

INFLUÊNCIA DA INFORMAÇÃO À *PRIORI* NA ESCOLHA DE MODELOS PARA AVALIAÇÃO GENÉTICA DE SUÍNOS

Choice of models on swine breeding program

Ricardo da Fonseca¹, Jaime Araujo Cobuci², Aldrin Vieira Pires³, Cláudio Vieira Araújo⁴,
André Ribeiro Correa da Costa⁵

RESUMO

Programas de melhoramento são atividades que se desenvolvem durante anos e, por isso, devem ser flexíveis ao ajuste às novas situações criadas por mudanças nas tendências de mercado, na situação econômica e aquelas causadas por aumento do volume e qualidade dos dados e, também, por novas técnicas propostas pela comunidade científica. O ajuste a essas últimas deve ser feito, principalmente, por meio da substituição e escolha do modelo mais adequado para a descrição do fenômeno, em um determinado cenário. Os dados de ganho de peso médio diário, de um programa de melhoramento de suínos, envolvendo as raças Duroc, Landrace e Large White, foram analisados por meio da teoria bayesiana, por meio de dois modelos candidatos. Foram simulados três níveis de informação *à priori*: informativa, pouco informativa e não informativa. O comportamento das curvas das distribuições *à posteriori* e as respectivas estimativas associadas a cada nível de informação *à priori* foram analisadas e comparadas. Os resultados indicam que no modelo mais simples, as amostras das três raças são suficientes para produzir estimativas que não são alteradas pela informação *à priori*. Com relação ao mais parametrizado, as estimativas, para a raça Duroc, são alteradas pelo conhecimento prévio e, nesse caso, deve se buscar a melhor representação possível da distribuição *à priori* para obtenção de estimativas que são mais adequadas, dado o estado de conhecimento atual do melhorista.

Termos para indexação: Função de verossimilhança, Informação *à priori*, Suínos, Teoria bayesiana.

ABSTRACT

Breeding programs are activities developed during several years and then must be flexible to adjust new situations created by changes in market trends, economic context and those caused by increase of amount and quality of data and scientific discoveries. The adjust to the latter must be done essentially by means of change and chose of the most suitable model to describe the phenomenon at specific scenario. Data of swine breeding program involving Duroc, Landrace and Large White races were analyzed through Bayesian theory to two candidate models. Three levels of *à priori* information were simulated: informative, little informative and non informative. The behavior of the *à posteriori* distributions curves and estimates associated of each level of information *à priori* were analyzed and compared. The results indicate that in the simplest model the samples of the three races are sufficient to yield estimates that aren't altered by *à priori* information. Duroc's estimates, with regard to complex model, are altered by *à priori* knowledge before data collecting and then should be tried the best representation of the *à priori* distribution to obtain suitable estimates given the actual knowledge of the breeder at that moment.

Index terms: Likelihood function, *à priori* information, Pig, Bayesian theory.

(Recebido para publicação em 30 de abril de 2004 e aprovado em 6 de junho de 2005)

INTRODUÇÃO

A escolha do modelo mais adequado para análise dos dados em um programa de melhoramento animal é uma das etapas mais difícil e importante. Uma vez que os programas de melhoramento se estendem por vários anos, durante o seu desenvolvimento mudanças significativas com relação ao volume de dados, características de importância econômica e efeitos de maior importância na descrição dos dados observados devem ocorrer. Para

adaptar aos novos cenários, mudanças nos modelos de análise devem ser feitas e normalmente mais de um modelo candidato existe, e somente um deverá ser escolhido.

Algumas técnicas e interpretações com base na teoria bayesiana não possuem análogos na teoria frequentista. No paradigma bayesiano as inferências acerca de parâmetros populacionais são obtidas da distribuição *à posteriori* dos respectivos parâmetros considerados como variáveis aleatórias. A função densidade de probabilidade *à posteriori* é proporcional a multiplicação da função de

¹ Zootecnista – Professor da Universidade Estadual Paulista/UNESP – Dracena, SP – ricardo@dracena.unesp.br

² Zootecnista – Professor da Universidade Federal do Rio Grande do Sul/UFRGS – Porto Alegre, RS.

³ Zootecnista – Professor da Faculdade de Ciências da Saúde de Vitória/FAESA – Vitória, ES.

⁴ Zootecnista – Professor da Universidade Federal Rural da Amazônia/UFRA – Belém, PA.

⁵ Zootecnista – Mestre em Zootecnia

verossimilhança pela função densidade de probabilidade *à priori* do parâmetro. A função de verossimilhança representa a contribuição dos dados para o conhecimento dos parâmetros e, a distribuição *à priori*, reflete a incerteza sobre os possíveis valores dos parâmetros antes da coleta dos dados.

Zellner (1971), citado por Gianola & Fernando (1986), distingue dois tipos de informação *à priori*: baseada nos dados e não baseada nos dados. No melhoramento animal informações *à priori* baseadas nos dados, com relação a herdabilidades e componentes de variância, por exemplo, são obtidas de estimativas associadas a conjuntos de dados anteriores. A informação *à priori* não baseada nos dados é resultado de considerações teóricas.

À medida que o número de dados aumenta, a distribuição *à posteriori* se aproxima da função de verossimilhança. Portanto, a informação *à priori* tende a perder importância relativa com o aumento do volume de dados, ou seja, o conhecimento *à priori* é relativamente mais importante quando os dados são escassos (GIANOLA & FERNANDO, 1986).

