

SELEÇÃO BAYESIANA DE MODELOS USANDO OS PRINCÍPIOS DOS EFEITOS ESPARSOS DA HIERARQUIA E DA HEREDITARIEDADE

Guilherme BIZ¹
Silvio Sandoval ZOCCHI¹
Roseli Aparecida LEANDRO¹

- RESUMO: No planejamento de experimentos para o ajuste de modelos polinomiais envolvendo k fatores principais e respectivas interações, é bastante comum a utilização dos fatoriais 2^k , 3^k ou frações dos mesmos. Para as análises dos resultados desses experimentos, frequentemente se considera o princípio da hereditariedade, ou seja, uma vez constatada uma interação significativa entre fatores, os fatores que aparecem nesta interação e respectivas interações devem também estar presentes no modelo. Neste trabalho, esse princípio é incorporado diretamente *a priori*, para um método de seleção de variáveis bayesiana, seguindo as idéias propostas por Chipman *et al.* (1997), porém com uma alteração dos valores sugeridos pelos autores para uma *priori*. Essa alteração, proposta neste trabalho, promove uma melhoria considerável na metodologia original. A metodologia é então ilustrada por meio da análise dos resultados de um experimento fatorial para a elaboração de biofilmes de amido originado da ervilha.
- PALAVRAS-CHAVE: Análise bayesiana; princípio da hereditariedade; princípio dos efeitos esparsos; princípio da hierarquia; ensaio fatorial.

1 Introdução

Em diversas áreas de pesquisa, deseja-se obter uma função matemática que relaciona uma variável com uma ou mais variáveis ou fatores de interesse ou mais especificamente, determinar um subconjunto de variáveis explicativas para melhor explicar a variável resposta. Para a seleção destas variáveis explicativas, geralmente

¹Universidade de São Paulo – USP, Escola Superior de Agricultura “Luiz de Queiroz”, Departamento de Ciências Exatas, Caixa Postal 9, CEP: 13418-900, Piracicaba, SP, Brasil. E-mail: gbiz@esalq.usp.br/sszocchi@esalq.usp.br/rleandr@esalq.usp.br

são usados métodos que, não contemplam alguns princípios, tais como os princípios: dos efeitos esparsos, de hierarquia e o da hereditariedade.

Princípio dos efeitos esparsos: o número de efeitos significativos é pequeno em relação ao número total de efeitos.

Princípio de hierarquia: efeitos de baixa ordem são mais prováveis de estarem presentes no modelo do que os efeitos de alta ordem; e a presença de efeitos de mesma ordem são igualmente prováveis.

Princípio da hereditariedade: uma vez constatada uma interação entre fatores significativa, os fatores que aparecem na mesma e suas interações devem também estar presentes no modelo.

Este trabalho utiliza a abordagem bayesiana considerando os princípios dos efeitos esparsos, de hierarquia e da hereditariedade na escolha da distribuição *a priori* considerando algumas modificações das idéias propostas por Chipman *et al.* (1997). A proposta de Chipman *et al.* (1997), em certas situações, leva necessariamente à inclusão de todos os efeitos principais em modelos polinomiais para superfícies de resposta. A metodologia é ilustrada por meio da análise de dados resultantes de um experimento para a elaboração de biofilme de amido originado da ervilha.

2 Métodos

Considere o modelo linear geral, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, em que \mathbf{y} é o vetor de observações com dimensão n ; \mathbf{X} é a matriz do delineamento, de dimensões $n \times (p+1)$; $\boldsymbol{\beta}$ é o vetor com dimensão $p+1$, incluindo o intercepto e demais p parâmetros correspondentes aos efeitos fatoriais e $\boldsymbol{\epsilon}$ é o vetor de erros aleatórios com distribuição normal multivariada $MN(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, em que $\mathbf{0}$ denota o vetor de zeros, com dimensão n e $\mathbf{I}_n = \text{diag}\{1, 1, \dots, 1\}$ é a matriz identidade.

Considerando-se a abordagem bayesiana, o princípio da hereditariedade pode ser incorporado diretamente na especificação da distribuição *a priori* considerando-se, para isso, um vetor $\boldsymbol{\delta}$ desconhecido de dimensão $p+1$, formado por 0's e 1's. Se o j -ésimo elemento, δ_j , é 1, significa que a presença do parâmetro β_j , ($j = 1, 2, \dots, p+1$) é significativa no ajuste do modelo; caso contrário, isto é, se δ_j é 0, não. Desta forma $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\delta}^T, \sigma^2)^T$ é o vetor de parâmetros a serem estimados, com dimensão $2p+3$.

Considere, inicialmente, que a distribuição *a priori* para o j -ésimo elemento de $\boldsymbol{\beta}$ seja a mistura de normais dada por:

$$\pi(\beta_j | \delta_j) = \begin{cases} N(0, \sigma^2 \tau_j^2) & \text{se } \delta_j = 0 \\ N(0, \sigma^2 (c_j \tau_j)^2) & \text{se } \delta_j = 1 \end{cases} \quad (1)$$

Neste caso, os hiperparâmetros c_j e τ_j devem ser especificados de maneira tal que, quando $\delta_j = 0$, β_j tenha um valor próximo a zero e quando $\delta_j = 1$, exista uma probabilidade elevada de $|\beta_j|$ ser significativamente diferente de zero.

