

Universidade Federal de Uberlândia

Faculdade de Matemática

Curso de Estatística

**ESTIMAÇÃO DE MEDIDAS DE DESEMPENHO DE
TESTES DIAGNÓSTICOS VIA INFERÊNCIA BAYESIANA**

Luana Amâncio Terra

Uberlândia - MG

Dezembro, 2016

Luana Amâncio Terra

**ESTIMAÇÃO DE MEDIDAS DE DESEMPENHO DE
TESTES DIAGNÓSTICOS VIA INFERÊNCIA BAYESIANA**

Trabalho de Conclusão de Curso apresentado à Faculdade de Matemática da Universidade Federal de Uberlândia como requisito para a obtenção do grau de bacharel em Estatística.

Orientador: Prof. Dr. José Waldemar da Silva

Uberlândia - MG

Dezembro, 2016

Luana Amâncio Terra

**ESTIMAÇÃO DE MEDIDAS DE DESEMPENHO DE
TESTES DIAGNÓSTICOS VIA INFERÊNCIA BAYESIANA**

Trabalho de Conclusão de Curso apresentado à Faculdade de Matemática da Universidade Federal de Uberlândia, perante a banca de examinadores abaixo, como requisito para obtenção do Grau de Bacharel em Estatística. Trabalho aprovado dia 01/12/2016.

Prof. Dr. José Waldemar da Silva - UFU
Orientador

Prof. Me. Danilo Guedes Junqueira Junior - UNITRI
Convidado

Prof. Dr. Ednaldo Carvalho Guimarães - UFU
Convidado

Prof. Dr. Janser Moura Pereira - UFU
Convidado

Uberlândia - MG

Dezembro, 2016

Resumo

A utilização de testes diagnósticos para identificar a presença ou ausência de uma determinada doença em um indivíduo é fundamental na área da saúde. O desempenho de testes diagnósticos é medido pela sua sensibilidade e especificidade. Estas medidas são obtidas ao se comparar os resultados do teste diagnóstico com um teste de referência, denominado Padrão Ouro, no qual classifica corretamente os indivíduos doentes e não doentes. Na ausência de um teste Padrão Ouro, as medidas de desempenho do teste precisam ser estimadas por métodos alternativos pois a quantidade de indivíduos doentes na população em estudo é desconhecida. O objetivo deste estudo foi apresentar a metodologia bayesiana proposta por JOSEPH, GYORKOS e COUPAL (1995) para estimar a sensibilidade (S) e a especificidade (E), consideradas medidas de desempenho de um teste diagnóstico, assim como a prevalência (P) da doença na população, nas situações de ausência de um teste Padrão Ouro. Além disso, teve-se como objetivo verificar a influência da distribuição *a priori* informativa, bem como a influência do tamanho amostral, nas estimativas *a posteriori*. Para atender tal objetivo, seis cenários foram considerados. Em três dos cenários, utilizaram-se distribuições *a priori* uniformes para os parâmetros com diferentes tamanhos amostrais, e nos outros três cenários, distribuições *a priori* betas também com diferentes tamanhos amostrais. Para ilustrar a metodologia em uma situação real, em dois dos seis cenários mencionados anteriormente, foram utilizados dados reais oriundos de um experimento no qual realizou-se o teste Cervical Comparativo para detectar a tuberculose bovina. A partir dos resultados obtidos, foi constatado que o tamanho da amostra não exerceu influência expressiva sobre os resultados, mas as estimativas foram mais precisas nos cenários em que foram consideradas distribuições *a priori* betas para os parâmetros S, E e P. Concluiu-se que a metodologia bayesiana é adequada para obter estimativas para os parâmetros de interesse, porém há necessidade de utilizar distribuições *a priori* com alguma informação para que o problema de identificabilidade seja resolvido.

Palavras-chave: *Inferência Bayesiana, Teste Diagnóstico, Sensibilidade, Especificidade.*

Abstract

The use of diagnostic tests to identify the presence or absence of a particular disease in an individual is essential in the area of health. The performance of diagnostic tests is measured by their sensitivity and specificity. These measures are obtained by comparing the results of the diagnostic test with a reference test, called Gold Standard, which correctly classifies both diseased and non-diseased individuals. In the absence of a Gold Standard test, the sensitivity and specificity need to be estimated by alternative methods since the number of diseased individuals in the population is unknown. The objective of the study was to introduce the Bayesian methodology proposed by JOSEPH, GYORKOS and COUPAL (1995) to estimate sensitivity (S), specificity (E) and disease prevalence (P), in the absence of a gold standard test. In addition, we aimed to compare the influence of an informative prior distribution and the sample size on the results. In order to meet the objective, six cases or situations were considered. In three of the cases, prior uniform distributions were used for the parameters with different sample sizes, and in the other three cases, prior beta distributions with different sample sizes. To illustrate the methodology in a real situation, in two of the six cases, mentioned previously, we used real data from an experiment in which the Comparative Cervical Test was performed to detect bovine tuberculosis. From the results obtained, the size of the sample did not have a significant influence on the results, but the estimates were more precise in cases which prior beta distributions were used for the parameters S, E and P. It was concluded that the Bayesian methodology is adequate to estimate the parameters, but there is a need to use prior distributions with some information so that the identifiability problem is solved.

Keywords: *Bayesian Inference, Diagnostic test, Sensibility, Especificity.*

Sumário

1	Introdução	1
2	Metodologia	4
3	Resultados	14
4	Conclusão	20

1 Introdução

O diagnóstico precoce de doenças é fundamental na área da saúde. Historicamente, ao se fazer um diagnóstico, a decisão médica era baseada nos sinais e sintomas do paciente (JAMA, 1992). Visando a melhoria de resultados, novos meios de diagnósticos que demonstrem evidências científicas vêm sendo estudados e desenvolvidos.

Teste diagnóstico é uma ferramenta utilizada para identificar a presença ou ausência de uma determinada doença em um indivíduo submetido ao teste. Além disso, testes diagnósticos são utilizados para avaliar a gravidade do quadro clínico do paciente e monitorar a resposta à uma intervenção clínica.

O teste diagnóstico ideal ou perfeito, podendo ser chamado de Padrão Ouro, é aquele que distingue com exatidão os doentes e os não-doentes, ou seja, o resultado do teste é positivo para a presença da doença e negativo para a ausência da doença. Entretanto, alguns meios de diagnósticos classificados como Padrão Ouro não são acessíveis a todos devido à algumas limitações, como por exemplo, alto custo de aplicação do teste e risco para o paciente (MARKLE, 2007).

Em alguns casos em que não é viável ou possível diagnosticar uma doença através de um teste Padrão Ouro, se faz necessária a substituição por um outro teste diagnóstico. Nesses casos, o teste Padrão Ouro pode ser utilizado como um teste de referência e tem a finalidade de avaliar o desempenho de outros testes.

As medidas de desempenho de um teste são definidas como sensibilidade (S) e especificidade (E). A sensibilidade corresponde à probabilidade de um teste diagnóstico identificar corretamente um indivíduo doente e a especificidade, à probabilidade de um teste diagnóstico identificar corretamente um indivíduo não doente.

Outras medidas importantes como a prevalência (P) da doença, o valor preditivo positivo (VPP) e o valor preditivo negativo (VPN) podem ser obtidas a partir de testes diagnósticos.

O VPP é a probabilidade do indivíduo estar doente dado que o resultado do teste foi positivo e o VPN é a probabilidade do indivíduo não estar doente dado que o resultado do teste foi negativo.

Os valores preditivos de um teste são determinados pelas medidas de desempenho (sensibilidade e especificidade) e pela prevalência da doença na população em estudo. É importante notar que o valor preditivo estimado para uma determinada população não

pode ser adotado por outra população sem considerar a prevalência da doença.

