UNIVERSIDADE FEDERAL DE MINAS GERAIS

JORIA MARTINHO GONÇALVES

SOLUÇÕES PARA O PROBLEMA DE SEPARAÇÃO QUASE-COMPLETA EM REGRESSÃO LOGÍSTICA

BELO HORIZONTE

2008

JORIA MARTINHO GONÇALVES

SOLUÇÕES PARA O PROBLEMA DE SEPARAÇÃO QUASE-COMPLETA EM REGRESSÃO LOGÍSTICA

Dissertação apresentada ao Programa de Pós-graduação em Estatística da Universidade Federal de Minas Gerais para obtenção do título de Mestre em Estatística.

Orientador: Prof. PhD. Enrico Antônio Colosimo
Co-orientadora: Profª. Drª. Rosângela Helena Loschi

BELO HORIZONTE
2008

Aos meus queridos pais,
Lourdinha e José Gonçalves

Agradecimentos

Agradeço a Deus, pela luz.

Aos meus pais, pelo apoio e pela dedicação, sem os quais esta conquista não seria possível.

A Rede Sarah e sua equipe, que me deram a oportunidade e o apoio para que eu me dedicasse ao mestrado.

Ao Maurício, que esteve comigo durante os momentos mais difíceis, me apoiando e me confortando.

Aos amigos, irmãos, cunhadas, sobrinhos e demais parentes, colegas do mestrado e do trabalho que me ajudaram e torceram tanto pela minha vitória. Em especial, ao colega Elias Kraisnki, pelas contribuições em estatística e recursos computacionais.

Ao Enrico, pelo carinho e pela disponibilidade para me orientar e apoiar na realização deste sonho.

À Rosângela, pela grande ajuda e pela chance de conhecer mais um pouco de estatística bayesiana.

Ao Sebastião Martins Filho, pela importante contribuição em análise de Bayes-empírica.

"We shall not cease from exploration,
And the end of all our exploring
Will be to arrive where we started,
And know the place for the first time.'

"Não cessaremos com a exploração,
E o fim de todo nosso explorar
Será chegar ao ponto onde começamos,
E conhecer o lugar pela primeira vez."

T.S. Eliot

Resumo

A regressão logística é o método estatístico, frequentemente, utilizado quando o objetivo do estudo é verificar a relação entre uma variável resposta dicotômica e covariáveis relacionadas a ela. Os parâmetros do modelo, usualmente, são estimados através do método de máxima verossimilhança e testes sobre estes parâmetros são construídos considerando as distribuições aproximadas dos estimadores. Isto significa que amostras grandes tornam-se necessárias para termos resultados mais confiáveis. Em estudos envolvendo dados binários, é frequente a presença de uma variável resposta cujo sucesso é pouco provável, ou seja, temos um evento raro, o que pode gerar uma amostra com dados esparsos. Nestes casos, podemos ter dados que se encaixam na classificação de separação quase-completa e esta situação está, frequentemente, associada à presença de uma covariável categórica. Neste caso, os estimadores de máxima verossimilhança não existem.

A inclusão de informações *a priori* sobre os parâmetros no problema pode trazer um ganho para a análise dos dados. O objetivo deste trabalho foi abordar o modelo de regressão logística binária para os casos de separação quase-completa via métodos bayesianos e bayesianos empíricos. Realizamos um estudo da especificação da distribuição *a priori* utilizando dados gerados com separação quase-completa e superposição. Para avaliar o efeito de distribuições *a priori* nas distribuições *a posteriori* dos parâmetros do modelo, utilizamos, como exemplo, os dados de um estudo apresentado em Colosimo, Franco e Couto (1995). Além disto, construímos uma distribuição *a priori* empírica para o modelo logístico usando os dados do exemplo e avaliamos se este tipo de especificação *a priori* traz algum ganho para a análise de dados com separação quase-completa. Os resultados foram comparados com a proposta de estimação por máxima verossimilhança penalizada.

Verificamos que a especificação da distribuição *a priori* é a chave para a apropriada utilização da estatística bayesiana. Com uma adequada definição da distribuição *a priori* podemos chegar a melhores resultados que com a estimação por máxima verossimilhança penalizada, no caso de separação quase-completa.

Palavras-chave: Regressão Logística; Separação quase-completa; Eventos Raros; Estatística bayesiana; Bayes empírico.

Abstract

Logistic regression is a statistic method, often used when the study's objective is to verify the relationship between a dichotomous outcome variable and a set of covariates related to it. The model parameters are usually estimated through the maximum likelihood method, and tests for such parameters are constructed taking into account the estimators approximate distributions. This means that large samples are required for more reliable results. In studies involving binary data, the presence of an outcome variable whose success is very unlikely is frequent, that is, it is a rare event, which may produce a sample with sparse data. In such cases we may have data which fit in the classification of quasi-complete separation, and that situation is often associated to the presence of a categorical covariate. In that case maximum likelihood estimators do not exist.

Including *a priori* information on the problem parameters may yield a gain for data analysis. The goal of the present study is to approach the binary logistic regression model for cases of quasi-complete separation by Bayesian and empirical Bayes methods. We carried out a study on the specification of *a priori* distribution employing data produced with quasi-complete separation and overlap. To assess the effect of *a priori* distributions on *a posteriori* distributions of the model parameters, we used, as an example, the data from a study presented in Colosimo, Franco and Couto (1995). In addition, we constructed an empirical *a priori* distribution for the logistic model using data from the example and verified whether that type of *a priori* specification produces any gain to the data analysis with quasi-complete separation. Results were compared with the estimation proposal by penalized maximum likelihood.

We observed that specification for *a priori* distribution is the key to a proper use of Bayesian statistics. With an adequate definition of *a priori* distribution we may proceed to better results than with penalized maximum likelihood estimation, in cases of quasi-complete separation.

Key-words: Logistic Regression; Quase-complete separation; Rare Events; Bayesian statistics; Empirical Bayes.

Lista de Figuras

Lista de Tabelas

Sumário

Capítulo 1

Introdução

Em muitos estudos na área de saúde, a variável de interesse, também conhecida como variável resposta, apresenta apenas duas categorias. Como por exemplo, podemos citar a remissão de uma doença (sim ou não), o resultado de um tratamento (bom ou ruim), entre outras. Variáveis deste tipo são classificadas como binárias ou dicotômicas.

Quando o objetivo do estudo é verificar a relação entre uma variável resposta dicotômica e variáveis explicativas ou covariáveis relacionadas a ela, a regressão logística é o método estatístico, frequentemente, utilizado. Os parâmetros do modelo, usualmente, são estimados através do método de máxima verossimilhança e testes sobre estes parâmetros são construídos considerando as distribuições assintóticas dos estimadores. Isto significa que amostras grandes tornam-se necessárias para termos resultados mais confiáveis.

Em estudos envolvendo dados binários é frequente a presença de uma variável resposta cujo sucesso é pouco provável de ocorrer, ou seja, temos um evento raro o que pode gerar uma amostra com dados esparsos. Neste caso, os estimadores de máxima verossimilhança podem não fornecer resultados satisfatórios para a estimação dos parâmetros ou podem não existir. Albert e Anderson (1984) identificaram as condições para existência dos estimadores de máxima verossimilhança em modelos cujo comportamento pode ser descrito via modelo logístico. Conjuntos de dados logísticos podem ser classificados em três categorias mutuamente exclusivas e exaustivas: dados com separação completa, separação quase-completa e superposição. Estimadores de máxima verossimilhança não existem para as duas primeiras categorias.

Não são raros os problemas reais que se encaixam na classificação de separação quase-completa. Segundo Nacle (2004), esta situação está, frequentemente, associada à existência de uma variável explicativa categórica. Se, numa tabela de contingência relacionando as variáveis explicativa e resposta, observarmos frequência nula em uma das caselas da tabela, diz-se que o conjunto de dados está na categoria de separação quase-completa. Um evento raro pode ocasionar a separação quase-completa no conjunto de dados.

Quando uma tabela, cruzando a variável resposta com uma covariável categórica, apresenta dois zeros em caselas discordantes diz-se que o conjunto de dados está na categoria de separação completa.

Segundo Heinze e Schemper (2002) as seguintes soluções são possíveis para tratarmos uma situação em que se observa separação completa ou separação quase-completa: omissão da covariável no modelo, utilização de uma função de ligação diferente da logit para o modelo de regressão logística, manipulação de dados, regressão logística exata e a modificação da função escore, sendo esta última recomendada por eles.

Uma vez que, em dados que apresentam separabilidade quase-completa, o estimador de máxima verossimilhança não existe, a inclusão de informações *a priori* sobre os parâmetros no problema pode trazer um ganho na análise dos dados. O objetivo deste trabalho é, então, abordar o modelo de regressão logística binária para os casos de separação quase-completa via métodos bayesianos e bayesianos empíricos. Inicialmente, a meta é avaliar o efeito nas estimativas *a posteriori* de distribuições *a priori* vagas e informativas para os parâmetros do modelo. Segundo Agresti (2006), a sensibilidade dos resultados a mudanças na especificação da distribuição *a priori* quando a informação é vaga é um problema para aqueles que preferem uma abordagem objetiva da análise de dados, mas é atrativa em relação a outros aspectos da abordagem bayesiana.

Também construiremos uma distribuição *a priori* empírica para o modelo logístico e avaliaremos se este tipo de especificação *a priori* traz algum ganho na análise de dados com separabilidade quase-completa.

Além disto, queremos comparar estes resultados com a proposta de estimação por verossimilhança penalizada recomendada por Heinze e Schemper (2002). Segundo Zorn (2005), o método de verossimilhança penalizada proposto por Firth (1993) fornece uma solução simples, válida e fácil de implementar em problemas de separabilidade. Ele não envolve manipulação arbitrária de dados nem modificações complicadas de modelos padrão. Ele, também, não altera a interpretação dos modelos e é disponível em pacotes estatísticos existentes. Ainda segundo Zorn (2002), talvez a melhor vantagem é que este procedimento é, assintoticamente, equivalente ao método de máxima verossimilhança no caso de amostras grandes e superior a ele no caso de amostras pequenas, onde a separabilidade é mais provável de ocorrer.

Este trabalho está organizado da seguinte forma: no Capítulo 2 é apresentado o exemplo que motivou este estudo. No Capítulo 3 são apresentados o modelo de regressão logística e a interpretação dos seus parâmetros, os estimadores de máxima verossimilhança e os critérios para classificar os dados logísticos, além dos estimadores de máxima verossimilhança penalizada. O Capítulo 4 introduz a análise bayesiana no modelo de regressão logística, a especificação da distribuição *a priori* e a análise de Bayes-empírica. No Capítulo 5 encontram-se a análise dos resultados para o banco de dados apresentado em Colosimo, Franco e Couto (1995), além de um estudo da especificação da distribuição *a priori* utilizando dados gerados com separação quase-completa e superposição. Finalmente, no Capítulo 6, encontram-se as conclusões desta dissertação.

Capítulo 2

Motivação

Colosimo, Franco e Couto (1995) analisaram um conjunto de dados formado por 102 pacientes submetidos à cirurgia de craniotomia no Hospital São Francisco em Belo Horizonte, MG, entre julho de 1991 e junho de 1992. A variável resposta N_1 considerada no estudo é a ocorrência (1) ou não (0) de meningite durante os 30 dias subseqüentes à realização da cirurgia. Duas covariáveis foram estudados para verificar se poderiam ser consideradas como fatores de risco para a ocorrência de meningite, a saber, a gravidade do caso, X_1, que foi categorizada em alta (1) e baixa (0) e o tempo (em horas) da cirurgia, denotada aqui por X_2. Os dados foram coletados pela equipe do controle de infecção e são apresentados na Tabela 2.1, onde N denota o número de observações em cada categoria.

Tabela 2.1 - Conjunto de dados de pacientes submetidos à craniotomia.

