  $r = 3$  e 4, mas para valores de  $\theta$  próximos de 0,5 ainda mostra certa instabilidade.

Observe agora as figuras 5.45 e 5.46 que retratam as proporções de cobertura dos intervalos de credibilidade e HPD para a distribuição *a priori* de Jeffreys. Note que, independente dos valores de  $r$  e de  $\theta$ , a proporção de cobertura para os dois intervalos bayesianos analisados é mais estável em relação às proporções de coberturas obtidas com a distribuição *a priori* de Bayes-Laplace (figuras 5.41 e 5.42), principalmente para  $r = 1$ .

Nessas figuras ainda é possível verificar que as proporções de cobertura obtidas pelos intervalos de credibilidade, apesar de apresentarem amplitudes médias maiores que as dos intervalos HPD, são mais próximas da cobertura nominal (figura 5.45 e 5.46).

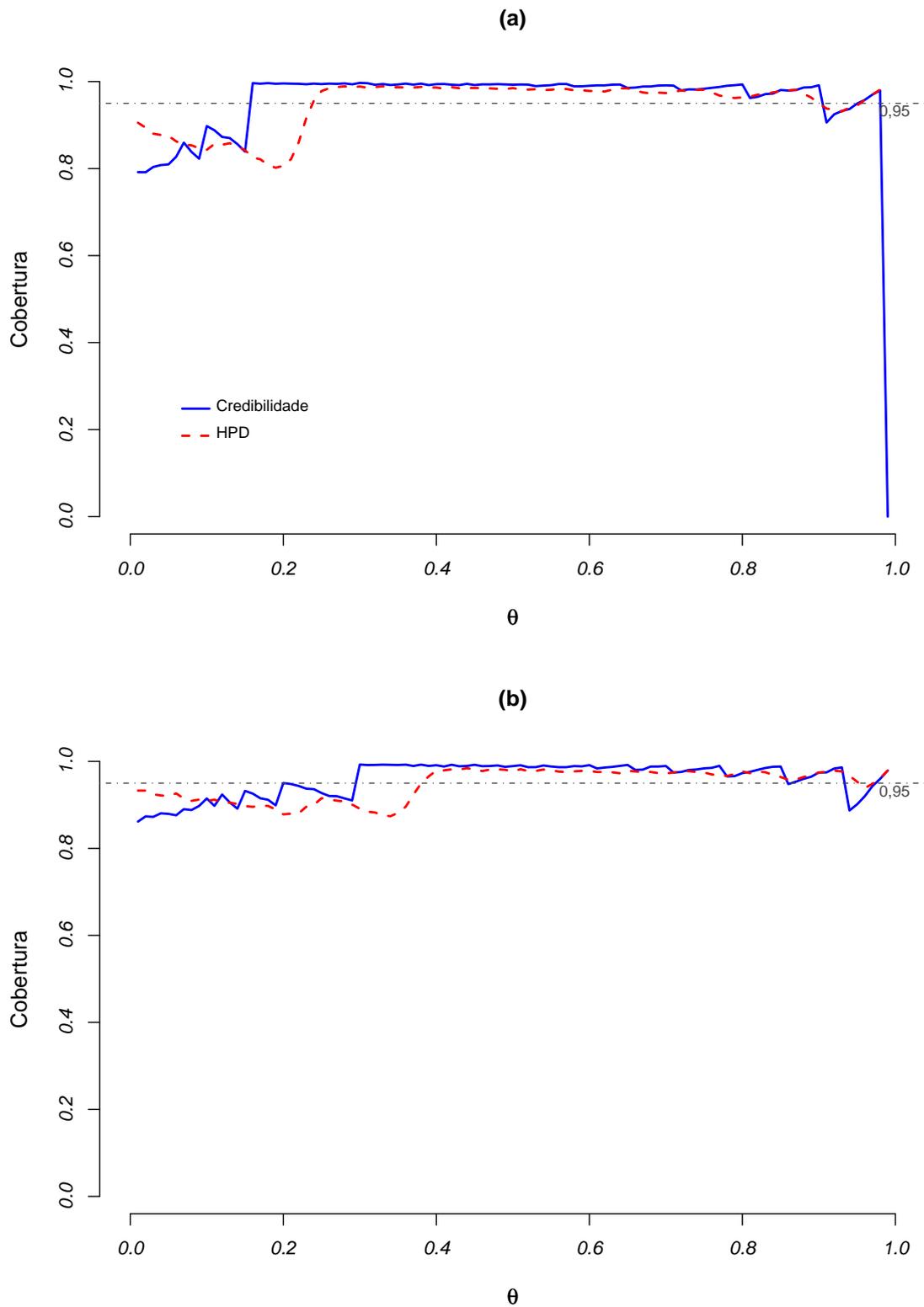


FIGURA 5.41: Proporções de cobertura dos intervalos de credibilidade e HPD com distribuição *a priori* de Bayes-Laplace para a probabilidade de sucesso do modelo Binomial Negativa: (a)  $r = 1$  e (b)  $r = 2$ .

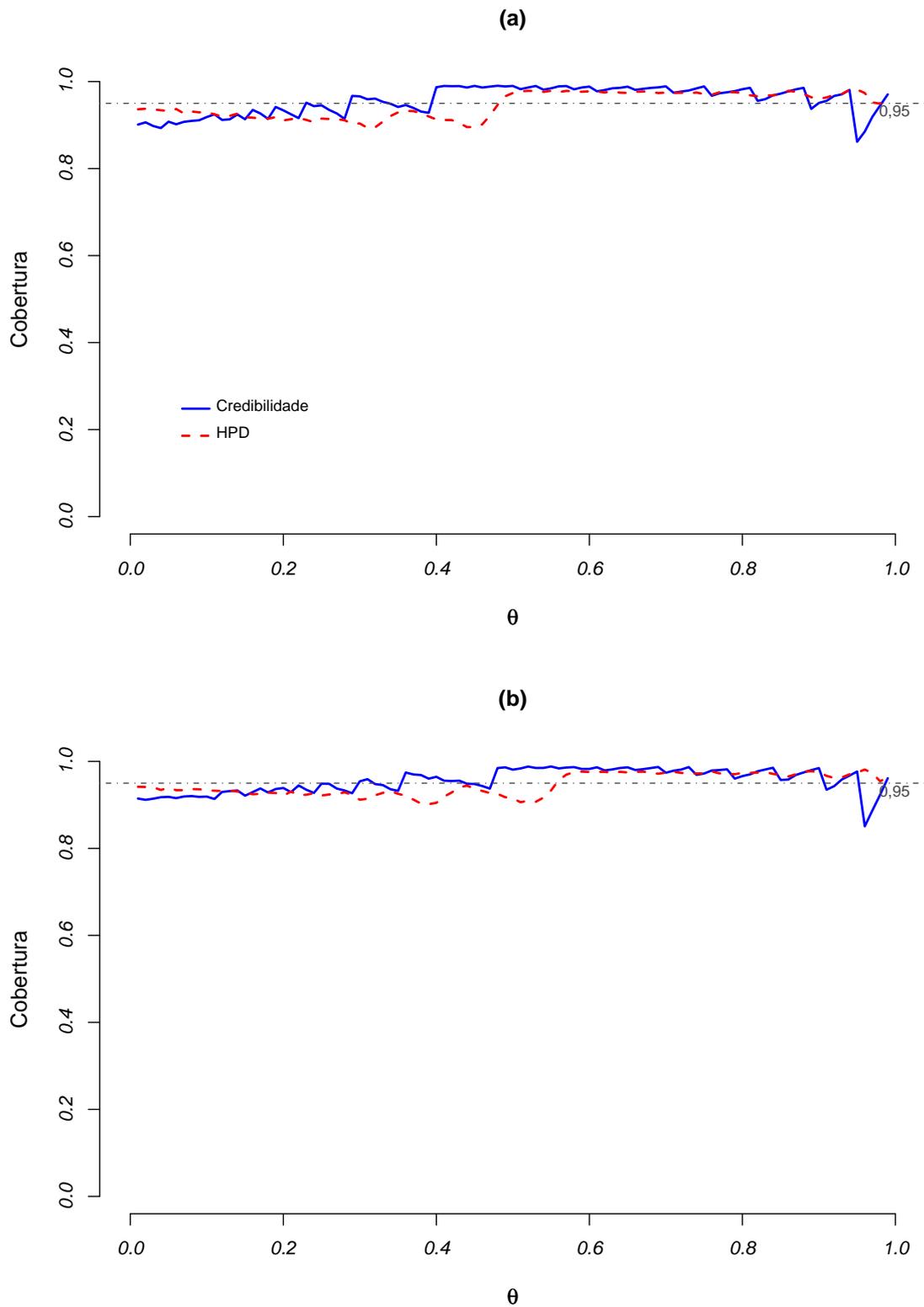


FIGURA 5.42: Proporções de cobertura dos intervalos de credibilidade e HPD com distribuição *a priori* de Bayes-Laplace para a probabilidade de sucesso do modelo Binomial Negativa: (a)  $r = 3$  e (b)  $r = 4$ .

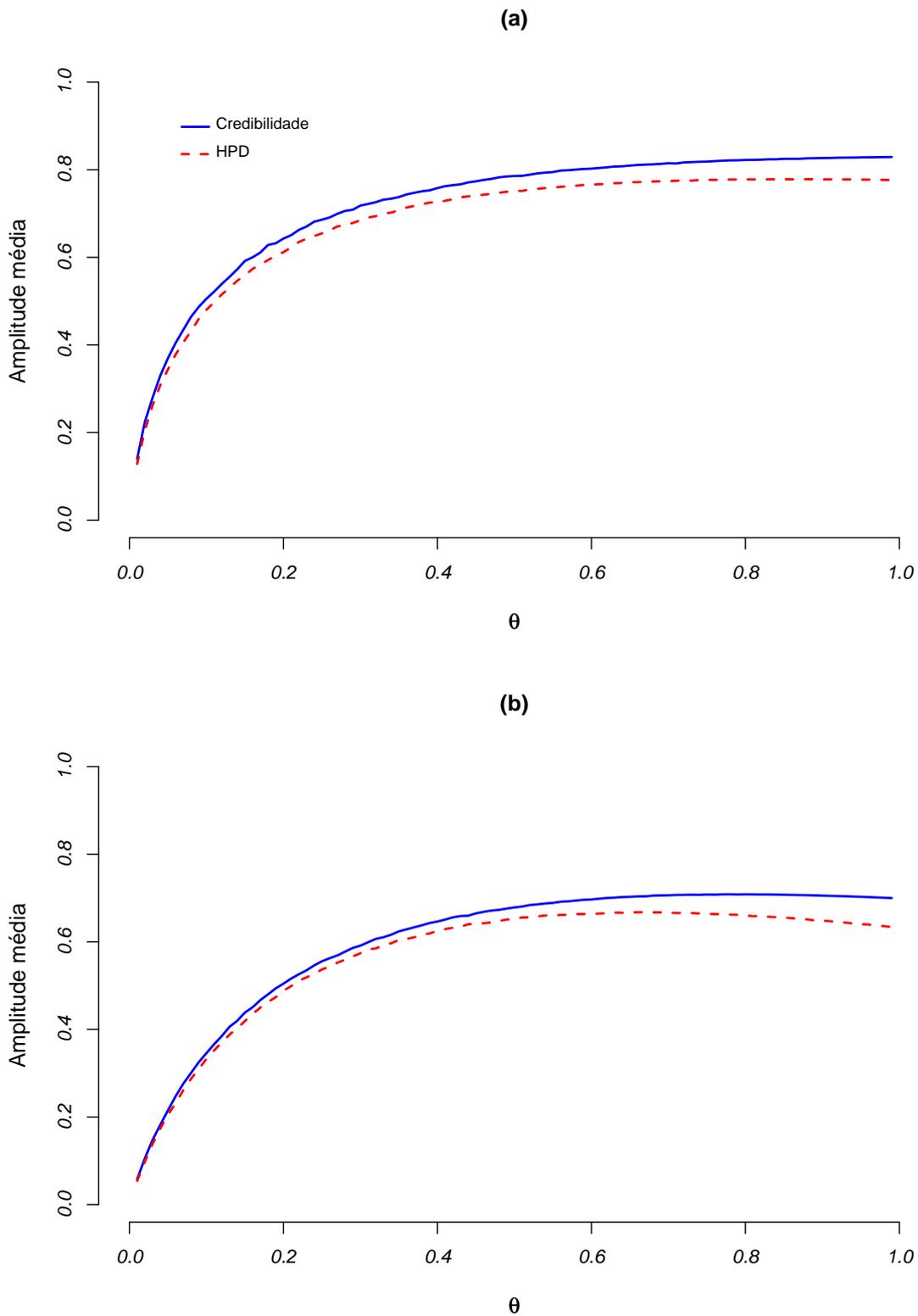


FIGURA 5.43: Amplitudes médias dos intervalos de credibilidade e HPD com distribuição *a priori* de Bayes-Laplace para a probabilidade de sucesso do modelo Binomial Negativa: (a)  $r = 1$  e (b)  $r = 2$ .

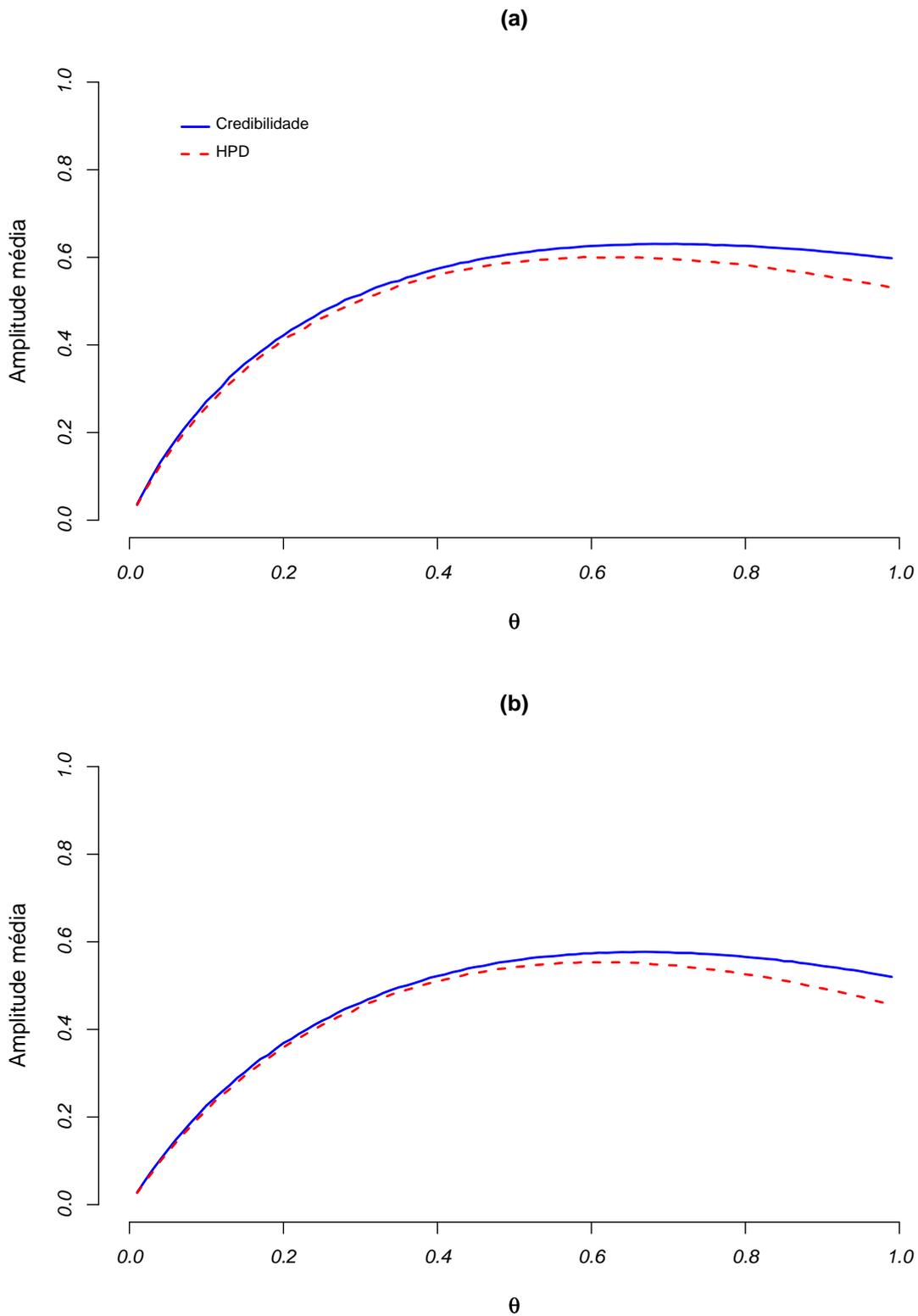


FIGURA 5.44: Amplitudes médias dos intervalos de credibilidade e HPD com distribuição *a priori* de Bayes-Laplace para a probabilidade de sucesso do modelo Binomial Negativa: (a)  $r = 3$  e (b)  $r = 4$ .