É desejável que os testes diagnósticos apresentem resultados satisfatórios pois os profissionais baseiam-se nesses resultados para decidir sobre o diagnóstico de alguma determinada doença. Para exemplificar, um teste com alta sensibilidade pode ser utilizado para diagnosticar doenças potencialmente graves e que precisam de tratamento imediato. Por outro lado, um teste com alta especificidade pode ser utilizado quando um resultado falso positivo pode ser lesivo. Portanto, é de grande importância a estimação adequada das medidas de desempenho de um teste.

Na ausência de um teste Padrão Ouro, não é possível obter os valores para a sensibilidade e especificidade pois a quantidade de indivíduos doentes na população em estudo é desconhecida. A sensibilidade seria determinada pela razão entre o número de indivíduos doentes com resultado positivo e o total de indivíduos doentes, enquanto a especificidade seria determinada pela razão entre o número de indivíduos não doentes com resultado negativo e o número total de indivíduos não doentes. Portanto, as medidas de desempenho do teste e a prevalência da doença precisam ser estimadas por métodos alternativos.

Na inferência clássica, os estimadores dos parâmetros são frequentemente obtidos através do método de máxima verossimilhança, enquanto na inferência bayesiana consideram-se, além da informação dos dados expressa através da função verossimilhança, outras informações adicionais como dados históricos ou opiniões de especialistas sobre o parâmetro, em outras palavras, informações prévias.

Para estimar os parâmetros de interesse P , S e E por meio do método da máxima verossimilhança é necessário a utilização de procedimentos numéricos e/ou restrições nos parâmetros da população, haja vista que a solução não é única. Outra possibilidade é a abordagem bayesiana proposta por JOSEPH, GYORKOS e COUPAL (1995). Na análise bayesiana as restrições são definidas por meio de distribuições *a priori* para os parâmetros, que combinadas com a função de verossimilhança através do Teorema de Bayes, gera a distribuição conjunta *a posteriori*. Toda inferência bayesiana é realizada a partir desta distribuição.

Em geral, a distribuição conjunta *a posteriori* tem forma complexa, tornando-se difícil a avaliação das quantidades de interesse. Usualmente, este problema é transferido para o caso univariado e, assim, solucionado calculando-se a distribuição marginal para cada parâmetro.

No entanto, a abordagem univariada não contorna a dificuldade de integração. Deste modo, métodos numéricos iterativos como a amostragem Gibbs (GELFAND et al., 1990) podem ser utilizados para aproximar a distribuição de interesse a partir das distribuições condicionais completas a *posteriori*. O Amostrador de Gibbs é um exemplo de algoritmo que utiliza o método de Monte Carlo via Cadeia de Markov ou, em inglês, *Markov Chain Monte Carlo* (MCMC).

A implementação do algoritmo Gibbs consiste em amostrar recursivamente e iterativamente de cada distribuição condicional completa a *posteriori*. Para efetuar a amostragem, é necessário que existam funções prontas, ou seja, já implementadas no software utilizado. Caso isso não ocorra, deve-se utilizar outras formas de amostragem, geralmente baseadas no algoritmo Metropolis-Hastings (METROPOLIS et al., 1953 e HASTINGS, 1970).

A distribuição estacionária atingida pelo algoritmo MCMC, implementada a partir das distribuições condicionais completas, é a distribuição conjunta a *posteriori*. Além disso, as cadeias geradas individualmente para cada parâmetro representam amostras de suas respectivas distribuições marginais, desde que a condição de estacionariedade seja atendida. Logo, há a necessidade de verificar a convergência do algoritmo.

Dentre os vários métodos de convergência existentes, podem ser citados os critérios de Geweke (GEWEKE, 1992) e Gelman e Rubin (GELMAN & RUBIN, 1992).

A ideia no método de Geweke é verificar a hipótese de igualdade, em média, da parte inicial da cadeia com a parte final. A padronização da diferença entre as médias, sob determinadas condições, segue uma distribuição normal padrão. Escores grandes, em módulo, levariam à rejeição da hipótese de igualdade e, assim, haveria indícios de falta de convergência.

Já no método de Gelman e Rubin, a ideia é comparar a variação total com a variação dentro das cadeias. Se houver convergência, a razão entre estas duas variações deveria ser um valor próximo de um, uma vez que a distribuição de equilíbrio seria atingida sempre. Portanto, as amostras obtidas, mesmo em momentos diferentes, seriam similares.

Não é objetivo, neste trabalho, apresentar critérios de convergências de forma detalhada. Uma avaliação criteriosa destes métodos de convergência pode ser encontrada em NOGUEIRA et al. (2004).

Análises gráficas também são utilizadas para diagnosticar convergência. Neste caso, observa-se o comportamento dos valores amostrados ao longo das iterações por meio de

um gráfico de linha no qual é possível identificar a não estacionariedade, ou, por meio de gráficos que mostram a similaridade de uma cadeia quando diferentes intervalos da mesma são utilizados.

Na ausência do teste Padrão Ouro, a inferência bayesiana é uma boa metodologia para estimar as medidas de desempenho de um teste diagnóstico e da prevalência da doença na população, haja vista que para todas as quantidades não observáveis, como os parâmetros P , S e E e as variáveis latentes (Y_1) , número de indivíduos doentes com resultado positivo, e (Y_2) , número de indivíduos doentes com resultado negativo, são atribuídas distribuições de probabilidade. Em outras palavras, estas quantidades são tratadas ou estudadas probabilisticamente. O Teste Cervical Comparativo utilizado para diagnosticar a tuberculose bovina (BRASIL, 2006), uma doença de caráter zoonótico, é um teste que pode ser avaliado utilizando-se a metodologia bayesiana.

Alguns exemplos de testes diagnósticos que já foram avaliados utilizando a metodologia bayesiana são o Teste Hemocult SENSE para o câncer colorretal (WALTER et al., 1991), o Ensaio Imunoenzimático (ELISA) para o diagnóstico da infecção pelo HIV (ALVORD et al., 1988) e testes de triagem para câncer de mama (CHAMBERLAIN et al., 1975).

Os objetivos deste trabalho foram mostrar como a metodologia bayesiana pode ser aplicada para estimar as medidas de desempenho de um teste diagnóstico e a prevalência da doença na população, na ausência de um teste Padrão Ouro, e verificar a influência da informação *a priori* e do tamanho amostral nas estimativas *a posteriori*.

2 Metodologia

Os resultados de um teste diagnóstico podem ser organizados conforme a Tabela 1,

Tabela 1: Possíveis resultados em um teste de diagnóstico.

Resultado do Teste	Padrão Ouro		Total
	Doente (D)	Não Doente (D^c)	
Positivo (Pos)	Y_1	$n_1 - Y_1$	n_1
Negativo (Neg)	Y_2	$n_2 - Y_2$	n_2
Total	$Y_1 + Y_2$	$N - (Y_1 + Y_2)$	N

em que, n_1 e n_2 são, respectivamente, as quantidades de resultados positivos e negativos obtidos a partir do teste diagnóstico, e $(Y_1 + Y_2)$ e $N - (Y_1 + Y_2)$ são, respectivamente, as quantidades de indivíduos doentes e não doentes para o teste Padrão Ouro. Outras

medidas importantes apresentadas na Tabela 1 são Y_1 , Y_2 , $n_1 - Y_1$ e $n_2 - Y_2$, as quais são definidas como Verdadeiro Positivo, Falso Negativo, Falso Positivo e Verdadeiro Negativo, respectivamente. As variáveis Y_1 e Y_2 são as mesmas apresentadas na seção anterior.

