

UM ALGORITMO PARA MODELOS DE LIMIAR USANDO AS DISTRIBUIÇÕES ACUMULADAS NORMAL E “T” DE STUDENT

José Waldemar da SILVA¹
Júlio Sílvio de Sousa BUENO FILHO²

- RESUMO: Dados categorizados podem ser modelados por meio de uma variável subjacente ou latente contínua (L), que relaciona as observações da variável resposta aleatória categórica (Y) à pertinência em intervalo correspondente. A função de ligação utilizada é a diferença entre valores de uma função acumulada nos limites de intervalo. Pode-se então estabelecer um modelo linear para a esperança condicional de L . O amostrador de Gibbs foi o primeiro algoritmo proposto para a obtenção de amostras das distribuições conjuntas a *posteriori*, usando a distribuição t de Student, para o modelo fixo. O mesmo algoritmo foi implementado por Sørensen et al. (1995) para modelos mistos e com distribuição normal para L . Neste trabalho ajustou-se um modelo misto em duas situações, na primeira utilizou-se a distribuição t de Student como função de ligação e na segunda a distribuição normal. As duas distribuições foram comparadas quanto à convergência. O estudo foi exemplificado com um experimento de análise sensorial com três concentrações de sacarose no processo de desidratação de frutos de banana da terra. A cor do produto foi avaliada em uma escala hedônica de nove pontos. Não houve evidência de diferença entre os tempos de convergência para as duas funções de ligação utilizadas. O modelo com função de ligação t de Student tem melhor ajuste, sendo mais provável segundo o fator de Bayes.
- PALAVRAS-CHAVE: Amostrador de Gibbs; modelos de limiar; modelos probit; ligação normal; ligação t.

1 Introdução

Muitos atributos na experimentação, são avaliados por meios subjetivos. Juízes, técnicos, provadores ou consumidores são escolhidos ou designados para

¹Universidade Federal de Uberlândia – UFU, Faculdade de Matemática, Caixa Postal 593, CEP: 38400-902, Uberlândia, Minas Gerais, Brasil. E-mail: zewaldemar@famat.ufu.br

²Universidade Federal de Lavras – UFLA, Departamento de Ciências Exatas, Caixa Postal 3037, CEP: 37200-000, Lavras, Minas Gerais, Brasil. E-mail: jssbueno@ufla.br

atribuir um conceito ao material experimental de acordo com suas preferências ou de acordo com algum regulamento.

Em geral, para a obtenção dos dados adota-se um conjunto de categorias ordinais nas quais os produtos possam ser classificados do pior ao melhor, de acordo com o julgamento de cada avaliador, diferentemente dos casos em que se tem uma variável contínua ou na forma de contagem.

Ainda é muito comum o uso da análise de variância sem a preocupação com a validade dos testes aplicados, apesar de os dados descritos não seguirem as pressuposições necessárias para o uso desta metodologia.

Outro ponto a ser destacado é a dificuldade ou impossibilidade encontrada, em muitos casos, para quantificar uma característica por meio de medição. Podem ser citados alguns exemplos típicos para esta situação, como a seleção para algumas características morfológicas de touros, no melhoramento genético de gado de corte ou leiteiro, tais como escore corporal e a classificação do animal quanto ao padrão racial, a avaliação do grau de dificuldade do parto de matrizes reprodutoras e o grau de satisfação do consumidor com relação ao sabor de um produto alimentício a ser lançado no mercado. Vários outros exemplos poderiam ser mencionados, dada a grande quantidade de características de importância econômica existentes que são subjetivas, categóricas e ordinais. Uma descrição do estado da arte na análise frequentista para este tipo de dados pode ser encontrado em Agresti (2002). Esta análise envolve o ajuste de modelos com parâmetros *threshold*.

Wright (1934) foi o primeiro a propor esta metodologia, no estudo da herança do número de dedos de *Cavia porcellus* (porquinhos da Índia). Ainda é muito comum, no entanto, aplicar a análise da variância usual, com ou sem transformação de dados, apesar de os dados não seguirem um modelo linear Gauss Markov normal (Steel e Torrie, 1980).

A literatura recente sobre análise de dados categorizados subjetivamente, sugere o ajuste de modelos com parâmetros *threshold* ou de limiar, no qual, uma variável latente, com distribuição contínua, é especificada de forma que uma resposta é observada em uma dada categoria se o valor desta variável está entre os limites que definem tal categoria.

Albert e Chib (1993) propuseram a utilização da amostragem Gibbs para obter aproximação das distribuições marginais dos parâmetros, em modelos *thresholds*. Sørensen et al. (1995) utilizaram a mesma metodologia, porém na área da Genética Quantitativa Animal em que o Modelo Linear Generalizado (GLM) correspondente usava uma função de ligação normal acumulada.

Podem ser citados outros exemplos de uso desta metodologia como os trabalhos de Heringstad et al. (2004), Vazquez et al. (2009) e Piepho e Kalka, (2002), entre outros. No primeiro ajustaram-se um modelo de limiar de forma multivariada para inferir sobre herdabilidade da mastite e correlação genética em rebanho leiteiro. No segundo confrontaram um modelo *threshold* e um modelo Poisson também em dados sobre mastite em gado de leite e observou que o modelo de limiar foi mais apropriado para identificar pais que provavelmente terão descendentes saudáveis. No terceiro fez-se o confronto de um modelo *threshold* e análise de variância e

destacaram a grande flexibilidade dos modelos de limiar, para o ajuste a dados categóricos ordinais.

A variável latente facilita a implementação computacional, pois, pode ser modelada de forma linear. Esta variável pode ser interpretada como o estímulo que deve ser aplicado a um indivíduo, objeto ou produto, para que este possa ser classificado em uma das K categorias. A área da análise sensorial dos alimentos é um exemplo disto pois, muitas características são difíceis ou impossíveis de serem quantificadas por meio de medida objetiva. Nestes casos, cada provador ou juiz classifica um produto em alguma categoria, por meio do seu sabor, cor, textura, etc. Estas variáveis são dotadas de uma escala contínua, porém, a execução do experimento torna-se mais fácil quando são fragmentadas em categorias.

A forte autocorrelação entre as estimativas, principalmente dos parâmetros de limiar, obtidos por meio da amostragem Gibbs, usando esta metodologia, é evidente e foi destacada por Sørensen et al. (1995) e Sørensen e Gianola (2004). Este problema pode ser solucionado obtendo uma cadeia muito grande para que se tenha uma boa representação da distribuição marginal do parâmetro.

A escolha de um bom modelo e a conseqüente maior qualidade no processo de inferência podem ser quantificados por alguns indicadores como a maior razão de verossimilhanças marginais dos modelos que se pretende comparar, a maior acurácia nas estimativas ou a melhor interpretabilidade de seus parâmetros.

O algoritmo para a análise de modelos mistos com função de ligação dada por uma distribuição normal acumulada foi desenvolvida por Sørensen et al. (1995). O mesmo algoritmo porém, para modelos fixos com função de ligação t de Student acumulada, é apresentado por Albert e Chib (1993).

Este trabalho teve como objetivo adaptar o algoritmo de Albert e Chib (1993) para os modelos mistos e contrastá-lo com o algoritmo de Sørensen et al. (1995) quanto às funções de ligação utilizadas e quanto à convergência. A aplicação da metodologia será feita em um experimento de análise sensorial de alimentos.

2 Material e métodos

Alguns atributos na análise sensorial de alimentos são avaliados por meio da escala hedônica (E. H.) de nove pontos, conforme Tabela 1.

Porém, a existência de um grande número de categorias para a classificação pode dificultar a decisão do provador quanto a qual nota atribuir ao produto. Além disso, o acréscimo do número de parâmetros de limiar estimados pode dificultar a convergência do algoritmo e sob o aspecto da representatividade, Kizilkaya et al. (2003) sugerem o agrupamento de categorias com poucas observações.

Os dados utilizados para a ilustração da metodologia proposta neste trabalho foram obtidos pelo Instituto Federal de Educação, Ciência e Tecnologia Goiano, campus de Rio verde-GO durante a realização da Exposição Agropecuária deste município no ano de 2006. Obteve-se a variável resposta por meio de notas atribuídas por 36 provadores a amostras de banana da terra desidratada. Os tratamentos avaliados nesta análise foram três diferentes concentrações de sacarose

Tabela 1 - Escala hedônica de nove pontos

Categorias	E. H. de 9 Pontos
Desgostei muitíssimo	1
Desgostei muito	2
Desgostei	3
Desgostei moderadamente	4
Nem gostei nem desgostei	5
Gostei moderadamente	6
Gostei	7
Gostei muito	8
Gostei muitíssimo	9

Fonte: Dutcosky (1996).

utilizadas no processo de desidratação osmótico, sendo o primeiro deles uma solução a 30%, o segundo uma solução a 40% e o último, solução a 50% de sacarose.