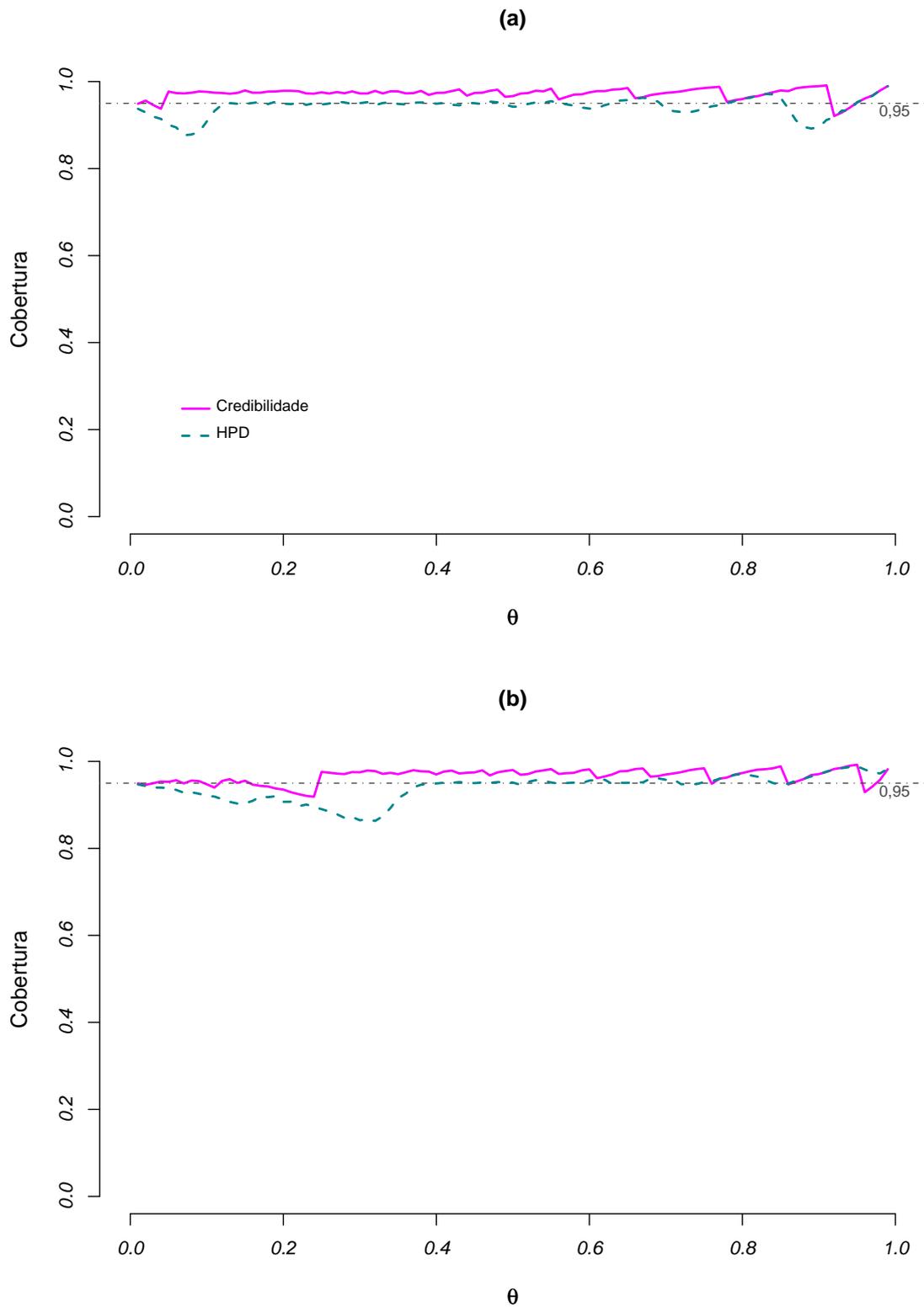


FIGURA 5.45: Proporções de cobertura dos intervalos de credibilidade e HPD com distribuição *a priori* de Jeffreys para a probabilidade de sucesso do modelo Binomial Negativa: (a)  $r = 1$  e (b)  $r = 2$ .

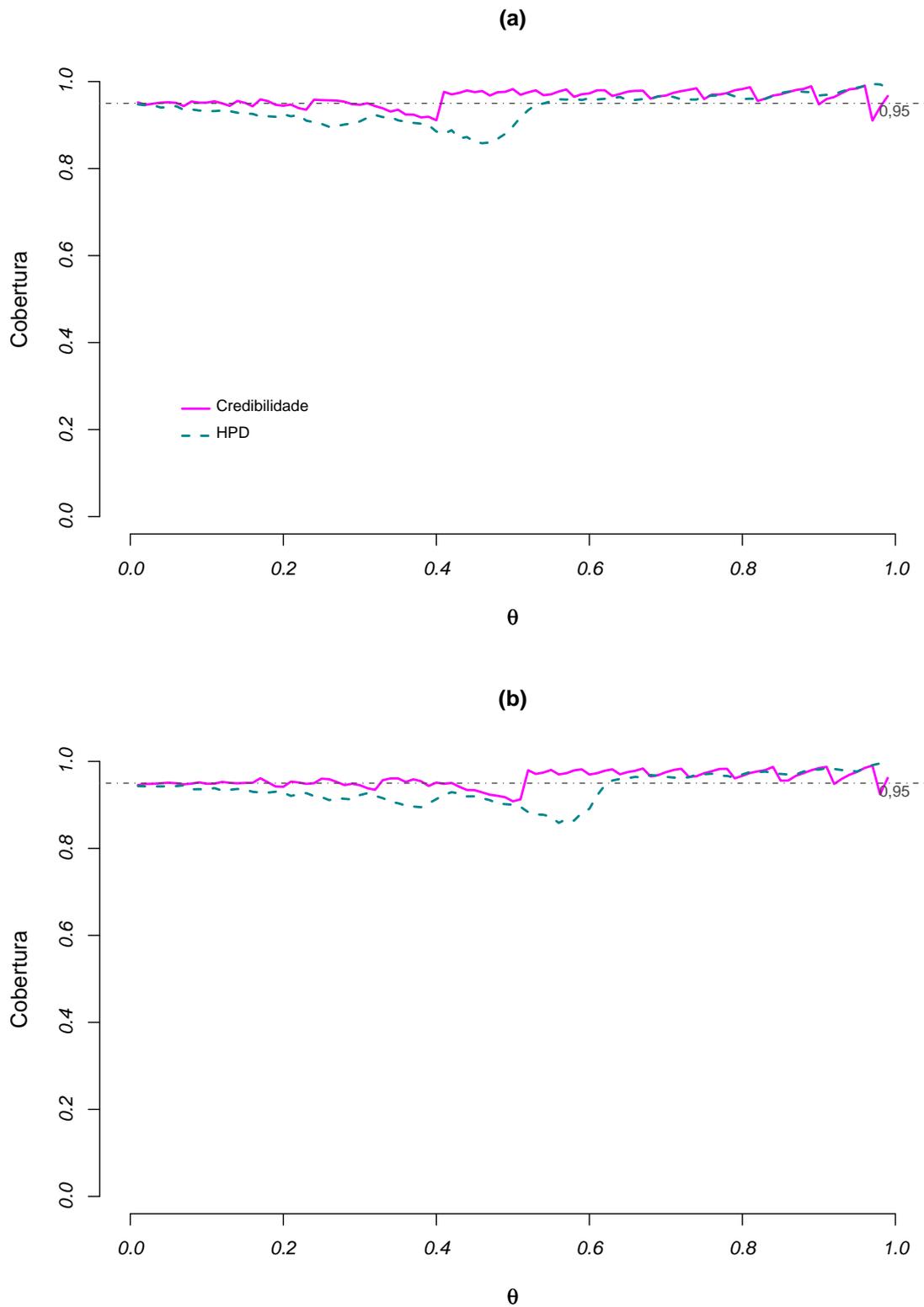


FIGURA 5.46: Proporções de cobertura dos intervalos de credibilidade e HPD com distribuição *a priori* de Jeffreys para a probabilidade de sucesso do modelo Binomial Negativa: (a)  $r = 3$  e (b)  $r = 4$ .

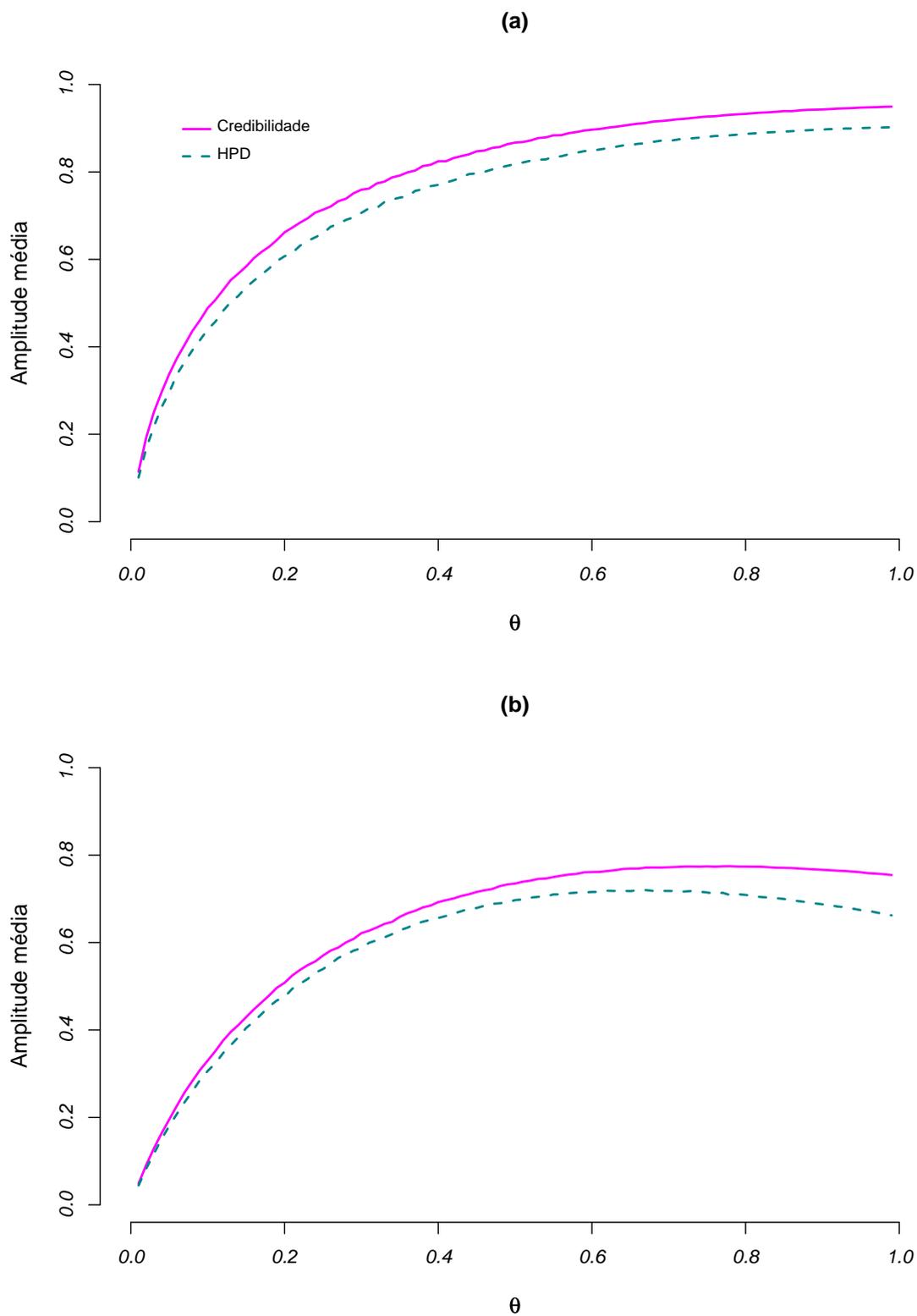


FIGURA 5.47: Amplitudes médias dos intervalos de credibilidade e HPD com distribuição *a priori* de Jeffreys para a probabilidade de sucesso do modelo Binomial Negativa: (a)  $r = 1$  e (b)  $r = 2$ .

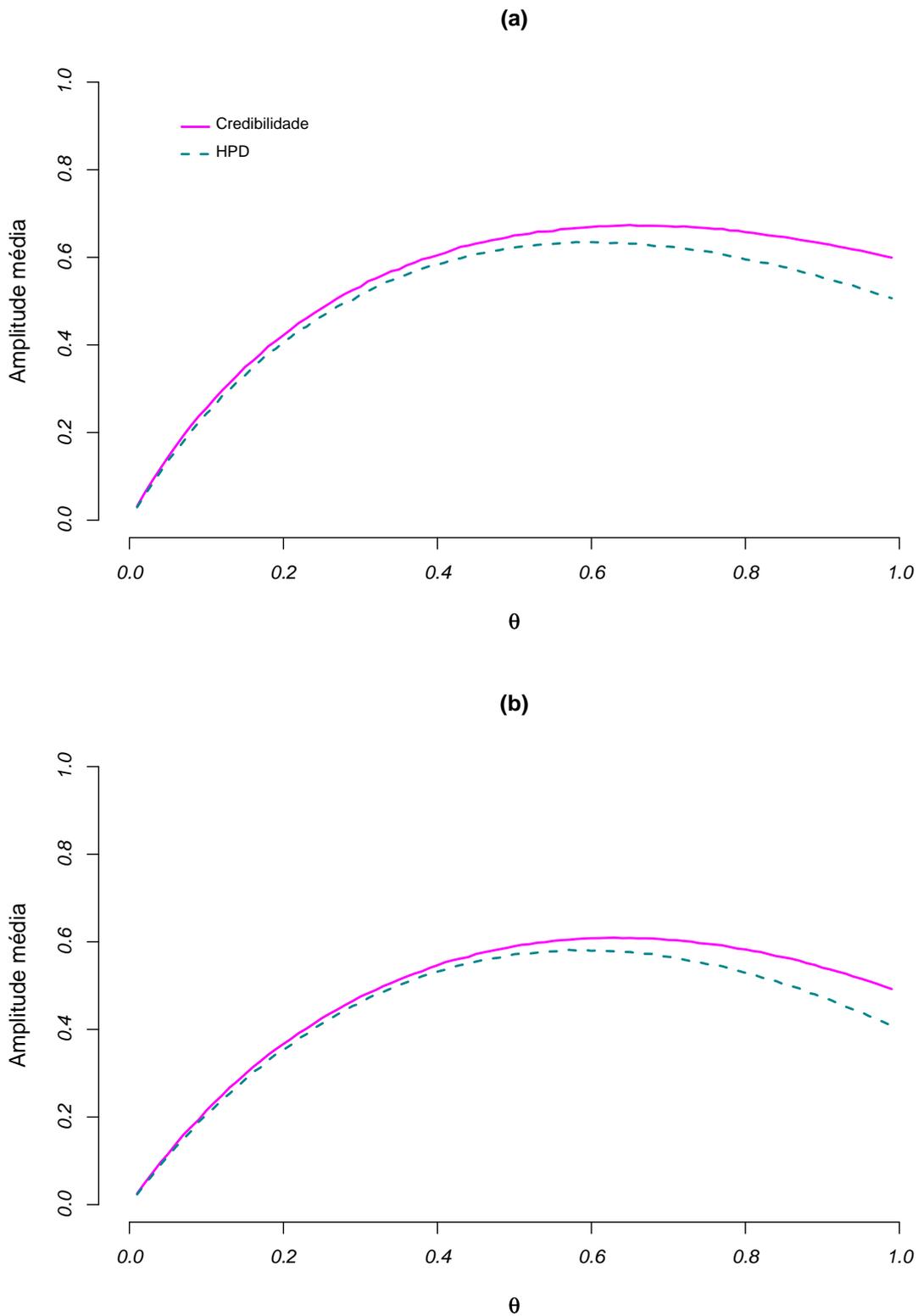


FIGURA 5.48: Amplitudes médias dos intervalos de credibilidade e HPD com distribuição *a priori* de Jeffreys para a probabilidade de sucesso do modelo Binomial Negativa: (a)  $r = 3$  e (b)  $r = 4$ .

### 5.2.2 Proporções de cobertura e amplitudes médias dos intervalos clássicos

As figuras 5.49 e 5.50 apresentam as proporções de cobertura dos intervalos de confiança obtidos pela aproximação com a distribuição Normal e exato de Clopper & Pearson.

Para o intervalo de confiança exato a proporção de cobertura é sobrestimada (já que deveria ser 95%) para  $\theta < 0,4$ ; para  $\theta > 0,4$  é subestimada, principalmente com o aumento de  $r$ . O que não é natural, pois esse método pelas figuras 5.51 e 5.48 apresenta amplitudes médias muito maiores para  $\theta > 0,5$ . Desta forma, o método se mostra insatisfatório para estimação intervalar da probabilidade de sucesso do modelo Binomial Negativa.

Nas figuras 5.49 e 5.50 os intervalos de confiança com aproximação pela distribuição Normal, que exigem grandes amostras para boa adequabilidade, apresentam-se insatisfatórios mesmo com o aumento de  $r$ , principalmente para  $\theta > 0,5$ , apesar de apresentarem amplitudes médias satisfatórias. Isso é naturalmente explicado, pois amplitudes pequenas implicam em intervalos curtos e, conseqüentemente, com tendência a apresentar baixa proporção de cobertura.

Quando  $r = i$ , o interesse é no número de ensaios até o  $i$ -ésimo sucesso, e para  $\theta \approx 0$ , tem-se probabilidade de sucesso próxima de zero. O tamanho da amostra é aleatório, pois não se conhece o número de ensaios até o  $i$ -ésimo sucesso. Considerando estas informações, pode-se dizer que este é o motivo pelo qual a proporção de cobertura para valores extremos inferiores de  $\theta$  é mais próxima da nominal e para valores de  $\theta$  próximos de 1 e  $r$  pequeno a cobertura é insatisfatória para a aproximação pela distribuição Normal.

Diante dos resultados e das condições do estudo de simulação, os métodos clássicos apresentaram-se muito menos apropriados que os métodos bayesianos. Sendo que entre os métodos bayesianos, é mais aconselhável a utilização da distribuição *a priori* de Jeffreys para a probabilidade de sucesso da Binomial Negativa com  $r$  fixo.