A partir dos resultados da Tabela 1, poder-se-ia obter as medidas de desempenho, sensibilidade e especificidade, do teste diagnóstico em avaliação através das expressões (1) e (2).

$$S = \frac{Y_1}{Y_1 + Y_2} \quad (1)$$

$$E = \frac{n_2 - Y_2}{N - (Y_1 + Y_2)} \quad (2)$$

Outras medidas como a prevalência, o valor preditivo positivo e o valor preditivo negativo também podem ser obtidas. Estas três medidas são apresentadas, respectivamente, nas expressões (3), (4) e (5).

$$P = \frac{Y_1 + Y_2}{N} \quad (3)$$

$$VPP = \frac{Y_1}{n_1} \quad (4)$$

$$VPN = \frac{n_2 - Y_2}{n_2} \quad (5)$$

Na ausência do teste Padrão Ouro, as variáveis aleatórias n_1 e n_2 são observáveis, sendo $N = n_1 + n_2$, porém as variáveis Y_1 e Y_2 , denominadas variáveis latentes, são não observáveis. Dado o exposto, as quantidades apresentadas anteriormente, (1), (2), (3), (4) e (5), não podem ser calculadas diretamente a partir dos dados amostrais.

As medidas de desempenho do teste, a prevalência da doença e os valores preditivos são representados em termos de probabilidade por:

$$S = P(Pos|D) \quad (6)$$

$$E = P(Neg|D^c) \quad (7)$$

$$P = P(D) \quad (8)$$

$$VPP = P(D|Pos) \quad (9)$$

$$VPN = P(D^c|Neg) \quad (10)$$

As probabilidades em (9) e (10) podem ser determinadas a partir das probabilidades em (6), (7) e (8), as quais correspondem aos parâmetros de desempenho do teste e à prevalência. Sob o enfoque bayesiano a estimação destes parâmetros é feita a partir de distribuições de probabilidade determinadas para os mesmos e não por meio das expressões (1), (2) e (3).

Toda a inferência bayesiana é realizada a partir da distribuição conjunta *a posteriori* e, esta, é determinada a partir do Teorema de Bayes, o qual possibilita “combinar” a informação dos dados expressa por meio da função de verossimilhança com a informação prévia sobre as quantidades de interesse expressa através da distribuição *a priori*. Portanto, é necessário definir a função de verossimilhança e a distribuição *a priori*.

Com o intuito de especificar a função de verossimilhança, observa-se que as medidas de interesse podem ser determinadas a partir de probabilidades de eventos descritos na Tabela 1. Considerando que os dados nesta tabela estão organizados em quatro caselas e que a probabilidade de sucesso em cada uma delas é como a seguir,

$$p_1 = P(Pos \cap D) = P(Pos|D)P(D) = SP ,$$

$$p_2 = P(Neg \cap D) = P(Neg|D)P(D) = [1 - P(Pos|D)]P(D) = (1 - S)P ,$$

$$p_3 = P(Pos \cap D^c) = P(Pos|D^c)P(D^c) = [1 - P(Neg|D^c)][1 - P(D)] = (1 - E)(1 - P) ,$$

$$p_4 = P(Neg \cap D^c) = P(Neg|D^c)P(D^c) = E[(1 - P(D))] = E(1 - P) ,$$

e, considerando ainda que os resultados das caselas são mutuamente exclusivos, que $Y_1 + Y_2 + (n_1 - Y_1) + (n_2 - Y_2) = N$ e que $\sum_{i=1}^4 p_i = 1$, conclui-se que os dados, ou seja, o número de resultados nas caselas, seguem uma distribuição Multinomial com parâmetros N e p , em que $p = [p_1, p_2, p_3, p_4]$. Com isso a distribuição de probabilidade Multinomial é dada por:

$$P(Y_1, Y_2, n_1 - Y_1, n_2 - Y_2) = \frac{N!}{Y_1!Y_2!(n_1 - Y_1)!(n_2 - Y_2)!} \times p_1^{Y_1} p_2^{Y_2} p_3^{n_1 - Y_1} p_4^{n_2 - Y_2} \quad (11)$$

Neste caso, a função de verossimilhança corresponde à distribuição Multinomial identificada, a qual é apresentada em (12):

$$\begin{aligned} L(\theta|N, n_1, n_2, Y_1, Y_2) &= \frac{N!}{Y_1!Y_2!(n_1 - Y_1)!(n_2 - Y_2)!} \\ &\times (SP)^{Y_1} [(1 - S)P]^{Y_2} [(1 - E)(1 - P)]^{n_1 - Y_1} [E(1 - P)]^{n_2 - Y_2} \\ &= \frac{N!}{Y_1!Y_2!(n_1 - Y_1)!(n_2 - Y_2)!} \\ &\times P^{Y_1 + Y_2} (1 - P)^{N - Y_1 - Y_2} S^{Y_1} (1 - S)^{Y_2} E^{n_2 - Y_2} (1 - E)^{n_1 - Y_1}, \end{aligned} \quad (12)$$

em que $\theta = [S, E, P]$.

Conforme TANNER e WONG (1987), o aumento de dados com o uso de variáveis latentes facilita a obtenção de distribuições *a posteriori*. Fundamentado nas informações sobre dados aumentados, os valores para Y_1 e Y_2 podem ser gerados a partir de suas distribuições condicionais, sendo elas:

$$Y_1|n_1, p_5 \sim \text{Binomial}(n_1, p_5) \quad \text{e}, \quad Y_2|n_2, p_6 \sim \text{Binomial}(n_2, p_6) \quad (13)$$

com

$$p_5 = P(D|Pos) = \frac{P(Pos|D)P(D)}{P(Pos|D)P(D) + P(Pos|D^c)P(D^c)} = \frac{PS}{PS + (1-P)(1-E)}$$

e,

$$p_6 = P(D|Neg) = \frac{P(Neg|D)P(D)}{P(Neg|D)P(D) + P(Neg|D^c)P(D^c)} = \frac{P(1-S)}{P(1-S) + (1-P)(E)}$$

Os parâmetros P, S e E são definidos no intervalo de 0 a 1, então uma distribuição a *priori* possível para esses parâmetros é a beta. Assumindo independência entre eles, a distribuição conjunta a *priori* pode ser escrita de forma proporcional, como a seguir:

$$\pi(P, S, E) = \pi(P)\pi(S)\pi(E) \propto P^{\alpha_p-1}(1-P)^{\beta_p-1} \times S^{\alpha_s-1}(1-S)^{\beta_s-1} \times E^{\alpha_e-1}(1-E)^{\beta_e-1} \quad (14)$$

Na distribuição conjunta a *priori* apresentada em (14), α_p , β_p , α_s , β_s , α_e , β_e são os parâmetros das distribuições a *priori* ou hiperparâmetros .

Assim, a distribuição conjunta a *posteriori*, que é proporcional ao produto da função de verossimilhança e da distribuição conjunta a *priori*, será como segue:

$$\begin{aligned} \pi(P, S, E|N, n_1, n_2, Y_1, Y_2) &\propto \frac{N!}{Y_1!Y_2!(n_1-Y_1)!(n_2-Y_2)!} \\ &\times P^{Y_1+Y_2}(1-P)^{N-Y_1-Y_2} S^{Y_1}(1-S)^{Y_2} E^{n_2-Y_2}(1-E)^{n_1-Y_1} \\ &\times P^{\alpha_p-1}(1-P)^{\beta_p-1} \times S^{\alpha_s-1}(1-S)^{\beta_s-1} \times E^{\alpha_e-1}(1-E)^{\beta_e-1} \\ &\propto \frac{N!}{Y_1!Y_2!(n_1-Y_1)!(n_2-Y_2)!} P^{Y_1+Y_2+\alpha_p-1}(1-P)^{N-Y_1-Y_2+\beta_p-1} \\ &\times S^{Y_1+\alpha_s-1}(1-S)^{Y_2+\beta_s-1} E^{n_2-Y_2+\alpha_e-1}(1-E)^{n_1-Y_1+\beta_e-1} \end{aligned} \quad (15)$$

Pelo Teorema de Bayes, a expressão (15) teria como denominador a integral tripla em

relação a P , S e E ,

$$\int_0^1 \int_0^1 \int_0^1 L(\theta|N, n_1, n_2, Y_1, Y_2) \pi(P, S, E) dP dS dE \quad (16)$$

porém esta foi desconsiderada pois seu resultado funciona apenas como uma constante normalizadora.