Os provadores incluíam crianças e adultos de ambos os sexos, não treinados. Empregou-se a escala hedônica de nove pontos, cujo ponto 1 correspondeu a “desgostei muitíssimo” e o 9 a “gostei muitíssimo” (Dutcosky, 1996). A característica analisada foi a cor da banana da terra desidratada. O delineamento experimental adotado foi em blocos ao acaso em que cada provador avaliou os três tratamentos e, assim, este constituía um bloco completo.

Seguindo recomendação de Kizilkaya et al. (2003) as categorias foram reagrupadas de tal forma que em cada uma tivesse no mínimo cinco observações. O esquema de reorganização está descrito na Tabela 2.

Tabela 2 - Esquema de reclassificação dos dados em cinco categorias

Categorias	E. H. de 9 Pontos*	E. H. de 5 Pontos
Desgostei muitíssimo	1	
Desgostei muito	2	
Desgostei	3	1
Desgostei moderadamente	4	
Nem gostei nem desgostei	5	
Gostei moderadamente	6	2
Gostei	7	3
Gostei muito	8	4
Gostei muitíssimo	9	5

(*)Fonte: Dutcosky (1996).

O conhecimento prévio sobre um fenômeno e a informação fornecida por experimentos podem ser combinados por meio do Teorema de Bayes, usando a seguinte expressão:

$$p(\theta|Y) \propto p(\theta)p(Y|\theta),$$

em que θ é o vetor de parâmetros, Y o vetor de observações, $p(\theta|Y)$ a distribuição conjunta a *posteriori* e $p(\theta)$ a distribuição a *priori*.

As inferências são realizadas a partir da distribuição conjunta a *posteriori* que em geral tem forma complexa e se torna difícil a avaliação das quantidades de interesse. O problema a ser resolvido é o cálculo da esperança a *posteriori* de alguma função $h(\theta)$,

$$E[h(\theta)|Y] = \int h(\theta)p(\theta|Y) d\theta. \quad (1)$$

A partir de N amostras do vetor θ , $(\theta_1, \theta_2, \dots, \theta_N)$ da distribuição a *posteriori* $p(\theta|Y)$ a integral em (1) é aproximada por meio da média,

$$\widehat{E}[h(\theta)|Y] = \frac{1}{N} \sum_{j=1}^N h(\theta_j).$$

Geralmente esse problema é transferido para o caso univariado e assim, ele seria solucionado calculando-se a distribuição marginal para cada parâmetro. No entanto, a abordagem univariada não contorna a dificuldade de integração para a obtenção da esperança a *posteriori*. Métodos numéricos como a amostragem Gibbs podem ser usados para aproximar a integral em (1) quando as distribuições condicionais completas a *posteriori* possuem formas algébricas fechadas (das quais se pode tomar amostras). Caso isso não ocorra, deve-se utilizar outras formas de amostragem, em geral baseadas no algoritmo Metropolis-Hastings (Metropolis et al., 1953 e Hastings, 1970).

Os modelos lineares generalizados (GLM) podem ser usados para estabelecer uma relação entre uma variável resposta categorizada ordinal e variáveis preditoras (McCullagh, 1980), sendo que, a função ligadora é alguma função de distribuição. Supondo que existam n variáveis, Y_1, Y_2, \dots, Y_n , classificadas em K categorias ordenadas, em que Y_i é tal que $p_{ik} = P(Y_i = k)$ para $k = 1, 2, \dots, K$ e $\sum_{k=1}^K p_{ik} = 1, i = 1, \dots, n$ e assumindo que as probabilidades acumuladas possam ser modeladas por um GLM, têm-se:

$$P(Y_i \leq k) = G(\gamma_k - w_i' \theta), k = 1, 2, \dots, K \quad (2)$$

em que w_i' é o vetor linha de incidência, ligando θ à i -ésima observação, G é uma função de distribuição de uma variável contínua que pode assumir várias formas, θ é o vetor de efeitos fixos e aleatórios e γ é o vetor de parâmetros *threshold* que divide a reta real em K intervalos disjuntos, da seguinte forma:

$$(\gamma_0, \gamma_1); [\gamma_1, \gamma_2); \dots; [\gamma_{K-1}, \gamma_K);$$

em que $\gamma_0 = -\infty$ e $\gamma_K = +\infty$.

Em geral adota-se uma origem para a variável latente fixando um dos parâmetros de limiar ou *threshold* e em geral, escolhe-se $\gamma_1^* = 0$. A motivação para $\gamma_1^* = 0$ é dada pela transformação $\gamma_k^* = \gamma_k - \gamma_1$. Assim, cada γ_k^* será a distância entre γ_k e γ_1 . A variável latente transformada será obtida por meio da expressão $L_i^* = L_i - \gamma_1, i = 1, \dots, n$ e assim,

$$y_i = k \text{ se } \gamma_{k-1} - \gamma_1 \leq L_i - \gamma_1 \leq \gamma_k - \gamma_1.$$

Os novos parâmetros (θ^*), de efeitos fixos e aleatórios, serão $\theta^* = \theta - \gamma_1$.

Na escala da variável latente, $L^* = W\theta^* + \epsilon$ em que, W é a matriz de incidência, θ^* o vetor de efeitos e ϵ o vetor de resíduos. Assumindo distribuição normal para a variável latente (L^*), então

$$(L^* | \theta^*, \sigma_e^2) \sim N(W\theta^*, I\sigma_e^2).$$

Sem perda de generalidade a variância residual pode ser constante, adotando-se por exemplo, $\sigma_e^2 = 1$ (Sörensen e Gianola, 2004).

A distribuição conjunta a *posteriori*, de todos os parâmetros e da variável latente resulta em:

$$\begin{aligned} p(\theta^*, L^*, \gamma^* | Y) &\propto p(Y | \theta^*, L^*, \gamma^*) p(\theta^*, \gamma^*) = p(Y | L^*, \gamma^*) p(\theta^*, L^*, \gamma^*) \\ &= p(Y | L^*, \gamma^*) p(\theta^*, L^*, \gamma^*) = p(Y | L^*, \gamma^*) p(L | \theta^*) p(\theta^*, \gamma^*) \\ &= p(Y | L^*, \gamma^*) \left[\prod_{i=1}^n p(L_i^* | \theta^*) \right] p(\theta^*, \gamma^*). \end{aligned} \quad (3)$$

A distribuição $p(Y | L^*, \gamma^*)$ em (3) é degenerada pois, a probabilidade de uma determinada observação pertencer a uma determinada categoria, dado o valor da variável latente e os limiares, é completamente especificada. Conforme notação de Albert e Chib (1993), $p(Y | L^*, \gamma^*)$ pode ser escrita como

$$p(Y | L^*, \gamma^*) = \prod_{i=1}^n \left[\sum_{k=1}^K I_{[\gamma_{k-1}^*, \gamma_k^*)}(L_i^*) I(Y_i = k) \right].$$

Albert e Chib (1993) apresentam um algoritmo para a modelagem de dados categorizados, no qual a função de ligação (G) em (2) corresponde à distribuição t de Student acumulada. O mesmo algoritmo é implementado por Sörensen et al. (1995) em que a função de ligação em (2) é a distribuição normal padrão acumulada considerando o modelo misto (efeitos fixos e aleatórios no modelo para a variável latente). Nesta modelagem, um parâmetro de limiar ou *threshold* (γ_k^*) tem distribuição condicional completa uniforme com domínio no intervalo $[\max(L^* | Y = k); \min(L^* | Y = k + 1)]$ em que L^* é o vetor de variáveis latentes.

O algoritmo desenvolvido por Sørensen et al. (1995) é descrito a seguir.

2.1 Algoritmo de Sørensen et al. (1995)

Supondo que a função G em (2) é dada por uma distribuição normal padrão acumulada ($\Phi(\cdot)$), então o modelo amostral é

$$p(Y_i = k | \theta^*, \gamma^*) = \Phi(\gamma_k^* - w_i' \theta^*) - \Phi(\gamma_{k-1}^* - w_i' \theta^*). \quad (4)$$

e usando uma variável latente L^* podemos reescrever (2) dados os parâmetros de limiar γ^* , θ^* e w_i como:

$$Y_i = k \text{ se } \gamma_{k-1}^* \leq L_i \leq \gamma_k^* \\ L_i^* \sim N(w_i' \theta^*, 1) \quad (5)$$

com $i = 1, 2, \dots, n$ e $k = 1, 2, \dots, K$.

Para a variável latente L^* , tem-se o modelo linear $L^* = W\theta^* + \epsilon$, em que $W = [X|Z]$, com X e Z correspondendo às matrizes de delineamento para os efeitos fixos e aleatórios, respectivamente; ϵ é o vetor de erros aleatórios e $\theta^{*'} = (\beta', u')$ é o vetor de efeitos fixos (β) e aleatórios (u). Quando não se tem efeitos aleatórios o modelo terá sua estrutura simplificada, com $W = X$ e $\theta^{*'} = \beta'$.