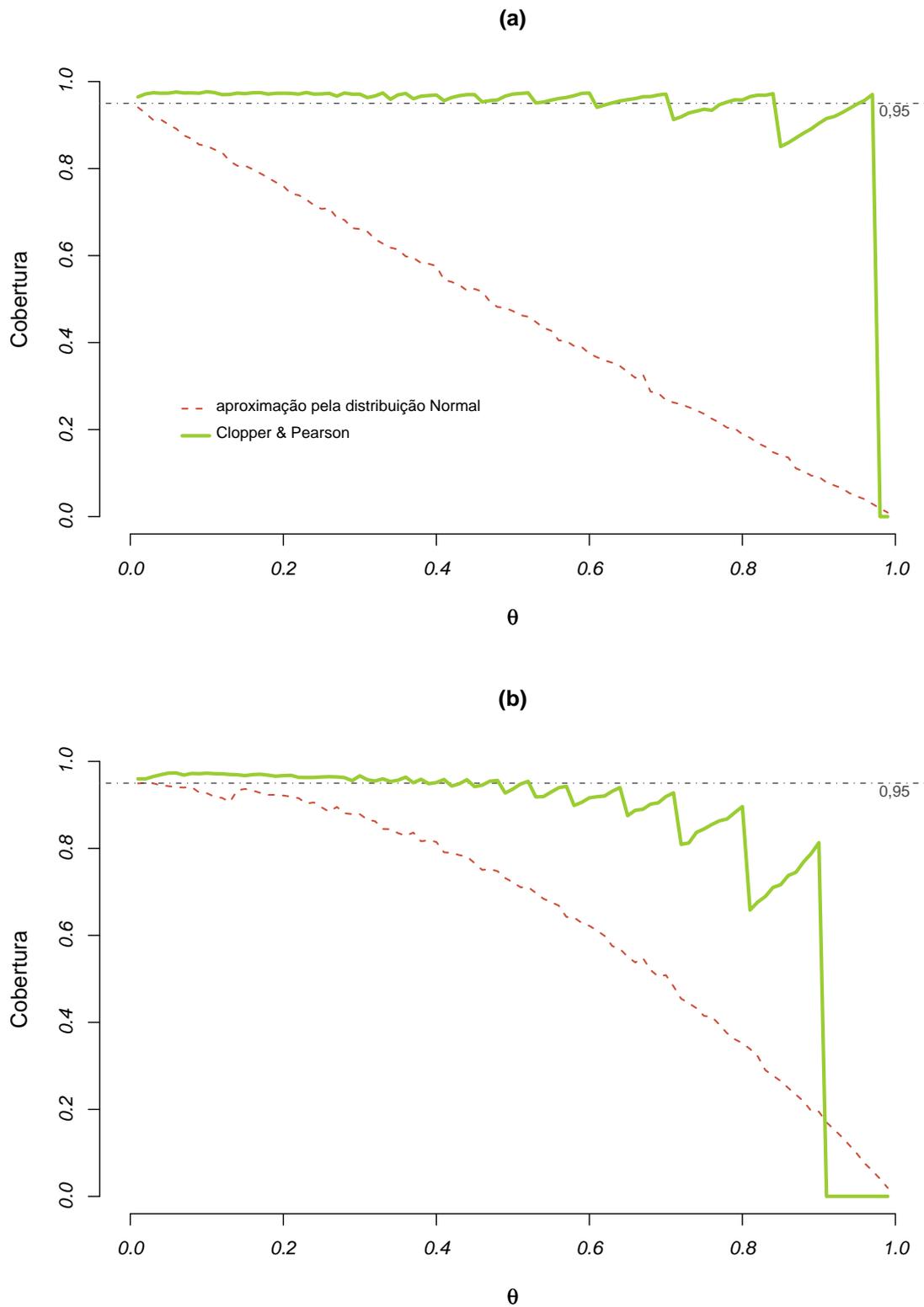


FIGURA 5.49: Proporções de cobertura dos intervalos clássicos para a probabilidade de sucesso do modelo Binomial Negativa: (a)  $r = 1$  e (b)  $r = 2$ .

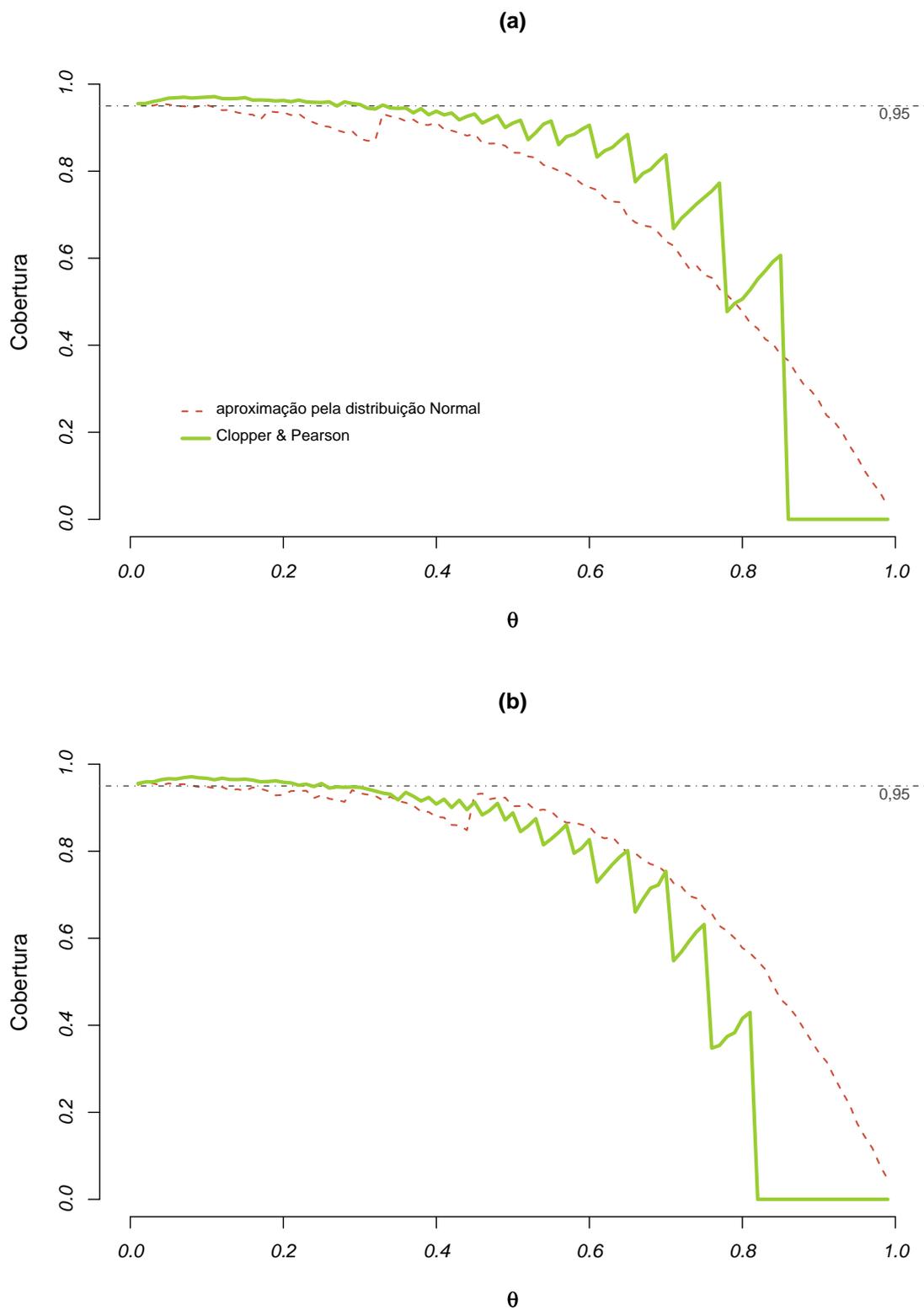


FIGURA 5.50: Proporções de cobertura dos intervalos clássicos para a probabilidade de sucesso do modelo Binomial Negativa: (a)  $r = 3$  e (b)  $r = 4$ .

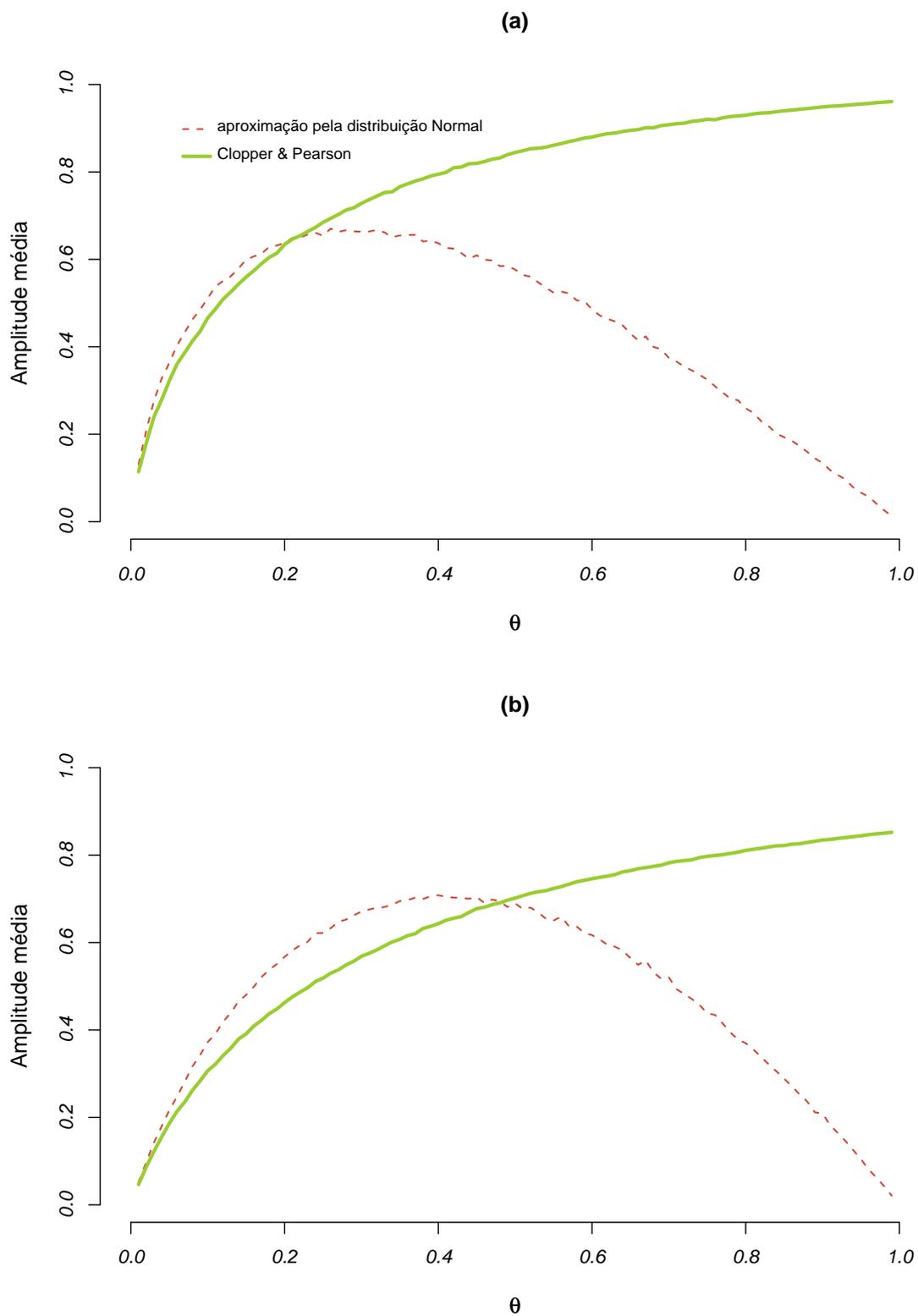


FIGURA 5.51: Amplitudes médias dos intervalos clássicos para a probabilidade de sucesso do modelo Binomial Negativa: (a)  $r = 1$  e (b)  $r = 2$ .

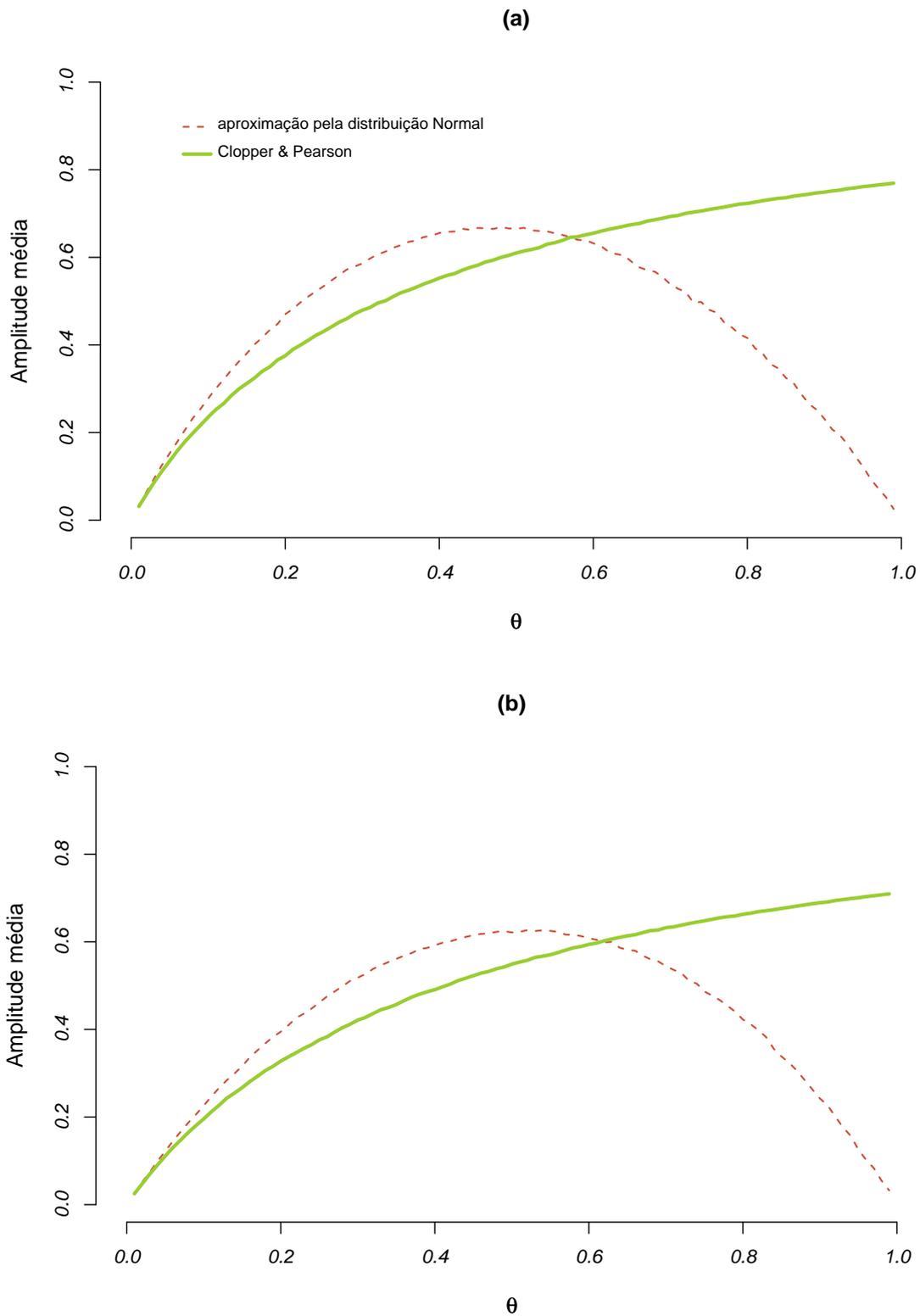


FIGURA 5.52: Amplitudes médias dos intervalos clássicos para a probabilidade de sucesso do modelo Binomial Negativa: (a)  $r = 3$  e (b)  $r = 4$ .

### 5.2.3 Consistência, viés e erro quadrático médio dos estimadores

Assim como visto para o modelo Binomial, nesta seção é de interesse confirmar via estudo de simulação se os estimadores pontuais considerados apresentam-se consistentes para estimar a probabilidade de sucesso do modelo Binomial Negativa. Os estimadores pontuais utilizados também foram o de máxima verossimilhança (EMV), a média e a moda da distribuição *a posteriori* de  $\theta$ , expostos no Capítulo 4.

As esperanças e variâncias dos estimadores pontuais estão apresentadas no Apêndice B, ressalta-se que o resultado analítico e numérico das séries infinitas foram obtidas por meio do sistema Maple.

Diferente do caso Binomial, agora o estimador de máxima verossimilhança é viciado para  $\theta$ , assim como os estimadores pontuais bayesianos. No entanto, observa-se pelos vícios e variâncias desses estimadores que os mesmos são consistente (ver Apêndice B). O estimador de máxima verossimilhança coincide com a moda da distribuição *a posteriori* de  $\theta$  ao considerar a distribuição *a priori* de Bayes-Laplace. Por esse motivo, nas figuras 5.57 e 5.58, 5.59 e 5.60 serão apresentados somente os valores dos vícios e erros quadráticos médios do estimador de máxima verossimilhança.

Pelas figuras 5.53 e 5.54 pode-se observar que o estimador de máxima verossimilhança, para todos os valores de  $r$  analisados, apresenta-se com sobrestimação, ou seja, a estimativa do parâmetro é superior ao verdadeiro valor da probabilidade de sucesso. Conjuntamente com as variâncias obtidas os erros quadráticos médios para o estimador de máxima verossimilhança só têm valores inferiores aos erros quadráticos médios da moda da distribuição *a posteriori* de  $\theta$  com distribuição *a priori* de Jeffreys (figuras 5.55 e 5.56, 5.59 e 5.60). No entanto, o estimador clássico apresenta-se consistente, pois à medida que  $r$  aumenta o vício tende a zero.

Nas figuras 5.53 e 5.54 podem ser verificados também os vícios obtidos no estudo de simulação para a média da distribuição *a posteriori* de  $\theta$ . Ambas as distribuições *a priori* consideradas apresentam comportamento dos vícios similares, com sobrestimação do valor “correto” para  $\theta < 0,5$  e subestimação do parâmetro para  $\theta > 0,5$ . Com o aumento de  $r$  observa-se que o vício do estimador pontual bayesiano tende a zero, ou seja, o estimador é consistente.

A média da distribuição *a posteriori* de  $\theta$  ao empregar a distribuição *a priori* de Jeffreys apresenta os menores vícios e erros quadráticos médios para  $\theta < 0,5$  com o aumento de  $r$  (figuras 5.53 e 5.54).

Agora observe as figuras 5.57, 5.58, 5.59 e 5.60 que apresentam os vício e erros quadráticos médios obtidos no estudo de simulação para o estimador de máxima verossimilhança e a moda da distribuição *a posteriori* de  $\theta$ . Note que a moda da distribuição *a posteriori* de  $\theta$  ao considerar a distribuição *a priori* de Bayes-Laplace (Seção 4.1.2) apresenta estimador comum ao de máxima verossimilhança.

Na figura 5.57 para  $r = 1$ , observe que a moda da distribuição *a posteriori* de  $\theta$  ao considerar a distribuição *a priori* de Jeffreys apresenta-se sobrestimando o valor “correto” do parâmetro, independente do valor assumido pela probabilidade de sucesso. Pois quando  $r = 1$ , a moda da distribuição *a posteriori* de  $\theta$  para essa distribuição *a priori* apresenta uma função da forma exponencial, ou seja, o ponto de máximo é o limite inferior do espaço paramétrico, assim o vício é dado por  $\beta(\theta) = -\theta$ . Na figura 5.59 também para  $r = 1$  e 2 são observados os maiores erros quadráticos médios dos estimadores da probabilidade de sucesso do modelo Binomial Negativa, que diminui drasticamente com  $r = 3$  e 4.