A amostragem Gibbs pode ser utilizada para obter amostras das quantidades não observáveis, P , S , E , Y_1 e Y_2 em (15), sendo necessário para tal, o conhecimento das distribuições condicionais completas a *posteriori* para cada uma destas quantidades. Apresenta-se, como demonstração, os cálculos para encontrar a distribuição condicional completa a *posteriori* para o parâmetro P :

Pelo Teorema de Bayes, temos que:

$$\begin{aligned} \pi(P|N, n_1, n_2, Y_1, Y_2, S, E, \alpha_p, \beta_p) &\propto \frac{\pi(P, S, E|N, n_1, n_2, y_1, y_2)}{\int_0^1 \pi(P, S, E|N, n_1, n_2, y_1, y_2) dP} \\ &\propto \frac{P^{Y_1+Y_2+\alpha_p-1} (1-P)^{N-Y_1-Y_2+\beta_p-1}}{\int_0^1 P^{Y_1+Y_2+\alpha_p-1} (1-P)^{N-Y_1-Y_2+\beta_p-1} dP} \quad (17) \\ &\propto \frac{P^{Y_1+Y_2+\alpha_p-1} (1-P)^{N-Y_1-Y_2+\beta_p-1}}{B(Y_1+Y_2+\alpha_p, N-Y_1-Y_2+\beta_p)} \end{aligned}$$

em que $B(Y_1+Y_2+\alpha_p, N-Y_1-Y_2+\beta_p)$ representa uma função beta com argumentos $Y_1+Y_2+\alpha_p$ e $N-Y_1-Y_2+\beta_p$. Portanto, a distribuição condicional completa a *posteriori* para P é uma distribuição beta com parâmetros $(Y_1+Y_2+\alpha_p, N-Y_1-Y_2+\beta_p)$.

De forma similar encontram-se as distribuições condicionais completas a *posteriori* para os parâmetros S e E .

As distribuições condicionais completas a *posteriori* para Y_1 , Y_2 , P , S , E , são apresentadas pelas expressões (18), (19), (20), (21) e (22), respectivamente.

$$Y_1|P, S, E, n_1 \sim \text{Binomial} \left(n_1, \frac{PS}{PS + (1-P)(1-E)} \right) \quad (18)$$

$$Y_2|P, S, E, n_2 \sim \text{Binomial} \left(n_2, \frac{P(1-S)}{P(1-S) + (1-P)E} \right) \quad (19)$$

$$P|Y_1, Y_2, \alpha_p, \beta_p, n_1, n_2 \sim \text{Beta} (Y_1 + Y_2 + \alpha_p, n_1 + n_2 - Y_1 - Y_2 + \beta_p) \quad (20)$$

$$S|Y_1, Y_2, \alpha_s, \beta_s \sim \text{Beta} (Y_1 + \alpha_s, Y_2 + \beta_s) \quad (21)$$

$$E|Y_1, Y_2, \alpha_e, \beta_e, n_1, n_2 \sim \text{Beta} (n_2 - Y_2 + \alpha_e, n_1 - Y_1 + \beta_e) \quad (22)$$

Escolhido os parâmetros $P^{(0)}$, $S^{(0)}$ e $E^{(0)}$ como valores iniciais arbitrários, o algoritmo de Gibbs é dado por:

1. Na etapa m , geram-se valores para as variáveis não observáveis Y_1 e Y_2 :

$$Y_1^{(m)} \sim \text{binomial} \left(n_1, \frac{P^{(m)}S^{(m)}}{P^{(m)}S^{(m)} + (1-P)^{(m)}(1-E)^{(m)}} \right)$$

$$Y_2^{(m)} \sim \text{binomial} \left(n_2, \frac{P^{(m)}(1-S)^{(m)}}{P^{(m)}(1-S)^{(m)} + (1-P)^{(m)}E^{(m)}} \right)$$

2. Na etapa $m + 1$, geram-se valores para P , S e E usando os valores de Y_1 e Y_2 amostrados na etapa anterior:

$$P^{(m+1)} | Y_1^{(m)}, Y_2^{(m)} \sim \text{Beta}(Y_1^{(m)} + Y_2^{(m)} + \alpha_p, n_1 + n_2 - Y_1^{(m)} - Y_2^{(m)} + \beta_p)$$

$$S^{(m+1)} | Y_1^{(m)}, Y_2^{(m)} \sim \text{Beta}(Y_1^{(m)} + \alpha_s, Y_2^{(m)} + \beta_s)$$

$$E^{(m+1)} | Y_1^{(m)}, Y_2^{(m)} \sim \text{Beta}(n_1 - Y_2^{(m)} + \alpha_e, n_1 - Y_1^{(m)} + \beta_e)$$

3. A amostra é gerada a partir de um grande número de repetições das etapas, onde a cada iteração utiliza-se os valores amostrados da iteração anterior.

Obtendo estimativas para P , S e E , os valores preditivos positivo e negativo podem ser determinados, inclusive em cada iteração do algoritmo Gibbs, pois eles se relacionam conforme apresentado em (23) e (24).

$$VPP = \frac{PS}{(PS) + (1-P)(1-E)} \quad (23)$$

$$VPN = \frac{(1 - P)E}{(1 - P)E + P(1 - S)} \quad (24)$$

Conforme visto anteriormente, a informação prévia é alocada na distribuição *a priori* e a especificação desta depende da determinação dos valores de seus parâmetros ou hiperparâmetros. Uma possível estratégia para o estabelecimento dos hiperparâmetros é por meio de estimativas intervalares prévias, indicadas por pesquisadores da área ainda que informalmente. Admitindo o ponto médio do intervalo como valor esperado *a priori* para o respectivo parâmetro, o qual tem distribuição beta com parâmetros α e β , então pode-se escrever,

$$\frac{L_1 + L_2}{2} = \frac{\alpha}{\alpha + \beta} \quad (25)$$

em que L_1 e L_2 são os valores que delimitam o intervalo inferiormente e superiormente, respectivamente.

Em (25) tem-se uma equação e duas incógnitas e logo a solução não é única, mas adotando uma estimativa conservadora para o desvio padrão igual a um quarto da amplitude do intervalo (TRIOLA, 2005), a equação (26) pode ser estabelecida e os hiperpâmetros determinados.

$$\frac{L_2 - L_1}{4} = \sqrt{\frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}} \quad (26)$$

Para atender os objetivos deste trabalho, seis cenários foram considerados. Em três deles, utilizaram-se distribuições *a priori* uniformes para os parâmetros com diferentes tamanhos amostrais, e nos outros três, distribuições *a priori* betas também com diferentes tamanhos amostrais. Para ilustrar a metodologia em uma situação real, em dois dos seis cenários mencionados anteriormente, foram utilizados dados reais oriundos de um experimento no qual realizou-se o teste Cervical Comparativo para detectar a tuberculose bovina (BRASIL, 2006). Foram utilizados 68 bovinos de aptidão leiteira provenientes de uma propriedade rural localizada na mesorregião do Triângulo Mineiro, Minas Gerais.

Todos animais estavam em lactação e tinham idade superior a 24 meses. Dentre os 68 animais, 4 apresentaram resultado positivo e 64 apresentaram resultado negativo. Para escolha dos hiperparâmetros para as distribuições *a priori*, foram considerados intervalos indicados por pesquisadores da área, sendo eles: $[0,0028 ; 0,0389]$ para a prevalência, $[0,7 ; 0,84]$ para a sensibilidade e $[0,92 ; 0,98]$ para a especificidade.