X_1	X_2	N_1	N	X_1	X_2	N_1	N
0	2,50	0	1	0	2,17	0	1
1	1,33	0	1	1	6,50	0	1
1	6,00	0	2	1	1,00	0	3
0	4,50	0	1	1	4,00	0	4
0	1,50	0	3	1	3,00	0	8
0	1,33	0	4	0	4,00	0	8
0	5,00	0	3	0	4,75	0	1
1	0,75	0	1	0	3,00	0	13
0	2,00	0	8	1	8,00	0	1
0	3,50	0	3	0	5,50	0	1
1	3,25	0	1	0	2,67	0	1
0	1,83	0	4	0	2,25	0	1
1	7,00	0	1	0	7,00	0	2
0	1,67	0	1	0	3,67	0	1
0	8,00	0	1	0	2,33	0	1
1	3,50	0	1	0	6,50	0	1
0	3,17	0	1	0	1,00	0	3
1	5,50	0	1	0	6,00	0	3
1	2,00	0	6	1	1,50	1	2
0	1,25	0	1	1	10,00	1	1

A Tabela 2.2 mostra a distribuição conjunta dos pacientes submetidos à craniotomia segundo a gravidade do caso e a ocorrência de meningite.

Tabela 2.2 - Distribuição dos pacientes segundo a gravidade do caso e a presença de meningite

	Ocorrência de meningite		
Gravidade	Sim	Não	Total
Baixa	0	68	68
Alta	2	32	34
Total	2	100	102

A ocorrência de meningite parece ser um evento raro, uma vez que, somente 1,96% dos pacientes a apresentaram. Além disto, todos os pacientes com meningite eram considerados pacientes de alta gravidade. Este fato sugere a existência de uma chance maior de pacientes graves contraírem meningite após a cirurgia. Este tipo de comportamento sugere a existência de separabilidade quase-completa dos dados. No que segue, serão mostradas algumas estratégias sugeridas para tratar este tipo de problema e sugeriremos algumas outras (ver Capítulo 5).

Capítulo 3

Modelo de Regressão Logística

O modelo de regressão logística é utilizado para determinar os fatores que estão associados com a ocorrência de um evento de interesse quando a variável resposta é binária. Segundo Hosmer e Lemeshow (2000), entre outras coisas, a partir do modelo de regressão logística é possível estimar a probabilidade da ocorrência deste evento para um indivíduo. Segundo Breiman et al.(1984), a técnica de Árvore de Classificação e Regressão (CART) é outra opção para analisarmos este tipo de dados, mas não trataremos desta técnica neste trabalho.

Assuma que Y é uma matriz de n variáveis independentes com Y_i~Bernoulli(θ_i) onde $Y_i = 1$ representa a ocorrência do evento de interesse com $i = 1,...,n$. Neste caso, $E(Y_i) = \theta_i = P(Y_i = 1)$, onde θ_i é a probabilidade de ocorrência do evento de interesse para o *i*-ésimo indivíduo. Denote por X_i a *i*-ésima linha da matriz de p variáveis explicativas e por β o vetor de ordem $(p+1)$ referente aos parâmetros a serem estimados:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ ... \\ Y_n \end{bmatrix},\ X = \begin{bmatrix} 1 & X_{11} & ... & X_{1p} \\ 1 & X_{21} & ... & X_{2p} \\ ... & ... & ... & ... \\ 1 & X_{n1} & ... & X_{np} \end{bmatrix} \text{ e } \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ ... \\ \beta_p \end{bmatrix}$$

Dados os valores das covariáveis X_i, o interesse está em determinar-se a probabilidade:

$$\theta_i = P(Y_i = 1) = \frac{e^{X_i'\beta}}{1 + e^{X_i'\beta}}. \quad (1)$$

A função logit é dada por:

$$\text{logit}(\theta_i) = \log\left(\frac{\theta_i}{1-\theta_i}\right) = X_i'\beta, \quad i = 1,...,n. \quad (2)$$

Esta função não é a única função de ligação que pode ser utilizada na regressão logística, mas sua principal vantagem é a facilidade de interpretação, uma vez que ela é o logaritmo da chance de ocorrência de um determinado evento. Outras funções de ligação utilizadas na regressão binária são a probit e a log-log (Hosmer e Lemeshow, 2000).

3.1 Interpretação dos parâmetros

Apesar de estarmos interessados nas estimativas dos coeficientes β, a interpretação dos seus valores não é tão simples pois depende dos valores das variáveis explicativas. Ao invés de interpretarmos estes coeficientes β diretamente, podemos fazer a interpretação através da razão das chances (odds ratio), que é dada por:

$$\psi_j = e^{\beta_j}, \quad j = 1,..., p.$$

Esta razão mede o quanto é mais provável a ocorrência do evento de interesse para um nível da covariável categórica j em relação a outro nível da mesma covariável, mantendo fixos os valores das outras covariáveis. Uma razão das chances $\psi = 1$ significa que o evento de interesse é tão provável para um nível da covariável, quanto para outro.

No nosso problema, por exemplo, ao utilizarmos a razão das chances para interpretarmos o coeficiente da covariável gravidade, $\psi = 1$ significa que a probabilidade de um paciente desenvolver meningite é a mesma tanto para pacientes com gravidade alta, quanto para pacientes com gravidade baixa.

3.2 Estimadores de Máxima Verossimilhança

A estimação dos parâmetros do modelo de regressão logística é, geralmente, feita usando o método de máxima verossimilhança. Os estimadores de máxima verossimilhança são os que maximizam a função de verossimilhança. Sob a suposição de independência dos valores de Y_i, $i = 1,...,n$, a função de verossimilhança é dada por:

$$L(\beta) = \prod_{i=1}^{n} \theta_i^{y_i} (1 - \theta_i)^{1-y_i}. \tag{3}$$

Maximizar a função de verossimilhança é equivalente a maximizar o logaritmo neperiano da mesma função, que pode ser escrito como :

$$l(\beta) = \ln L(\beta) = \sum_{i=1}^{n} (y_i \ln \theta_i + (1 - y_i) \ln(1 - \theta_i)). \quad (4)$$

Sob condições de regularidade, segundo Casella e Berger (2002), o máximo global da função $l(\beta)$ é encontrado, unicamente, pelas soluções da seguinte expressão:

$$\frac{\partial l(\beta)}{\partial \beta} = 0. \quad (5)$$

Os valores de β são obtidos pela solução do sistema de $(p+1)$ equações que fazem o vetor escore igual a zero, ou seja:

$$U_j(\beta) = \frac{\partial l(\beta)}{\partial \beta_j} = \sum_{i=1}^{n} x_{ij} (y_i - \theta_{ij}) = 0, \qquad j = 1,..., p+1. \quad (6)$$

Não existem soluções exatas para a expressão em (6). Então, em geral, são utilizados métodos numéricos iterativos – método de Newton-Raphson, por exemplo - para solucionar este sistema de equações (Casella e Berger, 2002) e, assim, encontrar os estimadores de máxima verossimilhança quando estes existirem.

3.3 Existência de estimadores de máxima verossimilhança em modelos de regressão logística

Neste capítulo, apresentaremos formalmente os conceitos de separação completa, separação quase-completa e superposição utilizados para classificar dados logísticos. Também apresentaremos resumidamente o modelo logístico e os estimadores de máxima verossimilhança para os parâmetros do modelo e discutiremos condições para a sua existência. Também discuremos o método de estimação baseado na verossimilhança penalizada proposto por Heinze e Schemper (2002).

3.3.1 Classificação de dados logísticos

Como citado anteriormente, Albert e Anderson (1984) mostraram que os dados logísticos podem ser classificados em três categorias mutuamente exclusivas e exaustivas: separação completa, separação quase-completa e superposição.

A seguir apresentaremos formalmente esta classificação. Para isto, consideremos as configurações possíveis dos n valores amostrais no espaço de observação $\Re^p$ e a partir destes valores definiremos cada uma das categorias citadas.

3.3.1.1 Separação Completa

Ocorre separação completa quando, baseado na informação de uma covariável ou combinação de covariáveis, pode-se predizer corretamente o valor de uma variável de interesse. Isto implica na existência de um vetor $\beta \in \Re^{p+1}$ pelo qual todos os n valores amostrais podem ser corretamente classificados entre $Y = 0$ ou $Y = 1$, tal que para todo $i \in E_j$, $j = 0,1$, tem-se

$$X_i^{'}\beta > 0, \quad i \in E_0,$$

$$X_i^{'}\beta < 0, \quad i \in E_1,$$

onde E_j é o conjunto de linhas identificadas da matriz X com valor de $Y = j$. A Figura 3.1(a) ilustra esta categoria de separação para $\Re^2$.

3.3.1.2 Separação Quase-Completa

Ocorre separação quase-completa quando, baseado na informação de uma covariável ou combinação de covariáveis, pode-se predizer perfeitamente os valores de pelo menos um grupo da variável de interesse, ou seja, $Y = 0$ ou $Y = 1$. A separação quase-completa implica na existência de um vetor $\beta \in \Re^{p+1}$ tal que, para todo $i \in E_j$, com $j = 0,1$

$$X_i^{'}\beta \geq 0, \quad i \in E_0,$$

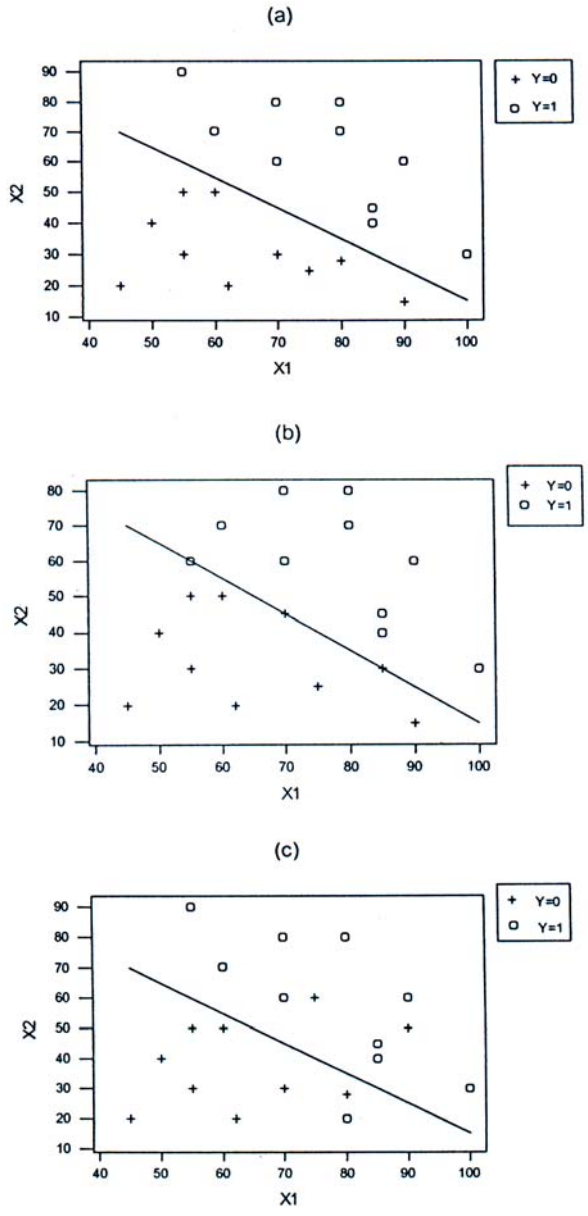
$$X_i^{'}\beta \leq 0, \quad i \in E_1,$$

com igualdade para, pelo menos, um valor de i. A Figura 3.1(b) ilustra esta categoria de separação para $\Re^2$.

3.3.1.3 Superposição

Se os dados não estão nas duas categorias anteriores, necessariamente, eles estão na categoria de superposição. A Figura 3.1(c) ilustra esta categoria de separação para $\Re^2$.

Figura 3.1 - Exemplo de conjunto de dados com separação completa (a), separação quase-completa (b) e superposição (c).



Esta categoria implica na existência de um vetor $\beta \in \Re^{p+1}$ tal que, para todo $i \in E_j$, com $j = 0,1$

$$X_i^{'}\beta < 0, \quad i \in E_0,$$

$$X_i^{'}\beta > 0, \quad i \in E_1,$$

onde E, é o conjunto de linhas identificadas da matriz X com valor de $Y = j$.