Para os demais valores de  $r$  considerados os vícios desse estimador bayesiano sobrestimam o valor do parâmetro, com o aumento de  $r$  os vícios tendem a zero confirmando a consistência do estimador.

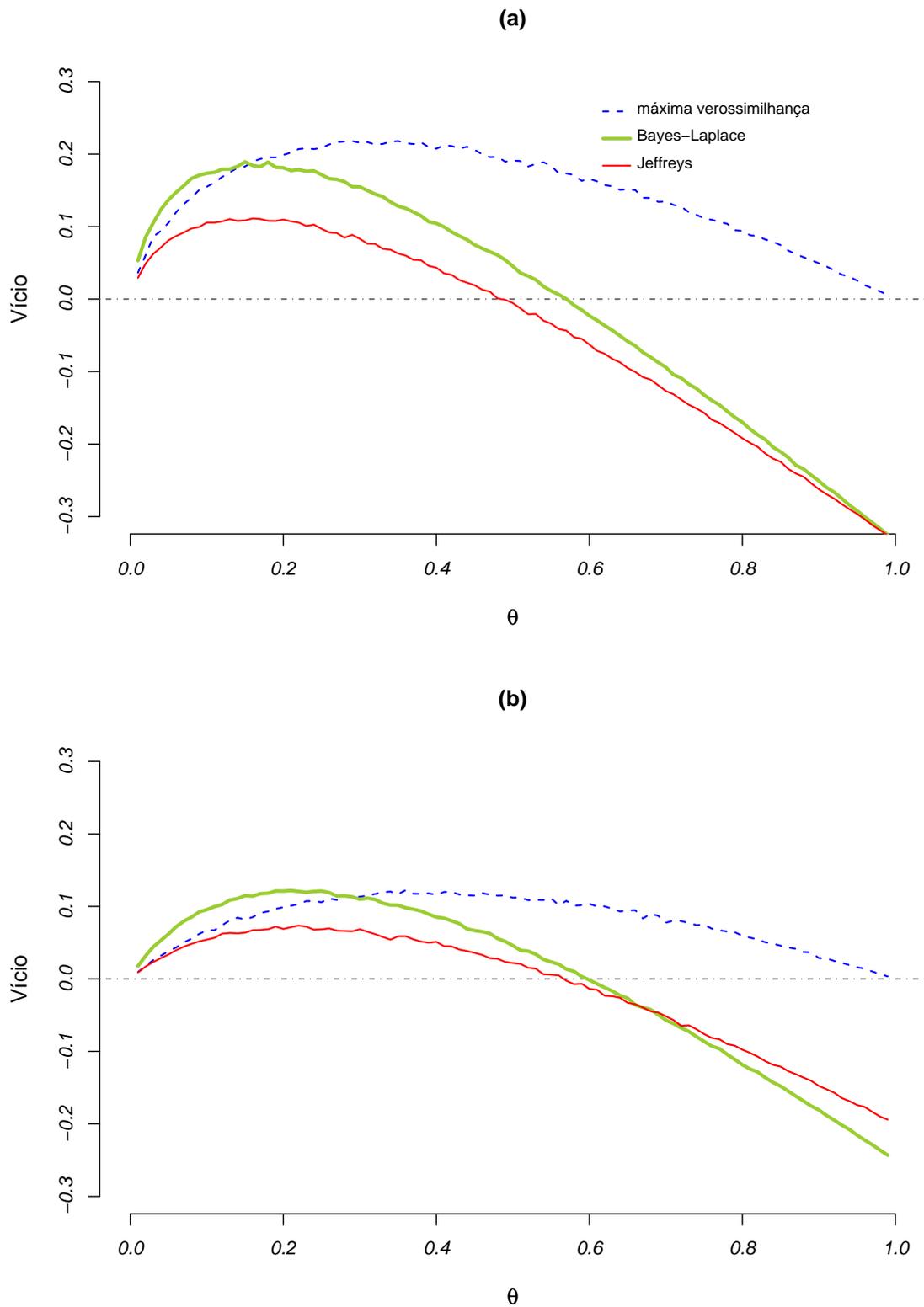


FIGURA 5.53: Vícius dos estimadores de máxima verossimilhança e da média da distribuição *a posteriori* para a probabilidade de sucesso para o modelo Binomial Negativa: (a)  $r = 1$  e (b)  $r = 2$ .

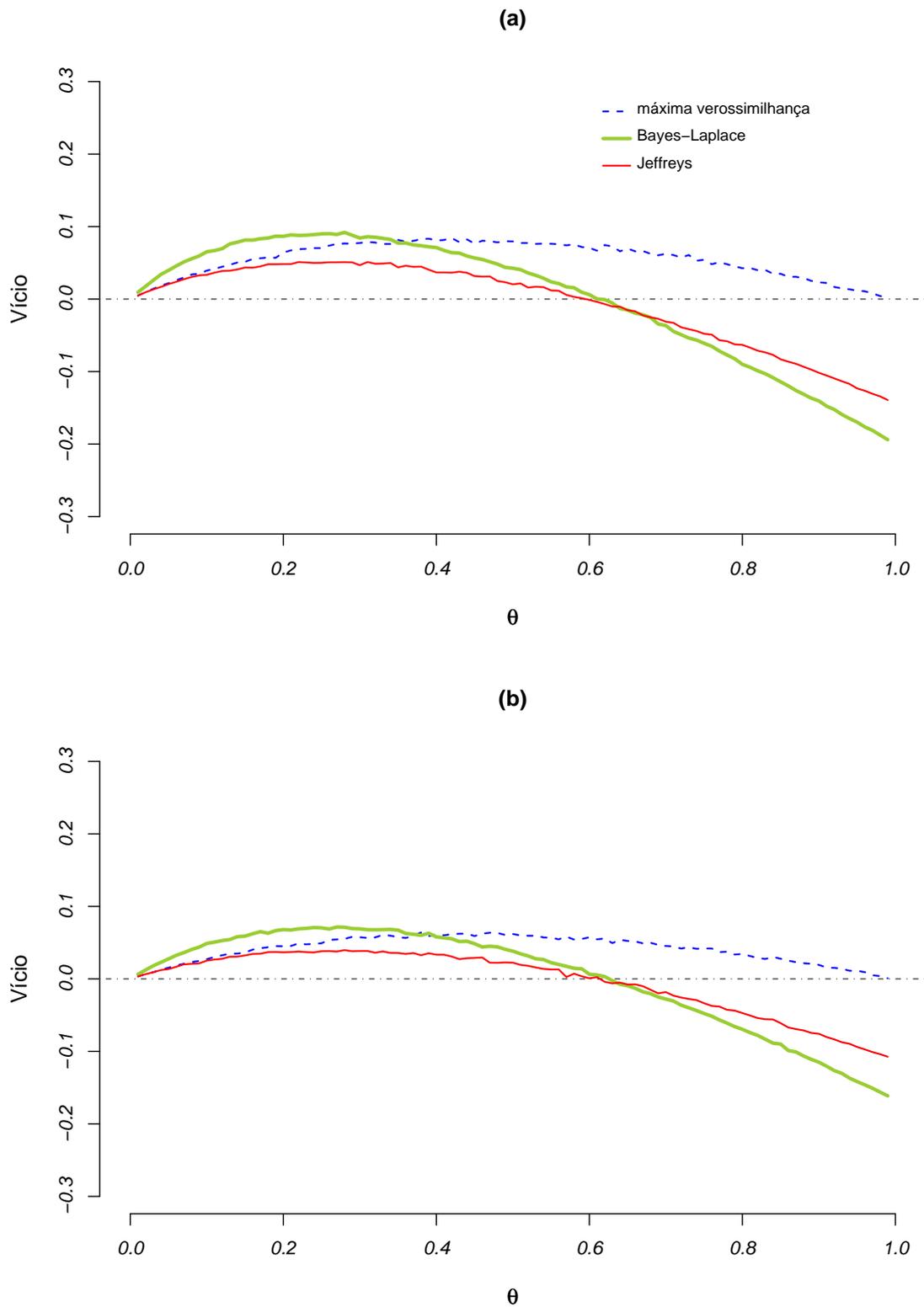


FIGURA 5.54: Vícius dos estimadores de máxima verossimilhança e da média da distribuição *a posteriori* para a probabilidade de sucesso para o modelo Binomial Negativa: (a)  $r = 3$  e (b)  $r = 4$ .

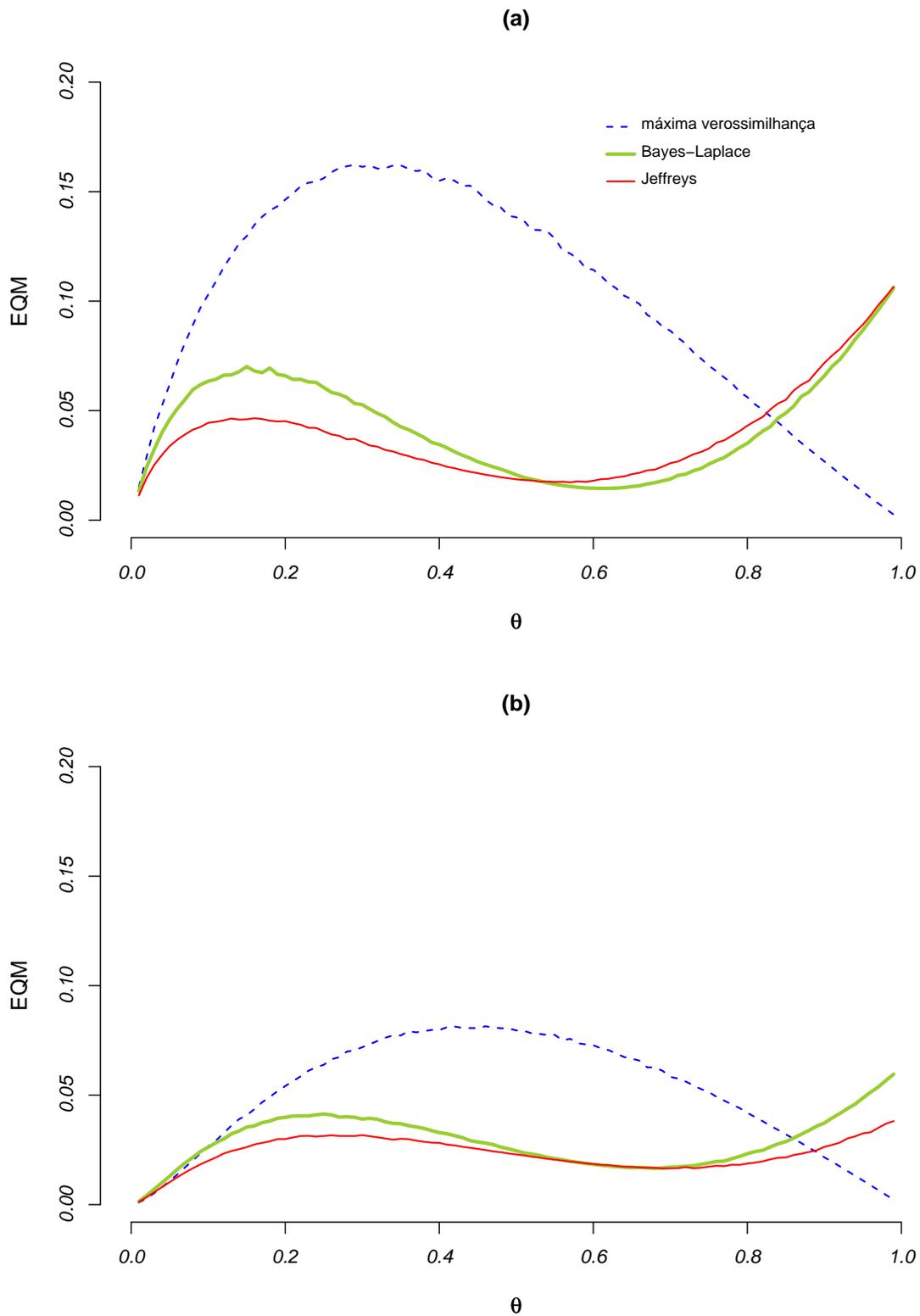


FIGURA 5.55: Erros quadráticos médios dos estimadores de máxima verossimilhança e da média da distribuição *a posteriori* para a probabilidade de sucesso para o modelo Binomial Negativa: (a)  $r = 1$  e (b)  $r = 2$ .

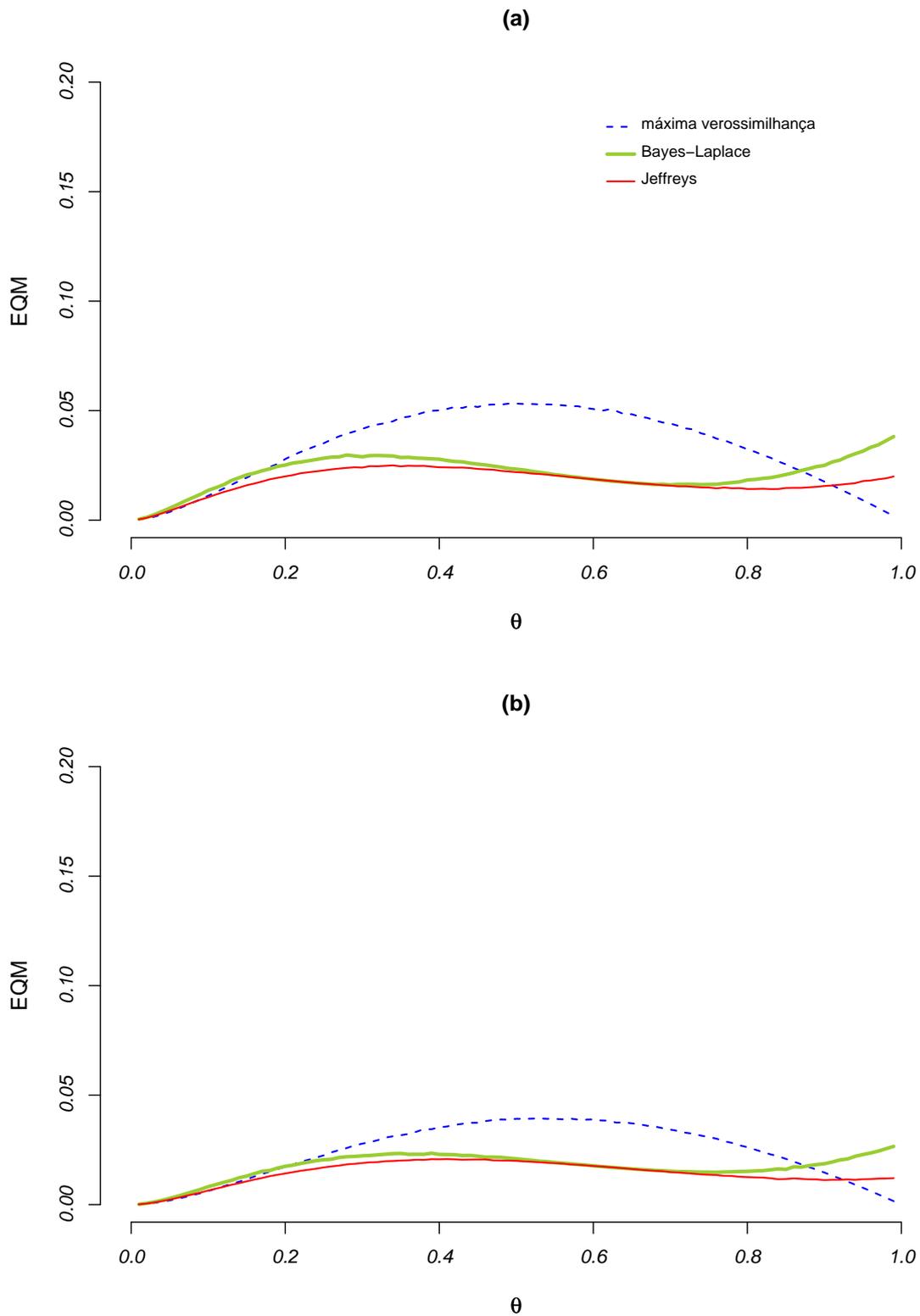


FIGURA 5.56: Erros quadráticos médios dos estimadores de máxima verossimilhança e da média da distribuição *a posteriori* para a probabilidade de sucesso para o modelo Binomial Negativa: (a)  $r = 3$  e (b)  $r = 4$ .