Assumindo que os elementos de Y são condicionalmente independentes dado θ^* , a densidade conjunta a *posteriori* de todos os parâmetros, incluindo a variável latente L^* , é dada como segue:

$$\begin{aligned} p(\theta^*, \gamma^*, L^* | Y) &\propto p(Y | L^*, \theta^*, \gamma^*) p(\theta^*, L^*, \gamma^*) \\ &= p(Y | L^*, \gamma^*) p(\theta^*, L^* | \gamma^*) \\ &= p(Y | L^*, \gamma^*) p(L^* | \theta^*) p(\theta^*, \gamma^*) \\ &\propto p(Y | L^*, \gamma^*) \left[\prod_{i=1}^n \exp\left(-\frac{1}{2}(L_i^* - w_i' \theta^*)^2\right) \right] I_{[\gamma_{k-1}^*, \gamma_k^*]}(L_i^*) \\ &\times \phi(\theta^*; \mathbf{0}, V). \end{aligned} \quad (6)$$

A informação a priori sobre (θ^*, γ) pode ser dada por:

$$p(\theta^*, \gamma) \propto p(\theta^*) \propto \exp\left\{-\frac{\tau}{2} \theta^{*'} \theta^*\right\}$$

a qual será vaga se o parâmetro de precisão τ assumir um valor pequeno, por exemplo, $\tau = 0,001$ ou em (6) fazendo $V = 1000I_m$, em que I_m é uma matriz identidade $m \times m$ e m representa a quantidade de parâmetros associados aos efeitos fixos e aleatórios.

Quando se tem um modelo linear misto, com $\theta' = (\beta', u')$, u é especificado por

$$u | \sigma_u^2 \sim N(\emptyset, A_q \sigma_u^2)$$

em que A é a matriz de correlação entre os efeitos aleatórios. Em modelos de Genética Quantitativa, por exemplo, A é um múltiplo da matriz de parentesco (Kempthorne, 1966).

Assim, a matriz V pode ser dada por

$$V = \begin{pmatrix} 1000I_p & \emptyset \\ \emptyset & \sigma_u^2 A_q \end{pmatrix}$$

em que σ_u^2 é a variância dos efeitos aleatórios, p e q correspondem, respectivamente, à quantidade de efeitos fixos e à quantidade de efeitos aleatórios, I_p é uma matriz identidade $p \times p$ e I_q é a matriz de correlação entre os elementos de u , $q \times q$.

No modelo misto,

$$p(\theta^*, \gamma) \propto p(\theta^*) \propto \left(\frac{1}{\sigma_u^2}\right)^{\binom{q}{2}} \exp\left\{-\frac{1}{2}\theta^{*'}V^{-1}\theta^*\right\}.$$

A *priori* para σ_u^2 pode ser dada por uma Gama Inversa,

$$p(\sigma_u^2) = (\sigma_u^2)^{-(a+1)} \exp\left\{\frac{-b}{\sigma_u^2}\right\}, \quad (7)$$

em que a e b são os parâmetros desta distribuição a *priori*, a qual tem desejáveis propriedades de conjugação com as distribuições normais.

E assim, a distribuição conjunta em (6) é reescrita como segue:

$$\begin{aligned} p(\theta^*, \gamma^*, L^*, \sigma_u^2 | Y) &\propto p(Y|L^*, \theta^*, \gamma^*, \sigma_u^2)p(\theta^*, L^*, \gamma^*, \sigma_u^2) \\ &= p(Y|\theta^*, \gamma^*)p(\theta^*, L^*, \gamma^*, \sigma_u^2) \\ &= p(Y|L^*, \gamma^*)p(L^*|\theta^*)p(\theta^*, \gamma^*, \sigma_u^2) \\ &\propto p(Y|L^*, \gamma^*) \left[\prod_{i=1}^n \exp\left(-\frac{1}{2}(L_i^* - w_i\theta^*)^2\right) \right] I_{[\gamma_{k-1}^*, \gamma_k^*]}(L_i^*) \\ &\times \phi(\theta^*; \mathbf{0}, V)p(\sigma_u^2). \end{aligned} \quad (8)$$

As inferências sobre cada parâmetro podem ser feitas com base em amostras da distribuição conjunta a *posteriori* (Gelfand e Smith, 1990), calculadas a partir de (6), se o interesse for no modelo fixo ou a partir de (8) se o interesse for no modelo misto, usando os métodos Monte Carlo via Cadeias de Markov (MCMC).

A metodologia de amostragem das distribuições condicionais foi originalmente desenvolvida por Sørensen et al. (1995) a partir de (8) e está apresentada também, de forma bem didática, em Sørensen e Gianola (2004). Nos parágrafos que se seguem estamos reproduzindo a maior parte desta apresentação.

É preciso notar que a distribuição $p(Y|L^*, \gamma^*)$ na expressão em (8) é degenerada, pois o conhecimento de L^* leva ao conhecimento exato de Y (Albert e Chib, 1993) e sua expressão é dada por,

$$p(Y|L^*, \gamma^*) = \prod_{i=1}^n \left\{ \sum_{k=1}^K I_{[\gamma_{k-1}^*, \gamma_k^*]}(L_i^*) I(Y_i = k) \right\}. \quad (9)$$

A partir da expressão anterior pode-se observar que a distribuição condicional completa de γ_k^* dado γ_{-k}^*, θ^* e L^* , em que γ_{-k}^* representa todos os parâmetros de limiar exceto γ_k , é dada por,

$$P(\gamma_k^* | \gamma_{-k}^*, L^*, Y) \propto \prod_{i=1}^n \left[I(Y_i = k) I_{[\gamma_{k-1}^*, \gamma_k^*]}(L_i^*) + I(Y_i = k+1) I_{[\gamma_k^*, \gamma_{k+1}^*]}(L_i^*) \right] \quad (10)$$

a qual pode ser vista como uma distribuição uniforme, conforme expressão a seguir

$$P(\gamma_k^* | \gamma_{-k}^*, L^*, Y) = \frac{1}{a-b}, \quad a > b \quad (11)$$

em que $a = \min((L^* | Y = k+1), \gamma_{k+1}^*)$ e $b = \max((L^* | Y = k), \gamma_{k-1}^*)$.

Com relação à variável latente, observa-se que L_i^* tem distribuição condicional completa a *posteriori* dada por,

$$P(L_i^* | \theta^*, Y_i = k, \gamma_{k-1}^*, \gamma_k^*) = \frac{\phi_{L_i^*}(w_i' \theta^*, 1)}{\Phi(\gamma_k^* - w_i' \theta^*) - \Phi(\gamma_{k-1}^* - w_i' \theta^*)}, \quad (12)$$

com $\gamma_{k-1}^* < L_i^* \leq \gamma_k^*$, $\Phi(\cdot)$ representando a distribuição normal padrão acumulada e $\phi_{L_i^*}(w_i' \theta^*, 1)$ a densidade normal com média $w_i' \theta^*$ e variância unitária.

O vetor de parâmetros θ^* tem distribuição condicional completa a *posteriori* dada por uma normal multivariada, ou seja,

$$\theta^* | Y, L^* \sim N(B^{-1} W' L^*, B^{-1}) \quad (13)$$

em que $B = V^{-1} + W' W$.

Para a variância dos efeitos aleatórios (σ_u^2), a distribuição condicional completa será

$$p(\sigma_u^2 | \theta^*, L^*, Y) = (\sigma_u^2)^{-(\frac{q}{2} + a + 1)} \exp \left\{ \frac{-1}{2\sigma_u^2} (u' u + 2b) \right\}, \quad (14)$$

que é o núcleo de uma distribuição Gamma Inversa com parâmetros como a seguir:

$$\sigma_u^2 | \theta^*, L^*, Y \sim GI \left(\frac{q+2a}{2}, \frac{u' u + 2b}{2} \right). \quad (15)$$

A amostragem Gibbs, segundo Sørensen et al. (1995), pode ser implementada a partir das distribuições condicionais completas dadas em (13), (15), (12) e (11), não necessariamente nesta ordem.

2.2 Algoritmo adaptado de Albert & Chib (1993)

Em geral a distribuição t de Student é considerada mais plástica para dados binários, mas no caso dos dados categorizados, é provável que tal distribuição possa assumir formas mais próximas tanto da distribuição normal quanto da logística, com diferentes gradações dadas pelo número de graus de liberdade. A distribuição t de Student aproxima-se da distribuição normal quando o número de graus de liberdade ν é grande ($\nu \rightarrow \infty$) e da distribuição logística quando ($\nu \rightarrow 0$). Na Figura 1 estão ilustradas as distribuições normal padrão, t de Student com 3 graus de liberdade e logística.