3.3.2 Identificação de separação e sua importância

Classificar os dados logísticos em uma das três categorias, pela definição, requer muito esforço. Santner e Duffy (1986) e Clarkson e Jenrick (1991) apresentaram procedimentos computacionais sofisticados para detectar se há separação nos dados. Na prática, duas alternativas simples para identificar a separação são:

- Monitorar a variância estimada dos coeficientes da regressão (Heinze e Schemper, 2002). Se observarmos variância grande para algum parâmetro estimado, há um indicativo de separação;
- Fazer uma tabela de contingência, cruzando a variável resposta com as covariáveis e verificar se existem caselas com valores observados iguais a zero (Nacle, 2004). O valor zero em uma casela indica separação quase-completa, dois zeros em caselas discordantes indicam separação completa.

Albert e Anderson (1984) provaram que quando temos um conjunto de dados nas categorias de separação completa ou quase-completa, a função de verossimilhança do modelo logístico é monótona e, portanto, estimadores de máxima verossimilhança não existem. Sendo assim, torna-se importante encontrar um procedimento eficiente para a estimação destes parâmetros na presença de separabilidade completa ou quase completa.

3.3.3 Estimadores de máxima verossimilhança para o exemplo de craniotomia

Levando-se em conta os dados de pacientes submetidos à craniotomia (veja em Colosimo, Franco e Couto (1995)), verifica-se, através da Tabela 2.2, que ocorre a separação quase-completa nos dados, pois uma das caselas tem valor nulo**.**

Fazendo a análise tradicional, tentamos estimar os coeficientes do modelo de regressão logística através do método de máxima verossimilhança. Utilizamos, para isto, o pacote livre R Project for Statistical Computing (R) desenvolvido por R Development Core Team (2006). Verificamos que há estimativa para todos os coeficientes do modelo. Mas nota-se pela Tabela 3.1 que apesar destes coeficientes terem sido estimados, o erro padrão da estimativa do coeficiente β_1 é muito grande. Isto mostra que não existe máximo da função de verossimilhança para o coeficiente β_1.

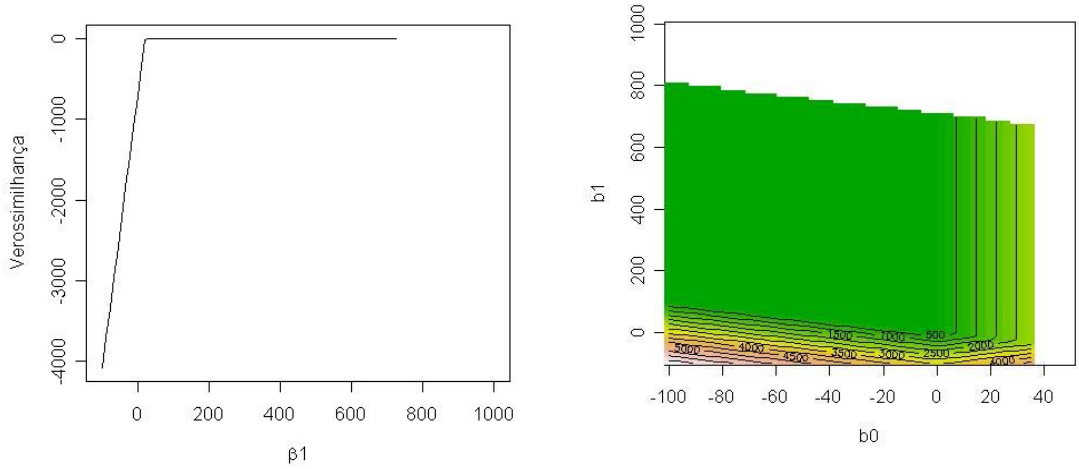
Tabela 3.1 – Estimadores de máxima verossimilhança para os coeficientes do modelo de regressão para os dados de craniotomia

Programa	Coeficiente	Estimativa	Erro padrão	Estatística de teste	p
R - glm	β_0	-0,8546	0,4599	-1,858	0,0632
	β_1	16,24	1.026,46	0,016	0,9874
	β_2	0,0314	0,1260	0,249	0,8034

Isto ocorre por causa da separação quase-completa que inviabiliza qualquer tentativa de estimação do coeficiente β_1 através do método de máxima verossimilhança usual.

Como se observa na Figura 3.2, não há um ponto único de máximo para a função de verossimlhança com relação ao coeficiente β_1.

Figura 3.2 – Função de verossimilhança dos dados de craniotomia



3.4 Máxima verossimilhança penalizada

Visando resolver o problema de existência dos estimadores de máxima verossimilhança na presença de separação, Heinze e Schemper (2002) sugerem a modificação da função escore para a estimação dos coeficientes do modelo de regressão logística.

Originalmente, essa proposta foi desenvolvida por Firth (1993) buscando reduzir o vício das estimativas de máxima verossimilhança em modelos lineares generalizados. Ela produz estimativas finitas para os parâmetros do modelo através da estimação por máxima verossimilhança penalizada.

As estimativas de máxima verossimilhança dos parâmetros da regressão são encontradas solucionando o sistema de equações do vetor escore, como visto em (3). No entanto, Firth (1993) sugere a estimação baseada nas equações escore modificadas dadas por:

$$U_j(\beta)^* \equiv U_j(\beta) + \frac{1}{2}\, traço\left[I(\beta)^{-1}\left\{ \frac{\partial I(\beta)}{\partial \beta_j} \right\} \right] = 0, \qquad j = 1,..., p+1.,$$

onde $I(\beta)^{-1}$ é a inversa da matriz de informação de Fisher avaliada em β. A função escore modificada $U_j(\beta)^*$ é relacionada à função de log-verossimilhança penalizada:

$$l(\beta)^* = l(\beta) + \frac{1}{2}\ln\left|I(\beta)\right|,$$

e à função de verossimilhança penalizada:

$$L(\beta)^* = L(\beta)\left|I(\beta)\right|^{\frac{1}{2}}.$$

A função de penalização $\left|I(\beta)\right|^{\frac{1}{2}}$ tem influência, assintoticamente, desprezível. Utilizando esta modificação, Firth (1993) mostrou que o vício das estimativas de máxima verossimilhança é removido.

Aplicando a idéia geral de Firth para o modelo logístico em (1), a equação escore em (6) é substituída pela equação escore modificada que é dada por:

$$U_j(\beta)^* = \sum_{i=1}^{n} x_{ij}\left\{y_i - \theta_{ij} + h_i\left(\frac{1}{2} - \theta_{ij}\right)\right\} = 0, \qquad j = 1,..., p+1,$$

onde h_i é o i-ésimo elemento da diagonal principal de matriz H:

$$H = W^{1/2}X(X^TWX)^{-1}X^TW^{1/2} \text{ e } W = diag\{\theta_{ij}(1-\theta_{ij})\}.$$

As estimativas podem ser obtidas iterativamente pelo método usual até a convergência ser obtida:

$$\beta^{(s+1)} = \beta^{(s)} + I^{-1}\left(\beta^{(s)}\right)U\left(\beta^{(s)}\right)^*,$$

onde (s) se refere à s-ésima iteração.

Três pacotes do R implementam a estimativa de máxima verossimilhança penalizada: o logistf, o brlr e o brglm. Todos eles corrigem o vício de estimação dos coeficientes do modelo de regressão logística, porém há algumas diferenças básicas. O brglm pode ser utilizado em modelos com outras funções de ligação, além do ”logit”, e ainda é mais eficiente computacionalmente. O pacote estatístico Statistical Analysis System (SAS) desenvolvido pelo SAS Institut Inc. (Cary, 1985) também implementa a estimativa de máxima verossimilhança penalizada.

Capítulo 4

Estatística bayesiana para o modelo de regressão logística

Neste capítulo, faremos uma breve descrição de alguns métodos bayesianos de inferência e construiremos uma distribuição *a priori,* via análise Bayes-empírica, para os parâmetros do modelo logístico.

4.1 O método bayesiano

Usando a abordagem bayesiana, inicialmente devemos eliciar a distribuição *a priori* para o vetor de coeficientes β, a qual será denotada por $\pi(\beta)$. As distribuições *a priori* obtidas a partir de $\pi(\beta)$ devem refletir o conhecimento prévio do pesquisador sobre estes coeficientes. A função de verossimilhança do modelo em (1), que resume a informação amostral sobre β, atualiza tal distribuição *a priori*, gerando-se assim uma distribuição atualizada para β. Esta distribuição é chamada de distribuição *a posteriori* e é obtida via teorema de Bayes como segue:

$$\pi(\beta \mid \underset{\sim}{x}) \propto \pi(\beta)L(\beta) \propto \pi(\beta)\prod_{i=1}^{n}\theta_i^{y_i}(1-\theta_i)^{1-y_i}.$$

A distribuição *a posteriori* reflete toda incerteza sobre β após a observação dos dados. Resumos desta distribuição tais como média, moda, mediana e variância, podem ser obtidos de forma habitual. Com podemos, também, realizar testes de hipóteses e intervalos de credibilidade.

Neste caso, não temos uma distribuição *a posteriori* com forma fechada. Desta forma, faz-se necessário utilizarmos métodos numéricos ou métodos MCMC (Markov Chain Monte Carlo) para obtermos uma estimativa da distribuição *a posteriori* e/ou de seus resumos. Neste trabalho, utilizamos o pacote estatístico WinBUGS (Lunn et al., 2000) para obtermos amostras das distribuições *a posteriori* de interesse. O WinBUGS utiliza métodos MCMC para a geração da amostra da distribuição *a posteriori*. Um número grande de amostras é gerado a partir de distribuições condicionais e, após a

convergência ter sido atingida, temos uma amostra da distribuição *a posteriori.* A partir desta amostra obtemos os resumos *a posteriori* desejados, tais como, média, mediana, desvio-padrão, intervalos de credibilidade. O WinBUGS também fornece uma estatística para a comparação de modelos – o Critério de Informação da função Deviance (DIC). Segundo Spiegelhalter et al. (2002), o DIC é uma generalização do Critério de Informação de Akaike (AIC). Assim como o observado para o AIC, um valor pequeno para o DIC indica boa adequabilidade do modelo, ou seja, indica que ele fornece boas estimativas para os coeficientes. Os DIC's referentes a diferentes modelos são comparáveis somente quando os mesmos dados observados são considerados na análise.

Uma vantagem de utilizarmos métodos bayesianos na análise do modelo com separação quase-completa é a possibilidade de existência de estimadores pontuais para β. Mesmo quando a distribuição *a priori* é a uniforme podemos utilizar a média ou a mediana *a posteriori* como estimadores pontuais, uma vez que, neste caso, a moda *a posteriori* é exatamente o estimador de máxima verossimilhança e, portanto, também não existirá em situações de separação.

4.2 Definição da distribuição *a priori*

A distribuição a priori deve refletir o grau de conhecimento inicial do pesquisador sobre os parâmetros do modelo. Quando o pesquisador tem informação sobre os coeficientes não trazida pelos dados, esta deve ser trazida para a análise através da distribuição *a priori* visando melhorar as estimativas. Quando não se tem tal informação, ou se tem e não se deseja utilizar-se dela, uma distribuição *a priori* não informativa deve ser utilizada e, neste caso deixa-se que a função de verossimilhança seja a principal responsável por trazer a informação sobre os coeficientes.

Diante do exposto na seção anterior, percebe-se que a especificação da distribuição *a priori* tem um papel fundamental no estudo do Modelo de Regressão Logística com separação. Galindo-Garre, Vermunt, e Bergsma (2004) afirmaram que, assumindo que não há informação prévia sobre a dependência entre os parâmetros do modelo, é conveniente assumir independência entre os coeficientes e adotar distribuições *a priori* normais univariadas para cada um deles. Eles utilizaram estatística bayesiana para suavizar as estimativas dos parâmetros da regressão logística, assumindo várias

distribuições *a priori* para estes parâmetros. As distribuições utilizadas foram: normais univariadas, Dirichlet, Jeffreys e Clogg-Eliasson. Congdon (2001) sugere o uso de distribuições normais com média zero e variância grande. Greenland (2001) afirma que distribuição *a priori* e verossimilhança podem ser aproximadas por normais multivariadas em casos de grandes amostras mas afirma que, no caso de dados esparsos, tais aproximações podem ser inadequadas. Neste caso, ele recomenda análise conjugada exata.

Neste trabalho, utilizaremos a abordagem de Galindo-Garre, Vermunt, e Bergsma (2004), ou seja, assumiremos independência entre os coeficientes e adotaremos distribuições *a priori* normais univariadas para cada um deles. Também utilizaremos a distribuição *a priori* empírica que introduziremos na próxima seção.