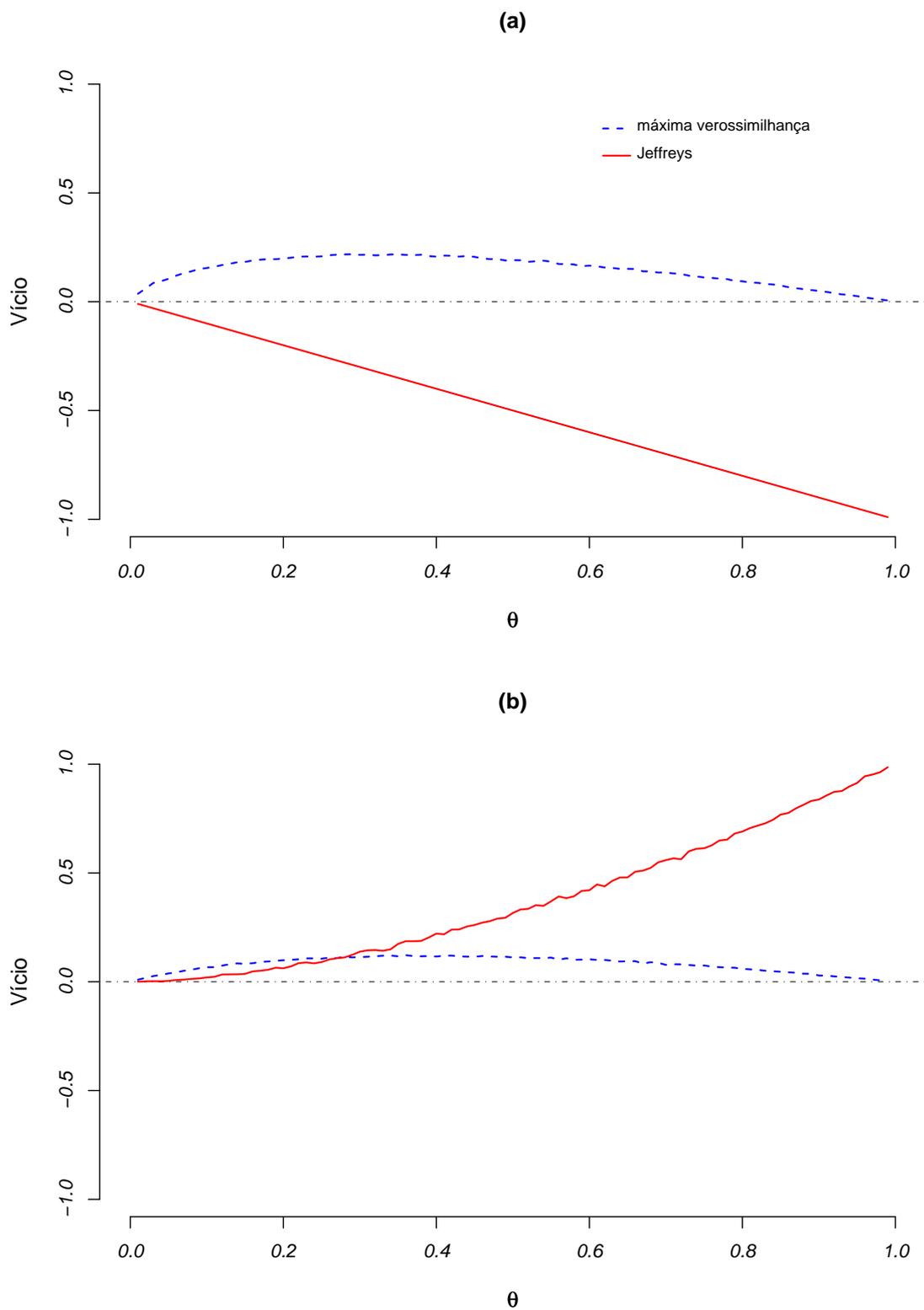


FIGURA 5.57: Vícios dos estimadores de máxima verossimilhança e da moda da distribuição *a posteriori* para a probabilidade de sucesso para o modelo Binomial Negativa: (a)  $r = 1$  e (b)  $r = 2$ .

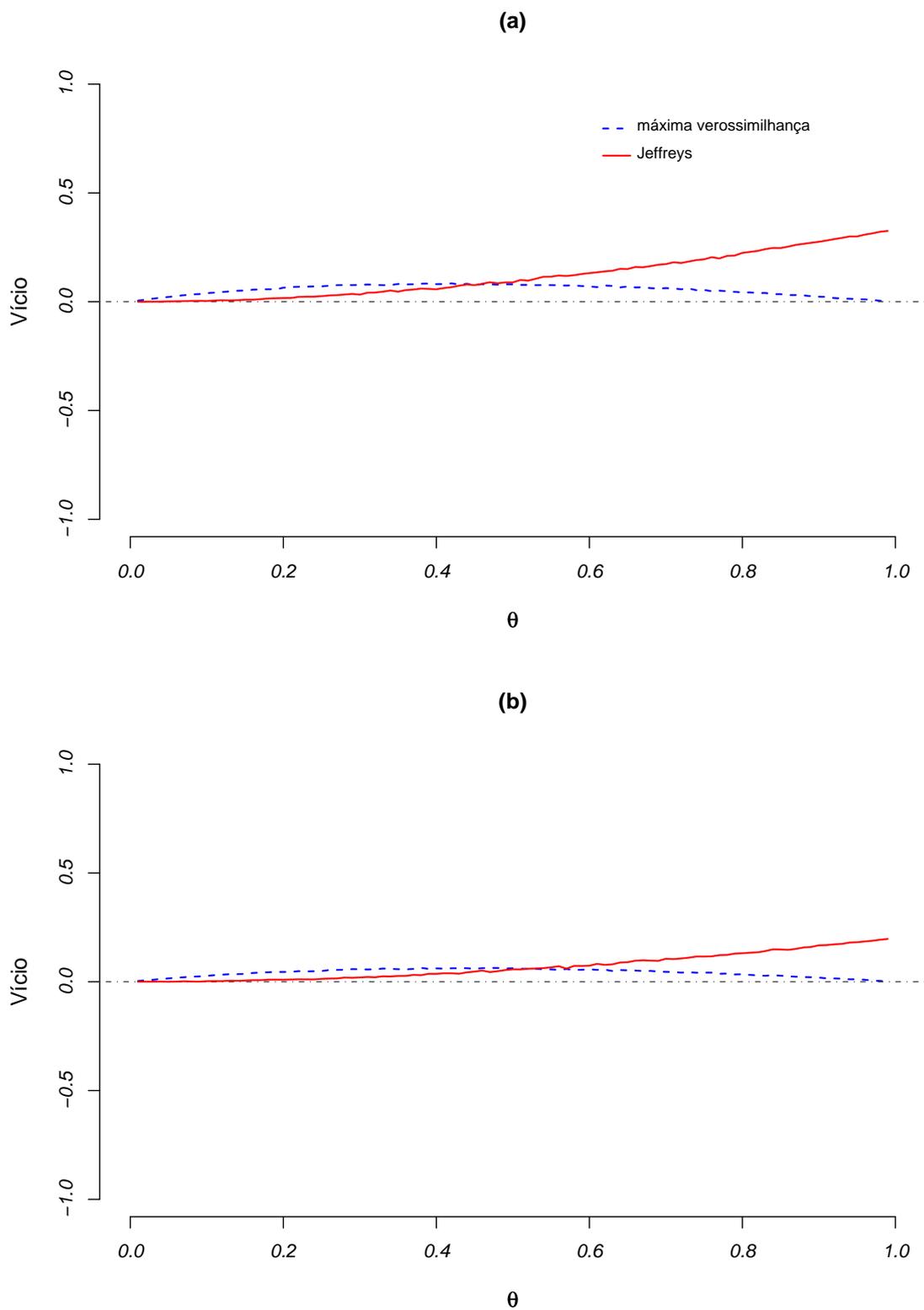


FIGURA 5.58: Vícius dos estimadores de máxima verossimilhança e da moda da distribuição *a posteriori* para a probabilidade de sucesso para o modelo Binomial Negativa: (a)  $r = 3$  e (b)  $r = 4$ .

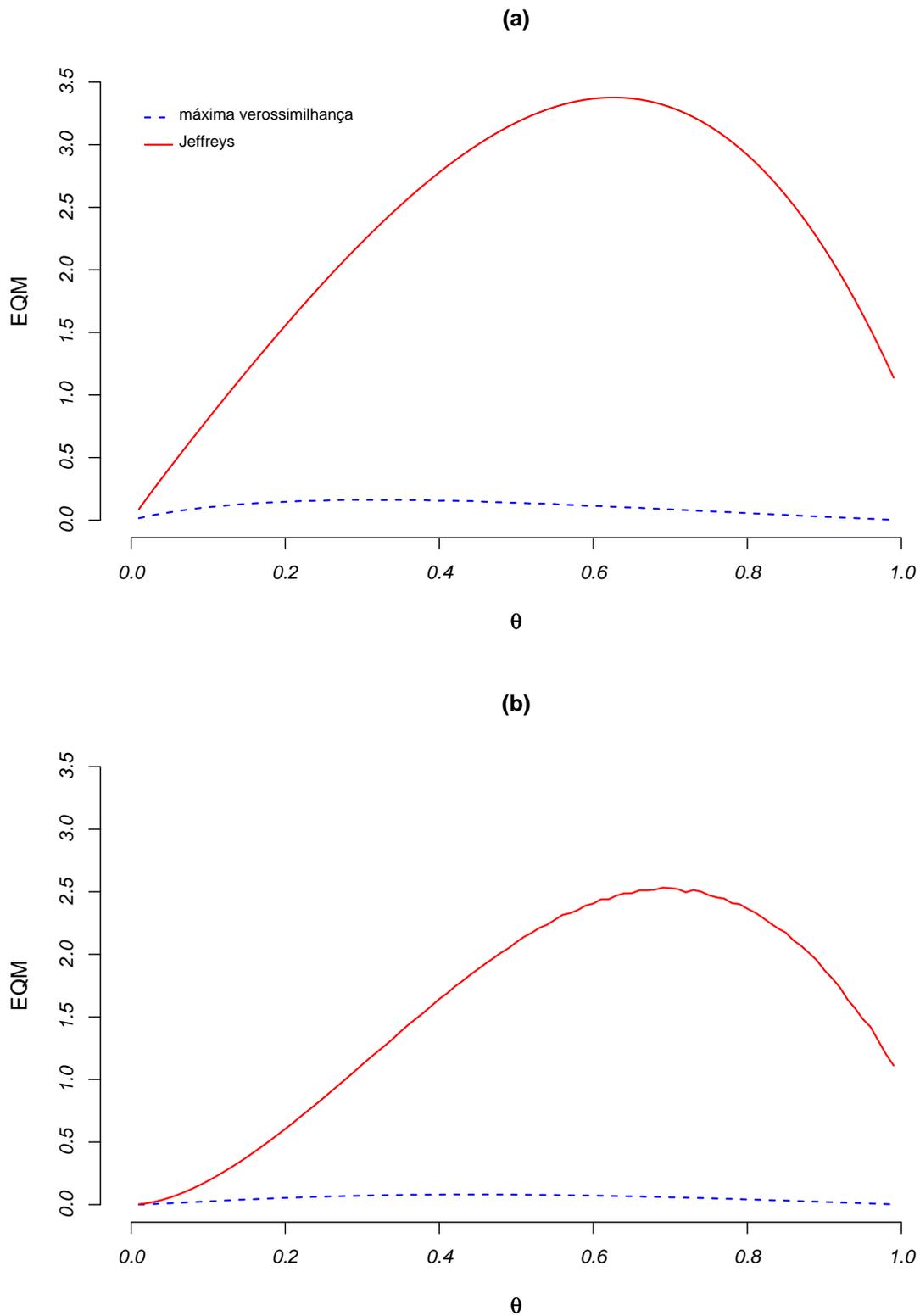


FIGURA 5.59: Erros quadráticos médios dos estimadores de máxima verossimilhança e da moda da distribuição *a posteriori* para a probabilidade de sucesso para o modelo Binomial Negativa: (a)  $r = 1$  e (b)  $r = 2$ .

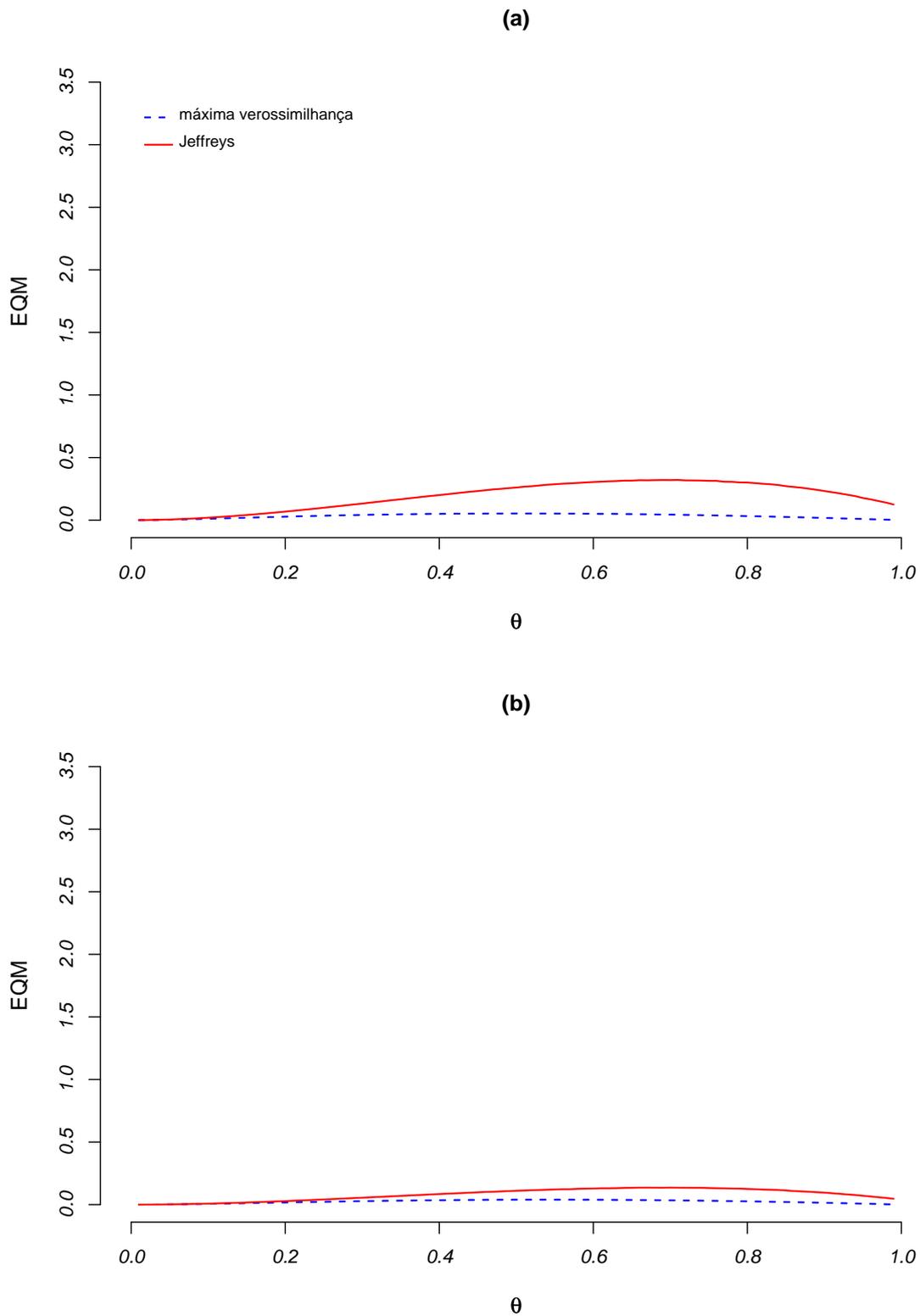


FIGURA 5.60: Erros quadráticos médios dos estimadores de máxima verossimilhança e da moda da distribuição *a posteriori* para a probabilidade de sucesso para o modelo Binomial Negativa: (a)  $r = 3$  e (b)  $r = 4$ .

## Capítulo 6

# Comparação de duas proporções correlacionadas

Nos Capítulos 3 e 4 a probabilidade de sucesso, denotada por  $\theta$ , foi modelada segundo as distribuições Binomial e Binomial Negativa. Esses modelos são baseados em ensaios de Bernoulli cujas respostas são binárias.

Agora o interesse é analisar a ocorrência de um evento que assume valores no conjunto  $\{(0, 0), (0, 1), (1, 0), (1, 1)\}$ . Um problema decorrente desses eventos, geralmente presentes quando se trata de mudança de comportamento ou mudança de opinião, é encontrar proporções correlacionadas (Irony *et al.*, 2000). Como em Irony *et al.* (2000), um exemplo é analisar a opinião de um grupo de indivíduos que apóiam um candidato à presidência antes e após um discurso.

Seja  $(T_1, T_2)$  uma variável aleatória que admite valores no conjunto  $\{(0, 0), (0, 1), (1, 0), (1, 1)\}$ , em que  $T_1$  e  $T_2$  representam as opiniões antes e depois do discurso, com distribuição de probabilidade conjunta dada por  $\theta_{t_1 t_2} = \mathbb{P}[T_1 = t_1, T_2 = t_2]$  e  $n_{t_1 t_2}$  as frequências observadas como na tabela 6.1.

Geralmente, para testar a significância estatística da mudança de opinião é empregada uma solução clássica por meio do teste de McNemar (1947), que consiste em um teste não-paramétrico adequado para tabelas  $2 \times 2$  de dados dicotômicos com amostras correlacionadas (Agresti, 2003, p. 411). É importante lembrar que a dependência entre as medidas causa dificuldade na obtenção de intervalos de confiança adequados, conforme

citam Irony *et al.* (2000).

TABELA 6.1: Distribuição de probabilidade conjunta de  $T_1$  e  $T_2$ .

	$T_2 = 0$	$T_2 = 1$	<i>Total</i>
$T_1 = 0$	$\theta_{00}(n_{00})$	$\theta_{01}(n_{01})$	$\theta_{0.}(n_{0.})$
$T_1 = 1$	$\theta_{10}(n_{10})$	$\theta_{11}(n_{11})$	$\theta_{1.}(n_{1.})$
<i>Total</i>	$\theta_{.0}(n_{.0})$	$\theta_{.1}(n_{.1})$	$1(n)$

A primeira análise bayesiana para esta situação é atribuída a Altham (1971), enquanto contribuições mais recentes incluem Irony *et al.* (2000) e Kateri *et al.* (2001) e outros autores mencionados em Agresti & Hitchcock (2005) e Agresti (2003, p. 442). Segundo Irony *et al.* (2000), para verificar a ocorrência de proporções correlacionadas, ou seja, a diferença entre as opiniões, podem ser consideradas as seguintes estratégias:

- $\theta_{01} = \theta_{10}$ .
- $\theta_{01} = \lambda(\theta_{01} + \theta_{10})$ , em que  $\lambda$  mede a proporção de indivíduos que mudaram de opinião.

O interesse desse capítulo é analisar o impacto da verossimilhança parcial (Altham, 1971) na percepção da ocorrência da mudança de opinião ao considerar os testes clássicos de McNemar (1947) e Binomial e no teste de significância genuinamente bayesiano (FBST) (Pereira *et al.*, 2008) para duas proporções correlacionadas.