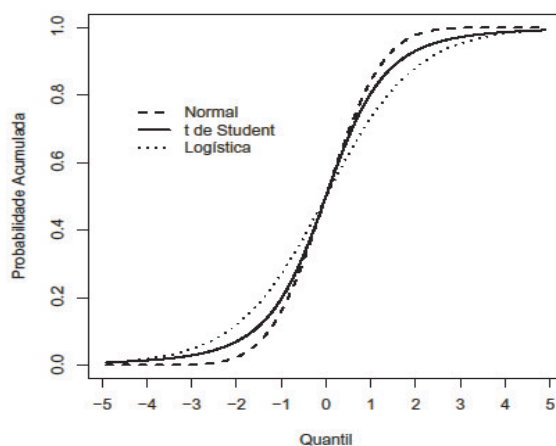


Figura 1 - Funções de distribuição normal, t de Student e logística de acordo com o quantil.

Para a utilização da distribuição t de Student acumulada em (2) é interessante escrevê-la em dois estágios (Paulino e Turkman e Murteira 2003, Sørensen e Gianola 2004), isto é, como a mistura de normais (17) com a gamma invertida (18) para a variância. Tendo em vista que a maioria das distribuições utilizadas na modelagem são da família exponencial, este procedimento propiciará facilidades algébricas e melhor tratabilidade analítica para as condicionais completas a posteriori.

Utilizando densidade acumulada de uma distribuição t de Student com ν graus de liberdade a probabilidade de Y pode ser modelada por meio da expressão (16).

$$P(Y_i = k | \theta^*, \nu, \gamma^*) = F_\nu(\gamma_k^* - w_i' \theta^*) - F_\nu(\gamma_{k-1}^* - w_i' \theta^*), \quad (16)$$

$$k = 1, 2, \dots, K.$$

Amostras da distribuição t de Student para a variável latente podem ser obtidas por um processo hierárquico de amostragem de distribuição normal com variância modificada. Desta forma, amostra-se inicialmente os parâmetros modificadores λ a partir da distribuição em (18) e em seguida, utilizando estes valores, amostra-se L de (17).

Os dois estágios para a especificação da distribuição t de Student são indicados nas duas expressões seguintes.

$$L_i^* | \theta^*, \lambda_i \sim N \left(w_i' \theta^*, \frac{1}{\lambda_i} \right), i = 1, 2, \dots, n \quad (17)$$

$$\lambda_i | \nu \sim \text{Gamma} \left(\frac{\nu}{2}, \frac{\nu}{2} \right). \quad (18)$$

Com esta especificação, o modelo amostral (16) pode ser reescrito da seguinte maneira:

$$P(Y_i = k | \theta^{**}, \lambda_i, \nu, \gamma^{**}) = \Phi \left(\frac{\gamma_k^{**} - w_i' \theta}{\frac{1}{\sqrt{\lambda_i}}} \right) - \Phi \left(\frac{\gamma_{k-1}^{**} - w_i' \theta}{\frac{1}{\sqrt{\lambda_i}}} \right), \quad (19)$$

$$k = 1, 2, \dots, K$$

A metodologia de Albert e Chib (1993) é adaptada neste caso, com a introdução de parâmetros de efeitos aleatórios na modelagem de L^* , conforme pode ser observado na expressão 20.

Admitindo *priori* vaga para θ^* , conforme especificado em 2.1, a posteriori conjunta de todos os parâmetros e a variável auxiliar ou latente é:

$$p(\theta^*, \gamma_k^*, L^*, \lambda, \sigma_u^2, \nu | Y) \propto \left[\prod_{i=1}^n \phi \left(L_i^*, w_i \theta^*, \frac{1}{\lambda_i} \right) I_{[\gamma_{k-1}^*, \gamma_k^*]}(L_i^*) \right]$$

$$\times \left[\prod_{i=1}^n \lambda_i^{\left(\frac{\nu}{2}\right)-1} \exp \left(-\frac{\lambda_i}{2} \nu \right) \right]$$

$$\times \phi(\theta^*, 0, V) p(\sigma_u^2) p(\nu) \quad (20)$$

em que $\lambda = \{\lambda_i\}_{i=1}^n$.

A matriz V , em (20) é dada na seção 2.1, a *priori* para σ_u^2 , é como em (7) e a *priori* para ν é $p(\nu) = 1/(1 + \nu)^2$, isto representa pequena probabilidade de ocorrência de valores altos para os graus de liberdade, já que, para ν grande a distribuição t de Student aproxima-se da distribuição normal.

As distribuições condicionais completas para cada parâmetro foram obtidas a partir da equação (20) e são as que seguem:

$$\theta^* | L^*, \lambda, \nu \sim N \left(M^{-1} W' R^{-1} L^*, M^{-1} \right) \quad (21)$$

em que $M = W' R^{-1} W + V^{-1}$ e R é uma matriz diagonal com seus elementos dados por λ_i^{-1} $i = 1, 2, \dots, n$ e,

$$p(\lambda_i | \lambda_{-i}, L^*, \theta, \nu) \propto \lambda_i^{\left(\frac{\nu+1}{2}\right)-1} \exp \left(-\frac{\lambda_i}{2} \left((L_i^* - w_i' \theta)^2 + \nu \right) \right) \quad (22)$$

a qual é o núcleo de uma distribuição Gama com parâmetros dados a seguir e λ_{-i} , denota todos os elementos de λ exceto λ_i .

$$\lambda_i | \lambda_{-i}, L^*, \nu, \theta \sim G \left(\frac{\nu + 1}{2}, \frac{1}{2} \left((L_i^* - w_i' \theta)^2 + \nu \right) \right) \quad (23)$$

e

$$L_i^* | L_{-i}^*, \theta, \nu, \lambda_i \sim N \left(w_i' \theta^*, \frac{1}{\lambda_i} \right), \quad \gamma_{k-1}^* < L_i^* < \gamma_k^*, \quad k = 1, 2, 3, \dots, K. \quad (24)$$

A distribuição condicional completa a *posteriori* para ν não tem forma fechada e, assim, torna-se necessário o uso do algoritmo Metropolis-Hastings para a amostragem desta distribuição:

$$p(\nu | \theta^*, L^*, \gamma^*) \propto \left(\frac{(\frac{\nu}{2})^{(\nu/2)}}{\Gamma(\nu/2)} \right)^n \left(\prod_{i=1}^n \lambda_i^{\frac{\nu}{2}-1} \exp \left(-\frac{\nu}{2} \lambda_i \right) \right). \quad (25)$$

As distribuições condicionais para os parâmetros σ_u^2 e γ_k são, respectivamente, dadas pelas expressões (15) e (11).

A implementação do Amostrador de Gibbs pode ser feita amostrando seguidamente das distribuições condicionais completas, dadas em (21), (15), (23), (25), (24) e (11).

2.3 Contrastes de interesse

Os três tratamentos do exemplo são níveis quantitativos igualmente espaçados do fator concentração de sacarose. Assim, o modelo de regressão linear simples pode ser ajustado com o modelo quadrático permitindo testar a falta de ajuste do modelo linear.

As médias de tratamentos podem ser estruturadas em dois contrastes ortogonais referentes aos efeitos de regressão linear e quadrático, quais sejam:

- $C_1' \beta = \frac{\beta_3 - \beta_1}{2}$, associado à regressão linear para efeito de porcentagem de sacarose; e
- $C_2' \beta = \beta_2 - \frac{\beta_3 + \beta_1}{2}$, associado à regressão quadrática.

Tais contrastes podem ser diretamente calculados na amostra da distribuição a posteriori conjunta.

2.4 Comparação de modelos

A comparação entre modelos pode ser realizada empregando o fator de Bayes (Gelman et al., 2003), o qual de forma geral é dado por,

$$B_{ij} = \frac{p(y|M_i)}{p(y|M_j)} = \frac{\int p(y|\theta_i, M_i) p_i(\theta_i|M_i) d\theta_i}{\int p(y|\theta_j, M_j) p_j(\theta_j|M_j) d\theta_j} \quad (26)$$

em que θ_i e θ_j são os vetores de parâmetros dos modelos M_i e M_j , respectivamente, com M_i e M_j dois modelos concorrentes. Se o fator ($B_{i,j}$) for maior do que 1 há evidências de que o modelo i seja melhor do que o modelo j .

Porém, nem sempre é possível obter as integrais indicadas em (26), alternativamente o método da amostragem por importância pode ser considerado para obter uma aproximação da marginal dos dados e, neste caso, a comparação dos modelos se torna mais fácil, pois basta obter uma estimativa para a média ou para a média harmônica da verossimilhança caso a função de importância ($g(\theta_i)$) seja, respectivamente, a distribuição a priori e a distribuição a posteriori dos parâmetros.