4.2.1 Análise de Bayes-empírica

Segundo Paulino, Turkman e Murteira (2003), a análise de Bayes-empírica utiliza os dados para especificar a distribuição *a priori* e, posteriormente, utiliza a análise bayesiana. Isso a torna "uma terceira via" entre os paradigmas bayesiano e frequentista, e, como tal, tem sido rejeitada ou, pelo menos, secundarizada pela grande maioria dos adeptos da Escola Bayesiana. Entretanto, a análise Bayes-empírica tem permitido ultrapassar as dificuldades de análises integralmente bayesianas de problemas complexos e produzido estimadores com boas propriedades frequentistas.

Voltando ao nosso problema, segundo Tsutakawa e Lin (1986), é mais fácil obtermos informações *a priori* sobre a probabilidade de sucesso $E(Y_i \mid x_i) = \theta_i$ e, consequentemente, seria mais fácil eliciarmos a distribuição *a priori* sobre tal probabilidade, do que obtermos algum conhecimento *a priori* sobre β que é um objeto que, em geral, não tem significado prático.

O método sugerido por Bedrick, Christensen e Johnson (1996) envolve eliciar a distribuição *a priori* para respostas médias correspondentes aos valores observados das covariáveis e, a partir desta distribuição inicial, induzir a uma distribuição *a priori* para os coeficientes da regressão β.

Como θ_i é uma probabilidade, portanto $0 \leq \theta_i \leq 1$, podemos assumir que, independentemente, cada $\theta_i \sim \text{Beta}(a_{1i}, a_{2i})$, isto é,

$$\pi(\theta_i) \propto \prod_{i=1}^{p} \theta_i^{a_{1i}-1}(1-\theta_i)^{a_{2i}-1}.$$

Estas independentes distribuições *a priori* médias condicionais induzem às seguintes distribuições *a priori* para os coeficientes β:

$$\pi(\beta) \propto \prod_{i=1}^{p} F(\tilde{x}_i^{'}\beta)^{a_{1i}-1}[1-F(\tilde{x}_i^{'}\beta)]^{a_{2i}-1} f(\tilde{x}_i^{'}\beta)$$

onde $F(.)$ é a função de distribuição de probabilidade com função de densidade de probabilidade $f(.)$. Para o modelo logístico, $f(.) = F(.)(1-F(.))$.

A idéia do Bayes-empírico é escolher alguns pontos, ou seja, configurações de valores das variáveis do problema, e atribuir uma probabilidade de ocorrência a cada configuração selecionada. O número de pontos escolhidos deve ser igual ao número de coeficientes do modelo e sua escolha deve ser feita com base nos valores que ocorrem com maior freqüência.

A probabilidade de ocorrência de cada conjunto de pontos deve ser definida de forma a refletir o conhecimento *a priori* que o pesquisador tem sobre o assunto e deve ser diferente o suficiente para garantir a independência das probabilidades de cada ponto selecionado.

A média da probabilidade de ocorrência de cada conjunto de pontos é $E(\theta_i) = \dfrac{a_{1i}}{a_{1i}+a_{2i}}$ e, através do seu valor, definido pelo pesquisador, encontra-se a relação entre a_{1i} e a_{2i}. Definidos os valores de a_{1i} e a_{2i}, obtém-se as distribuições *a priori* beta para os θ_i e, a partir destas, encontram-se distribuições *a priori* beta para os coeficientes do modelo, β.

Neste trabalho, aproximações das distribuições *a posteriori* de β_i são obtidas utilizando-se o WinBUGS.

Capítulo 5

Resultados

Neste capítulo, analisaremos o banco de dados de pacientes submetidos à craniotomia, descrito no Capítulo 2, utilizando os estimadores de máxima verossimilhança penalizada. Utilizaremos o método bayesiano com distribuições *a priori* normais para três conjuntos de dados, um deles gerado com superposição, outro com separação quase-completa e outro de pacientes submetidos à craniotomia. Posteriormente, utilizaremos a análise Bayes-empírica para os dados de craniotomia.

Para a análise dos dados, consideramos os pacotes estatísticos R e SAS e, para a análise via métodos bayesianos, utilizamos o WinBugs.

5.1 Estimação por máxima verossimilhança penalizada

Utilizando os dados do exemplo de craniotomia, foram estimados os coeficientes do modelo de regressão logística utilizando máxima verossimilhança penalizada. Os resultados são mostrados na Tabela 5.1.

Tabela 5.1 – Resultados da estimação por máxima verossimilhança penalizada para cada programa

				I.C. de 95%			
Programa	Coeficiente	Estimativa	Erro padrão	Limite inferior	Limite superior	Estatística de teste	p
SAS	β_0	-6,1428	1,7066	-9,4878	-2,7978	-	<0,0001
	β_1	2,1689	1,5294	-0,4739	7,1210	-	0,1114
	β_2	0,35655	0,2377	-0,12885	0,88558	-	0,1373
R - brglm	β_0	-6,1428	1,7066	-	-	-3,5990	0,000319
	β_1	2,1689	1,5294	-	-	1,4180	0,156164
	β_2	0,3565	0,2377	-	-	1,5000	0,13362
R - logistf	β_0	-6,1428	1,7066	-11,2652	-3,4235	25,8137	<0,0001
	β_1	2,1689	1,5294	-0,4739	7,1210	2,5337	0,111441
	β_2	0,3565	0,2377	-0,1289	0,8856	2,2077	0,137326

Através dos coeficientes estimados pelos modelos no SAS, no brglm e no logistf, verificamos que as covariáveis não são significativas para o modelo. Isto é, a gravidade do caso do paciente que realizou craniotomia e o tempo de duração da sua cirurgia (em horas) não influenciam na ocorrência de meningite.

5.2 Análise bayesiana com distribuição *a priori* normal

Nesta seção, utilizaremos dois bancos de dados gerados a partir do modelo de regressão logística. Um deles na situação de superposição e outro na situação de separação quase-completa. Vamos estudar a especificação da distribuição *a priori* normal na estimação dos coeficientes, através da distribuição *a posteriori*, comparando com os verdadeiros valores dos parâmetros. Além disto, estudaremos a especificação da distribuição *a priori* normal na estimação dos coeficientes da regressão usando o banco de dados de pacientes submetidos à craniotomia. Comparamos também os resultados com as estimativas de máxima verossimilhança penalizada.

5.2.1 Análise dos dados gerados

Inicialmente, geramos dois bancos de dados, de tamanho 100, a partir do modelo em (1), assumindo que Y e X são variáveis dicotômicas. Para este fim, assumimos $\beta_0 = -3$ e $\beta_1 = 5$. No primeiro deles, assumimos superposição e no outro consideramos uma situação com separação quase-completa. Note que os valores assumidos para os parâmetros, levam à zero a probabilidade condicional de que $Y = 1$, dado $x = 0$, isto é, tem-se que $P[Y = 1 \mid x = 0] \cong 0$.

Para a análise bayesiana, consideramos diferentes distribuições *a priori* normais univariadas para β_0 e β_1 ambas centradas em zero e no verdadeiro valor do parâmetro e com as variâncias variando entre 1 e 1.000. Isto é, definimos distribuições *a priori* mais e menos informativas e temos como objetivo avaliar a influência destas especificações nas estimativas *a posteriori*.

5.2.1.1 Situação de superposição

Os dados gerados na situação de superposição são mostrados na Tabela 5.2.

Tabela 5.2 - Tabela de contingência de Y versus X gerados na situação de superposição

	Y		
X	0	1	Total
0	49	1	50
1	4	46	50
Total	53	47	100

Para este caso, as estimativas de máxima verossimilhança são $\hat{\beta}_0 = -3{,}892$ e $\hat{\beta}_1 = 6{,}334$ com erros padrão das estimativas de 1,01 e 1,137, respectivamente. As estimativas de máxima verossimilhança penalizada são $\hat{\beta}_0 = -3{,}497$ e $\hat{\beta}_1 = 5{,}832$ com erros padrão das estimativas de 0,837 e 0,974, respectivamente. As estimativas de máxima verossimlhança penalizada são mais próximas dos valores reais ($\beta_0 = -3$ e $\beta_1 = 5$) e tem erros-padrão menores que as estimativas de máxima verossimlhança.

A Tabela 5.3 mostra a média e a mediana e também o desvio padrão *a posteriori* para β_0 e β_1 para várias especificações de distribuições *a priori*. Notamos que, entre os modelos com distribuições *a priori* centradas em zero, ou seja, em que, *a priori,* não se está estimando bem os parâmetros (já que $\beta_0 = -3$ e $\beta_1 = 5$), as melhores estimativas (média e mediana) são obtidas quando assumimos uma distribuição *a priori* com variância 10. Este modelo é o que tem as estimativas mais próximas das estimativas de máxima verossimilhança penalizada ($\hat{\beta}_0 = -3{,}497$ e $\hat{\beta}_1 = 5{,}832$). Apesar disto, o DIC aponta como o melhor modelo aquele em que a distribuição *a priori* tem variância 25. Nota-se ainda que o modelo indicado pelo DIC foi o que forneceu estimativas dos coeficientes mais próximas das estimativas de máxima verossimilhança ($\hat{\beta}_0 = -3{,}892$ e $\hat{\beta}_1 = 6{,}334$). É perceptível também que, exceto nos casos em que a distribuição *a priori* é muito

concentrada em torno de zero, as estimativas *a posteriori* tendem a subestimar β_0 e superestimar β_1. Nos casos onde as distribuições *a priori* revelam grande incerteza inicial sobre os parâmetros, a subestimação e a superestimação são ainda maiores. Acontece a mesma coisa com os desvios padrão das estimativas. Quanto maior a incerteza da distribuição *a priori* maior fica a incerteza *a posteriori*.

Como esperado, quando comparamos os modelos com distribuições *a priori* centradas nos verdadeiros valores dos parâmetros, as estimativas *a posteriori* são melhores quando a certeza *a priori* é grande. Isto também foi o indicado pelo DIC, que foi menor para modelo com a variância 1. Este também é o melhor modelo ajustado e foi o que produziu as melhores estimativas para os parâmetros. Note que, neste caso, β_0 é subestimado e β_1 é superestimado.

Tabela 5.3 - Estimativas de parâmetros dos coeficientes da regressão logística para dados simulados na situação superposição.

Distribuição *a priori*	Resultados *a posteriori* para β_0		Resultados *a posteriori* para β_1		DIC
	Média (desvio padrão)	Mediana	Média (desvio padrão)	Mediana	
N(0;1)	-1,93 (0,3669)	-1,914	3,688 (0,493)	3,684	50,670
N(0;10)	**-3,474 (0,7892)**	**-3,386**	**5,862 (0,9201)**	**5,78**	**41,540**
N(0;25)	**-3,894 (0,9683)**	**-3,787**	**6,372 (1,099)**	**6,263**	**41,423**
N(0;1000)	-4,424 (1,268)	-4,211	6,955 (1,365)	6,804	41,783
N(parâmetro;1)	**-3,303 (0,5337)**	**-3,28**	**5,655 (0,6166)**	**5,654**	**40,753**
N(parâmetro;25)	-4,237 (1,074)	-4,106	6,776 (1,197)	6,673	41,476

5.2.1.2 Situação de separação quase-completa

Os dados gerados na situação de separação quase-completa são mostrados na Tabela 5.4.

Tabela 5.4 - Tabela de contingência de Y versus X gerados na situação de separação quase-completa

	Y		
X	0	1	Total
0	50	0	50
1	3	47	50
Total	53	47	100

Neste caso, não existem as estimativas de máxima verossimilhança. As estimativas de máxima verossimilhança penalizada são $\hat{\beta}_0 = -4{,}615$ e $\hat{\beta}_1 = 7{,}223$ com erros padrão das estimativas de 1,435 e 1,540, respectivamente. Perceba que há uma subestimação de β_0 e superestimação de β_1.