A distribuição *a priori* para σ^2 considerada é a gama invertida,

$$\sigma^2 \sim IG(\nu/2, \nu\lambda/2), \quad (2)$$

sendo ν e λ seus hiperparâmetros, cuja escolha será abordada posteriormente.

Finalmente, como δ especifica um determinado modelo, *a priori* para δ especifica a priori para cada modelo possível e é dada por:

$$\pi(\delta) = \prod_{j=1}^{p+1} p_j^{\delta_j} (1-p_j)^{1-\delta_j}, \quad (3)$$

em que $p_j = Prob(\delta_j = 1)$, considerando independências entre os componentes de δ , a saber, δ_j , ($j = 1, 2, \dots, p+1$). Esta consideração, no entanto, não é apropriada para modelos com interações. Neste caso, deve-se usar uma *priori* hierárquica para δ , na qual é incorporado o princípio da hereditariedade. Considere, para fins de ilustração, que há três fatores, A , B e C , com dois níveis cada e que apenas os efeitos principais e as suas interações duplas devem fazer parte do modelo. Neste caso, tem-se que $\delta = (\delta_{Int}, \delta_A, \delta_B, \delta_C, \delta_{AB}, \delta_{AC}, \delta_{BC})$ e baseando-se no princípio de hereditariedade, a significância de AB depende da significância de A e de B , ou seja, se nenhum dos efeitos principais é significativo no ajuste do modelo, então a interação parece menos plausível.

Considerando-se o princípio da hereditariedade a incerteza relativa ao parâmetro δ será modelada assumindo que os termos de mesma ordem são independentes. Com isso, os valores de δ associados aos termos de primeira ordem, δ_A, δ_B e δ_C , são independentes e os associados aos termos de segunda ordem, δ_{AB}, δ_{AC} e δ_{BC} , também o são. Assim, o princípio da hereditariedade assume que o termo com interação depende apenas dos termos que o formam e no caso, $Prob(\delta_{AB}|\delta_A, \delta_B, \delta_C) = Prob(\delta_{AB}|\delta_A, \delta_B)$. Logo, tem-se que:

$$Prob(\delta) = Prob(\delta_{Int})Prob(\delta_A)Prob(\delta_B)Prob(\delta_C) \\ \times Prob(\delta_{AB}|\delta_A, \delta_B)Prob(\delta_{AC}|\delta_A, \delta_C)Prob(\delta_{BC}|\delta_B, \delta_C). \quad (4)$$

Este princípio vai ser definido, então, por meio da forma como as interações dependem dos termos que as formam. No caso em questão, a probabilidade da interação AB ser ativa ou importante, ou seja, de $\delta_{AB} = 1$, dados δ_A e δ_B , pode ser escrita como:

$$Prob(\delta_{AB} = 1|\delta_A, \delta_B) = \begin{cases} p_{00} & \text{se } (\delta_A, \delta_B) = (0, 0) \\ p_{01} & \text{se } (\delta_A, \delta_B) = (0, 1) \\ p_{10} & \text{se } (\delta_A, \delta_B) = (1, 0) \\ p_{11} & \text{se } (\delta_A, \delta_B) = (1, 1) \end{cases}, \quad (5)$$

em que p_{00}, p_{01}, p_{10} e p_{11} são constantes entre 0 e 1 que refletem a intensidade da hereditariedade. Convém ressaltar que no caso de se ter fatores quantitativos e modelos de segunda ordem ou superior, esta forma de se especificar a distribuição *a priori* também pode ser utilizada.

No caso em questão a distribuição conjunta *a posteriori* é dada por:

$$f(\theta | \mathbf{y}) = f(\mathbf{y}|\beta, \sigma^2, \delta)\pi(\beta, \sigma^2, \delta) \\ = f(\mathbf{y}|\beta, \sigma^2)\pi(\beta|\sigma^2, \delta)\pi(\sigma^2)\pi(\delta). \quad (6)$$

No entanto, como a distribuição *a posteriori* dada em (6) não é conhecida, são utilizadas técnicas MCMC (*Markov Chain Monte Carlo*), especificamente o amostrador de Gibbs (Gelman *et al.*, 1995), visto que todas as distribuições condicionais completas *a posteriori* são conhecidas, para a obtenção de uma amostra da distribuição conjunta *a posteriori*.

A função de verossimilhança, por sua vez, considera que $(\mathbf{y}|\boldsymbol{\beta}, \sigma^2) \sim MN(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n)$, ou seja, uma distribuição normal multivariada com esperança $\mathbf{X}\boldsymbol{\beta}$ e matriz de variância e covariâncias $\sigma^2\mathbf{I}_n$, que tem a forma:

$$f(\mathbf{y}|\boldsymbol{\beta}, \sigma^2) \propto \sigma^{-n} \exp \left\{ \frac{-(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{2\sigma^2} \right\}, \quad (7)$$

em que n é o número de observações.