As duas formas de estimar a marginal dos dados, apresentadas a seguir, são descritas segundo Sørensen e Gianola (2004). A distribuição marginal dos dados pode ser expressa como

$$\begin{aligned} p(y|M_i) &= \frac{\int p(y|\theta_i, M_i)p(\theta_i|M_i) d\theta_i}{\int p(\theta_i|M_i) d\theta_i} \\ &= \frac{\int p(y|\theta_i, M_i) \frac{p(\theta_i|M_i)}{g(\theta_i)} g(\theta_i) d\theta_i}{\int \frac{p(\theta_i|M_i)}{g(\theta_i)} g(\theta_i) d\theta_i} \end{aligned} \quad (27)$$

o denominador em (27) reescrito, como a seguir, pode ser interpretado como a esperança de $p(\theta_i|M_i)/g(\theta_i)$.

$$\int \frac{p(\theta_i|M_i)}{g(\theta_i)} g(\theta_i) d\theta_i = \lim_{m \rightarrow \infty} \left[\frac{1}{m} \sum_{j=1}^m \frac{p(\theta_i^{[j]}|M_i)}{g(\theta_i^{[j]})} \right] \quad (28)$$

em que $p(\theta_i^{[j]}|M_i)$ é a densidade a priori para o modelo i avaliada no j -ésimo valor amostrado para θ .

O numerador em (27) de forma análoga ao denominador pode ser escrito como:

$$\begin{aligned} &\int p(y|\theta_i, M_i) \frac{p(\theta_i|M_i)}{g(\theta_i)} g(\theta_i) d\theta_i \\ &= \lim_{m \rightarrow \infty} \left[\frac{1}{m} \sum_{j=1}^m p(y|\theta_i^{[j]}, M_i) \frac{p(\theta_i^{[j]}|M_i)}{g(\theta_i^{[j]})} \right] \end{aligned} \quad (29)$$

em que $p(y|\theta_i^{[j]}, M_i)$ é avaliada no j -ésimo valor obtido da distribuição de importância.

Assim, fazendo $\omega_i = p(\theta_i^{[j]}|M_i) / g(\theta_i^{[j]})$, para m grande a distribuição marginal dos dados é estimada por meio da razão entre (29) e (28) como a seguir:

$$\hat{p}(y|M_i) = \frac{\sum_{j=1}^m \omega_i^{[j]} p(y|\theta_i^{[j]}, M_i)}{\sum_i \omega_i^{[j]}}. \quad (30)$$

Se a função de importância for a distribuição a *priori*, então cada um dos pesos (ω_i) será igual a 1 e a distribuição marginal dos dados (30) resultará em:

$$\widehat{p}(y|M_i) = \frac{1}{m} \sum_{j=1}^m p\left(y|\theta_i^{[j]}, M_i\right), \quad (31)$$

com $\theta_i^{[j]}$ retirado da distribuição a priori.

A vantagem do uso da distribuição a *priori* como função de importância é a facilidade de obtenção de uma amostra de θ e a desvantagem é que, em geral, com estes parâmetros, encontra-se valores muito pequenos para a $p(y|\theta)$.

Por outro lado, se a função de importância for a distribuição a *posteriori* ($p(\theta_i|y, M_i)$), obtém-se maior probabilidade para os dados e, além disso, não há necessidade do conhecimento da forma desta distribuição, conforme a expressão em (33). A desvantagem é a instabilidade numérica para o estimador da distribuição marginal dos dados ($\widehat{p}(y|M_i)$), mas este problema pode ser contornado utilizando a escala logarítmica.

$$\begin{aligned} \omega_i &= \frac{p(\theta_i|M_i)}{g(\theta_i)} = \frac{p(\theta_i|M_i)}{p(\theta_i|y, M_i)} \\ &= \frac{p(\theta_i|M_i)}{\frac{p(y|\theta_i, M_i)p(\theta_i|M_i)}{p(y|M_i)}} = \frac{p(y|M_i)}{p(y|\theta_i, M_i)} \end{aligned} \quad (32)$$

e substituindo a expressão à direita da última igualdade em (30) e observando que $p(y|M_i)$ não depende do somatório, obtêm-se:

$$\begin{aligned} \widehat{p}(y|M_i) &= \frac{\sum_{j=1}^m \frac{p(y|M_i)}{p(y|\theta_i^{[j]}, M_i)} p\left(y|\theta_i^{[j]}, M_i\right)}{\sum_{j=1}^m \frac{p(y|M_i)}{p(y|\theta_i^{[j]}, M_i)}} \\ &= \frac{m}{\sum_{j=1}^m \frac{1}{p(y|\theta_i^{[j]}, M_i)}} = \left[\frac{1}{m} \sum_{j=1}^m \frac{1}{p(y|\theta_i^{[j]}, M_i)} \right]^{-1}. \end{aligned} \quad (33)$$

O estimador da distribuição marginal dos dados em (33) é a média harmônica dos valores da verossimilhança calculada em cada ciclo j do processo MCMC.

A interpretação do fator de Bayes segundo a escala dada por Jeffreys (1961) é apresentada na Tabela 3. Com esta escala é possível visualizar, dentre dois modelos em comparação, a intensidade com a qual um é superior ao outro.

2.5 Caracterização das amostras e estudo da convergência

Os algoritmos foram implementados em **R** (R Development Core Team, 2009) e permitem analisar quaisquer estruturas de covariância para os efeitos aleatórios.

Tabela 3 - Escala para interpretação do fator de Bayes

B_{ij}	Evidência favorável a M_i
< 1	negativa (favorável a M_j)
de 1 a 3	duvidosa
de 3 a 10	substancial
de 10 a 30	forte
de 30 a 100	muito forte
> 100	decisiva

Fonte: Jeffreys (1961).

Por meio da distribuição conjunta a *posteriori* em cada uma das situações obteve-se cadeias com 10000 iterações para cada parâmetro. Como um dos objetivos deste trabalho foi verificar a convergência no algoritmo nas duas funções de ligação utilizadas, para este fim, não foi considerada uma amostra válida para a realização de inferências e sim, uma cadeia de tamanho suficiente para o estudo da convergência em cada caso. Para a aplicação do critério de Gelman & Rubin foram consideradas duas cadeias para cada parâmetro.

Para comparar as modelagens por meio da distribuição t de Student e por meio da distribuição normal, foi necessário o uso da amostra válida, pois, além da velocidade de convergência dos algoritmos, o interesse também consistiu em constatar a melhor modelagem (distribuição t de Student ou normal). A escolha do melhor modelo foi realizada por meio do fator de Bayes, estimado utilizando a média harmônica conforme a expressão em (33).

Para a realização das inferências foi considerada uma amostra final de tamanho 5000, obtida a partir de uma cadeia com 255000 iterações, em que as 5000 primeiras foram descartadas para a eliminação do efeito do valor inicial arbitrário e armazenado um valor a cada 50 para que estes fossem não correlacionados.

Essas cadeias foram utilizadas também para a estimação dos contrastes anteriormente citados.

A verificação da convergência foi realizada por meio de gráficos para as cadeias dos parâmetros, pelos critérios de Raftery e Lewis (1992) e de Gelman e Rubin (1992).

Foi calculada a autocorrelação para todos os parâmetros no processo amostral. O cálculo dos valores de autocorrelações para diferentes intervalos (autocorrelação de *lag k*) entre os dados das cadeias de Markov foi feito como se segue:

$$r_k = \frac{E[(X_t - \bar{X})(X_{t+k} - \bar{X})]}{\sqrt{E[(X_t - \bar{X})^2]E[(X_{t+k} - \bar{X})^2]}}$$

em que X_t é o valor amostrado na iteração t , X_{t+k} o valor amostrado na iteração $t + k$ e \bar{X} é a média de todos os valores amostrados.

Estão apresentadas autocorrelações calculadas para os parâmetros correspondentes à média do tratamento e ao parâmetro *threshold* que apresentaram maior problema de autocorrelação nos casos em que o algoritmo foi implementado com distribuição normal e t de Student para a variável latente.

3 Resultados e discussão

Os resultados obtidos, conforme Figuras 2 e 3, mostram que as estimativas para os parâmetros de limiar, principalmente, possuem uma autocorrelação muito forte, dificultando o processo de convergência tanto com função de ligação normal quanto t de Student acumuladas. Esta análise é coerente com discussões apresentadas por Sørensen et al. (1995), Sørensen e Gianola (2004), Cowles (1996) e Nandram e Chen (1996), pois, estes autores destacaram o mesmo problema quando utilizaram o algoritmo de Albert e Chib (1993).

Para os contrastes entre tratamentos com o uso desse algoritmo foi verificado que estes convergem rapidamente (Figura 4).

A representatividade das categorias quanto à quantidade de elementos tem relação direta com o processo de convergência. A redução do número de categorias de nove para cinco, conforme sugestão de Kizilkaya et al. (2003), aumentou a eficiência do processo de amostragem quanto à convergência. As análises com nove categorias não são aqui apresentadas.

As estimativas de médias de tratamentos, também, apresentaram forte autocorrelação entre seus valores (Figura 2), indicando a necessidade de uma cadeia longa para a obtenção de amostras adequadas para fins de inferências.