A influência da informação prévia e do tamanho amostral pode ser verificada por meio da precisão das estimativas *a posteriori*. A precisão afeta diretamente a amplitude da estimativa intervalar, isto é, quanto menor a amplitude do intervalo estimado, maior será a precisão. Outras ferramentas que também podem ser utilizadas para comparar os cenários são o Critério de Informação de Akaike Corrigido (AICc) (BOZDOGAN, 1987),

$$AICc = -2\log L(\hat{\theta}) + 2k \left(\frac{n}{n - k - 1} \right) \quad (27)$$

e o Critério de Informação Bayesiano (BIC) (SCHWARZ, 1978),

$$BIC = -2\log L(\hat{\theta}) + p\log(n). \quad (28)$$

Na expressão (27), k é a quantidade de parâmetros no modelo, n é o tamanho da amostra e $L(\hat{\theta})$ é a verossimilhança calculada utilizando uma estimativa de θ . Na expressão (28), n e $L(\hat{\theta})$ são definidos como em (27). Um modelo, dentre dois em comparação, será melhor se apresentar menor AICc ou menor BIC, dependendo do critério em uso.

Os seis cenários mencionados anteriormente estão descritos a seguir:

Cenário 1: Dados reais com 4 resultados positivos (n_1) e 64 resultados negativos (n_2), considerando distribuição *a priori* beta com hiperparâmetros $\beta = 1$ e $\alpha = 1$ para a sensibilidade, especificidade e prevalência da doença.

Cenário 2: Dados reais com 4 resultados positivos (n_1) e 64 resultados negativos (n_2), considerando distribuição *a priori* beta com hiperparâmetros $\beta_p = 244,4$ e $\alpha_p = 5,2$ para a prevalência, $\beta_s = 33,0$ e $\alpha_s = 110,5$ para a sensibilidade, e $\beta_e = 10,5$ e $\alpha_e = 199,6$ para a especificidade.

Cenário 3: Dados simulados com 40 resultados positivos (n_1) e 640 resultados negativos (n_2), considerando distribuição a *priori* beta com hiperparâmetros $\beta = 1$ e $\alpha = 1$ para a sensibilidade, especificidade e prevalência da doença.

Cenário 4: Dados simulados com 40 resultados positivos (n_1) e 640 resultados negativos (n_2), considerando distribuição a *priori* beta com hiperparâmetros $\beta_p = 244,4$ e $\alpha_p = 5,2$ para a prevalência, $\beta_s = 33,0$ e $\alpha_s = 110,5$ para a sensibilidade, e $\beta_e = 10,5$ e $\alpha_e = 199,6$ para a especificidade.

Cenário 5: Dados simulados com 400 resultados positivos (n_1) e 6400 resultados negativos (n_2), considerando distribuição a *priori* beta com hiperparâmetros $\beta = 1$ e $\alpha = 1$ para a sensibilidade, especificidade e prevalência da doença.

Cenário 6: Dados simulados com 400 resultados positivos (n_1) e 6400 resultados negativos (n_2), considerando distribuição a *priori* beta com hiperparâmetros $\beta_p = 244,4$ e $\alpha_p = 5,2$ para a prevalência, $\beta_s = 33,0$ e $\alpha_s = 110,5$ para a sensibilidade, e $\beta_e = 10,5$ e $\alpha_e = 199,6$ para a especificidade.

Os tamanhos amostrais dos casos simulados foram escolhidos propositalmente de forma a manter a proporção de resultados positivos do caso real.

Os hiperparâmetros dos cenários 2, 4 e 6, ditos casos com prioris informativas, foram determinados conforme metodologia apresentada anteriormente na página 11. As distribuições a *priori* não informativas são as distribuições betas com hiperparâmetros $\beta = 1$ e $\alpha = 1$, as quais correspondem à distribuições uniformes definidas no intervalo de 0 a 1.

Utilizou-se o amostrador de Gibbs para obter amostras de cada parâmetro nos seis cenários. Para a realização das inferências, foi considerada uma amostra final de tamanho 9.000 obtida a partir de uma cadeia com 1.000.000 iterações, em que as 100.000 primeiras foram descartadas para a eliminação do efeito do valor inicial arbitrário e armazenado um valor a cada 100 para que estes fossem não correlacionados. A convergência foi verificada para todos os parâmetros através dos critérios de convergência Geweke (1992) e Gelman e Rubin (1992). Todas as análises foram realizadas com o auxílio do software R (R

DEVELOPMENT CORE TEAM, 2016).

Os resultados para cada medida de interesse serão apresentados por meio de médias, intervalos de máxima densidade *a posteriori*, em inglês, *Highest Posterior Density* (HPD), e por meio de gráficos de densidades. O intervalo HPD é o menor intervalo possível para uma dada probabilidade *a posteriori*.

3 Resultados

As estimativas são consideradas válidas se houver convergência do algoritmo Gibbs. Por meio dos critérios de Geweke e Gelman e Rubin, verificou-se convergência para as variáveis latentes Y_1 e Y_2 , e para os parâmetros P , S e E , conforme apresentado na Tabela 2.

Tabela 2: Resultados dos critérios de convergência de Geweke e Gelman e Rubin (1992).

Critério de Convergência		Y_1	Y_2	P	S	E
Cenário 1	Geweke ¹	1,3981	0,2061	0,7073	0,1099	1,2271
	Gelman ²	1	1	1	1	1
Cenário 2	Geweke	0,6286	1,4233	0,3209	-1,7002	0,8452
	Gelman	1	1	1	1	1
Cenário 3	Geweke	1,02754	0,04241	0,55318	1,06840	-1,52521
	Gelman	1	1	1	1	1
Cenário 4	Geweke	0,88633	-0,06158	0,47174	1,24134	-0,54046
	Gelman	1	1	1	1	1
Cenário 5	Geweke	0,3019	-1,3262	0,8381	1,3751	1,1788
	Gelman	1	1	1	1	1
Cenário 6	Geweke	-1,5352	-0,6619	-1,3035	-0,9775	-0,8734
	Gelman	1	1	1	1	1

¹A hipótese de convergência não é rejeitada se o valor da estatística (z) pertencer ao intervalo $[-1,96; 1,96]$.

²A hipótese de convergência não é rejeitada se o limite superior é próximo de 1.

Gráficos de linha, também, foram utilizados para diagnosticar a convergência dos parâmetros. Neste caso, a convergência para P , S e E , nos seis cenários, foi verificada devido a ocorrência dos valores amostrados em intervalos constantes ao longo das iterações, conforme Figuras 1, 2 e 3.

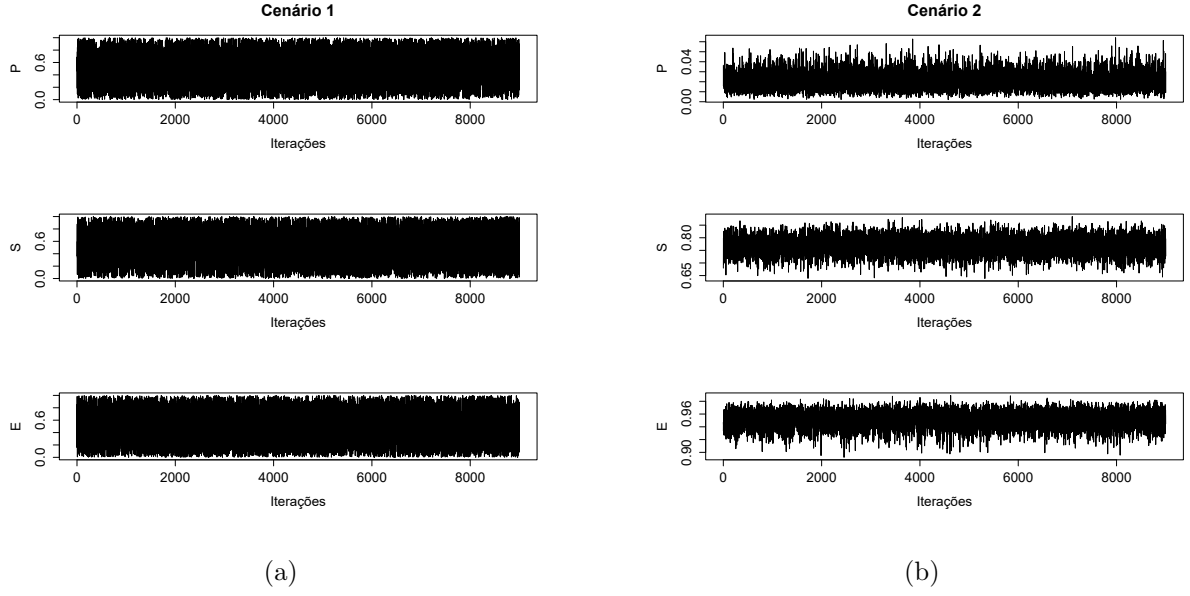


Figura 1: Comportamento das trajetórias das cadeias ao longo das iterações para os parâmetros P , S e E para os cenários 1 e 2.