Na análise bayesiana (Tabela 5.5), percebemos que, salvo para o caso em que a distribuição *a priori* para ambos os parâmetros é uma normal padrão, a qual fornece as melhores estimativas, em todos os outros casos há uma subestimação de β_0 e superestimação de β_1. Da mesma forma que observamos no caso com superposição, quanto maior a variância *a priori* menor a estimativa de β_0 e maior a estimativa de β_1. No entanto, como mostrado na Tabela 5.5, quando aumentamos a variância a priori o DIC diminui levando-nos a avaliações contraditórias. Pelo DIC, concluímos que quanto menos informativas são as distribuições *a priori*, melhor o ajuste do modelo. Mas observem que as estimativas e seus desvios padrão aumentam muito conforme definimos distribuições *a priori* menos informativas.

Perceba que, quando a variância *a priori* tende para infinito os estimadores bayesianos também crescem muito. Ou seja, quando escolhemos uma distribuição *a priori* não informativa, esta é dominada pelos dados, os quais dão a maior contribuição no cálculo da distribuição *a posteriori*.

Comparando os resultados bayesianos com as estimativas de máxima verossimilhança penalizada, verificamos que a análise bayesiana forneceu melhores resultados, exceto com o modelo centrado no valor do parâmetro e com variância maior.

Tabela 5.5 - Estimativas de parâmetros dos coeficientes da regressão logística para dados simulados com separação quase-completa.

Distribuição *a priori*	Resultados *a posteriori* para β_0		Resultados *a posteriori* para β_1		DIC
	Média (desvio padrão)	Mediana	Média (desvio padrão)	Mediana	
N(0;1)	**-2,053 (0,3852)**	**-2,029**	**3,918 (0,5069)**	**3,909**	**40,811**
N(0;10)	-4,231 (1,063)	-4,128	6,917 (1,156)	6,819	27,605
N(0;25)	-5,407 (1,631)	-5,109	8,182 (1,733)	7,957	26,348
N(0;1000)	-18,68 (10,26)	-17,28	21,59 (10,31)	20,14	24,906
N(parâmetro;1)	**-3,539 (0,5756)**	**-3,506**	**6,07 (0,6589)**	**6,075**	**28,220**
N(parâmetro;25)	-6,458 (2,069)	-6,051	9,215 (2,164)	8,872	25,694

5.2.2 Pacientes submetidos à craniotomia

Os dados do exemplo publicado em Colosimo, Franco e Couto (1995) foram apresentados na Tabela 2.2. Cruzando a resposta presença de meningite com a covariável gravidade do paciente percebeu-se que há separação quase-completa neste bando de dados. Também aqui foram realizadas análises bayesianas utilizando distribuições *a priori* normais com média zero e variâncias diferentes para os coeficientes do modelo, como mostrado na Tabela 5.6.

Como esperado, verificamos que quando utilizamos distribuições *a priori* menos informativas, as estimativas e seus desvios padrão tendem para infinito. Além disto, os valores dos DIC's diminuem cada vez mais.

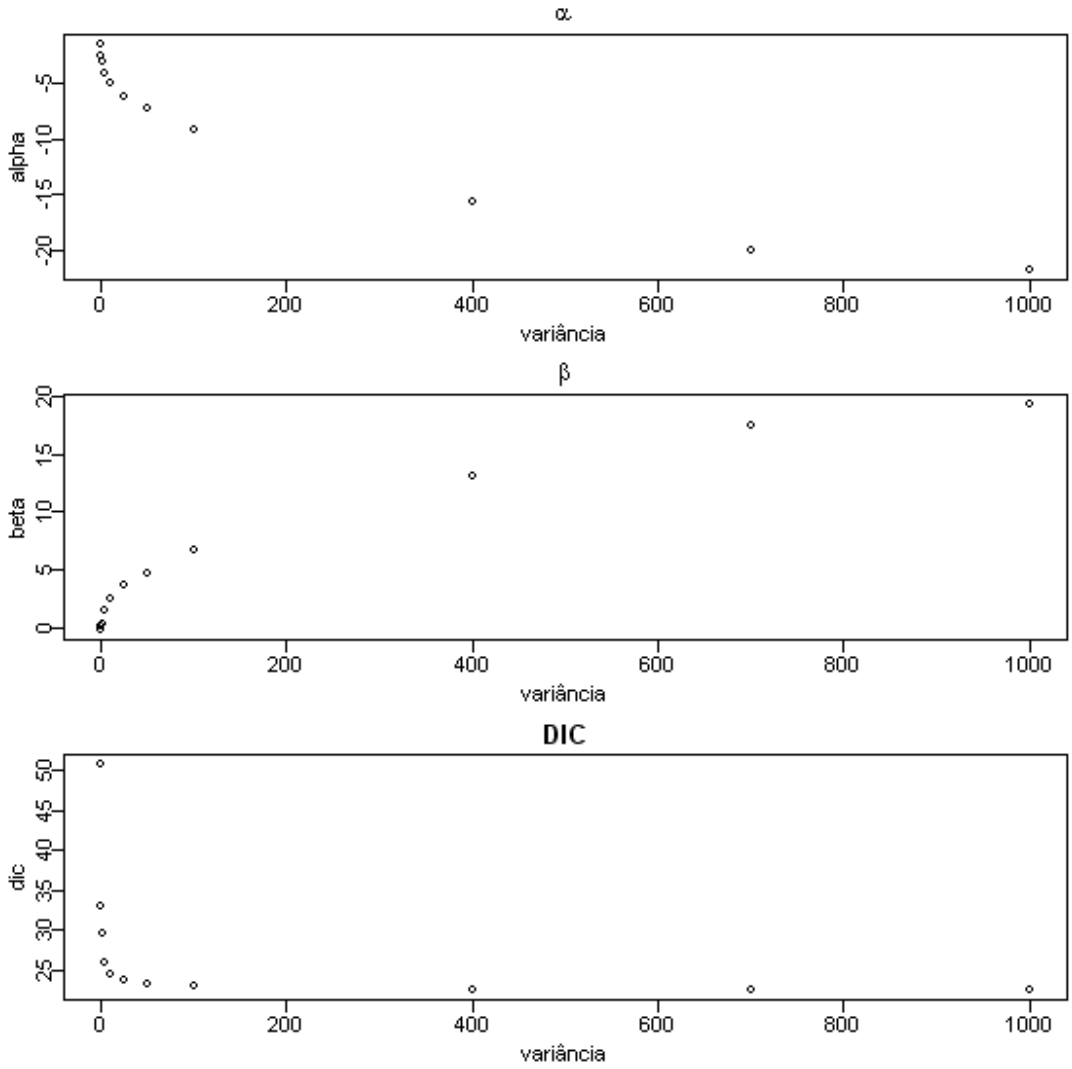
As distribuições *a posteriori* que ficaram com as médias mais próximas das estimativas por máxima verossimlhança penalizada são as que tiveram as distribuições *a priori* com variância 25 para β_0 com variância 10 para β_1.

Tabela 5.6 - Estimativas de parâmetros dos coeficientes da regressão logística para dados de craniotomia.

Distribuição *a priori*	Resultados *a posteriori* para β_0		Resultados *a posteriori* para β_1		DIC
	Média (desvio padrão)	Mediana	Média (desvio padrão)	Mediana	
N(0;0,1)	-1,489 (0,206)	-1,486	-0,2364 (0,2714)	-0,229	50,809
N(0;0,5)	-2,542 (0,3582)	-2,526	0,05712 (0,5109)	0,06874	32,998
N(0;1)	-3,009 (0,4559)	-2,984	0,3755 (0,6517)	0,3916	29,605
N(0;4)	-4,104 (0,7661)	-4,022	1,485 (0,9787)	1,462	25,858
N(0;10)	-5,029 (1,159)	-4,93	**2,425 (1,294)**	**2,363**	24,516
N(0;25)	**-6,175 (1,713)**	**-5,925**	3,596 (1,833)	3,411	23,663
N(0;50)	-7,352 (2,413)	-7,018	4,761 (2,476)	4,481	23,275
N(0;100)	-9,236 (3,727)	-8,538	6,735 (3,761)	6,077	22,927
N(0;400)	-15,64 (8,214)	-13,99	13,11 (8,227)	11,71	22,570
N(0;700)	-19,99 (11,34)	-17,74	17,47 (11,35)	15,31	22,504
N(0;1000)	-21,84 (13,17)	-19,03	19,34 (13,19)	16,59	22,481

Os gráficos da Figura 5.1 mostram os valores das estimativas de β_0, β_1 e também os valores dos DIC's para cada distribuição *a priori* variando os valores da variância *a priori*.

Figura 5.1 - Gráficos das estimativas de β_0 , β_1 e DIC para modelos com distribuições *a priori* normais com média zero e vários valores para a variância.



5.3 Análise usando distribuições *a priori* Bayes-empírica

Como visto na seção anterior, o problema para estimar os coeficientes da regressão na presença de separação pode ser sanado se existir uma quantidade razoável de informação inicial que gere alguma distribuição *a priori* bastante informativa. Caso esta informação seja escassa, o que nos levaria a eliciar uma distribuição *a priori* não informativa, o problema de estimação dos parâmetros do modelo logístico permaneceria.

Ou seja, o enfoque bayesiano para este tipo de problema poderá levar a uma solução adequada apenas em situações muito particulares em que a informação *a priori* exista e seja forte o bastante para não ser tão influenciada pela informação trazida pelos dados, o que não acontece em muitos casos.

Diante deste problema uma alternativa que pode ser atrativa é o uso de métodos bayesianos empíricos para a construção da distribuição a priori para os parâmetros do modelo logístico.

No que segue, foi utilizada a análise de Bayes-empírica para os dados do exemplo de craniotomia. Para construir a distribuição *a priori* para os parâmetros do modelo logístico, digo β_i, foram especificadas, subjetivamente, as distribuições *a priori* beta para os θ_i, ou seja, assumiu-se que $\theta_i \sim$ Beta (a_{1i}, a_{2i}), e foram selecionados três pontos distintos (x_1, x_2) do conjunto de dados, a saber, $(x = 0, x_2 = 2)$, $(x_1 = 1, x_2 = 3)$ e $(x_1 = 1, x_2 = 10)$. Estes pontos foram selecionados por serem representativos em relação ao conjunto de dados. Aqui, também, tem-se como objetivo avaliar a influência das especificações *a priori* nas inferências *a posteriori.*

A Tabela 5.7 mostra as 9 especificações *a priori* para θ_i que serão consideradas neste estudo. Note que alguns destes casos pressupõem a existência de muita informação *a priori* gerando distribuições *a priori* muito informativas para cada θ_i – Caso 1, por exemplo – e outros a quase inexistência de uma informação a priori – Caso 6, por exemplo – o que nos leva a construir distribuições pouco informativas para cada θ_i.

Das distribuições *a priori* beta para os θ_i, foram encontradas as distribuições *a priori* para os coeficientes β_i. O mesmo procedimento foi realizado com outros três pontos selecionados para o mesmo conjunto de dados. São eles: $(x = 0, x_2 = 8)$, $(x_1 = 0, x_2 = 3)$ e $(x_1 = 1, x_2 = 10)$. Nestes casos, foram consideradas as mesmas distribuições *a priori* para cada θ_i mostradas na Tabela 5.7.

Tabela 5.7 – Distribuições *a priori* para cada θ_i.

Caso	θ_i	a_{1i}	a_{2i}	Caso	θ_i	a_{1i}	a_{2i}	Caso	θ_i	a_{1i}	a_{2i}
	θ_1	2	198		θ_1	2	198		θ_1	2	198
1	θ_2	5	45	4	θ_2	0,1	0,9	7	θ_2	5	45
	θ_3	5	5		θ_3	5	5		θ_3	0,1	0,1
	θ_1	0,1	9,9		θ_1	0,1	9,9		θ_1	1	1
2	θ_2	0,1	0,9	5	θ_2	5	45	8	θ_2	0,1	0,9
	θ_3	1	1		θ_3	5	5		θ_3	1	1
	θ_1	2	198		θ_1	1	1		θ_1	0,1	9,9
3	θ_2	5	45	6	θ_2	1	1	9	θ_2	1	1
	θ_3	1	1		θ_3	1	1		θ_3	1	1

Para se ter uma idéia do efeito destas escolhas *a priori* para os θ_i e dos pontos selecionados da amostra nas distribuições *a priori* de β_i, histogramas foram construídos para as distribuições *a priori* de cada β_i, assim como foram avaliadas suas médias e variâncias de acordo com as escolhas dos valores de a_{1i} e a_{2i}. As Figuras 5.2 a 5.4 mostram exemplos dos casos 1, 2 e 6 para a primeira escolha de pontos $(x = 0, x_2 = 2)$, $(x_1 = 1, x_2 = 3)$ e $(x_1 = 1, x_2 = 10)$. O caso 1 é um caso onde as distribuições *a priori* são mais informativas para todos os β_i. No caso 2, já existe bastante informação *a priori* para β_0 e também para β_1, mas menos informação *a priori* sobre os β_2, usando, neste caso, a distribuição uniforme. No caso 6, utilizamos a distribuição *a priori* uniforme para todos os β_i. Verificamos que o coeficiente β_2 sempre tem a distribuição mais concentrada. As distribuições *a priori* para β_0 e β_1, no caso 2, ficaram com uma grande variabilidade.