O número de iterações descartadas para eliminar o efeito do valor inicial arbitrário, isto é, o *Burn-in* (B) e o salto entre duas iterações consecutivas para que a amostra final obtida fosse considerada não correlacionada, ou seja, o *Thinning* (T) foi obtido por meio do critério de Raftery & Lewis (Tabela 4). Os maiores valores de B (descarte) e T (salto) foram, respectivamente, 248 e 31 para o parâmetro γ_2 , no modelo com função de ligação normal acumulada e 510 e 24 para γ_3 e σ_u^2 no modelo com função de ligação t de Student acumulada.

Foram realizadas análises para os dados distribuídos em nove categorias em que os valores de B e T considerados foram, respectivamente, de 5000 e 100 em uma cadeia de 505000 e, portanto, obtendo uma amostra final de 5000 valores e, ainda assim, não obteve-se convergência. Nesta análise foi utilizado o algoritmo de Sørensen et al. (1995) em que a variável latente tem distribuição normal.

O maior valor do fator R (Gelman e Rubin, 1992) foi de 1,2584054 também para γ_2 no modelo normal e de 1,2233068 para γ_3 no modelo t de Student. Segundo os autores para que haja convergência o valor de \hat{R} deve estar próximo de 1. No entanto, segundo os mesmos autores, um valor próximo de 1,2 pode ser satisfatório para muitos casos. Se $\hat{R} \approx 1$, então o valor inicial arbitrário da cadeia não exerce mais efeito sobre as iterações (Paulino e Turkman e Murteira, 2003). É conveniente ressaltar que o valor \hat{R} é calculado usando a segunda metade de cada cadeia.

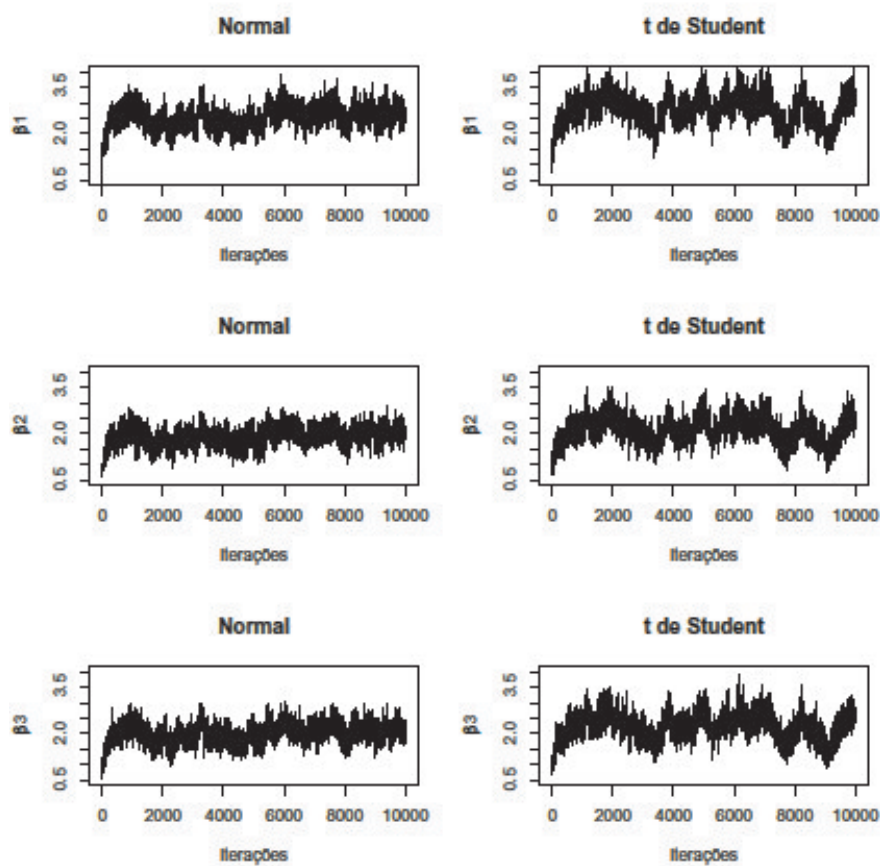


Figura 2 - Cadeias para médias de tratamentos no algoritmo com funções de ligação normal e t de Student, acumuladas.

Em várias outras amostras obtidas esses valores apresentaram outra configuração, isto é, eles possuem oscilação. Porém com um valor de B de alguns poucos milhares de iterações e um valor de T de algumas poucas dezenas seria o suficiente para atender às exigências indicadas pelo critério de Raftery e Lewis em todas as situações.

O fato de considerar para a função G em (2) a distribuição t de Student acumulada, não altera o tempo para a convergência do algoritmo em relação ao modelo normal (Figuras 2 e 3) e os valores de B, T e \hat{R} também, não diferem expressivamente do caso com função de ligação normal acumulada. Neste caso, o modelo é mais complexo, pois há acréscimo de parâmetros visto que a distribuição

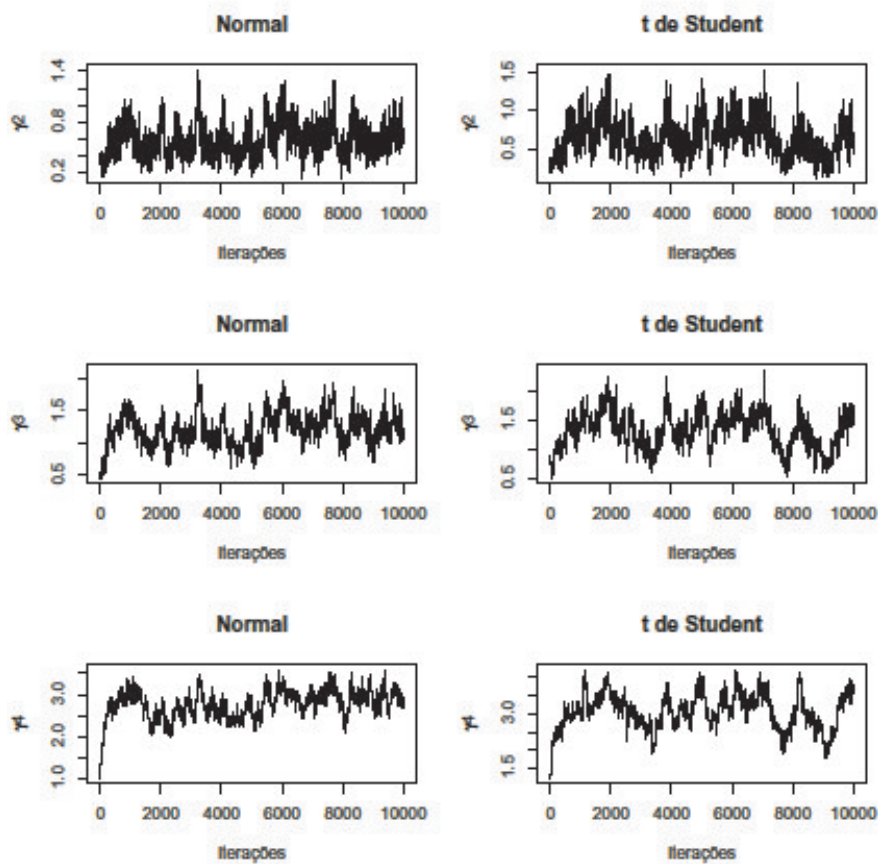


Figura 3 - Cadeias para os parâmetros de limiar no algoritmo nos modelos normal e t de Student.

t de Student foi expressa por meio da mistura de uma normal e uma Gama para cada um dos parâmetros λ_i , associado às observações. A distribuição Gama para o parâmetro λ_i é dada em função do grau de liberdade ν , o qual não tem uma distribuição “fechada” para a implementação computacional, sendo necessário o uso do algoritmo Metropolis-Hastings para a amostragem de seus valores e isto contribui para a complexidade do algoritmo.