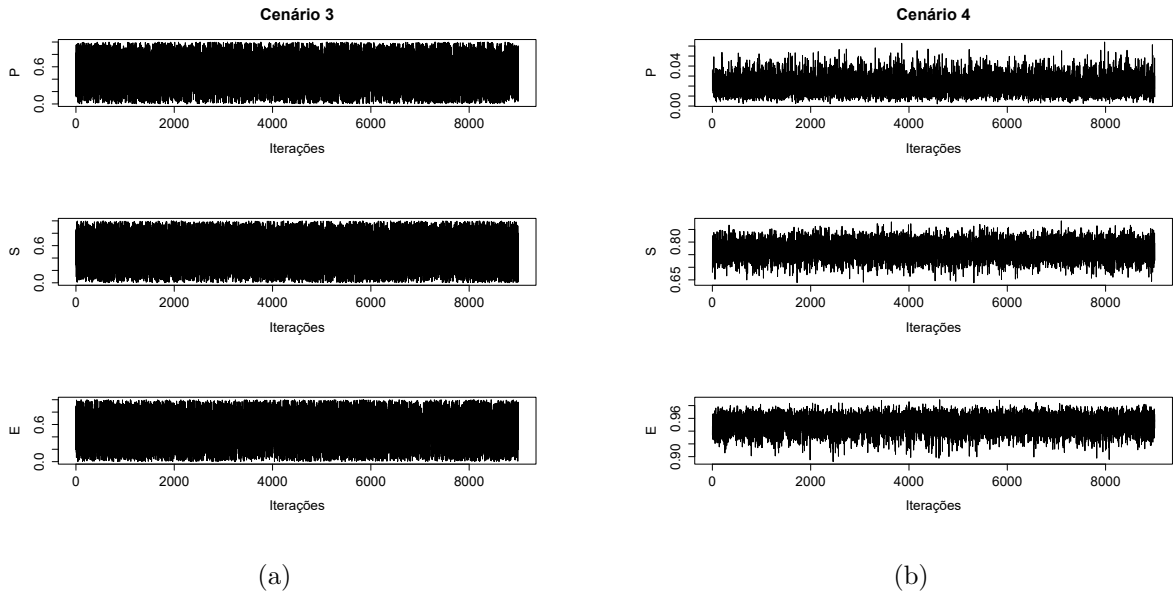


Figura 2: Comportamento das trajetórias das cadeias ao longo das iterações para os parâmetros P , S e E para os cenários 3 e 4.

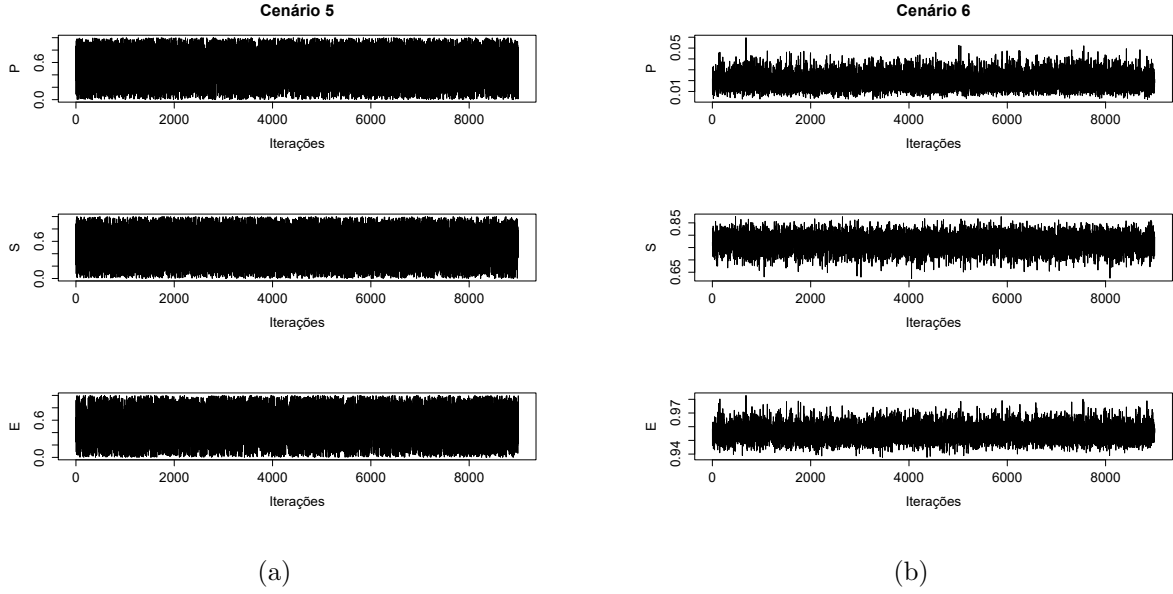


Figura 3: Comportamento das trajetórias das cadeias ao longo das iterações para os parâmetros P, S e E para os cenários 5 e 6.

As estimativas obtidas para cada cenário estão demonstradas nas Tabelas 3, 4, 5, 6, 7 e 8. Além disso, os gráficos das densidades estimadas para os parâmetros P, S e E são apresentados nas Figuras 4, 5 e 6.

Tabela 3: Estimativas pontual e intervalar *a posteriori*, para o cenário 1.

Parâmetro	Média	HPD	
		Limite Inferior	Limite Superior
y1	2,0064	0,0000	4,0000
y2	31,8011	0,0000	63,0000
P	0,5033	0,0415	0,9901
S	0,4936	0,0097	0,9566
E	0,4967	0,0391	0,9858

Tabela 4: Estimativas pontual e intervalar *a posteriori*, para o cenário 2.

Parâmetro	Média	HPD	
		Limite Inferior	Limite Superior
y1	0,9999	0,0000	3,0000
y2	0,3335	0,0000	1,0000
P	0,0207	0,0053	0,0379
S	0,7699	0,6999	0,8356
E	0,9511	0,9251	0,9758
VPP	0,2517	0,0757	0,4410
VPN	0,9949	0,9901	0,9988

Tabela 5: Estimativas pontual e intervalar a *posteriori*, para o cenário 3.

Parâmetro	Média	HPD	
		Limite Inferior	Limite Superior
y1	19,8837	0,0000	39,0000
y2	319,2620	0,0000	624,0000
P	0,5000	0,0126	0,9606
S	0,5022	0,0543	0,9984
E	0,4997	0,0339	0,9823

Tabela 6: Estimativas pontual e intervalar a *posteriori*, para o cenário 4.

Parâmetro	Média	HPD	
		Limite Inferior	Limite Superior
y1	10,2365	1,0000	19,0000
y2	3,1825	0,0000	7,0000
P	0,0201	0,0054	0,0360
S	0,7701	0,7037	0,8364
E	0,9541	0,9369	0,9711

Tabela 7: Estimativas pontual e intervalar a *posteriori*, para o cenário 5.

Parâmetro	Média	HPD	
		Limite Inferior	Limite Superior
y1	197,0215	0,0000	391,0000
y2	3222,4931	0,0000	6246,0000
P	0,4962	0,0487	0,9984
S	0,5024	0,0497	0,9983
E	0,4993	0,0026	0,9524

Tabela 8: Estimativas pontual e intervalar a *posteriori*, para o cenário 6.