A distribuição *a priori* para $\boldsymbol{\beta}$, dada em (1), pode ser escrita, alternativamente, como a normal multivariada $(\boldsymbol{\beta}|\sigma^2, \boldsymbol{\delta}) \sim MN(\mathbf{0}, \sigma^2\mathbf{D}_\boldsymbol{\delta})$ e tem a forma:

$$\pi(\boldsymbol{\beta}|\sigma^2, \boldsymbol{\delta}) \propto \sigma^{-(p+1)} \exp \left\{ \frac{-\boldsymbol{\beta}^T \mathbf{D}_\boldsymbol{\delta}^{-1} \boldsymbol{\beta}}{2\sigma^2} \right\}, \quad (8)$$

em que $\mathbf{D}_\boldsymbol{\delta}$ é uma matriz diagonal dada por: $\mathbf{D}_\boldsymbol{\delta} = \text{diag}\{(c_1^{\delta_1} \tau_1)^2, \dots, (c_{p+1}^{\delta_{p+1}} \tau_{p+1})^2\}$.

A distribuição *a priori* para σ^2 , que é uma gama invertida dada em (2) e a distribuição *a priori* para $\boldsymbol{\delta}$ é uma multinomial com 2^{p+1} valores, dada em (3).

A distribuição condicional completa para $\boldsymbol{\beta}$ é calculada por meio do termos de (6) que envolvem $\boldsymbol{\beta}$, obtendo $f(\boldsymbol{\beta}|\sigma^2, \boldsymbol{\delta}, \mathbf{y}) \propto f(\mathbf{y}|\boldsymbol{\beta}, \sigma^2)\pi(\boldsymbol{\beta}|\sigma^2, \boldsymbol{\delta})$. Como a distribuição *a priori* para $\boldsymbol{\beta}$ dada por (8) e a função de verossimilhança (7) seguem distribuições normais multivariadas, então a distribuição condicional completa para $\boldsymbol{\beta}$ também segue uma distribuição normal multivariada, dada por:

$$(\boldsymbol{\beta}|\sigma^2, \boldsymbol{\delta}, \mathbf{y}) \sim MN(\sigma^{-2}\mathbf{A}_\boldsymbol{\delta}\mathbf{X}^T\mathbf{y}, \mathbf{A}_\boldsymbol{\delta}), \quad (9)$$

em que $\mathbf{A}_\boldsymbol{\delta} = \sigma^2(\mathbf{X}^T\mathbf{X} + \mathbf{D}_\boldsymbol{\delta}^{-1})^{-1}$.

Analogamente, a distribuição condicional completa para σ^2 é obtida deixando de fora o termo $\pi(\boldsymbol{\delta})$ da Equação 6; portanto, obtém-se que a distribuição condicional completa para σ^2 é uma gama invertida dada por:

$$(\sigma^2|\boldsymbol{\beta}, \boldsymbol{\delta}, \mathbf{y}) \sim IG\left(\frac{1}{2}(n+p+1+\nu), \frac{1}{2}[\nu\lambda + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \boldsymbol{\beta}^T \mathbf{D}_\boldsymbol{\delta}^{-1} \boldsymbol{\beta}]\right). \quad (10)$$

Para a distribuição condicional completa para $\boldsymbol{\delta}$, basta tirar de (6) os termos que não dependem de $\boldsymbol{\delta}$, obtendo $f(\boldsymbol{\delta}|\boldsymbol{\beta}, \sigma^2, \mathbf{y}) \propto \pi(\boldsymbol{\beta}|\sigma^2, \boldsymbol{\delta})\pi(\boldsymbol{\delta})$. Sabe-se que a distribuição condicional completa para cada δ_j é uma Bernoulli, mas é necessário determinar a probabilidade de δ_j ser igual a 1 e a distribuição condicional completa para $\delta_j = 1$ tem a forma $f(\delta_j|\boldsymbol{\delta}_{-j}\boldsymbol{\beta}, \sigma^2, \mathbf{y}) \propto \pi(\boldsymbol{\beta}|\sigma^2, \delta_j, \boldsymbol{\delta}_{-j})\pi(\delta_j, \boldsymbol{\delta}_{-j})$, em que

$\delta_{-j} = (\delta_1, \delta_2, \dots, \delta_{j-1}, \delta_{j+1}, \dots, \delta_{p+1})$. Então,

$$\begin{aligned} Pr(\delta_j = 1 | \delta_{-j}, \beta, \sigma^2, \mathbf{y}) &= \\ &= \frac{\pi(\beta | \delta_j = 1, \delta_{-j}, \sigma^2) \pi(\delta_j = 1, \delta_{-j})}{\pi(\beta | \delta_j = 1, \delta_{-j}, \sigma^2) \pi(\delta_j = 1, \delta_{-j}) + \pi(\beta | \delta_j = 0, \delta_{-j}, \sigma^2) \pi(\delta_j = 0, \delta_{-j})} \\ &= \frac{\pi(\delta_j = 1, \delta_{-j})}{\pi(\delta_j = 1, \delta_{-j}) + \pi(\delta_j = 0, \delta_{-j}) \pi(\beta | \delta_j = 0, \delta_{-j}, \sigma^2) / \pi(\beta | \delta_j = 1, \delta_{-j}, \sigma^2)}. \end{aligned} \quad (11)$$