Os critérios de convergência aplicados e comentados anteriormente e principalmente os valores de B e T apontados pelo critério de Raftery e Lewis (1992) indicam a necessidade de cadeias pequenas em contraposição ao que se pode observar por meio das Figuras 2 e 3, pois verifica-se forte dependência entre as estimativas

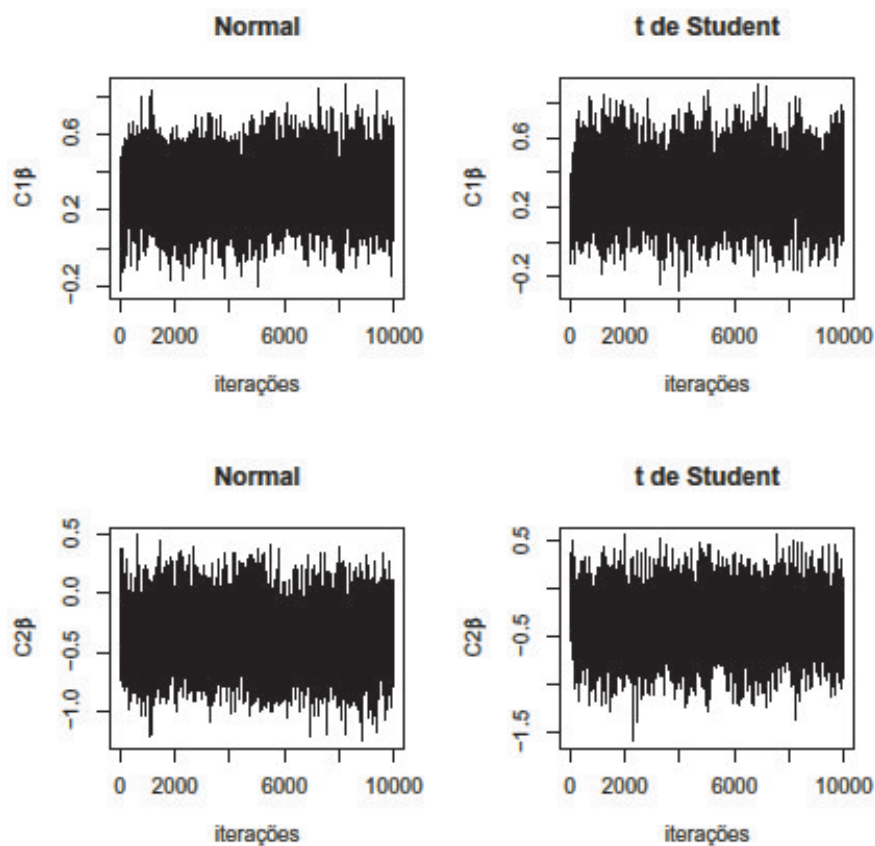


Figura 4 - Cadeias para os contrastes associados à regressão linear e ao desvio de regressão no algoritmo com funções de ligação normal e t de Student, acumuladas.

obtidas em cada iteração. Na Tabela 5 está ilustrada a dependência da média do tratamento e do parâmetro de limiar que apresentaram maior autocorrelação, tanto no caso em que a função de ligação foi a distribuição normal acumulada quanto no caso em que essa função foi a t de Student acumulada, conforme Tabela 5. A justificativa para a apresentação da autocorrelação de apenas dois parâmetros se deve ao fato de que se a convergência não acontecer para apenas algumas das quantidades estimadas é suficiente para ilustrar a ineficiência do algoritmo.

Observa-se, ainda, que existe forte autocorrelação para esses parâmetros principalmente para *lag* 50 e *lag* 100 e no modelo t de Student verifica-se forte

Tabela 4 - Valores de *Burn-in* (B) e *Thinning* (T) obtidos por meio do diagnóstico de Raftery & Lewis, para as médias de tratamentos (β), variância de blocos (σ_u^2) e graus de liberdade (ν), segundo a distribuição da variável latente

Parâmetro	Normal		t	
	B	T	B	T
β_1	54	9	48	12
β_2	45	9	60	10
β_3	40	8	66	11
γ_2	45	9	50	10
γ_3	248	31	117	13
γ_4	209	19	510	10
σ_u^2	9	3	96	24
ν	-	-	20	5

Tabela 5 - Autocorrelação entre as estimativas para β_1 e para γ_3 no caso em que a função de ligação foi a distribuição normal acumulada e para β_2 e γ_2 quando a função de ligação foi a distribuição t de Student acumulada

Distribuição	Parâmetro	lag		
		50	100	200
Normal	β_1	0,1761	0,1039	0,0474
	γ_3	0,2264	0,1405	0,0637
t	β_2	0,3273	0,2451	0,1849
	γ_2	0,3420	0,2980	0,2330

autocorrelação também no *lag* 200. Estes resultados são coerentes com as cadeias ilustradas nas Figuras 2 e 3.

Conforme valores de ν justifica-se o uso da distribuição t de Student neste algoritmo (Figura 5), pois a maioria das estimativas foram abaixo de 10. No entanto, a dependência entre as estimativas para os parâmetros é maior neste caso. Este problema é solucionado adotando um grande valor de T.

Estudo de convergência na amostra válida para inferências, isto é, nas cadeias sem o efeito do valor inicial arbitrário e estimativas não autocorrelacionadas, foi realizado por meio do critério de Gelman e Rubin (1992). Este estudo, apresentou pequenas variações em torno do valor 1 e, portanto, considera-se que houve convergência, mesmo neste algoritmo, no qual foram verificadas dificuldades de convergência a partir do estudo inicial com 10000 iterações. Maiores valores de B e T dentre todas as cadeias foram, respectivamente, de 6 e 2 quando foi utilizada a distribuição normal padrão como função de ligação e de 16 e 5 quando se usou a

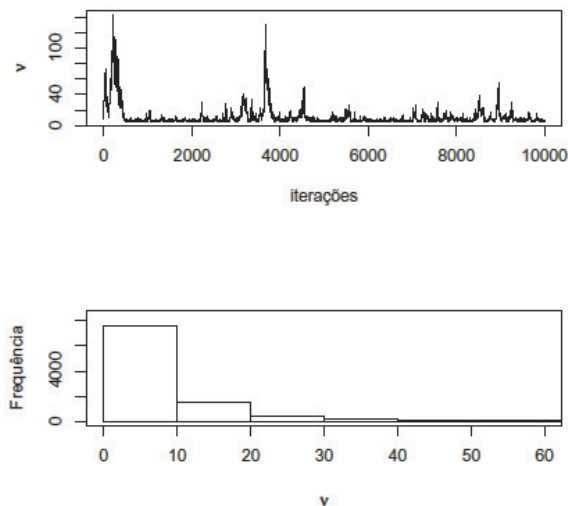


Figura 5 - Cadeia e distribuição das estimativas de ν no algoritmo AC.

distribuição t de Student na cadeia referente aos graus de liberdade (ν).

Para a escolha da melhor modelagem foi utilizado o quociente entre os valores obtidos a partir da expressão (33) com o uso da função de ligação, a distribuição normal padrão acumulada e da distribuição t de Student acumulada. A aproximação ao fator de Bayes, dada pela razão mencionada foi de 0,005862994 e tomando o inverso deste valor verifica-se, a partir da Tabela 3, que há forte evidência em favor do modelo usando a função ligadora t de Student.

3.1 Comparações com a análise de variância e considerações gerais

A análise de variância usual, para os dados sem transformação e com transformação logarítmica, indica que a porcentagem de sacarose influencia significativamente a nota atribuída à cor da banana da terra desidratada (Tabela 6) o que não ocorreu quando foi aplicada a transformação de Box-Cox. Foi detectado efeito linear significativo por meio do teste F, a 5% de probabilidade, apenas no caso da transformação logarítmica.

O poder da análise reduziu, no caso em que foi usada a transformação de Box-Cox, pois, nesta análise não foi detectado influência significativa dos tratamentos na variável cor (Tabela 6). O valor de λ para a transformação de Box-Cox foi de 3,5.

Tabela 6 - Quadro-resumo da análise da variância para a variável cor sem transformação, com transformação logarítmica e com transformação Box-Cox

Fonte Variação	GL	Tipo de transformação					
		Sem		Logarítmica		Box-Cox	
		<i>F</i>	valor p	<i>F</i>	valor p	<i>F</i>	valor p
Bloco	35	3,18	0,0000	3,19	0,0000	3,18	0,0000
Tratamento	2	3,19	0,0471	3,80	0,0272	2,38	0,0999
Efeito linear	1	3,58	0,0626	4,06	0,0476	2,86	0,0954
Desvio de regressão	1	2,81	0,0984	3,53	0,0644	1,90	0,1721

Em todos os três casos, conforme valor p da Tabela 6, a falta de ajuste foi não significativo. Isto é coerente com o contraste $C_2\beta$ estimado a partir das cadeias para a média de cada tratamento pois, o HPD para este contraste não inclui o valor zero e pode-se concluir então que o desvio de regressão é não significativo.

A hipótese de normalidade dos resíduos, segundo o teste de Shapiro-Wilk não foi aceita em nenhum dos casos. As probabilidades de significância foram respectivamente 0,0004421, 0,000007276 e 0,001082 nas análises com os dados sem transformação, com transformação logarítmica e com transformação de Box-Cox. A ausência de normalidade dos resíduos pode ser observada também, na Figura 6.

Em resumo, nenhuma das análises de variância pode ser considerada válida, mas a mais próxima de ser válida, com transformação Box-Cox, apresenta conclusões semelhantes às dos algoritmos propostos.

Conclusões

A distribuição t de Student é uma função de ligação mais flexível em relação à distribuição normal e propiciou melhor ajuste segundo o fator de Bayes.

Quanto à convergência não há distinção entre o uso das distribuições normal e t de Student para a variável latente.

A análise bayesiana de parâmetros de limiar é recomendável na modelagem de dados discretos ordinais provenientes da análise sensorial de alimentos.