Parâmetro	Média	HPD	
		Limite Inferior	Limite Superior
y1	102,6739	27,0000	181,0000
y2	30,9349	6,0000	58,0000
P	0,0197	0,0062	0,0348
S	0,7694	0,7034	0,8421
E	0,9553	0,9446	0,9684

As Tabelas 3 e 4 apresentam as estimativas pontuais e intervalares das quantidades para os cenários 1 e 2, respectivamente, nos quais as amostras possuem o mesmo tamanho (64 resultados negativos e 4 resultados positivos). No cenário 1, em que consideram-se distribuições a *priori* não informativas para a sensibilidade, especificidade e a prevalência, observou-se que as estimativas intervalares abrangem quase todo o domínio dos parâmetros e das variáveis latentes indicando forte falta de precisão. Para Y_2 , por exemplo,

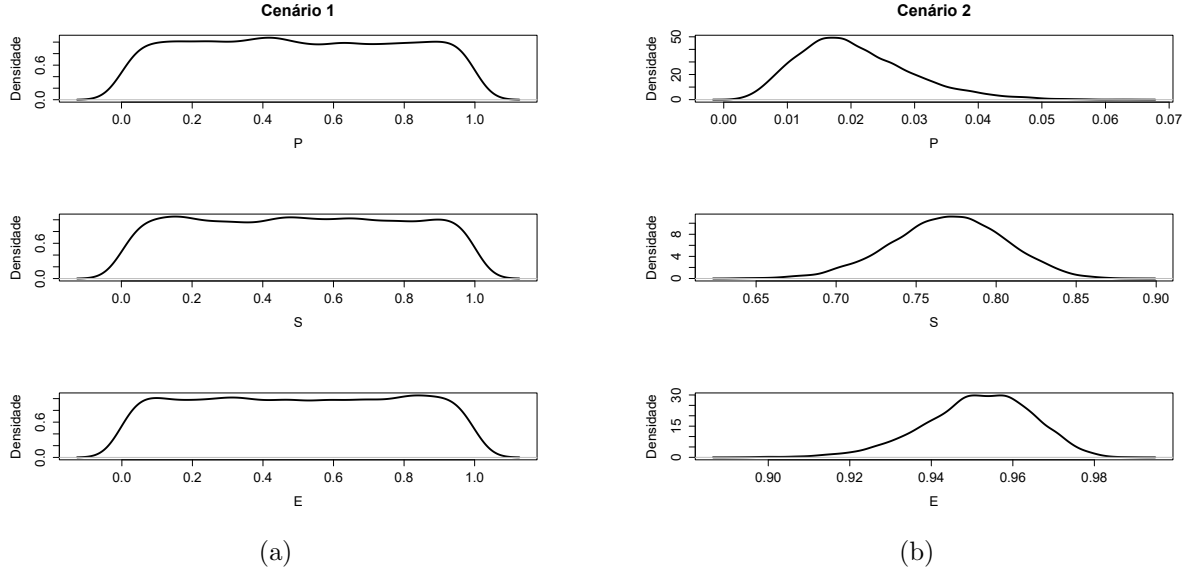


Figura 4: Gráficos das densidades estimadas para os parâmetros P, S e E para os cenários 1 e 2.

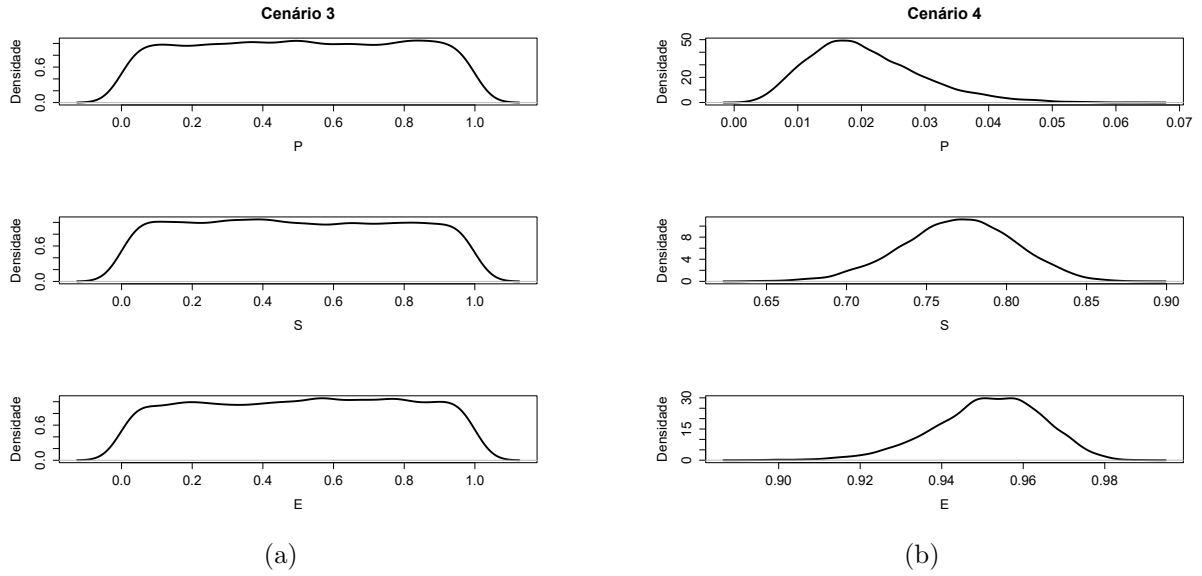


Figura 5: Gráficos das densidades estimadas para os parâmetros P, S e E para os cenários 3 e 4.

obteve-se um intervalo HPD de 0 a 63, porém a quantidade de resultados falso-positivos deveria ser um valor baixo. Para os parâmetros, a falta de precisão também pode ser visualizada através da Figura 4(a) contendo as densidades estimadas para P, S e E. Para o cenário 2, os intervalos são menores, indicando uma melhor precisão, que também pode ser observada na Figura 4(b). Outro aspecto importante a destacar é a discrepância entre as estimativas pontuais para a prevalência, sensibilidade e especificidade nos dois cenários;

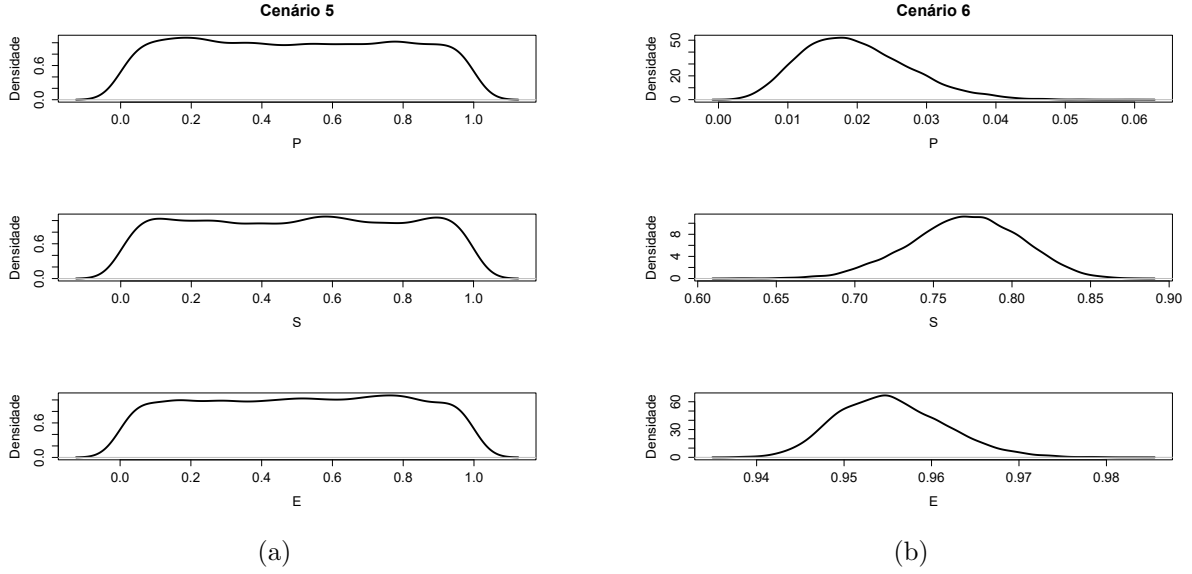


Figura 6: Gráficos das densidades estimadas para os parâmetros P, S e E para os cenários 5 e 6.