Concluídas as distribuições condicionais completas é implementado o amostrador de Gibbs no Software R (2007). E enfim fazer a estimação do vetor de parâmetros θ sobre as amostras obtidas no amostrado de Gibbs. Neste trabalho, $\theta = (\beta_0, \beta_1, \dots, \beta_p, \sigma^2, \delta_1, \delta_2, \dots, \delta_{p+1})^T$ e a implementação do amostrador de Gibbs resulta em uma amostra para cada parâmetro.

Para implementar o amostrador de Gibbs, no entanto, deve-se especificar os hiperparâmetros τ e c da *priori* para β , e ν e λ da *priori* para σ^2 que tem distribuição gama invertida dada em (10). Deve-se, ainda especificar os p_{00} , p_{01} , p_{10} e p_{11} adotados em (5) para especificação da *priori* para δ .

Box e Meyer (1986) aconselham usar $c_j = 10$, pois indica um grande efeito para os β_j dado $\delta_j = 1$.

Para a escolha de τ_j , por sua vez, George e McCulloch (1993) sugerem utilizar $\tau_j = \tau_j^*$, sendo τ_j^* dado por $\tau_j^* = \frac{\Delta y}{3 \Delta x_j}$, ($j = 1, \dots, p + 1$) em que $\Delta \mathbf{y}$ representa uma pequena mudança na variável resposta \mathbf{y} e Δx_j uma grande mudança em x_j . Isto é feito porque se $\delta_j = 0$, então uma grande mudança em x_j é improvável de acarretar mais do que uma pequena mudança em \mathbf{y} . Os autores definem Δx_j e $\Delta \mathbf{y}$ como $\Delta x_j = \max(x_{j1}, x_{j2}, x_{j3}, \dots, x_{jn}) - \min(x_{j1}, x_{j2}, x_{j3}, \dots, x_{jn})$ e $\Delta \mathbf{y} = \frac{\sqrt{Var(y_1, y_2, y_3, \dots, y_n)}}{5}$.

Deve-se salientar que a escolha de τ pode influenciar diretamente nos resultados, e conseqüentemente, na escolha de um bom modelo. Segundo George e McCulloch (1993), uma maneira de verificar se há essa influência, é atribuir a τ , três valores diferentes dados por $\frac{\tau^*}{2}$, τ^* e $2\tau^*$.

Considere, agora, a escolhas dos hiperparâmetros ν e λ . Quando o número de covariáveis é próximo ao número de observações, *a priori* para σ resulta em um valor muito pequeno para σ resultando em muitos termos considerados ativos no modelo. Uma maneira de corrigir esta deficiência é escolhendo uma *priori* informativa para σ . A escolha de ν é frequentemente próxima de 2, indicando uma *priori* não informativa. Para λ , a escolha depende de experimento para experimento uma vez que depende da escala da variável resposta. Assim uma boa afinação para λ resulta em modelos que não têm muitos e nem poucos termos. Convém salientar que estes hiperparâmetros têm que ser escolhidos e afinados de maneira correta, pois podem influenciar nos resultados.

Quanto à escolha das constantes que refletem a intensidade da hereditariedade, Chipman *et al.* (1997), consideram empiricamente que, para uma hereditariedade fraca, p_{00} é bastante pequena, isto é, menor do que $p_{10} = p_{01} = 0,1$ e $p_{11} = 0,25$. Para uma hereditariedade fraca rigorosa, sugerem $p_{00} = 0$ e para a hereditariedade

forte a interação somente será significativa se ambos os termos que os formam forem significativos, sugerindo $p_{00} = p_{10} = p_{01} = 0$ e $p_{11} = 0,25$. Resumidamente, tais constantes são apresentadas na Tabela 1. A escolhas dessas probabilidades pelos autores foram feitas de modo a considerar que na *priori* para δ sejam incluídos os princípios apresentados anteriormente, ou seja, de modo que: (i) o número de termos importantes em um experimento fatorial seja pequeno; (ii) os efeitos de ordem menor têm maior chance de serem importantes e (iii) caso um efeito principal seja significativo, as probabilidades de as interações que incluem este efeito principal sejam significativas aumenta.