## 6.1 Modelo Trinomial

O modelo estatístico para o problema em questão deve alocar quatro possíveis situações  $\{T_1 = t_1, T_2 = t_2\}$ , em que  $t_1, t_2 = 0, 1$ , em um número fixo  $n$  de vezes, ou seja, classificar as possíveis mudanças, ou não, de cada indivíduo entrevistado sem dispensar a dependência entre  $T_1$  e  $T_2$ . Ou seja, segue um modelo Trinomial. Observe a tabela 6.1, note que o valor de  $n$  é fixo (amostra na qual são analisadas as opiniões antes e após o discurso), ou seja, o par de variáveis aleatórias  $(T_1, T_2)$  são responsáveis por alocar em cada uma das quatro posições da tabela os  $n$  pares de informações. Se a proporção de respostas

de três dos quatro pares forem determinadas a quarta pode ser obtida, subtraindo do total a soma dessas. Assim, sua verossimilhança pode ser expressa em termos de apenas três parâmetros  $\theta_{t_1 t_2} = (\theta_{00}, \theta_{01}, \theta_{10}, 1 - \theta_{00} - \theta_{01} - \theta_{10})$  como segue o modelo Trinomial:

$$h(n_{t_1 t_2} | \theta_{t_1 t_2}) = \binom{n}{n_{00}, n_{01}, n_{10}, n - (n_{00} + n_{01} + n_{10})} \theta_{00}^{n_{00}} \theta_{01}^{n_{01}} \theta_{10}^{n_{10}} \times [1 - (\theta_{00} + \theta_{01} + \theta_{10})]^{n - (n_{00} + n_{01} + n_{10})} \quad (6.1)$$

em que  $0 \leq \theta_{t_1 t_2} \leq 1$ , para  $t_1, t_2 = 0, 1$  e  $\sum_{t_1, t_2=0,1} \theta_{t_1 t_2} = 1$ . O termo

$$\binom{n}{n_{00}, n_{01}, n_{10}, n - (n_{00} + n_{01} + n_{10})} = \frac{n!}{n_{00}! n_{01}! n_{10}! (n - (n_{00} + n_{01} + n_{10}))!}$$

descreve o *coeficiente multinomial*, conforme Casella & Berger (2002, p. 181).

Conforme mencionado no Capítulo 3, a partir do conhecimento sobre o vetor paramétrico  $\theta_{t_1 t_2}$  pode-se definir uma família paramétrica de densidades. Em geral essa abordagem facilita a análise, um caso muito considerado é o de distribuições *a priori* conjugadas, ou seja, a distribuição *a priori* e a distribuição *a posteriori* pertencem a mesma classe de distribuições havendo apenas atualização no hiperparâmetros.

Observe que a família de distribuições Dirichlet é naturalmente conjugada com a distribuição Multinomial. Desta forma, ao considerar a distribuição *a priori* Dirichlet com hiperparâmetros  $a_{t_1 t_2} = (a_{00}, a_{01}, a_{10}, a_{11})$ , denotada por  $\text{Di}(a_{00}, a_{01}, a_{10}, a_{11})$ , e a função de verossimilhança  $L_{n_{t_1 t_2}}(\theta_{t_1 t_2}) = h(n_{t_1 t_2} | \theta_{t_1 t_2})$  é obtida a distribuição *a posteriori* que segue uma distribuição Dirichlet com hiperparâmetros atualizados  $A_{t_1 t_2} = (A_{00}, A_{01}, A_{10}, A_{11})$ , em que  $A_{t_1 t_2} = a_{t_1 t_2} + n_{t_1 t_2}$ , de acordo com o Teorema de Bayes:

$$\begin{aligned} h(\theta_{t_1 t_2} | n_{t_1 t_2}) &\propto L_{n_{t_1 t_2}}(\theta_{t_1 t_2}) h(\theta_{t_1 t_2}) \\ &\propto [\theta_{00}^{n_{00}} \theta_{01}^{n_{01}} \theta_{10}^{n_{10}} (1 - (\theta_{00} + \theta_{01} + \theta_{10}))^{n_{00} + n_{01} + n_{10}}] \\ &\quad \times [\theta_{00}^{a_{00}-1} \theta_{01}^{a_{01}-1} \theta_{10}^{a_{10}-1} (1 - (\theta_{00} + \theta_{01} + \theta_{10}))^{a_{00} + a_{01} + a_{10} - 1}] \\ &\propto \theta_{00}^{A_{00}-1} \theta_{01}^{A_{01}-1} \theta_{10}^{A_{10}-1} (1 - (\theta_{00} + \theta_{01} + \theta_{10}))^{A_{00} + A_{01} + A_{10} - 1}. \end{aligned}$$

A distribuição *a priori* sob a hipótese de não mudança é induzida pela *priori* Dirichlet, conforme acima exposto. Note que os hiperparâmetros da distribuição *a priori* precisam ser elicitados.

Agresti & Hitchcock (2005) citam algumas propostas de distribuições *a priori* para o vetor paramétrico  $\theta_{t_1 t_2}$  do modelo Multinomial. Como o contexto analisado neste estudo é de não-informação, foi considerada a distribuição *a priori* Dirichlet como uma *Uniforme* para cada  $\theta_{t_1 t_2}$ , pois para cada um dos hiperparâmetros foi atribuído  $a_{t_1 t_2} = 1$ , proposto por Good (1980), utilizado em Irony *et al.* (2000). Outra distribuição *a priori* investigada foi a distribuição *a priori* de Jeffreys (coincidente com a distribuição *a priori* de referência) (Agresti & Hitchcock, 2005).

Assim, ao considerar a proposta de Good (1980) tem-se como distribuição *a posteriori* de  $\theta_{t_1 t_2} | n_{t_1 t_2}$  um modelo  $\text{Di}(1 + n_{00}, 1 + n_{01}, 1 + n_{10}, 1 + n_{11})$ . Para se obter a distribuição *a priori* de Jeffreys é necessário calcular o logaritmo da verossimilhança:

$$\begin{aligned} \ln h(n_{t_1 t_2} | \theta_{t_1 t_2}) &= \ln \binom{n}{n_{00}, n_{10}, n_{01}, n_{11}} + n_{00} \ln \theta_{00} \\ &\quad + n_{01} \ln \theta_{01} + n_{10} \ln \theta_{10} + n_{11} \ln \theta_{11}. \end{aligned}$$

Como a informação de Fisher, pela Definição 2.2, é obtida a partir da esperança matemática da derivada segunda, calculamos

$$-\mathbb{E} \left( \frac{\partial^2 \ln h(N_{t_1 t_2} | \theta_{t_1 t_2})}{\partial \theta'_{t_1 t_2} \theta_{t_1 t_2}} \right) = \begin{bmatrix} \frac{n}{\theta_{00}} & 0 & 0 & 0 \\ 0 & \frac{n}{\theta_{01}} & 0 & 0 \\ 0 & 0 & \frac{n}{\theta_{10}} & 0 \\ 0 & 0 & 0 & \frac{n}{\theta_{11}} \end{bmatrix}.$$

Portanto, a distribuição *a priori* de Jeffreys, conforme Definição 2.3, é dada por

$$h(\theta_{t_1 t_2}) = \sqrt{\frac{n^4}{\theta_{00} \theta_{01} \theta_{10} \theta_{11}}} \propto \sqrt{\frac{1}{\theta_{00} \theta_{01} \theta_{10} (1 - (\theta_{00} + \theta_{01} + \theta_{10}))}}.$$

Logo, a distribuição *a priori* de Jeffreys segue uma distribuição  $\text{Di} \left( \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2} \right)$  (Agresti & Hitchcock, 2005). Assim, a distribuição *a posteriori* de  $\theta_{t_1 t_2} | n_{t_1 t_2}$  satisfaz  $\text{Di} \left( \frac{1}{2} + n_{00}, \frac{1}{2} + n_{01}, \frac{1}{2} + n_{10}, \frac{1}{2} + n_{11} \right)$ .

Altham (1971) sugeriu uma função de verossimilhança, alternativa ao modelo acima exposto, que será discutido na próxima seção.

## 6.2 Verossimilhança parcial da distribuição Trinomial

Observe nas medidas para verificar a diferença entre opiniões (Irony *et al.*, 2000) que os parâmetros de interesse são  $\theta_{01}$  e  $\theta_{10}$  da distribuição Trinomial (equação 6.1). Altham (1971) e Irony *et al.* (2000) reescreveram essa verossimilhança ao considerar  $d = n_{01} + n_{10}$  e  $\eta = \theta_{01} + \theta_{10}$ , assim:

$$f(n_{01}, n_{10}, n - d | \theta_{01}, \theta_{10}, \eta) = \binom{n}{n_{01}, n_{10}, n_{00} + n_{11}} \theta_{01}^{n_{01}} \theta_{10}^{n_{10}} (\theta_{00} + \theta_{11})^{n_{00} + n_{11}},$$

ou seja,

$$\begin{aligned} f(n_{01}, n_{10}, n - d | \theta_{01}, \theta_{10}, \eta) &= \binom{n}{n_{01}, n_{10}, n - (n_{01} + n_{10})} \theta_{01}^{n_{01}} \theta_{10}^{n_{10}} \\ &\times (1 - (\theta_{01} + \theta_{10}))^{n - (n_{01} + n_{10})}. \end{aligned} \quad (6.2)$$

Ao substituir  $n_{01} + n_{10}$  por  $d$  e  $\theta_{01} + \theta_{10}$  por  $\eta$ , na igualdade (6.2) tem-se

$$f(n_{01}, n_{10}, n - d | \theta_{01}, \theta_{10}, \eta) = \binom{n}{n_{01}, n_{10}, n - d} \theta_{01}^{n_{01}} \theta_{10}^{n_{10}} (1 - \eta)^{n - d}. \quad (6.3)$$

Assim, ao multiplicar e dividir a igualdade (6.3) por  $\frac{(\theta_{01} + \theta_{10})^{n_{01} + n_{10}}}{(\theta_{01} + \theta_{10})^{n_{01} + n_{10}}}$  é obtido

$$\begin{aligned} f(n_{01}, n_{10}, n - d | \theta_{01}, \theta_{10}, \eta) &= \binom{n}{n_{01}, n_{10}, n - d} \frac{\theta_{01}^{n_{01}}}{(\theta_{01} + \theta_{10})^{n_{01}}} \frac{\theta_{10}^{n_{10}}}{(\theta_{01} + \theta_{10})^{n_{10}}} \\ &\times (\theta_{01} + \theta_{10})^{n_{01} + n_{10}} (1 - \eta)^{n - d}. \end{aligned}$$

Lembre-se que uma das estratégias para verificar a mudança de opinião é dada por  $\lambda = \frac{\theta_{01}}{\theta_{01} + \theta_{10}}$  (Irony *et al.*, 2000), portanto  $(1 - \lambda) = \frac{\theta_{10}}{\theta_{01} + \theta_{10}}$ , então

$$f(n_{01}, n_{10}, n - d | \theta_{01}, \theta_{10}, \eta) = \binom{n}{n_{01}, n_{10}, n - d} \lambda^{n_{01}} (1 - \lambda)^{n_{10}} \eta^d (1 - \eta)^{n - d} \quad (6.4)$$

ou seja,

$$f(n_{01}, n_{10}, n - d | \theta_{01}, \theta_{10}, \eta) = \binom{n}{d} \eta^d (1 - \eta)^{n - d} \binom{d}{n_{01}} \lambda^{n_{01}} (1 - \lambda)^{n_{10}}. \quad (6.5)$$

Portanto, conclui-se que

$$f(n_{01}, n_{10}, n - d | \theta_{01}, \theta_{10}, \eta) = f_0(d | \eta) f_1(n_{01} | \lambda, d). \quad (6.6)$$

Nas equações (6.4) e (6.5) utilizou-se a identidade

$$\binom{n}{n_{01}, n_{10}, n-d} = \binom{n}{d} \binom{d}{n_{01}}, \quad (6.7)$$

cuja demonstração esta no no Apêndice A.

A função de verossimilhança (6.6) é fatorada em dois termos, em que um deles depende do parâmetro de interesse  $\lambda$  (que será chamado de verossimilhança parcial de  $\lambda$ ) e o outro depende do parâmetro  $\eta$ . Para testar as hipóteses  $H_0 : \theta_{01} = \theta_{10}$  contra  $H_1 : \theta_{01} \neq \theta_{10}$ , apenas o segundo fator é necessário. Como  $d$  é uma estatística suficiente para o parâmetro  $\eta$  e ancilar para  $\lambda$  não existe perda de informação ao considerar apenas  $f_1$  para estimar  $\lambda$ , segundo Basu (1977).

## 6.3 Teste de Hipóteses

### 6.3.1 Testes de Hipóteses Clássicos

Para testar as hipóteses  $H_0 : \theta_{01} = \theta_{10}$  contra  $H_1 : \theta_{01} \neq \theta_{10}$ , ou seja, ao testar a mudança de opinião pelo modelo Trinomial, é considerado geralmente o teste clássico de McNemar (1947). Para medir a proporção de indivíduos que mudaram de opinião (considerar a medida  $\lambda$  para verificação da mudança de opinião) o modelo Trinomial é reduzido ao modelo Binomial, por meio da fatoração sugerida por Altham (1971). O que torna o teste de McNemar não aplicável mais sim o teste Binomial. Collani & Dräger (2001, cap. 11) comentam que ambos os testes são baseados na distribuição Binomial, embora com metodologias distintas.

Segundo Basu (1977) não há perda de informação ao considerar a verossimilhança parcial de  $\lambda$  diante do uso do modelo Trinomial. Note que pela metodologia clássica o teste de McNemar e o teste Binomial precisam ser aplicados para testar a mudança de opinião, o que impede a avaliação dessa afirmação.

#### Teste de McNemar

Para testar as hipóteses  $H_0 : \theta_{01} = \theta_{10}$  contra  $H_1 : \theta_{01} \neq \theta_{10}$ , McNemar (1947) propôs um teste qui-quadrado baseado na distribuição Binomial. A estatística desse teste

depende apenas da classificação dos casos nas diferentes categorias para duas observações. Considera que as observações de  $n_{00}$  e  $n_{11}$  da diagonal principal, da tabela 6.1, são irrelevantes para inferir sobre  $\theta_{01}$  e  $\theta_{10}$  (Agresti, 2007, p. 245). A estatística do teste é dada por

$$Z = \frac{(n_{01} - n_{10})^2}{n_{01} + n_{10}},$$

que, sob  $H_0$ , tem distribuição qui-quadrado com 1 grau de liberdade. Esta estatística  $Z$  pode ser reescrita com a correção de descontinuidade

$$Z^* = \frac{(|n_{01} - n_{10}| - 1)^2}{n_{01} + n_{10}}.$$

As frequências marginais não são homogêneas se a estatística de teste for significativa. Quando  $n_{01}$  ou  $n_{10}$  são pequenos ( $n_{01} + n_{10} < 20$ ), então não é apropriada a aproximação para distribuição qui-quadrado.

### Teste Binomial

Pela fatoração proposta por Altham (1971) a verossimilhança parcial da distribuição Trinomial pôde ser fatorada em dois termos, dois quais apenas um dos dois fatores é necessário para testar  $H_0 : \theta_{01} = \theta_{10}$  contra  $H_1 : \theta_{01} \neq \theta_{10}$  conforme Basu (1977), sendo este

$$f_1(n_{01}|\lambda, d) = \binom{d}{n_{01}} \lambda^{n_{01}} (1 - \lambda)^{n_{10}} \quad (6.8)$$

Conforme mencionado no início desse capítulo,  $\lambda = \frac{\theta_{01}}{\theta_{01} + \theta_{10}}$  e sob  $H_0 : \theta_{01} = \theta_{10}$ , assim é análogo testar  $H_0 : \lambda = 0,5$  contra  $H_1 : \lambda \neq 0,5$ . Assim, um teste de significância exato pode ser obtido para calcular o nível descritivo baseado no modelo Binomial (Collani & Dräger, 2001, p. 313), ao considerar a igualdade (6.8) sob  $H_0$  tem-se

$$f_1(n_{01}|\lambda, d) = \binom{d}{n_{01}} \lambda^{n_{01} + n_{10}}$$

Assim, o nível descritivo ( $p_v$ ) é obtido por

$$p_v = \sum_{k=0}^{\min(n_{01}, n_{10})} f_1(n_{01}|\lambda, d) + \sum_{k=\max(n_{01}, n_{10})}^{n_{01} + n_{10}} f_1(n_{01}|\lambda, d).$$

### 6.3.2 Uma medida de evidência completamente bayesiana

Para testar  $H_0 : \theta_{01} = \theta_{10}$  contra  $H_1 : \theta_{01} \neq \theta_{10}$  foi utilizado o FBST. Seja

$$\Theta = \left\{ \boldsymbol{\theta} : 0 \leq \theta_{t_1 t_2} \leq 1, \sum_{t_1, t_2=0,1} \theta_{t_1 t_2} = 1 \right\}$$

e  $\Theta_0 \subset \Theta$  satisfazendo a restrição  $\theta_{01} = \theta_{10}$ . Testar as hipóteses acima é equivalente a testar  $H_0 : \boldsymbol{\theta} \in \Theta_0$  contra  $H_1 : \boldsymbol{\theta} \in \Theta_1 = \Theta - \Theta_0$ .

Como a dimensão de  $\Theta_0$  é menor que a dimensão de  $\Theta$ , a hipótese nula é conhecida como hipótese precisa. Em geral, esta hipótese é difícil de ser verificada no caso clássico. Neste sentido, Pereira *et al.* (2008) sugeriram uma medida de evidência alternativa a favor de  $H_0$  totalmente bayesiana conhecida como *Full Bayesian Significance Test* (FBST). Esta medida é definida como  $ev = 1 - \mathbb{P}(\boldsymbol{\theta} \in T^* | \{n_{t_1 t_2}\}) = 1 - \int_{T^*} h(\boldsymbol{\theta}_{t_1 t_2} | n_{t_1 t_2}) d\boldsymbol{\theta}_{t_1 t_2}$ , em que  $T^* = \left\{ \boldsymbol{\theta} \in \Theta : h(\boldsymbol{\theta}_{t_1 t_2} | n_{t_1 t_2}) > \sup_{\Theta_0} h(\boldsymbol{\theta}_{t_1 t_2} | n_{t_1 t_2}) \right\}$  e  $h(\boldsymbol{\theta}_{t_1 t_2} | n_{t_1 t_2})$  é a distribuição *a posteriori* de  $\boldsymbol{\theta}_{t_1 t_2} | n_{t_1 t_2}$ .

## 6.4 Exemplos

**Exemplo I** (Pereira *et al.*, 2008; McNemar, 1947; Irony *et al.*, 2000): Dois professores, Ed e Joe, do Departamento de Odontologia, avaliaram o desempenho de 224 estudantes na preparação de uma restauração dental. Cada estudante foi avaliado por dois professores, no qual (A) significa aprovação e (F) reprovação. Com o intuito de verificar se os professores são igualmente exigentes, o Departamento realizou uma pesquisa cujos dados estão resumidos na seguinte tabela.