Figura 5.2 – Histogramas das distribuições *a priori* para β_0, β_1 e β_2 no caso 1

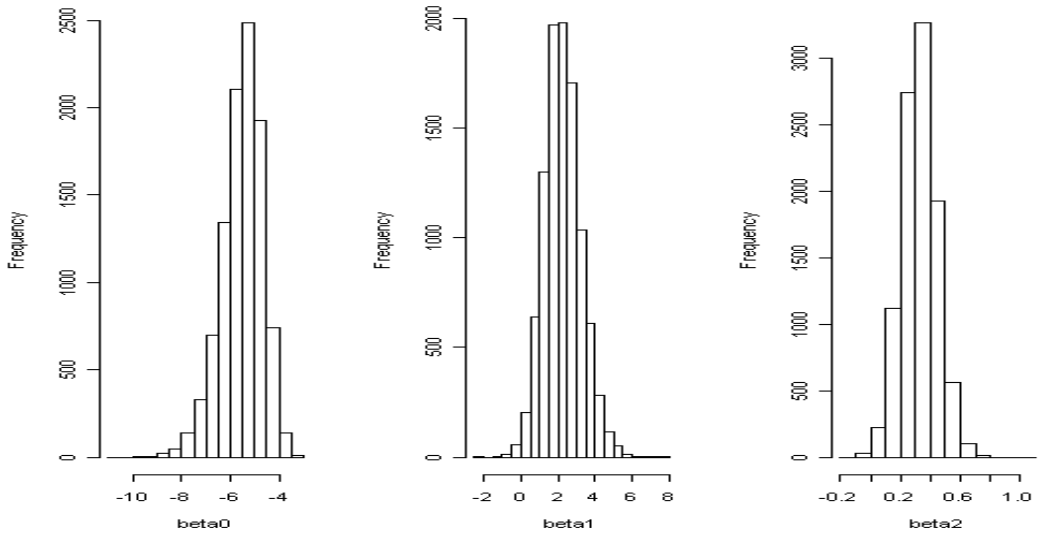


Figura 5.3 – Histogramas das distribuições *a priori* para β_0, β_1 e β_2 no caso 2

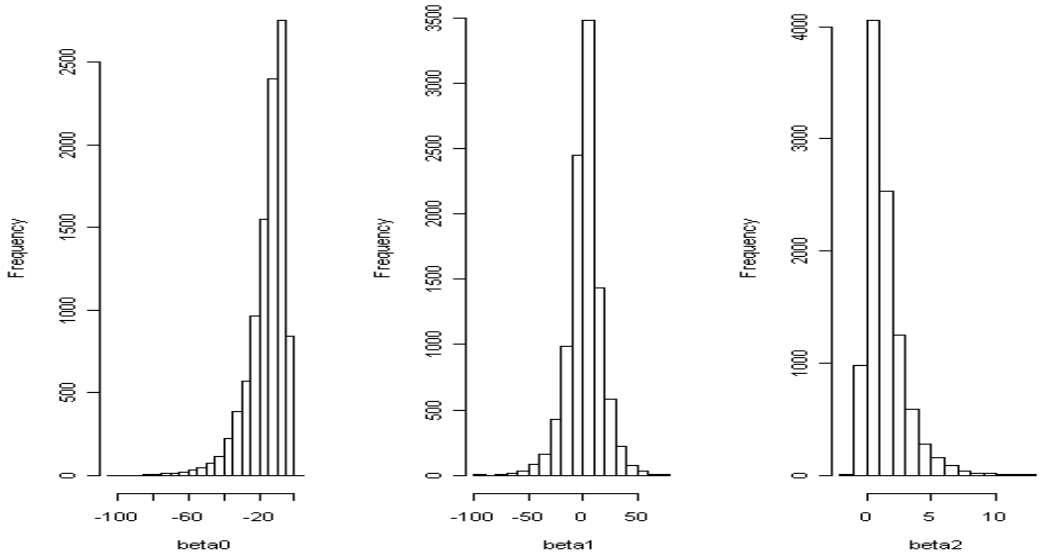
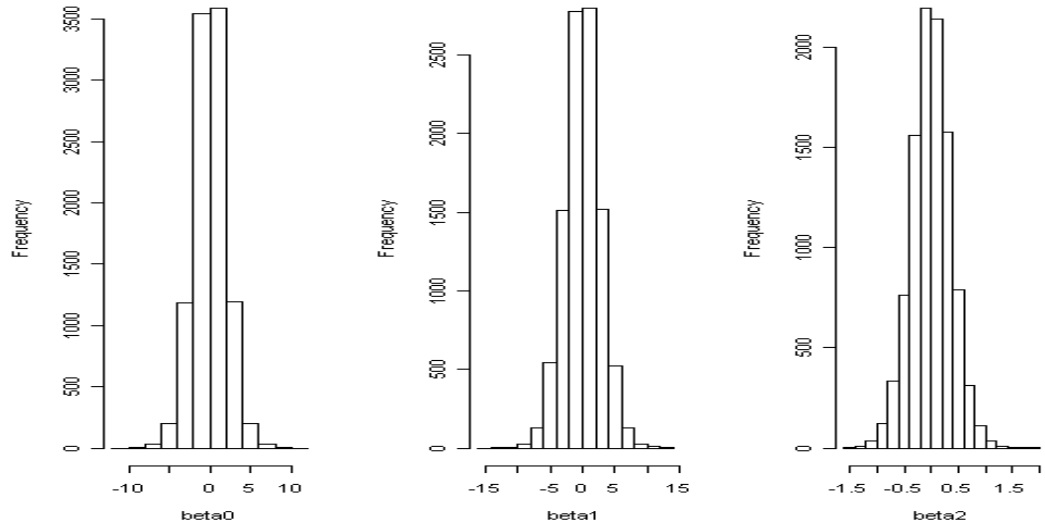


Figura 5.4 – Histogramas das distribuições *a priori* para β_0, β_1 e β_2 no caso 6



A Tabela 5.8 mostra as médias e variâncias para as distribuições *a priori* para β_i nos casos 1, 2 e 6. Veja que a distribuição *a priori* do caso 1 tem os resultados mais próximos das estimativas de máxima verossimilhança penalizada (β_0 =-6,14, β_1 =2,17 e β_2 =0,36).

Tabela 5.8 – Resultados das distribuições *a priori* para β_0, β_1 e β_2 nos casos 1,2 e 6

Caso	Resultados	β_0	β_1	β_2
Caso 1	Média	-5,5277	2,2562	0,3261
	Variância	0,7168	0,9844	0,0139
Caso 2	Média	-15,364	1,6371	1,3731
	Variância	108,15	235,64	2,163
Caso 6	Média	0,0124	-0,0058	-0,0017
	Variância	3,0759	7,5926	0,1324

Considerando as distribuições *a priori* para os β_i, utilizamos o WinBUGS para obter as distribuições *a posteriori* para estes coeficientes. Um total de 110.000 iterações para cada caso foi considerado e, após a convergência ter sido atingida, as 10.000 primeiras interações foram descartadas como período de “burn-in” de 10.000. Algumas

medidas das distribuições *a priori* para β_i são mostradas na Tabela 5.9. São elas as médias, os desvios padrão, as medianas e os DIC's.

Tabela 5.9 – Resumos *a posteriori* de β_i para os 9 casos e para cada conjunto de pontos selecionados.

		(X_1=0, X_2=2), (X_1=1, X_2=3) e (X_1=1, X_2=10)				(X_1=0, X_2=8), (X_1=0, X_2=3) e (X_1=1, X_2=10)			
Caso	β_i	média	desvio padrão	mediana	DIC	média	desvio padrão	mediana	DIC
	β_0	-6,004	0,861	-5,918		-4,887	0,826	-4,849	
	β_1	2,464	0,918	2,396		2,288	0,596	2,285	
1	β_2	0,332	0,104	0,331	15,710	0,152	0,109	0,149	18,210
	β_0	-16,680	10,120	-13,750		-12,360	5,350	-11,130	
	β_1	11,490	10,090	8,538		6,979	5,142	5,650	
2	β_2	0,485	0,244	0,471	15,130	0,509	0,250	0,492	15,230
	β_0	-5,832	0,918	-5,755		-3,719	0,882	-3,670	
	β_1	2,493	0,920	2,429		1,235	0,820	1,260	
3	β_2	0,264	0,162	0,269	16,160	-0,032	0,132	-0,033	20,370
	β_0	-6,360	0,955	-6,269		-9,154	1,730	-9,006	
	β_1	1,695	1,184	1,691		3,695	0,946	3,642	
4	β_2	0,445	0,154	0,436	16,150	0,538	0,196	0,519	**14,440**
	β_0	-16,010	10,040	-13,040		-4,532	0,725	-4,491	
	β_1	12,370	10,040	9,397		0,913	0,723	0,916	
5	β_2	0,350	0,108	0,349	**14,140**	0,306	0,116	0,304	19,700
	β_0	-5,870	1,498	-5,699		-6,133	1,393	-5,999	
	β_1	2,255	1,442	2,100		0,649	1,166	0,644	
6	β_2	0,285	0,188	0,284	17,930	0,530	0,210	0,519	18,960
	β_0	-5,625	0,986	-5,559		-3,165	0,852	-3,114	
	β_1	2,527	0,932	2,459		0,588	0,955	0,640	
7	β_2	0,181	0,220	0,194	16,850	-0,132	0,141	-0,132	22,420
	β_0	-6,219	1,582	-6,040		-7,710	1,944	-7,489	
	β_1	1,788	1,507	1,666		1,375	1,501	1,288	
8	β_2	0,367	0,209	0,362	17,510	0,650	0,255	0,630	17,760
	β_0	-16,100	10,070	-13,150		-6,630	1,700	-6,452	
	β_1	12,080	10,060	9,110		2,166	1,404	2,040	
9	β_2	0,359	0,204	0,354	15,250	0,369	0,204	0,362	16,450

As médias das distribuições *a posteriori* em alguns casos são bem diferentes que em outros e em alguns casos o desvio padrão foi muito grande. O menor DIC para a primeira escolha de pontos foi o do caso 5, o qual teve desvios padrão elevados para β_0 e β_1, assim como os valores das médias e medianas. Para a segunda escolha de pontos, o menor DIC foi o do caso 4.

Intervalos de credibilidade percentílicos para os coeficientes β_i foram construídos e são mostrados na Figura 5.5. Veja que nos casos 2, 5 e 9 do primeiro conjunto de pontos os intervalos para β_0 ficaram muito grandes.

Já no segundo conjunto de pontos, o intervalo para β_0 do caso 2 é que ficou muito grande, embora não tanto quanto os intervalos do primeiro conjunto de pontos.