Tabela 1 - Sugestões de valores para dos p_{00} , p_{01} , p_{10} e p_{11} da distribuição *a priori* de δ propostos por Chipman *et al.* (1997) e novos valores propostos neste trabalho (*) para uma hereditariedade forte

Hereditariedade	$p_1 = \dots = p_{p+1}$	p_{00}	$p_{01} = p_{10}$	p_{11}
Fraca	0,25	$0 < p_{00} < 0,1$	0,1	0,25
Fraca-rigorosa	0,25	0	0,1	0,25
Forte(*)	0,25	0	0,05	0,25
Forte	0,25	0	0	0,25

Convém ressaltar que ao optar pela hereditariedade forte, ou seja, adotar os valores $p_{10} = p_{01} = 0$, e *a priori* $\pi(\delta_j = 0, \delta_{-j})$ será, conseqüentemente, igual a 0 para os efeitos principais. Com isso o denominador da Equação 11 fica igual ao numerador, fazendo com que a probabilidade dos efeitos principais sejam iguais a 1. Assim, quando se utiliza hereditariedade forte a distribuição condicional completa para δ_j apresentada na Equação 11 tem, para todos os fatores principais, probabilidades iguais a 1, e os fatores principais sempre serão considerados no modelo. No entanto, segundo o princípio da hereditariedade, caso uma interação seja significativa, ao menos um dos efeitos principais deve ser significativo e não que todos os efeitos principais sejam necessariamente significativos ou que todos os efeitos principais devam estar no modelo. De modo a corrigir esse problema na especificação da *priori*, propõe-se neste trabalho novas probabilidades para a hereditariedade forte, denominada, aqui, nova hereditariedade forte, utilizar,

$$Prob(\delta_{AB}|\delta_A, \delta_B) = \begin{cases} 0 & \text{se } \delta_A = \delta_B = 0 \\ 0,05 & \text{se } \delta_A \neq \delta_B \\ 0,25 & \text{se } \delta_A = \delta_B = 1 \end{cases} . \quad (12)$$

A fim de evidenciar a importância da escolha da *priori* para δ , considere o modelo polinomial com $k = 3$ fatores e interações duplas, dado por:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + e. \quad (13)$$

Nota-se que, nesse caso, $\beta^T = [\beta_0, \beta_1, \beta_2, \beta_3, \beta_{12}, \beta_{13}, \beta_{23}]$. A Tabela 2 apresenta a probabilidade de cada termo do modelo vai ter para as diferentes hereditariedade

usadas, o número de modelos possíveis e o índice de diversidade H de Shannon (1948) (usado para medir a diversidade dos possíveis modelos) e na Figura 1 são apresentados os 14 modelos mais prováveis para as diferentes probabilidades usadas.

Tabela 2 - Probabilidades *a priori* de cada coeficiente do modelo 13 ser importante de acordo com a intensidade da hereditariedade, cujos valores são apresentados na Tabela 1; Números k de possíveis modelos e índices de diversidade de Shannon H dos possíveis modelos

Coef.	Hereditariedade				Coef.	Hereditariedade			
	Fra.	Fra.Rig.	Forte*	Forte		Fra.	Fra.Rig.	Forte*	Forte
β_0	0,250	0,250	0,250	0,250	β_{13}	0,059	0,053	0,034	0,016
β_1	0,250	0,250	0,250	0,250	β_{23}	0,059	0,053	0,034	0,016
β_2	0,250	0,250	0,250	0,250	k	128	90	90	36
β_3	0,250	0,250	0,250	0,250	H	2,815	2,720	2,587	2,355
β_{12}	0,059	0,053	0,034	0,016					

Na Tabela 2 pode-se notar que as escolhas dos valores para *a priori* δ apresentadas na Tabela 1 fazem com que as probabilidade dos parâmetros serem importantes ou ativos seja sempre menor ou igual a 0,25 expressando deste modo, o princípio dos efeitos esparsos. Nota-se, ainda que os efeitos principais têm sempre probabilidades superiores aos dos efeitos de interações duplas, expressando, por sua vez, o princípio da hierarquia. Finalmente, por meio da comparação dos gráficos da Figura 1, nota-se que, com o aumento da intensidade da hereditariedade, a presença de efeitos de interação dupla traz, como consequência a presença de seus efeitos principais formadores, que é a essência do princípio da hereditariedade.

Uma vez especificada *as prioris* e obtidas as amostras para cada parâmetro é necessário obter os valores centrais das distribuições de cada um deles. Para os parâmetros de β pode-se utilizar a média e mediana para σ , uma vez que sua distribuição condicional completa é uma gama invertida e é assimétrica. No entanto, com relação à δ_j , utiliza-se o vetor δ que indica cada modelo, assim, a moda dos vetores é que indica qual dos modelos deve-se usar.

3 Resultados e discussão

Para ilustrar a metodologia sugerida considere os resultados apresentados na Tabela 3 de um experimento para a formulação de biofilmes de amido de ervilha (*Pisum sativum*) conduzido por Manoel Divino da Matta Junior, no Departamento de Ciência e Tecnologia de Alimentos da ESALQ-USP, em Piracicaba, São Paulo, em 2009. Neste, as diferentes formulações utilizadas constavam de: 3, 4 ou 5% de amido; 0, 0,05 ou 0,1% de goma e 0, 10 ou 20% de glicerol em combinações dadas seguindo o esquema fatorial 2^3 com 5 repetições do ponto central em um delineamento