TABELA 6.2: Freqüências observadas quanto a aprovação (A) ou reprovação (F) de 224 estudantes na preparação de uma restauração, avaliados pelos professores Joe e Ed.

	Joe		<i>Total</i>
	A	F	
Ed			
A	62	41	103
F	25	96	121
<i>Total</i>	87	137	224

**Exemplo II** (Irony *et al.*, 2000): Uma amostra de indivíduos questionados sobre apoiar um presidente. Após um discurso do presidente são questionados novamente. Quem vai mudar de opinião?

TABELA 6.3: Freqüências observadas em relação a opinião de 100 indivíduos quanto ao apoio de um presidente antes e após um discurso.

Antes	Depois		<i>Total</i>
	Sim	Não	
Sim	20	17	37
Não	10	53	63
<i>Total</i>	30	70	100

### 6.4.1 Procedimentos clássicos

TABELA 6.4: Resultado da aplicação dos procedimentos clássicos para os dados do exemplo I e do exemplo II.

	Exemplo I	Exemplo II
Teste McNemar	$Z^* = 3,409$ e $p_v = 0,065$	$Z^* = 1,333$ e $p_v = 0,248$
Teste Binomial	$p_v = 0,064$	$p_v = 0,248$

Estes dois testes clássicos mostraram níveis descritivos muito próximos em ambos os conjuntos de dados, conforme observa-se na tabela 6.4. No exemplo I (tabela 6.2), os professores são igualmente exigentes ao nível de significância de 5%. No exemplo II (tabela 6.3), após o discurso não há mudança de opinião dos indivíduos ao nível de significância de 5%.

### 6.4.2 Medida de evidência bayesiana

#### Modelo Trinomial

Para testar  $H_0 : \theta_{01} = \theta_{10}$  contra  $H_1 : \theta_{01} \neq \theta_{10}$ , seja  $\theta^*$  o estimador de máxima verossimilhança sob  $H_0$  dado por  $\theta^* = \left( \frac{n_{00}}{n}, \frac{n_{01} + n_{10}}{2n}, \frac{n_{01} + n_{10}}{2n}, \frac{n_{11}}{n} \right)$ .

Pela Seção 6.1 ao considerar a distribuição *a priori* sugerida por Good (1980) tem-se a distribuição *a posteriori* que satisfaz  $\text{Di}(n_{00} + 1, n_{01} + 1, n_{10} + 1, n_{11} + 1)$ .

Assim, ao avaliar  $h(\theta_{t_1 t_2} | n_{x,y})$  em  $\theta^*$  temos que  $h(\theta^* | n_{x,y}) = 621,770$ , no qual  $T = \{\theta : h(\theta_{t_1 t_2} | n_{x,y}) > 621,770\}$ , para o exemplo I (tabela 6.2). Para o exemplo II (tabela 6.3) ao considerar também a distribuição *a priori* sugerida por Good (1980) e a atualização a partir dos dados tem-se  $h(\theta^* | n_{x,y}) = 622,465$ , no qual  $T = \{\theta : h(\theta_{t_1 t_2} | n_{x,y}) > 622,465\}$ .

O processo foi repetido para a distribuição *a priori* de Jeffreys para os dois conjuntos de dados. Utilizando integração numérica pelo R (R, 2008) encontrou-se os resultados da tabela 6.5.

TABELA 6.5: Resultado da aplicação da medida de evidência bayesiana para os dados do exemplo I e do exemplo II no modelo Trinomial.

	Distribuição <i>a priori</i>	Distribuição <i>a posteriori</i>	Evidência sob $H_0$
Exemplo I	Good (1980)	$\text{Di}(63, 42, 26, 97)$	$ev = 0,263$
	Jeffreys	$\text{Di}(63, 5; 42, 5; 26, 5; 97, 5)$	$ev = 0,260$
Exemplo II	Good (1980)	$\text{Di}(21, 18, 11, 54)$	$ev = 0,601$
	Jeffreys	$\text{Di}(21, 5; 18, 5; 11, 5; 54, 5)$	$ev = 0,573$

Ao observar as evidências sob  $H_0$  obtidas para os dados do exemplo I pode-se afirmar que a medida de evidência não mostra sensibilidade em relação à distribuição *a priori* empregada, pois ao considerar qualquer uma delas a medida de evidência foi muito próxima. Como a evidência é uma probabilidade e esta mais próxima de 0, então os dados sugerem aceitar a hipótese de que os professores não foram igualmente exigentes. Para o exemplo II os métodos clássico e bayesiano concordam que depois do discurso não houve mudança de opinião dos indivíduos e também não foram obtidas medidas de evidência sob  $H_0$  sensíveis a distribuição *a priori*.

### Verossimilhança parcial de $\lambda$

Observe que diante da verossimilhança parcial de  $\lambda$  o caso Trinomial é reduzido ao caso Binomial ( $\text{Bin}(n_{01} + n_{10}, \lambda)$ ), explorado no Capítulo 3, e ao considerar a distribuição *a priori* de Bayes-Laplace para testar  $H_0 : \lambda = 0,5$  contra  $H_1 : \lambda \neq 0,5$  tem-se que a

distribuição *a posteriori* é  $\text{Beta}(n_{01}, n_{10})$ . Ao aplicar o FBST para este modelo foram obtidas as evidências expostas na tabela 6.6. Ao considerar a distribuição *a priori* de Jeffreys para testar as mesmas hipóteses acima expostas para esse modelo obteve-se distribuição *a posteriori*  $\text{Beta}(n_{01} + 0, 5, n_{10} + 0, 5)$  e as evidências também seguem na tabela 6.6.

TABELA 6.6: Resultado da aplicação da medida de evidência bayesiana para os dados do exemplo I e do exemplo II na Verossimilhança parcial de  $\lambda$ .

	Distribuição <i>a priori</i>	Distribuição <i>a posteriori</i>	Evidência sob $H_0$
Exemplo I	Bayes-Laplace	$\text{Beta}(41, 25)$	$ev = 0,043$
	Jeffreys	$\text{Beta}(41, 5; 25, 5)$	$ev = 0,044$
Exemplo II	Bayes-Laplace	$\text{Beta}(17, 10)$	$ev = 0,172$
	Jeffreys	$\text{Beta}(17, 5; 10; 5)$	$ev = 0,159$

Novamente, utilizando a verossimilhança parcial o FBST não é sensível em relação às distribuições *a priori* não-informativas, mas mostrou forte sensibilidade em relação ao método de fatoração.

Tanto para o exemplo I como para o exemplo II observou-se que a medida de evidência foi sensível em relação a parametrização utilizada. Por exemplo para o exemplo I ao considerar o modelo Trinomial foi obtido  $ev = 0,263$  utilizando distribuição *a priori* Uniforme e para a verossimilhança parcial de  $\lambda$  (Binomial) foi obtido  $ev = 0,043$  para essa mesma distribuição *a priori*. Concluindo, a discordância entre os professores que está evidente nos dados foi enfatizada com a parametrização utilizada. Entretanto, isso não ocorreu com o método clássico, inclusive ao fixar um nível de significância de 5% afirma não haver diferença entre as exigências.

No exemplo II novamente observa-se que para a verossimilhança parcial e para o modelo Trinomial FBST não é sensível em relação as distribuições *a priori* não-informativas, mas mostrou-se forte sensibilidade em relação ao método de fatoração.

# Capítulo 7

## Considerações Finais

Com os estudos desenvolvidos nesse texto pôde-se observar que a Inferência Bayesiana Objetiva para a probabilidade de sucesso dos modelos Binomial e Binomial Negativa apresentaram boas propriedades freqüentistas, analisadas a partir da probabilidade de cobertura e amplitude média dos intervalos bayesianos e através de estimadores pontuais, dos quais foram explorados estudos de consistência, viés e erro quadrático médio.

Para o modelo Binomial, conforme exposto no Capítulo 3, era esperado que os métodos bayesianos considerando a distribuição *a priori* de Bayes-Laplace, que é não-informativa para a probabilidade de sucesso, coincidente com a distribuição *a priori* menos favorável com função de perda quadrática ponderada, apresentassem propriedades freqüentistas satisfatórias, verificado no estudo de simulação apresentado no Capítulo 5.

No estudo de simulação verificou-se que a distribuição *a priori* menos favorável para o modelo Binomial com perda ponderada apresentou propriedades freqüentistas melhores que quando foi utilizada a função de perda quadrática. Observou-se também que a construção da distribuição *a priori* menos favorável para o modelo Binomial depende da escolha adequada da função de perda.

Para estimação intervalar da probabilidade de sucesso para o modelo Binomial os métodos bayesianos apresentaram-se, pelas análises consideradas, mais adequados. O intervalo de credibilidade (Seção 3.2.1) utilizando a distribuição *a priori* de Bayes-Laplace seria o mais indicado.

Apesar das dificuldades encontradas para obtenção da distribuição *a priori* menos

favorável para a probabilidade de sucesso do modelo Binomial Negativa, foi possível estender ao modelo Binomial Negativa os estudos realizados para o modelo Binomial.

No estudo de simulação do modelo Binomial Negativa verificou-se que as probabilidades de coberturas dos intervalos de credibilidade foram estáveis enquanto que as probabilidades de cobertura dos intervalos de confiança foram extremamente subestimadas para probabilidade de sucesso maior que 0,5.

É importante salientar que a metodologia bayesiana desse trabalho foi baseada em distribuições a priori não informativas e sua qualidade foi avaliada por meio de conceitos freqüentistas (IBO).

Tanto para o modelo Binomial como para o Binomial Negativa os estimadores bayesianos analisados confirmaram sua consistência no estudo de simulação mostrado no Capítulo 5, para as distribuições *a priori* de Bayes-Laplace e de Jeffreys. No entanto, para o modelo Binomial a média e a moda da distribuição *a posteriori* para a probabilidade de sucesso com distribuição *a priori* menos favorável com função de perda quadrática não confirma sua consistência no estudo de simulação realizado, pois com o aumento de  $n$  o vício tende mais lentamente a zero.

Como proposta futura pretendemos analisar intervalos de confiança para a probabilidade de sucesso dos modelos Binomial e Binomial Negativa mais competitivos em relação aos utilizados nesse estudo.

Na etapa final do trabalho observou-se que o procedimento FBST para o modelo Trinomial e sua respectiva função de verossimilhança parcial é sensível para a parametrização adotada. Como proposta pretendemos verificar a sensibilidade da parametrização através de amostras simuladas do modelo Trinomial.

# Referências

- (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Abramowitz, M. & Stegun, A. (1970). *Handbook of Mathematical Functions*. Dover, New York.
- Agresti, A. (2003). *Categorical Data Analysis*. Wiley, New Jersey.
- Agresti, A. (2007). *An Introduction to Categorical Data Analysis*. Wiley, New Jersey.
- Agresti, A. & Coull, B. A. (1998). Approximate Is Better than Exact for Interval Estimation of Binomial Proportions. *The American Statistician*, pages 119–126.
- Agresti, A. & Hitchcock, D. B. (2005). Bayesian Inference for Categorical Data Analysis. *Statistical Methods and Applications*, pages 297–330.
- Altham, P. M. E. (1971). The Analysis of Matched Proportions. *Biometrika*, pages 561–576.
- Basu, D. (1977). On the Elimination of Nuisance Parameters. *Journal of the American Statistical Association*, pages 355–366.
- Bayarri, M. J. & Berger, J. O. (2004). The Interplay of Bayesian and Frequentist Analysis. *Statistical Science*, pages 58–80.
- Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York.
- Berger, J. O. (2006). The Case for Objective Bayesian Analysis. *Bayesian Analysis*, pages 385–402 and 457–464.
- Bernardo, J. M. (1989). The Geometry of Asymptotic Inference: Comment: On Multivariate Jeffreys' Priors. *Statistical Science*, pages 227–229.
- Bernardo, J. M. (2005). Reference Analysis. *Handbook of Statistics*, pages 17–90.
- Bernardo, J. M. & Smith, A. F. M. (2000). *Bayesian Theory*. Wiley, Chichester.
- Brown, L. D., Cai, T. T. & DasGupta, A. (2001). Interval Estimation for a Binomial Proportion. *Statistical Science*, pages 101–133.

- Brown, L. D., Cai, T. T. & DasGupta, A. (2002). Confidence Intervals for a Binomial Proportion and Asymptotic Expansions. *The Annals of Statistics*, pages 160–201.
- Casella, G. & Berger, R. L. (2002). *Statistical Inference*. Duxbury, Pacific Grove, second edition.
- Chen, M. & Shao, Q. (1998). Monte Carlo Estimation of Bayesian Credible and HPD Intervals. *Journal of Computational and Graphical Statistics*, pages 69–92.
- Chen, X., Zhou, K. & Aravena, J. L. (2008). Explicit Formula for Constructing Binomial Confidence Interval with Guaranteed Coverage Probability. *Communications in Statistics Theory and Methods*, pages 1173–1180.
- Clopper, C. J. & Pearson, E. S. (1934). The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial. *Biometrika*, pages 404–413.
- Collani, E. & Dräger, K. (2001). *Binomial Distribution Handbook for Scientists and Engineers*. Birkhäuser, Boston.
- Cox, D. R. (2006). *Principles of Statistical Inference*. Cambridge, New York.
- DeGroot, M. H. (1970). *Optimal Statistical Decisions*. McGraw-Hill, New York.
- Gelman, A., Carlin, J., Stern, H. & Rubin, D. (2004). *Bayesian Data Analysis*. Chapman & Hall, London.
- Good, I. J. (1980). Some History of the Hierarchical Bayesian Methodology. *Bayesian Statistics: Proceedings of the First International Meeting held in Valencia*, pages 489–504.
- Irony, T. Z., Pereira, C. A. d. B. & Tiwari, R. C. (2000). Analysis of Opinion Swing: Comparison of Two Correlated Proportions. *The American Statistician*, pages 57–62.
- Kateri, M., Papaioannou, T. & Dellaportas, P. (2001). Bayesian Analysis of Correlated Proportions. *The Indian Journal of Statistics*, pages 270–285.
- Lehmann, E. L. & Casella, G. (1998). *Theory of Point Estimation*. Springer-Verlag, New York.
- McNemar, Q. (1947). Note on the Sampling Error of the Difference Between Correlated Proportions or Percentages. *Psychometrika*, pages 153–157.
- Neyman, J. (1977). Frequentist Probability and Frequentist Statistics. *Synthese*, pages 97–131.
- Paulino, C. D., Turkman, A. A. & Murteira, B. (2003). *Estatística Bayesiana*. Fundação Calouste Gulbenkian, Lisboa.
- Pereira, C. A. d. B., Stern, J. M. & Wechsler, S. (2008). Can a Significance Test Be Genuinely Bayesian? *Bayesian Analysis*, pages 79–100.
- Yang, R. & Berger, J. O. (1997). A Catalog of Noninformative Priors. *Tech. Rep.*, pages ISDS 97–42.

# Apêndice A

## Demonstrações

As demonstrações que compreendem este Apêndice foram omitidas durante o texto principal por serem muito extensas.

### A.1 Demonstração do Lema 3.1

Deve-se mostrar que

$$\sum_{k=0}^x \binom{n}{k} \theta^k (1-\theta)^{n-k} = (n-x) \binom{n}{x} \int_0^{1-\theta} t^{n-x-1} (1-t)^x dt, \quad (\text{A.1})$$

conforme citam Casella & Berger (2002, p. 82).

Para demonstrar essa igualdade serão derivados ambos os lados da mesma. Primeiramente considere

$$\frac{\partial}{\partial \theta} \left[ \sum_{k=0}^x \binom{n}{k} \theta^k (1-\theta)^{n-k} \right] = \sum_{k=0}^x \frac{\partial}{\partial \theta} \binom{n}{k} \theta^k (1-\theta)^{n-k},$$

obtendo desta forma

$$\underbrace{- \sum_{k=0}^x \binom{n}{k} (n-k) \theta^k (1-\theta)^{n-k-1}}_{(a)} + \underbrace{\sum_{k=0}^x \binom{n}{k} k \theta^{k-1} (1-\theta)^{n-k}}_{(b)}.$$

Ao desenvolver os somatórios acima observa-se que a  $(k+1)$ -ésima parcela de (b) é igual à  $k$ -ésima parcela de (a) e a primeira parcela de (b) é zero, restando apenas a parcela (a) quando  $k = x$ . Assim, conclui-se que

$$\sum_{k=0}^x \frac{\partial}{\partial \theta} \binom{n}{k} \theta^k (1-\theta)^{n-k} = - \binom{n}{x} (n-x) \theta^x (1-\theta)^{n-x-1}.$$

Agora vejamos a derivação do lado direito da igualdade A.1, dado por

$$\frac{\partial}{\partial \theta} \left[ (n-x) \binom{n}{x} \int_0^{1-\theta} t^{n-x-1} (1-t)^x dt \right],$$

ou seja,

$$(n-x) \binom{n}{x} \int_0^{1-\theta} \frac{\partial}{\partial \theta} t^{n-x-1} (1-t)^x dt = -(n-x) \binom{n}{x} t^{n-x-1} (1-t)^x \Big|_{t=0}^{1-\theta},$$

logo tem-se que

$$(n-x) \binom{n}{x} \int_0^{1-\theta} t^{n-x-1} (1-t)^x dt = -(n-x) \binom{n}{x} (\theta)^x (1-\theta)^{n-x-1}.$$

Assim, ao substituir as expressões obtidas em A.1 e multiplicar ambos os lados por (-1), verifica-se que

$$(n-x) \binom{n}{x} (\theta)^x (1-\theta)^{n-x-1} = (n-x) \binom{n}{x} (\theta)^x (1-\theta)^{n-x-1}. \quad \blacksquare$$

## A.2 Demonstração do Lema 3.2

Seja  $U$  a função densidade de probabilidade da variável  $F$  com  $p$  e  $q$  graus de liberdade dada por

$$h(u|p, q) = \frac{\Gamma\left(\frac{p+q}{2}\right)}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{q}{2}\right)} \left(\frac{p}{q}\right)^{p/2} \frac{u^{(p-2)/2}}{\left(1 + \frac{p}{q}u\right)^{(p+q)/2}}.$$

e seja  $Y_1 = \frac{cU}{1+cU} \sim \text{Beta}\left(\frac{p}{2}, \frac{q}{2}\right)$ , no qual  $c = \frac{p}{q}$ .