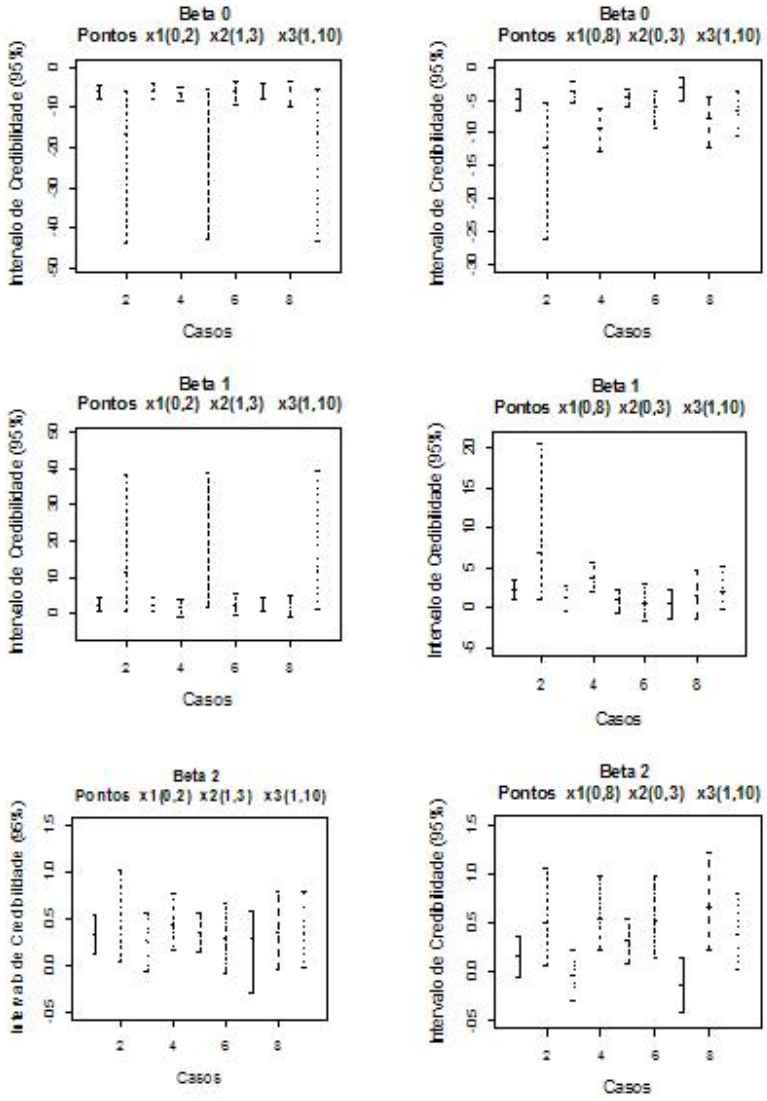
Quando observamos a primeira escolha de pontos, verificamos novamente os mesmos casos 2, 5 e 9 nos quais os intervalos para β_1 também ficaram muito grandes. E, verificamos novamente que, no caso 2, o intervalo para β_1 ficou grande para a segunda escolha de pontos.

Verificamos ainda que os intervalos de credibilidade para β_2 se comportam bem para todos os casos. Ou seja, parece que ele não é sensível à especificação das distribuições *a priori*. O coeficiente β_2 é o coeficiente da variável X_2, tempo de cirurgia, que não está associado à covariável que gera a separação.

No primeiro conjunto de pontos, as piores situações, isto é, os intervalos muito grandes, estão associadas à falta de informação para θ_1, com uma distribuição *a priori* beta (0,1; 9,9). No segundo conjunto de pontos, o intervalo grande também está no caso 2 que tem a mesma distribuição para θ_1.

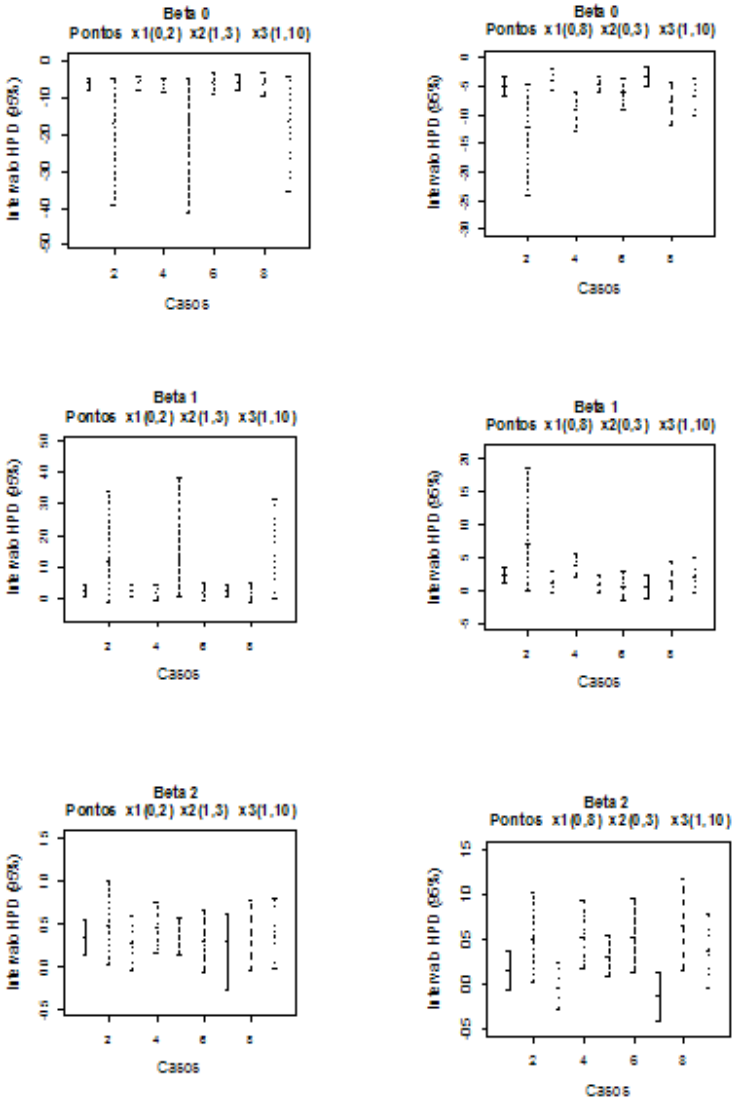
Parece então que o problema é a definição da beta (0,1; 9,9) como distribuição *a priori* para θ_1. Até quando utilizamos a distribuição uniforme, beta (1; 1), para θ_1 não ocorreu este problema.

Figura 5.5 – Intervalos de credibilidade percentílicos para as 9 distribuições *a posteriori* para β_0, β_1 e β_2.



Como mostra a Figura 5.6, fizemos também os intervalos de mais alta densidade, HPD, para os nove casos das duas escolhas de pontos. Os intervalos HPD são mais curtos que os intervalos percentílicos. Mas apesar disto, eles não ficaram tão diferentes.

Figura 5.6 – Intervalos HPD para as 9 distribuições *a posteriori* para β_0, β_1 e β_2.



Capítulo 6

Conclusões

Assim como Colosimo, Franco e Couto (1995), verificamos que as variáveis X_1, gravidade do caso, e X_2, tempo de cirurgia não influenciam na ocorrência de meningite como complicação da cirurgia de craniotomia. Para entendermos este resultado, devemos notar que dos pacientes que tiveram meningite um deles ficou 10h em cirurgia e o outro apenas 1h30. Um teve um tempo muito grande e outro teve o tempo muito menor e, apesar desta diferença, ambos desenvolveram meningite. Com relação à covariável gravidade da doença, muitos pacientes (32) tinham gravidade alta e não desenvolveram meningite.

Ao utilizar inferência bayesiana para estimar os parâmetros do modelo de regressão logística verificamos que é muito importante a especificação da distribuição *a priori* para os coeficientes quando temos um caso de separação quase-completa. Se tivermos alguma informação inicial que possa ser utilizada para eliciar a distribuição *a priori,* a distribuição *a posteriori* será melhor estimada. Mas, se não tivermos esta informação inicial, a utilização de distribuições não informativas fazem com que os dados com separação se sobressaiam e persistam os problemas de estimação.

Quanto à utilização da análise Bayes-empírica, verificamos que ela é uma boa saída quando não se tem informação suficiente para definir uma distribuição *a priori* mais informativa, o que não ajuda muito no caso de separação. Com base nos próprios dados, a análise Bayes-empírica é uma forma de chegar a uma distribuição *a priori* mais informativa, que possibilita resultados melhores.

Utilizamos a estimação por máxima verossimilhança penalizada e verificamos que quando há informação *a priori* para utilizar nas estimativas bayesianas, essas são melhores. Mas quando as distribuições *a priori* são não informativas, as estimativas por máxima verossimilhança penalizada podem ser melhores.

Análises futuras deverão ser realizadas para avaliar a especificação das distribuições *a priori* em outros exemplos. Como outros autores, verificamos que esta é a chave para a boa utilização da estatística bayesiana. Com uma boa definição da

distribuição *a priori* poderemos chegar a melhores resultados que com a estimação por máxima verossimilhança penalizada, no caso de separação quase-completa.

Referências

Albert A., Anderson, J.A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika*, 71:1-10.

Agresti, A. (2006?) *Bayesian Inference for Categorical Data Analysis: A Survey*. URL: http://www.stat.ufl.edu/~aa/cda/bayes.pdf.

Bedrick, E.J., Christensen R., Johnson, W. (1996). A New Perspective on Priors for Generalizes Linear Models. *Journal of the American Statistical Association*, 91:1450-1460.

Breiman, L., Friedman, J.H., Olshen, R.A. Stone, C.J. (1984). *Classification and Regression Trees*. Monterey: Wadsworth and Brooks/Cole.

Cary, N.C. (1985). *Statistical analysis system: guide for personal computer – version 6*, North Carolina: SAS Institut Inc. URL: http://www.sas.com.

Casella G., Berger R.L. (2002). *Statistical Inference*. USA: Duxbury, Thomson Learning.

Clarkson, D.B., Jenrick, R.I. (1991). Computing extended maximum likelihood estimates for linear parameter models. *Journal of the Royal Statistical Society*. Series B; 53:417-426.

Colosimo, E.A., Franco, G.C., Couto, B.M. (1995). The Logistic Regression Model and Rare Events. *Estadística*, 47, 148, 149:1-16.

Congdon, P. (2001). *Bayesian Statistical Modelling*. Chichester, UK: John Wiley.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80:27-38.

Galindo-Garre, F., Vermunt, J.K., Bergsma, W.P. (2004). Bayesian Posterior Estimation of Logit Parameters With Small Samples. *Sociological Methods & Research*, 33:88-117.

Greenland, S. (2001). Putting Background Information About Relative Risks into Conjugate Prior Distributions. *Biometrics*, 57:663-670.

Heinze, G., Schemper M. (2002). A solution to the problem of separation in logistic regression. *Statistics in Medicine*, 21:2409-2419.

Hosmer, D.W., Lemeshow, S. (2000). *Applied Logistic Regression*. New York: John Wiley.

Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing;* 10:325-337. URL: http://www.mrc-bsu.cam.ac.uk/bugs.

Nacle, D.P. (2004). *Estimadores de Máxima Verossimilhança em Modelos de Regressão Logística na Situação de Separação Quase-Completa*. Belo Horizonte: Dissertação de Mestrado (Departamento de Estatística) – UFMG.

Paulino, C.D., Turkman, M.A.A., Murteira, B. (2003). *Estatística Bayesiana*. Lisboa: Fundação Galouste Gulbenkian,

R Development Core Team (2006). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. ISBN 3-900051-07-0, URL: http://www.R-project.org.

Santner, T.J., Duffy, D.E.A. note on A. Albert and J.A. (1986). Anderson's conditions for the existence of maximum likelihood estimates in logistic regression models. *Biometrika*, 73:755-758.

Silvapulle, M.J. (1981). On the Existence of Maximum Likelihood Estimates for the Binomial Response Models. *Journal of the Royal Statistical Society*, Series B, 43:310-313.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society*, Series B, 64: 583-640.

Tsutakawa, R.K., Lin, H.Y. (1986). Bayesian estimation of item response curves, *Psychometrica*, 51:251-267.

Zorn, C. (2002). A Solution to Separation in Binary Response Models. *Political Analysis*, 13:157-170.