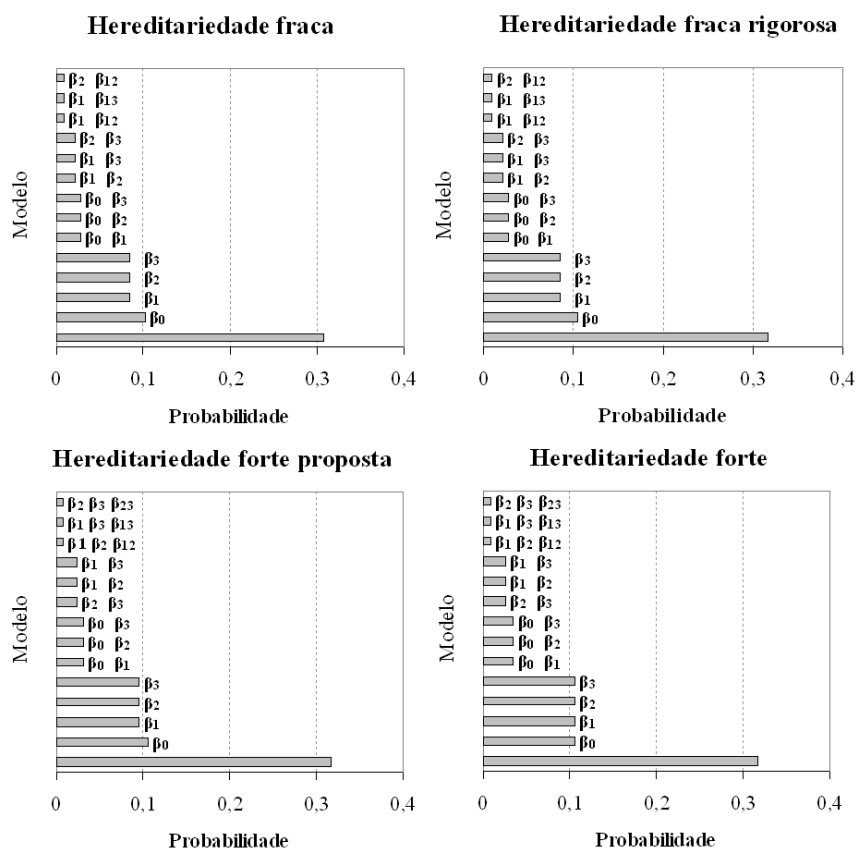


Figura 1 - Probabilidades *a priori* dos 14 modelos mais prováveis de acordo com a intensidade da hereditariedade.

inteiramente ao acaso. Para o presente trabalho, selecionou-se a variável resposta alongamento sob força máxima, em mm. As variáveis explanatórias X_1 = porcentagem de amido, X_2 = porcentagem de goma e X_3 = porcentagem de glicerol.

Para analisar os dados considere o modelo $y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_{12}x_1x_2 + \beta_{13}x_1x_3 + \beta_{23}x_2x_3 + \epsilon$ em que $x_1 = \frac{2X_1-8}{2}$, $x_2 = \frac{2X_2-0,1}{0,1}$ e $x_3 = \frac{2X_3-20}{20}$ e os seguintes valores para os hiperparâmetros: $\tau = 0,2053$, $c = 10$, $\lambda = 0,015$ e $\nu = 2, 2$.

O algoritmo de Gibbs, por sua vez, é iniciado com todos os termos ativos, ou seja, o vetor δ com todos os elementos iguais a 1, sendo geradas 50000 amostras de θ , descartando as 1000 primeiras amostras (“burn-in”=1000) e armazenando-se somente uma a cada dez amostras (“thin”=10).

Tabela 3 - Formulações de diferentes biofilmes de ervilha e respectivos alongamentos sob força máxima, em mm

Formulação Am (Trat)	Go x_1	Gli x_2	Along x_3	Along y	Formulação Am (Trat)	Go x_1	Gli x_2	Along x_3	Along y
1	-1	-1	-1	4,7458	8	1	1	1	25,8150
2	-1	-1	1	11,1002	9	0	0	0	4,6190
3	-1	1	-1	5,6792	9	0	0	0	3,8922
4	-1	1	1	10,7002	9	0	0	0	5,9042
5	1	-1	-1	-	9	0	0	0	4,9750
6	1	-1	1	12,7502	9	0	0	0	5,5378
7	1	1	-1	7,1750					

Na Figura 2 é apresentado o gráfico dos índices de diversidade (H) dos modelos *a posteriori* em função do valor de τ , seguindo a intensidade da hereditariedade. Nesta figura vê-se que, independente da intensidade da hereditariedade, a utilização de $\tau = \tau^*$ leva a uma menor diversidade dos modelos *a posteriori*. Além disso, o aumento da intensidade da hereditariedade leva a uma menor diversidade de modelos *a posteriori*, o que é também evidenciado pela diminuição dos valores de k apresentados na Tabela 4.