no cenário 2 as medidas de desempenho são maiores e a prevalência da doença é menor. Resumindo o cenário 2, situação real com distribuições *a priori* informativas, temos que a prevalência da tuberculose bovina na população em estudo é de 2%, a sensibilidade de 77% e a especificidade de 95%, aproximadamente. Além disso, os valores preditivos positivo e negativo foram calculados para este cenário, e nos mostram que se o resultado do teste for positivo, a probabilidade do animal estar realmente doente é de aproximadamente 25%, e se o resultado do teste for negativo, a probabilidade do animal não ter a doença é de aproximadamente 99%. Nos casos em que a prevalência da doença na população é baixa, o valor preditivo positivo tende a ser menor e o valor preditivo negativo tende a ser maior. Em contrapartida, os valores preditivos positivo e negativo tendem a ser maior e menor, respectivamente, para populações com alta prevalência da doença. Estas observações podem ser feitas a partir de outros trabalhos, como por exemplo, Pinho (2006).

Os valores preditivos positivo e negativo foram obtidos, propositalmente, apenas para o cenário 2, pois este apresenta uma situação real com distribuições *a priori* informativas. Para os outros cenários, os valores preditivos positivo e negativo podem ser obtidos de forma similar.

As observações realizadas na comparação do cenário 1 (Tabela 3) com o cenário 2 (Tabela 4) são similares quando se comparam o cenário 3 (Tabela 5) com o cenário 4

(Tabela 6), e o cenário 5 (Tabela 7) com o cenário 6 (Tabela 8).

Para os cenários que utilizaram a mesma distribuição *a priori* para P, S e E, observou-se que o tamanho da amostra não exerceu influência expressiva sobre os resultados pois as estimativas pontuais e intervalares estão bastante próximas.

A Tabela 9 apresenta os valores de AICc e BIC para os cenários em estudo.

Tabela 9: Valores de AICc e BIC para cada cenário.

Critério	Cenário					
	1	2	3	4	5	6
AICc	38,19	8,81	291,79	11,41	∞	14,35
BIC	37,32	7,93	994,25	13,87	∞	19,84

Observou-se que os valores de AICc e BIC são menores para os cenários que utilizaram distribuições *a priori* informativas, ou seja, estes cenários são considerados de melhor desempenho. Além disso, observou-se que os valores de AICc e BIC aumentaram com o aumento da amostra, dentre os cenários que utilizam a mesma distribuição *a priori* para os parâmetros, concluindo-se assim, que o aumento do tamanho amostral não melhora o desempenho dos resultados.

A partir das estimativas, pode-se afirmar que o desempenho do teste diagnóstico foi melhor nos cenários em que foram consideradas distribuições *a priori* beta para os parâmetros P, S e E.

4 Conclusão

A metodologia utilizada para inferência é um ponto importante no estudo de desempenho de testes diagnósticos pois diferentes resultados podem ser obtidos a partir de um mesmo experimento. Ao se fazer uma análise, pode-se concluir pela ineficiência de um teste diagnóstico quando de fato a metodologia de análise foi inadequadamente aplicada.

A escolha das distribuições *a priori* deve ser feita de forma criteriosa, haja vista que a partir dos resultados obtidos verificou-se que estas interferem fortemente nas estimativas *a posteriori*.

Por outro lado, a informação prévia surge como uma boa alternativa, uma vez que na sua ausência o problema de estimação não teria solução.

Neste ponto, observa-se que a informação prévia, traduzida em distribuição de probabilidade, está para a informação bayesiana assim como as restrições nos estimadores

de máxima verossimilhança estão para a inferência clássica, pois sem elas, as medidas de interesse em testes diagnósticos na ausência de um teste Padrão Ouro também não teria solução.

Referências

- [1] ALVORD, W. et al. *A method for predicting individual HIV infection status in the absence of clinical information*. AIDS Research and Human Retroviruses, New York, v. 4, p. 295-304, 1988.
- [2] BOZDOGAN, H. *Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions*. Psychometrika. n. 52, p. 345-370, 1987.
- [3] BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Secretaria de Defesa Agropecuária. Departamento de Defesa Animal. Manual Técnico do Programa Nacional de Controle e Erradicação da Brucelose e Tuberculose - PNCEBT: legislação. Brasília, 2006.
- [4] CHAMBERLAIN, J.; GINKS, S.; ROGERS, P.; NATHAN, B. E.; PRICE, J. L., and BURN, I. *Validity of clinical examination and mammography as screening tests for breast cancer*. Lancet, 2, 1026-1030, 1975.
- [5] GELFAND, A. E.; HILLS, S. E.; RACINE-POON, A.; SMITH, A. F. M. *Illustration of Bayesian inference in normal data models using Gibbs sampling*. Journal of the American Statistical Association, v. 85, n. 41, p.972-985, 1990.
- [6] GELMAN, A.; RUBIN, D. et al. *Bayesian Data Analysis*. England: Chapman & Hall, 1997.
- [7] GELMAN, A.; RUBIN, D. *Inference from iterative simulation using multiple sequences*. Statistical Science, v. 7, p. 457-511, 1992.
- [8] GEWEKE, J. *Evaluating the accuracy of sampling-based approaches to calculating posterior moments*. In Bayesian Statistics 4, p. 169-193, 1992.
- [9] HASTINGS, W. K. *Monte Carlo sampling methods using markov chains and their applications*. Biometrika 57(1), 97-109. 1970
- [10] JAMA. Evidence-Based Medicine Working Group. *Evidence-based medicine. A new approach to teaching the practice of medicine*. JAMA, v. 268, n. 17, 1992.

- [11] JOSEPH, L.; GYORKOS, T.W.; COUPAL, L. *Bayesian estimation of disease prevalence and the parameters of diagnostic test in the absence of a good standard*. American Journal of Epidemiology, v. 141, n. 3, p. 263-272; Fev. 1995.
- [12] MARKLE W.; FISHER M., SMEGO R, eds. *Understanding Global*. New York: McGraw-Hill; 2007.
- [13] METROPOLIS, N. ; ROSENBLUTH, A. W.; ROSENBLUTH, M. N.; TELLER, A. H.; TELLER E. *Equation of state calculations by fast computing machines*. J. Chem. Phys. 21(6), 1087-1092; 1953
- [14] NOGUEIRA, D. A. et al. *Avaliação de critérios de convergência para o método de Monte Carlo via Cadeias de Markov*. Revista Brasileira de Estatística, Rio de Janeiro, v. 65, n. 24, p. 59-88, 2004.
- [15] PINHO E. *Estimação Bayesiana para medidas de desempenho de testes diagnósticos*. Universidade Federal de São Carlos, 2006.
- [16] SCHWARZ, G. *Estimating the dimensional of a model*. Annals of Statistics, Hayward, v. 6, n. 2, p. 461-464, Mar. 1978.
- [17] TANNER, M. A.; WONG, W. H. *The Calculation of Posterior by Data Augmentation*. Journal of the American Statistical Association, v. 82 (398), p. 528- 540, 1987.
- [18] TRIOLA, M. F. *Introdução à Estatística*. 9ª edição. Editora: LTC. 2005.
- [19] WALTER, S. FROMMER, D. COOK, R. *The estimation of sensitivity and specificity in colorectal cancer screening methods*. Cancer Detection Prevention, [S.l.], v. 15, p. 465-469, 1991.
- [20] R DEVELOPMENT CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: <<http://www.R-project.org>>. Acesso em: 02 de Jul. 2016.