Apêndice: Programas utilizados

Programa em R para regressão logística e gráficos de verossimilhança

```
lvlogis <- function(beta, x, y, s){
 pred <- x%*%beta
 prob <- exp(pred)/(1+exp(pred))
 -sum(sapply(1:length(y), function(i)
   dbinom(y[i], 1, prob[i], log=T)))
}

dad <- read.table("E:/dissertação/craniotomia3.txt", dec=",")
summary(dad)

covar <- cbind(1, dad[,2])

summary(glm(V1~V3, family=binomial, data=dad))
optim(c(-1, 0), lvlogis, x=covar, y=dad[,3])

res1 <- optim(c(0, 0), lvlogis, x=covar, y=dad[,3], hessian=T)
res1
solve(res1$hess)
sqrt(diag(solve(res1$hess)))

b0 <- seq(-100,50,leng=51)
b1 <- seq(-100,1000,leng=101)

vero <- sapply(b1, function(b) sapply(b0, function(a)
  lvlogis(c(a, b), covar, dad[,1])))
dim(vero)

jpeg("F:/dissertação/veroB0B1.jpg")
image(b0, b1, vero, col=terrain.colors(51))
contour(b0, b1, vero, add=T)
dev.off()
res1$val

vb0 <- sapply(b0, function(b)
  lvlogis(c(b, res1$par[2]), covar, dad[,1]))
jpeg("F:/dissertação/veroB0dados.jpg")
plot(b0, -vb0, type="l", xlab=expression(paste(beta, 0)),
     ylab="Verossimilhança")
dev.off()

vb1 <- sapply(b1, function(b)
  lvlogis(c(res1$par[1], b), covar, dad[,1]))

jpeg("F:/dissertação/veroB1dados.jpg")
plot(b1, -vb1, type="l", xlab=expression(paste(beta, 1)),
     ylab="Verossimilhança")
dev.off()

res1$par
```

Programa em R para gerar distribuições *a priori* para β_i

```
# Para x1=(0,2), x2=(1,3), x3=(1,10)

tetha1<-rbeta(10000,0.1,9.9)
tetha2<-rbeta(10000,1,1)
tetha3<-rbeta(10000,1,1)

L1<-log(tetha1/(1-tetha1))
L2<-log(tetha2/(1-tetha2))
L3<-log(tetha3/(1-tetha3))

beta2<-(L3-L2)/7
beta0<-L1-2*beta2
beta1<-(L2-L1)-beta2

media<- c("beta0"= mean(beta0), "beta1"= mean(beta1), "beta2"= mean(beta2))
media
var<- c("beta0"= var(beta0), "beta1"= var(beta1), "beta2"= var(beta2))
var

par(mfrow=c(1,3))
hist(beta0)
hist(beta1)
hist(beta2)

# Para x1=(0,8), x2=(0,3), x3=(1,10)

tetha1<-rbeta(10000,2,198)
tetha2<-rbeta(10000,0.1,0.9)
tetha3<-rbeta(10000,5,5)

L1<-log(tetha1/(1-tetha1))
L2<-log(tetha2/(1-tetha2))
L3<-log(tetha3/(1-tetha3))

beta2<-(L1-L2)/5
beta0<-L1-8*beta2
beta1<-(L3-L1)-2*beta2

media<- c("beta0"= mean(beta0), "beta1"= mean(beta1), "beta2"= mean(beta2))
media
var<- c("beta0"= var(beta0), "beta1"= var(beta1), "beta2"= var(beta2))
var

par(mfrow=c(1,3))
hist(beta0)
hist(beta1)
hist(beta2)
```

Programa para gerar distribuições *a posteriori* para β_i no WinBUGS

```
#MODELO

model
{
#CASO 1
#alpha1<-2
#gama1<-198
#alpha2<-5
#gama2<-45
#alpha3<-5
#gama3<-5


  for(i in 1: N)
  {
  y[i] ~ dbin(p[i], n[i])
  logit(p[i])<- beta0+beta1*x1[i]+beta2*x2[i]
   }


#Priori (trick using  zeros to construct a prior for beta vector)

#Pontos  x1(0,2)  x2(1,3)   x3(1,10)
#theta0<- pow((exp(beta0+2*beta2)), alpha1)/ (pow(1+(exp(beta0+2*beta2)), (alpha1+gama1)))
#theta1<- pow((exp(beta0+beta1+3*beta2)), alpha2)/ (pow(1+(exp(beta0+beta1+3*beta2)),
(alpha2+gama2)))
#theta2<- pow((exp(beta0+beta1+10*beta2)), alpha3)/ (pow(1+(exp(beta0+beta1+10*beta2)),
(alpha3+gama3)))

#Pontos  x1(0,8)  x2(0,3)   x3(1,10)
theta0<- pow((exp(beta0+8*beta2)), alpha1)/ (pow(1+(exp(beta0+8*beta2)), (alpha1+gama1)))
theta1<- pow((exp(beta0+3*beta2)), alpha2)/ (pow(1+(exp(beta0+3*beta2)), (alpha2+gama2)))
theta2<- pow((exp(beta0+beta1+10*beta2)), alpha3)/ (pow(1+(exp(beta0+beta1+10*beta2)),
(alpha3+gama3)))

beta0~dflat()  # 'flat' prior
beta1~dflat()  # 'flat' prior
beta2~dflat()  # 'flat' prior

#priori <- theta0*theta1*theta2*39402*9534420*630              #CASO 1

zeros <- 0
zeros~dpois(phi)  # likelihood is exp(−phi)
phi<- -log(priori)  # expression (desired prior for beta )

}

#DADOS

list(
y=c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1),
n=c(1,1,2,1,3,4,3,1,8,3,1,4,1,1,1,1,1,1,6,1,1,1,3,4,8,8,1,13,1,1,1,1,2,1,1,1,3,3,2,1),
x1=c(0,1,1,0,0,0,0,1,0,0,1,0,1,0,0,1,0,1,1,0,0,1,1,1,1,0,0,0,1,0,0,0,0,0,0,0,0,0,1,1),
```

```
x2=c(2.50, 1.33, 6.00, 4.50, 1.50, 1.33, 5.00, 0.75, 2.00, 3.50, 3.25, 1.83, 7.00, 1.67, 8.00, 3.50,
3.17, 5.50, 2.00, 1.25, 2.17, 6.50, 1.00, 4.00, 3.00, 4.00, 4.75, 3.00, 8.00, 5.50, 2.67, 2.25, 7.00,
3.67, 2.33, 6.50, 1.00, 6.00, 1.50, 10.00), N = 40)


#VALORES INICIAIS

#VALORES INICIAIS PARA A CADEIA 1
list(beta0 = 0, beta1=0, beta2=0)

#VALORES INICIAIS PARA A CADEIA 2
list(beta0 = -1, beta1=0, beta2=0)
```

Programa no R para gerar intervalos de credibilidade das distribuições *a posteriori* para β_0

```
# CRANIOTOMIA (PARA BETA 0)

# Pontos  x1(0,2)  x2(1,3)   x3(1,10)
# ==================================

#Caso 1
plot(c(0,10),c(-49,0), type='n',xlim=range(c(0.5,9.5)),cex.lab=1.2, xlab="Casos", ylab="Intervalo de
Credibilidade (95%)")
segments(1,-7.932,1,-4.566,lty=1)
lines(c(0.90, 1.1), c(-7.932, -7.932))
lines(c(0.90, 1.1), c(-4.566, -4.566))
lines(c(0.95, 1.05), c(-6.004, -6.004))
title("Beta 0 \n Pontos  x1(0,2)  x2(1,3)   x3(1,10)")


#Caso 2
segments(2,-43.60,2,-5.798,lty=2)      #Caso 2, percentil 2.5, percentil 97.5
lines(c(1.9,2.1), c(-43.6,-43.6))      #percentil 2.5
lines(c(1.9, 2.1), c(-5.798,-5.798))   #percentil 97.5
lines(c(1.95, 2.05), c(-16.68,-16.68))   # mean


#Caso 3
segments(3,-7.857,3,-4.253,lty=3)
lines(c(2.9, 3.1), c(-7.857,-7.857))
lines(c(2.9, 3.1), c(-4.253,-4.253))
lines(c(2.95, 3.05), c(-5.832, -5.832))

#Caso 4
segments(4,-8.479,4,-4.754,lty=4)
lines(c(3.9, 4.1), c(-8.479,-8.479))
lines(c(3.9, 4.1), c(-4.754,-4.754))
lines(c(3.95, 4.05), c(-6.360,-6.360))

#Caso 5
segments(5,-42.75,5,-5.517,lty=5)
```

```
lines(c(4.9, 5.1), c(-42.75,-42.75))
lines(c(4.9, 5.1), c(-5.517,-5.517))
lines(c(4.95, 5.05), c(-16.01, -16.01))

#Caso 6
segments(6,-9.312,6,-3.436,lty=6)
lines(c(5.9, 6.1), c(-9.312,-9.312))
lines(c(5.9, 6.1), c(-3.436,-3.436))
lines(c(5.95, 6.05), c(-5.870,-5.870))

#Caso 7
segments(7,-7.758,7,-3.878,lty=7, lwd=2)
lines(c(6.9, 7.1), c(-7.758,-7.758))
lines(c(6.9, 7.1), c(-3.878,-3.878))
lines(c(6.95, 7.05), c(-5.625,-5.625))

#Caso 8
segments(8,-9.81,8,-3.626, lty=8, lwd=2)
lines(c(7.9, 8.1), c(-9.81,-9.81))
lines(c(7.9, 8.1), c(-3.626,-3.626))
lines(c(7.95, 8.05), c(-6.219,-6.219))

#Caso 9
segments(9,-43.19,9,-5.426,lty=9, lwd=2)
lines(c(8.9, 9.1), c(-43.19,-43.19))
lines(c(8.9, 9.1), c(-5.426,-5.426))
lines(c(8.95, 9.05), c(-16.1,-16.1))
```

Programa no R para gerar estimativas de máxima verossimilhança penalizada através do pacote brglm

```
cran<-read.table("craniotomia.txt",h=T)
attach(cran)

cran2 <- rbind(cbind(case=1, cran[rep(1:40, cran$r), 3:4]),
       cbind(case=0, cran[rep(1:40, cran$n-cran$r), 3:4]))

require(brglm)

fit <- brglm(case ~ .,family=binomial, data=cran2)

summary(fit)


# If for the data set we would like to test the specific hypothesis ßx1 = 2, ßx2 = 0,
# we do as follows:

logistftest(case ~ ., cran2, test = ~ x1 + x2 - 1, values = c(2, 0))


# If -1 is not included in the formula, the intercept would be tested, too!

logistftest(case ~ ., cran2, test = ~ x1 + x2, values = c(1, 2, 0))
```

```
#To test only the intercept specify test = ~ - .or test = 1

logistftest(case ~ ., cran2, test = 1) # para testar intercept = 0
logistftest(case ~ ., cran2, test = 1, values = c(2))  # para testar intercept = 2


#Testing the overall null hypothesis of ßi = 0 would be performed by the following call:

logistftest(case ~ ., data=cran2)


# This function plots the profile likelihood of a specific parameter.
# In our example we get the profile of the parameter ßdia as follows:

logistfplot(case ~ ., data=cran2, which= ~ x1 - 1)  # plot of the x1
logistfplot(case ~ ., data=cran2, which= ~ x2 - 1)  # plot of the x2
logistfplot(case ~ ., data=cran2, which= ~ 1)       # plot of the intercept
```

Programa no R para gerar estimativas de máxima verossimilhança penalizada através do pacote logistf

```
cran<-read.table("craniotomia.txt",h=T)
attach(cran)
cran2 <- rbind(cbind(case=1, cran[rep(1:40, cran$r), 3:4]),
       cbind(case=0, cran[rep(1:40, cran$n-cran$r), 3:4]))

require(logistf)

fit <- logistf(case ~ ., data=cran2)

summary(fit)


# If for the data set we would like to test the specific hypothesis ßx1 = 2, ßx2 = 0,
# we do as follows:

logistftest(case ~ ., cran2, test = ~ x1 + x2 - 1, values = c(2, 0))



# If -1 is not included in the formula, the intercept would be tested, too!

logistftest(case ~ ., cran2, test = ~ x1 + x2, values = c(1, 2, 0))


#To test only the intercept specify test = ~ - .or test = 1

logistftest(case ~ ., cran2, test = 1) # para testar intercept = 0
logistftest(case ~ ., cran2, test = 1, values = c(2))  # para testar intercept = 2


#Testing the overall null hypothesis of ßi = 0 would be performed by the following call:
```

```
logistftest(case ~ ., data=cran2)


# This function plots the profile likelihood of a specific parameter.
# In our example we get the profile of the parameter ßdia as follows:

logistfplot(case ~ ., data=cran2, which= ~ x1 - 1)  # plot of the x1
logistfplot(case ~ ., data=cran2, which= ~ x2 - 1)  # plot of the x2
logistfplot(case ~ ., data=cran2, which= ~ 1)       # plot of the intercept
```