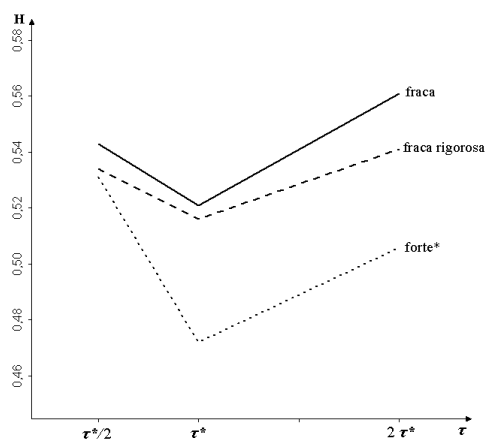


Figura 2 - Índice H de diversidade dos modelos *a posteriori*, segundo o valor de τ e a intensidade da hereditariedade: fraca, fraca rigorosa e forte proposta neste trabalho (forte*).

A Tabela 4 mostra, dentre outras informações, as probabilidades dos modelos *a posteriori* mais prováveis quando são usados diferentes valores de τ , e diferentes intensidades da hereditariedade. Nesta fica evidente a influência da intensidade da hereditariedade especialmente quando se considera $\tau = 2\tau^*$, pois, neste caso, para as hereditariedades fraca e fraca rigorosa, o parâmetro β_1 (ou termo Am) não está incluído nos modelos com maior probabilidade, o que acontece quando se adota a hereditariedade forte proposta neste trabalho. No entanto, quando se usa $\tau = \tau^*$, o modelo 6 apresentou, para todas as hereditariedades, maior probabilidade, embora, ainda assim, possa se observar a diminuição da probabilidade do modelo 5 e o aumento da probabilidade do modelo 6 com o aumento da intensidade da hereditariedade, o que evidencia a influência de tal princípio.

Concluindo, quando se usa $\tau = \tau^*$, valor recomendado por George e McCulloch (1993), independente da intensidade da hereditariedade, chega-se ao modelo de equação $y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_{12}x_1x_2 + \beta_{13}x_1x_3$, que inclui todos os efeitos principais formadores dos efeitos de interação dupla.

A Tabela 5 apresenta as probabilidades *a posteriori* de cada termo ser ativo no modelo, para as diferentes hereditariedade e valores de τ . Nestas pode-se ver que o termo Gli (Glicose) apresenta a maior probabilidade de ser ativo independente da hereditariedade ou valor de τ adotados. Como era de se esperar, com o aumento da intensidade da hereditariedade, os termos principais têm suas probabilidades de serem ativos aumentadas, o que é mais facilmente notado para o termo Am (Amido).

Conforme sugerido por George e McCulloch (1993), a estimação dos parâmetros mostrada a seguir, considera apenas para o valor de $\tau = \tau^*$ e a hereditariedade proposta neste trabalho.

As estimativas Bayesianas dos parâmetros, em que, os β_j são dadas pela média e σ é dada pela mediana, são $\beta_0 = 7,18$, $\beta_1 = 0,89$, $\beta_2 = 3,18$, $\beta_3 = 5,90$, $\beta_{12} = 2,38$, $\beta_{13} = 2,45$, $\beta_{23} = -0,09$ e $\sigma = 3,89$. Para realizar os diagnósticos das cadeias foram usados os testes de Gelman e Rubin (1992), Geweke (1992) e Raftery e Lewis (1992), na qual foram realizados pelos pacotes BOA(2009) e CODA, em que, para todos os parâmetros as cadeias apresentaram ser estacionárias e convergiram. Esta conclusão foram obtidas, pois pelos teste de Geweke todos valores estão entre uma normal padronizada, e pelo teste de Raftery e Lewis, todos os fatores de dependência são menores que cinco.

Tabela 4 - Probabilidades *a posteriori* de alguns modelos mais prováveis, considerando-se diferentes valores de τ , hereditariedade fraca, fraca rigorosa (fr. rig.) e forte proposta neste trabalho (forte*), números de diferentes modelos (*k*) e índices *H* de diversidade de Shannon (1948), ressaltando-se, em negrito, o modelos mais provável em cada uma das situações

Modelo	$\tau = \frac{\tau^*}{2}$			$\tau = \tau^*$			$\tau = 2\tau^*$		
	fraca	fr.rig.	forte*	fraca	fr.rig.	forte*	fraca	fr.rig.	forte*
1 - Int,Am,Go,Gli	0,118	0,129	0,122	0,104	0,109	0,150	0,088	0,096	0,171
2 - Int,Am,Go,Gli,Am×Gli	0,088	0,092	0,097	0,062	0,061	0,195	0,039	0,048	0,084
3 - Int,Am,Go,Gli,Am×Go	0,087	0,088	0,086	0,055	0,053	0,097	0,042	0,044	0,082
4 - Int,Am,Go,Gli,Go×Gli	0,050	0,053	0,049	0,023	0,026	0,039	0,013	0,012	0,026
5 - Int,Go,Gli,Am×Go,Am×Gli	0,046	0,048	0,052	0,131	0,128	0,049	0,108	0,118	0,055
6 - Int,Am,Go,Gli,Am×Go,Am×Gli	0,131	0,122	0,129	0,186	0,185	0,274	0,104	0,103	0,196
7 - Int,Am,Go,Gli,Am×Go,Am×Gli,Go×Gli	0,048	0,052	0,054	0,057	0,051	0,089	0,026	0,021	0,041
<i>k</i>	82	62	59	81	62	54	97	61	55
<i>H</i>	0,543	0,534	0,531	0,521	0,516	0,472	0,561	0,541	0,506