Ao isolar  $U$  da função de  $y_1 = \frac{cU}{1+cU}$ , tem-se  $U = \frac{y_1}{c(1-y_1)} = g^{-1}(y_1)$ . Assim:

$$\frac{\partial g^{-1}(y_1)}{\partial y_1} = -\frac{1}{c} \left[ \frac{1(y_1-1) - y_1}{(y_1-1)^2} \right] = \frac{1}{c(y_1-1)^2}.$$

Ao substituir  $c$ , na expressão anterior, obtemos

$$\frac{\partial g^{-1}(y_1)}{\partial y_1} = \left(\frac{q}{p}\right) \frac{1}{(y_1-1)^2}.$$

Por transformação de variáveis, tem-se que  $f_{y_1}(y_1) = f_u(g^{-1}(y_1)) \left| \frac{\partial g^{-1}(y_1)}{\partial y_1} \right|$ . Então

$$\begin{aligned}
f_{y_1}(y_1) &= \frac{\Gamma\left(\frac{p+q}{2}\right)}{\Gamma(p/2)\Gamma(q/2)} \left[\frac{p}{q}\right]^{p/2} \\
&\quad \times \left[\frac{y_1}{\frac{p}{q}(1-y_1)}\right]^{(p/2)-1} \left[1 + \left(\frac{p}{q}\right) \frac{y_1}{\frac{p}{q}(1-y_1)}\right]^{-(p+q)/2} \frac{1}{(p/q)(1-y_1)^2} \\
&= \frac{1}{B(y_1 + \alpha, n - y_1 + \beta)} \frac{\frac{y_1 + \alpha}{n - y_1 + \beta} \left[-\frac{y_1}{y_1 - 1}\right]^{y_1 + \alpha - 1}}{\left[1 - \frac{y_1}{y_1 - 1}\right]^{\alpha + \beta + n}} \\
&\quad \times \left(\frac{n - y_1 + \beta}{y_1 + \alpha}\right) \frac{1}{(y_1 - 1)^2} \\
&= \frac{1}{B(p/2, q/2)} y_1^{(p/2)-1} (1 - y_1)^{-(p/2)+1} \left(\frac{1 - y_1 + y_1}{1 - y_1}\right)^{-\frac{p+q}{2}} (1 - y_1)^{-2} \\
&= \frac{1}{B(p/2, q/2)} y_1^{(p/2)-1} (1 - y_1)^{(q/2)-1},
\end{aligned}$$

em que  $B$  descreve a função Beta. Portanto, verifica-se que a variável transformada é tal que  $Y_1 \sim \text{Beta}\left(\frac{p}{2}, \frac{q}{2}\right)$ . ■

### A.3 Demonstração do Lema 4.1

Considere os eventos  $\{\omega = (\omega_1, \omega_2, \dots) : \omega_i = 1 \text{ ou } 0, \text{ para } i \in \mathbb{N}\}$ .

Seja  $A = \{\omega : X(\omega) \leq r - 1\}$  e  $B = \{\omega : Y(\omega) \leq n - r\}$ , em que

$$X(\omega) = \sum_{i=1}^n \omega_i \quad \text{e} \quad Y(\omega) = \inf \left\{ n : \sum_{i=1}^n \omega_i = r \right\}.$$

Observe que se  $A$  ocorre, então nos  $n$  primeiros ensaios ocorrem mais que  $(n - r)$  insucessos. Logo, o número de insucessos até o  $r$ -ésimo sucesso é maior que  $(n - r)$ , ou seja, o evento  $B$  ocorre.

Reciprocamente, se  $B$  ocorre, então ocorrem mais que  $(n - r)$  insucessos antes do  $r$ -ésimo sucesso. Logo, nos  $n$  primeiros ensaios de Bernoulli ocorreram menos que  $r$  sucessos, ou seja, o evento  $A$  ocorre.

Portanto,  $A = B$  e conseqüentemente,  $\mathbb{P}(A) = \mathbb{P}(B)$ . ■

## A.4 Demonstração da igualdade 6.7, do Capítulo 6

Deve-se mostrar que

$$\binom{n}{n_{01}, n_{10}, n-d} = \binom{n}{d} \binom{d}{n_{01}}. \quad (\text{A.2})$$

Primeiramente será desenvolvida a parcela do esquerdo da igualdade (A.2), obtendo assim

$$\binom{n}{n_{01}, n_{10}, n-d} = \binom{n}{n_{01}, n_{10}, n - (n_{01} + n_{10})} = \frac{n!}{n_{01}! n_{10}! (n-d)!}.$$

Ao desenvolver o lado direito da igualdade (A.2), tem-se

$$\binom{n}{d} \binom{d}{n_{01}} = \frac{n!}{(n-d)! d!} \frac{d!}{(d-n_{01})! n_{01}!} = \frac{n!}{n_{01}! n_{10}! (n-d)!}.$$

Portanto,

$$\binom{n}{n_{01}, n_{10}, n-d} = \binom{n}{d} \binom{d}{n_{01}}. \quad \blacksquare$$

# Apêndice B

## Esperança e Variância dos estimadores pontuais

### Distribuição Binomial

Lembrando que o estimador de máxima verossimilhança para a probabilidade de sucesso da distribuição Binomial dado por  $\frac{\mathbf{X}}{n}$ , sua esperança e variância são

$$\mathbb{E}\left(\frac{\mathbf{X}}{n}\right) = \frac{\mathbb{E}(\mathbf{X})}{n} = \theta \quad \text{e} \quad \text{Var}\left(\frac{\mathbf{X}}{n}\right) = \frac{\text{Var}(\mathbf{X})}{n} = \frac{\theta(1-\theta)}{n}.$$

A média da distribuição *a posteriori* de  $\theta$  é um estimador bayesiano para a probabilidade de sucesso do modelo Binomial, dado por  $\frac{\mathbf{X} + \alpha}{\alpha + \beta + n}$ . Sua esperança e variância são

$$\mathbb{E}\left(\frac{\mathbf{X} + \alpha}{\alpha + \beta + n}\right) = \frac{\mathbb{E}(\mathbf{X}) + \alpha}{\alpha + \beta + n} = \frac{n\theta + \alpha}{\alpha + \beta + n}$$

e

$$\text{Var}\left(\frac{\mathbf{X} + \alpha}{\alpha + \beta + n}\right) = \frac{\text{Var}(\mathbf{X} + \alpha)}{(\alpha + \beta + n)^2} = \frac{\text{Var}(\mathbf{X})}{(\alpha + \beta + n)^2} = \frac{n\theta(1-\theta)}{(\alpha + \beta + n)^2}.$$

O outro estimador bayesiano considerado foi a moda da distribuição *a posteriori* de  $\theta$  dado por  $\frac{\mathbf{X} + \alpha - 1}{\alpha + \beta + n - 2}$ . Sua esperança e variância são

$$\mathbb{E}\left(\frac{\mathbf{X} + \alpha - 1}{\alpha + \beta + n - 2}\right) = \frac{\mathbb{E}(\mathbf{X}) + \alpha - 1}{\alpha + \beta + n - 2} = \frac{n\theta + \alpha - 1}{\alpha + \beta + n - 2}$$

e

$$\text{Var}\left(\frac{\mathbf{X} + \alpha - 1}{\alpha + \beta + n - 2}\right) = \frac{\text{Var}(\mathbf{X} + \alpha - 1)}{(\alpha + \beta + n - 2)^2} = \frac{\text{Var}(\mathbf{X})}{(\alpha + \beta + n - 2)^2} = \frac{n\theta(1-\theta)}{(\alpha + \beta + n - 2)^2}.$$

### Distribuição Binomial Negativa

O estimador de máxima de verossimilhança para a probabilidade de sucesso do modelo Binomial Negativa é dado por  $\frac{r}{\mathbf{X}}$  e sua esperança e variância são dadas por

$$\begin{aligned} \mathbb{E}\left(\frac{r}{\mathbf{X}}\right) &= r\mathbb{E}\left(\frac{1}{\mathbf{X}}\right) = r \sum_{x=r}^{\infty} \frac{1}{x} \binom{x-1}{r-1} \theta^r (1-\theta)^{x-r} \\ &= \theta^r \frac{\Gamma(r+1)}{[\Gamma(r)]^2} \sum_{i=1}^{\infty} \frac{(1-\theta)^i}{i!} \frac{[\Gamma(r+i)]^2}{\Gamma(r+1+i)} \end{aligned} \quad (\text{B.1})$$

$$= \frac{r\theta^r}{\Gamma(r)} \sum_{i=1}^{\infty} \frac{(1-\theta)^i}{i!} \frac{\Gamma(r+i)}{r+i}, \quad (\text{B.2})$$

em que (B.1) descreve a série hipergeométrica de Gauss (Abramowitz & Stegun, 1970, p. 556).

Para obtenção da variância desse estimador é necessário conhecer o segundo momento, ou seja,

$$\begin{aligned} \mathbb{E}\left(\frac{r^2}{\mathbf{X}^2}\right) &= r^2\mathbb{E}\left(\frac{1}{\mathbf{X}^2}\right) = r^2 \sum_{x=r}^{\infty} \frac{1}{x^2} \binom{x-1}{r-1} \theta^r (1-\theta)^{x-r} \\ &= \theta^r \frac{[\Gamma(r+1)]^2}{[\Gamma(r)]^3} \sum_{i=1}^{\infty} \frac{(1-\theta)^i}{i!} \frac{[\Gamma(r+i)]^3}{[\Gamma(r+1+i)]^2} \end{aligned} \quad (\text{B.3})$$

$$= \frac{r^2\theta^r}{\Gamma(r)} \sum_{i=1}^{\infty} \frac{(1-\theta)^i}{i!} \frac{\Gamma(r+i)}{(r+i)^2}, \quad (\text{B.4})$$

em que (B.3) descreve a série hipergeométrica de Gauss (Abramowitz & Stegun, 1970, p. 556).

A variância é obtida por

$$\text{Var}\left(\frac{r}{\mathbf{X}}\right) = \mathbb{E}\left(\frac{r^2}{\mathbf{X}^2}\right) - \left[\mathbb{E}\left(\frac{r}{\mathbf{X}}\right)\right]^2.$$

Ao considerar os resultados de (B.2) e (B.4) tem-se

$$\text{Var}\left(\frac{r}{\mathbf{X}}\right) = \frac{r^2\theta^r}{\Gamma(r)} \sum_{i=1}^{\infty} \frac{(1-\theta)^i}{i!} \frac{\Gamma(r+i)}{(r+i)^2} - \left(\frac{r\theta^r}{\Gamma(r)} \sum_{i=1}^{\infty} \frac{(1-\theta)^i}{i!} \frac{\Gamma(r+i)}{r+i}\right)^2.$$

Os estimadores bayesianos considerados são a média e moda da distribuição *a posteriori*. A média da distribuição *a posteriori* de  $\theta$  é dada por  $\frac{r+\alpha}{\alpha+\beta+\mathbf{X}}$ . Por

conveniência, considere  $\alpha + \beta = \gamma$ . Sua esperança e variância são

$$\begin{aligned} \mathbb{E} \left( \frac{r + \alpha}{\gamma + \mathbf{X}} \right) &= (r + \alpha) \mathbb{E} \left( \frac{1}{\gamma + \mathbf{X}} \right) = (r + \alpha) \sum_{x=r}^{\infty} \frac{1}{\gamma + x} \binom{x-1}{r-1} \theta^r (1-\theta)^{x-r} \\ &= \frac{(r + \alpha) \theta^r}{\gamma + r} \frac{\Gamma(1 + \gamma + r)}{\Gamma(r) \Gamma(r + \gamma)} \sum_{i=1}^{\infty} \frac{(1-\theta)^i}{i!} \frac{\Gamma(r+i) \Gamma(\gamma+r+i)}{\Gamma(1 + \gamma + r + i)}, \end{aligned} \quad (\text{B.5})$$

em que (B.5) descreve a série hipergeométrica de Gauss (Abramowitz & Stegun, 1970, p. 556).

Para calcular a variância da média *a posteriori* é necessário conhecer

$$\begin{aligned} \mathbb{E} \left( \frac{r + \alpha}{\gamma + \mathbf{X}} \right)^2 &= (r + \alpha)^2 \mathbb{E} \left( \frac{1}{\gamma + \mathbf{X}} \right)^2 = (r + \alpha)^2 \sum_{x=r}^{\infty} \left( \frac{1}{\gamma + x} \right)^2 \binom{x-1}{r-1} \theta^r (1-\theta)^{x-r} \\ &= \frac{(r + \alpha)^2 \theta^r}{(\gamma + r)^2} \frac{[\Gamma(1 + \gamma + r)]^2}{\Gamma(r) [\Gamma(r + \gamma)]^2} \sum_{i=1}^{\infty} \frac{(1-\theta)^i}{i!} \frac{\Gamma(r+i) [\Gamma(\gamma+r+i)]^2}{[\Gamma(1 + \gamma + r + i)]^2} \end{aligned} \quad (\text{B.6})$$

$$= \frac{(r + \alpha)^2 \theta^r}{\Gamma(r)} \sum_{i=1}^{\infty} \frac{(1-\theta)^i}{i!} \frac{\Gamma(r+i)}{(\gamma + r + i)^2}, \quad (\text{B.7})$$

em que (B.6) descreve a série hipergeométrica de Gauss (Abramowitz & Stegun, 1970, p. 556).

Assim, pelos resultados (B.5) e (B.7), a variância desse estimador é dada por

$$\begin{aligned} \text{Var} \left( \frac{r + \alpha}{\gamma + \mathbf{X}} \right) &= \mathbb{E} \left( \frac{r + \alpha}{\gamma + \mathbf{X}} \right)^2 - \left[ \mathbb{E} \left( \frac{r + \alpha}{\gamma + \mathbf{X}} \right) \right]^2 \\ &= \frac{(r + \alpha)^2 \theta^r}{\Gamma(r)} \sum_{i=1}^{\infty} \frac{(1-\theta)^i}{i!} \frac{\Gamma(r+i)}{(\gamma + r + i)^2} \\ &\quad - \left( \frac{(r + \alpha) \theta^r}{\gamma + r} \frac{\Gamma(1 + \gamma + r)}{\Gamma(r) \Gamma(r + \gamma)} \sum_{i=1}^{\infty} \frac{(1-\theta)^i}{i!} \frac{\Gamma(r+i) \Gamma(\gamma+r+i)}{\Gamma(1 + \gamma + r + i)} \right)^2. \end{aligned}$$

Para o cálculo da esperança e variância da moda da distribuição *a posteriori*, obtida pela função  $\frac{r + \alpha - 1}{\alpha + \beta + \mathbf{X} - 2}$ , considere  $\psi = \gamma - 2$ . Assim,

$$\begin{aligned} \mathbb{E} \left( \frac{r + \alpha - 1}{\psi + \mathbf{X}} \right) &= (r + \alpha - 1) \mathbb{E} \left( \frac{1}{\psi + \mathbf{X}} \right) = (r + \alpha - 1) \sum_{x=r}^{\infty} \frac{1}{\psi + x} \binom{x-1}{r-1} \theta^r (1-\theta)^{x-r} \\ &= \frac{(r + \alpha - 1) \theta^r}{\psi + r} \frac{\Gamma(1 + \psi + r)}{\Gamma(r) \Gamma(r + \psi)} \sum_{i=1}^{\infty} \frac{(1-\theta)^i}{i!} \frac{\Gamma(r+i) \Gamma(\psi+r+i)}{\Gamma(1 + \psi + r + i)}, \end{aligned} \quad (\text{B.8})$$

em que (B.8) descreve a série hipergeométrica de Gauss (Abramowitz & Stegun, 1970, p. 556).

Analogamente ao cálculo efetuado para a média da distribuição *a posteriori* tem-se que

$$\begin{aligned}
\mathbb{E} \left( \frac{r + \alpha - 1}{\psi + \mathbf{X}} \right)^2 &= (r + \alpha - 1)^2 \mathbb{E} \left( \frac{1}{\psi + \mathbf{X}} \right)^2 \\
&= (r + \alpha - 1)^2 \sum_{x=r}^{\infty} \left( \frac{1}{\psi + x} \right)^2 \binom{x-1}{r-1} \theta^r (1-\theta)^{x-r} \\
&= \frac{(r + \alpha - 1)^2 \theta^r [\Gamma(1 + \psi + r)]^2}{(\psi + r)^2 \Gamma(r) [\Gamma(r + \psi)]^2} \sum_{i=1}^{\infty} \frac{(1-\theta)^i \Gamma(r+i) [\Gamma(\psi + r + i)]^2}{i! [\Gamma(1 + \psi + r + i)]^2} \\
&= \frac{(r + \alpha - 1)^2 \theta^r}{\Gamma(r)} \sum_{i=1}^{\infty} \frac{(1-\theta)^i \Gamma(r+i)}{i! (\psi + r + i)^2}, \tag{B.9}
\end{aligned}$$

em que (B.9) descreve a série hipergeométrica de Gauss (Abramowitz & Stegun, 1970, p. 556).

Então, ao aplicar os resultados (B.8) e (B.9), a variância da moda da distribuição *a posteriori* é

$$\begin{aligned}
\text{Var} \left( \frac{r + \alpha - 1}{\psi + \mathbf{X}} \right) &= \mathbb{E} \left( \frac{r + \alpha - 1}{\psi + \mathbf{X}} \right)^2 - \left[ \mathbb{E} \left( \frac{r + \alpha - 1}{\psi + \mathbf{X}} \right) \right]^2 \\
&= \frac{(r + \alpha - 1)^2 \theta^r}{\Gamma(r)} \sum_{i=1}^{\infty} \frac{(1-\theta)^i \Gamma(r+i)}{i! (\psi + r + i)^2} \\
&\quad - \left( \frac{(r + \alpha - 1) \theta^r}{\psi + r} \frac{\Gamma(1 + \psi + r)}{\Gamma(r) \Gamma(r + \psi)} \sum_{i=1}^{\infty} \frac{(1-\theta)^i \Gamma(r+i) \Gamma(\psi + r + i)^2}{i! \Gamma(1 + \psi + r + i)} \right).
\end{aligned}$$