Agradecimentos

À Universidade Federal de Lavras – UFLA, ao Programa de Pós-Graduação em Estatística e Experimentação Agropecuária e ao Instituto Federal Goiano - Campus de Rio Verde-GO.

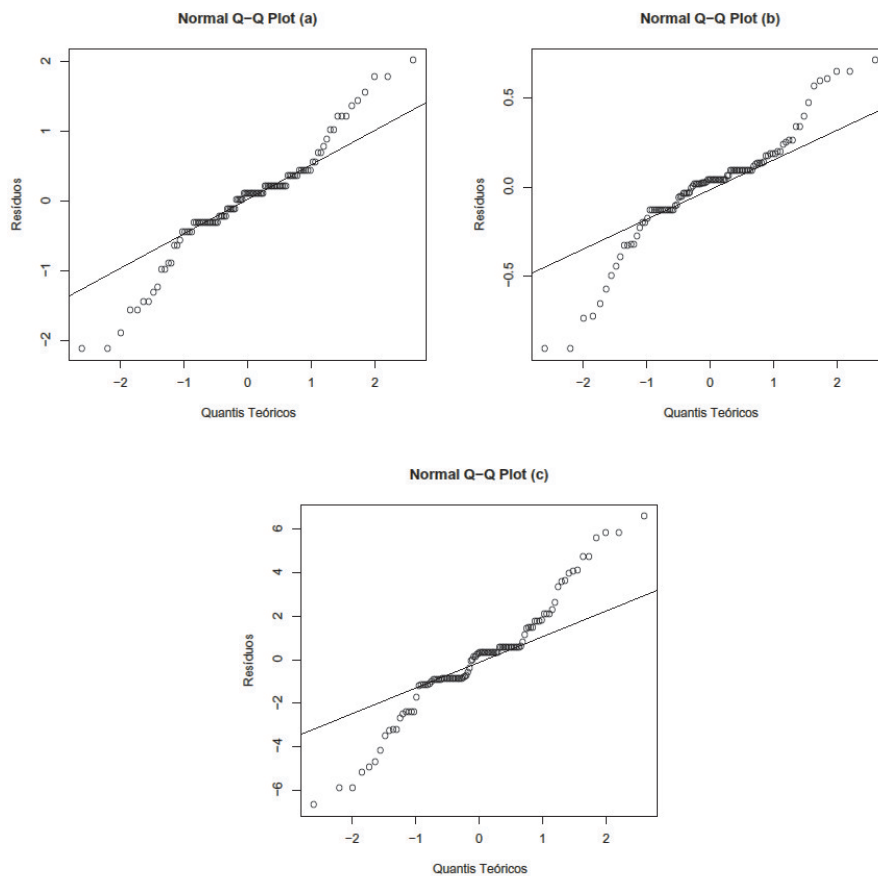


Figura 6 - Gráfico de resíduos a partir do modelo ajustado para a variável nota atribuída à cor da banana da terra desidratada, sem transformação (a), com transformação logarítmica (b) e com transformação de Box-Cox (c).

SILVA, J. W.; BUENO FILHO, J. S. S. An algorithm for threshold models using cumulative Gaussian or cumulative Student's "t" distributions. *Rev. Bras. Biom.*, São Paulo, v.28, n.3, p.59-83, 2010.

- **ABSTRACT:** *Categorical data can be modeled by a underlying continuous variable (L) that relates realizations of random response variable (Y categories) to the belonging of a linked interval. The link function used is the difference on (cumulated) distribution functions evaluated in the interval limits. This can be translated in a linear model for the conditional expectation of L . The Gibbs sampler was the first algorithm proposed*

to get samples from the joint posterior distribution in fixed effects models. The same algorithm was modified by Sørensen et al. (1995) to mixed effects models with normal distribution for the random effects. In this paper we fit a mixed model assuming both distributions for link functions (cumulative normal and cumulative Student t). Convergence and fit properties from both distributions were compared. A sensory analysis example was used in which three sugar solutions with different concentrations were used to dehydrate a landrace of banana fruits. Color of the product was evaluated using a nine points hedonic scale. There was no evidence for differences in convergence times from both link functions. Cumulative t link function has presented the best fit, being the more likely model according to the Bayes factor.

■ KEYWORDS: Gibbs sampler; normal link; probit models; threshold models; t link.

Referências

- AGRESTI, A. *Categorical data analysis*. 2.ed. New York: John Wiley & Sons, 2002. 710p.
- ALBERT, J. H.; CHIB, S. Bayesian analysis of binary and polychotomous response data. *J. Am. Stat. Assoc.*, Washington, v.88, n.422, p.669-679, 1993.
- COWLES, M. K. Accelerating Monte Carlo Markov chain convergence for cumulative link generalized linear models. *Stat. Comput.*, London, v.6, n.2 p.101-111, 1996.
- DUTCOSKY, S. D. *Análise sensorial de alimentos*. Curitiba:Champagnat, 1996. 123p. GELMAN, A.; RUBIN, D. B. Inference from iterative simulation using multiple sequences. *Stat. Sci.*, Hayward, v.7, n.4, p.457-72, 1992.
- GELFAND, A. E.; SMITH A. F. M. Sampling-based approaches to calculating marginal densities. *J. Am. Stat. Assoc.*, Washington, v.85, n.410, p.398-409, 1990.
- GELMAN, A.; CARLIN, J. B.; STERN, H. S.; RUBIN, D. B. *Bayesian data analysis*. London: Chapman and Hall, 2003. 668p.
- HASTINGS, W. K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, London, v.57, n.1, p.97-109, 1970.
- HERINGSTAD, B.; CHANG, Y. M.; GIANOLA, D.; KLEMETSDAL, G. Multivariate threshold model analysis of clinical mastitis in multiparous Norwegian dairy cattle. *J. Dairy Sci.*, Champaign, v.87, n.9, p.3038-3046, 2004.
- JEFFREYS, H. *Theory of probability*. Oxford: Clarendon Press, 1961. 470p.
- KEMPTHORNE, O. *An introduction to genetic statistic*. New York: John Wiley and Sons Inc., 1966. 545p.
- KIZILKAYA, K.; CARNIER, P.; ALBERA, A.; BITTANTE, G.; TEMPELMAN, R. J. Cumulative t -link threshold models for genetic analysis of calving ease scores. *Genet. Sel. Evol.*, Les Ulis, v.35, p.489-512, 2003.

- McCULLAGH, P. Regression models for ordinal data. *J. R. Stat. Soc.*, Oxford, v.42, n.2, p.109-142, 1980.
- METROPOLIS, N.; ROSENBLUTH, A. W.; ROSENBLUTH, M. N.; TELLER, A. H.; TELLER, E. Equations of state calculations by fast computing machines. *J. Chem. Phys.*, Woodbury, v.21, n.6, p.1087-1092, 1953.
- NANDRAM, B.; CHEN, M. H. Reparameterizing the generalized linear model to accelerate gibbs sampler convergence. *J. Stat. Comput. Simul.*, Abingdon, v.54, p.129-144, 1996.
- PAULINO, C. D.; TURKMAN, M. A. A.; MURTEIRA, B. *Estatística Bayesiana*, Lisboa: Fundação Calouste Gulbenkian, 2003, 446p.
- PIEPHO H. P; KALKA E. Threshold models with fixed and random effects for ordered categorical data. *Food Qual. Prefer.*, New York, v.14, n.5, p.343-357, 2003.
- RAFTERY, A. L.; LEWIS, S. Comment: one long run with diagnostics: implementation strategies for Markov chain Monte Carlo. *Stat. Sci.*, Hayward, v.7, n.4, p.493-497, 1992.
- 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 jul. 2009.
- SÖRENSEN, D.; GIANOLA, D. *Likelihood, bayesian and MCMC methods in quantitative genetics*. New York: Springer, 2004. 740p.
- SÖRENSEN, D. A.; ANDERSEN S.; GIANOLA D.; KORSGAARD I. Bayesian inference in threshold models using Gibbs sampling. *Genet. Sel. Evol.*, Les Ulis, v.27, n.3, p.229-249, 1995.
- STEEL, R. G. D.; TORRIE, J. H. *Principles and procedures of statistics: a biometrical approach*. 2.ed. New York: McGraw-Hill, 1980. 633p.
- VAZQUEZ, A. I., WEIGEL, K. A., GIANOLA, D., BATES, D. M., PEREZ-CABAL, M. A., ROSA, G. J. M., CCHANG, Y. M. Poisson versus threshold models for genetic analysis of clinical mastitis in US Holsteins. *J. Dairy Sci.*, Champaign, v.92, p.5239-5247, 2009.
- WRIGHT, S. An analysis of variability in number of digits in an inbred strain of guinea pigs. *Genetics*, Pittsburgh, v.19, p.506-536, 1934.

Recebido em 15.05.2010.

Approved after revised in 01.09.2010.