Tabela 5 - Probabilidades *a posteriori* dos termos considerados no modelo serem ativos, considerando-se diferentes valores de τ e herditariedades fraca, fraca rigorosa (fr.rig.) e forte proposta neste trabalho (forte*)

Termo	Parâmetro	$\tau = \frac{\tau^*}{2}$			$\tau = \tau^*$			$\tau = 2\tau^*$		
		fraca	fr.rig.	forte*	fraca	fr.rig.	forte*	fraca	fr.rig.	forte*
Intercepto	β_0	0,956	0,951	0,957	0,954	0,965	0,963	0,808	0,820	0,831
Am	β_1	0,779	0,793	0,926	0,657	0,657	0,875	0,565	0,580	0,812
Go	β_2	0,859	0,867	0,954	0,875	0,901	0,959	0,823	0,857	0,932
Gli	β_3	0,970	0,971	0,987	0,973	0,981	0,992	0,951	0,965	0,986
Am×Go	β_{12}	0,485	0,492	0,503	0,553	0,570	0,596	0,425	0,438	0,459
Am×Gli	β_{13}	0,493	0,477	0,501	0,572	0,569	0,607	0,437	0,448	0,470
Go×Gli	β_{23}	0,286	0,289	0,285	0,213	0,220	0,214	0,137	0,142	0,137

Conclusões

A especificação da *priori* para δ , sugerida por Chipman *et al.* (1997), apresentou problemas quando se considera a hereditariedade forte, e a alteração, proposta neste trabalho para esta *priori*, corrige os problemas apresentados. O novo nível de hereditariedade, proposto neste trabalho, não força os efeitos principais a estarem sempre presentes nos modelos, corrigindo a hereditariedade forte proposta por Chipman *et al.* (1997).

Como considerações adicionais, tem-se que, mesmo para situações em que há mais parâmetros do que observações, a metodologia poderia ser naturalmente utilizada sem quaisquer alterações.

Agradecimentos

Agradecemos à Coordenadoria para o Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela bolsa de doutorado do primeiro autor e aos revisores pelas sugestões que contribuíram para o aprimoramento do artigo.

BIZ, G.; ZOCCHI, S. S.; LEANDRO, R. A. Bayesian model selection using the sparse effects, hierarchy and heredity principles. *Rev. Bras. Biom.*, São Paulo, v.29, n.2, p.355-368, 2011.

■ **ABSTRACT:** *In experimental planning for adjustment of polynomials models involving k main factors and their interactions, it is frequent to adopt the 2^k , 3^k designs or its fractions. Furthermore, it is not unusual, when analysing the results of such experiments, to consider the heredity principle. In other words, once detected a significant interaction between factors, the factors that appear in this interaction and respective interactions should also be present in the model. In this work, this principle is incorporated directly in the prior, following the ideas proposed by Chipman *et al.* (1997), but changing some of the hyperparameters. What improves considerably the original methodology. Finally the methodology is illustrated by the analysis of the results of an experiment for the elaboration of pea starch biofilms.*

■ **KEYWORDS:** *Bayesian analysis; heredity principle; sparse effects principle; hierarchy principle; factorial design.*

Referências

BOX, G. E. P.; MEYER, R. D. An analysis for unreplicated fractional factorials. *Technometrics*, Baltimore, v.28, p.11-18, 1986.

CHIPMAN, H.; HAMADA, M.; WU, C. F. J. A bayesian variable-selection approach for analyzing designed experiments with complex aliasing. *Technometrics*, Baltimore, v.39, p.372-381, 1997.

- GELMAN, A.; CARLIN, J. B.; STERN, H. S.; RUBIN, D. B. *Bayesian data analysis*. London: Chapman and Hall, 1995. 526p.
- GELMAN, A.; RUBIN, D. B. Inference from iterative simulation using multiple sequence. *Statistical Science*, Hayward, v.7, n.4, p.457-511, 1992.
- GEORGE, E. I.; MCCULLOCH, R. E. Variable selection via Gibbs sampling. *J. Am. Stat. Assoc.*, Washington, v.88, n.423, p.881-889, 1993.
- GEWEKE, J. *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*. Oxford: Oxford University Press, 1992. p.625-631.
- R. Development Core Team. R Foundation for Statistical Computing. *R: A language and environment for statistical computing*, Vienna, Austria. ISBN 3-900051-07-0, URL. Disponível em: <<http://R-project.org>>. Acesso em: 27 jun. 2007.
- RAFTERY, A. L.; LEWIS, S. *How many iterations in the Gibbs sampler?*. Oxford: Oxford University, 1992. p.763-774.
- SHANNON, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.*, New York, v.27, p.379-423, 623-656, 1948.
- SMITH, B. J. *BOA: Bayesian Output Analysis Program for MCMC*. Disponível em: <<http://www.public-health.uiowa.edu/boa>>. Acesso em: 16 maio 2009.

Recebido em 10.03.2011.

Aprovado após revisão em 30.08.2011.