Wang et al. (1994) analisaram a influência da informação fornecida *à priori* sobre as estimativas de componentes de variância e contrastes em um experimento de resposta à seleção para tamanho de leitegada em suínos da raça Landrace. Quatro tipos de distribuições *à priori* foram usadas. Diferentes estimativas e desvios padrão foram encontrados conforme a informação *à priori* utilizada. Segundo esses autores, diferentes distribuições *à priori* devem ser testadas no mesmo conjunto de dados; se as conclusões fornecidas são muito diferentes, os resultados das análises devem ser interpretados com cuidado, uma vez que a função de verossimilhança não tem força e é influenciada por qualquer informação *à priori*.

Além disso, o mesmo conjunto de dados pode fornecer informação suficiente para produzir uma função de verossimilhança forte (produz as mesmas estimativas independente do tipo de informação *à priori*) para um modelo menos parametrizado e fraca (estimativas são dependentes da informação *à priori*) para um modelo mais

parametrizado. Dessa forma, um auxílio para escolha do modelo poderia vir da observação do comportamento das estimativas produzidas por um determinado modelo com diferentes tipos de informação *à priori*. Se para um modelo mais complexo as estimativas variam muito de acordo com a informação *à priori*, então um modelo mais simples deve produzir predições mais consistentes.

Dessa forma, por meio da observação da influência da informação *à priori*, tem-se como objetivo, a escolha do melhor modelo entre dois candidatos, utilizando-se dados provenientes de um programa de melhoramento de suínos.

MATERIAL E MÉTODOS

Dados e Modelos

Os registros de ganho de peso médio diário (dos 70 aos 140 dias) utilizados neste estudo, são provenientes de suínos das raças Duroc, Landrace e Large White (Tabela 1), de uma granja localizada no município de Chapecó, Estado de Santa Catarina. E são referentes ao período de 1995 a 1997.

Dois modelos foram utilizados para descrever as informações da característica ganho de peso diário: $y = Xb + Za + e$ (1) e $y = Xb + Z_1a + Z_2c + e$ (2). O primeiro foi constituído pelo efeito de grupo contemporâneo formado pelas classes de rebanho-ano-estação-sexo; da covariável linear da idade à saída da creche; dos efeitos genético aditivo de animal e residual. No segundo modelo, foi considerado todos os efeitos do primeiro modelo, acrescido do efeito de ambiente comum (leitegada). Em que, y é o vetor de observações da característica ganho de peso médio diário; X , a matriz de incidência associando os efeitos de grupo contemporâneo (18 níveis) e da covariável, b , o vetor de efeitos fixos de grupo; Z , a matriz de incidência associando os efeitos genéticos aditivos; a , o vetor de efeitos genéticos aditivos diretos; Z_1 , a matriz de incidência associando os efeitos genéticos aditivos; Z_2 , a matriz de incidência associando os efeitos devido ao ambiente comum de leitegada; c , o vetor de efeitos devido ao ambiente comum da leitegada; e , o vetor de efeito residual.

TABELA 1 – Número de registros, número de animais na matriz dos coeficientes de parentesco (A^{-1}), médias e seus respectivos desvios padrão.

Raça	Número de registros	Animais na A^{-1}	Média (g)	Desvio padrão
Duroc	1.810	6.835	800,55	89,77
Landrace	3.143	8.651	825,25	112,07
Large White	4.894	11.184	858,78	110,04

As suposições acerca da distribuição dos componentes dos modelos foram:

$$\mathbf{a} / \mathbf{A}, \sigma_a^2 \sim N(\mathbf{0}, \mathbf{A} \sigma_a^2), \quad \mathbf{e} / \sigma_e^2 \sim N(\mathbf{0}, \mathbf{I} \sigma_e^2)$$

$$\mathbf{c} / \sigma_c^2 \sim N(\mathbf{0}, \mathbf{I} \sigma_c^2)$$

em que, \mathbf{A} é a matriz do numerador de coeficientes de parentesco dos animais e, \mathbf{I} uma matriz identidade.

Distribuições “à priori”

Assim, para o efeito de grupo contemporâneo, um prior não informativo foi atribuído: $f(\mathbf{b}) \propto \text{uniforme}$. Para os componentes de variância, foi utilizada a distribuição qui-quadrado escalonada invertida:

$$f(\sigma_i^2 / v_i, s_i^2) \propto (\sigma_i^2)^{-\frac{v_i-1}{2}} \exp\left(-\frac{1}{2} v_i s_i^2 / \sigma_i^2\right) \quad i = a, c, e.$$

em que, v_i = parâmetro de forma ou grau de confiança; s_i^2 = parâmetro de escala.

Para os valores genéticos aditivos, assumiu-se a distribuição normal:

$$f(a_i / \mu_a, \sigma_{a_i}^2) = \frac{1}{2\pi^{\frac{1}{2}} \sqrt{\sigma_{a_i}^2}} \exp\left\{-\frac{1}{2} \left[(a_i - \mu_a) \frac{1}{\sigma_{a_i}^2} (a_i - \mu_a) \right]\right\}$$

em que, μ_a = média dos valores genéticos aditivos; $\sigma_{a_i}^2$ = componente de variância genético aditivo. A distribuição normal, também foi considerada para o efeito residual.

$$f(e_i / \mu_e, \sigma_{e_i}^2) = \frac{1}{2\pi^{\frac{1}{2}} \sqrt{\sigma_{e_i}^2}} \exp\left\{-\frac{1}{2} \left[(e_i - \mu_e) \frac{1}{\sigma_{e_i}^2} (e_i - \mu_e) \right]\right\}$$

em que, μ_e = média dos valores residuais; $\sigma_{e_i}^2$ = componente de variância residual.

Para o efeito comum de leitegada a seguinte distribuição normal foi ajustada:

$$f(c_i / \mu_c, \sigma_{c_i}^2) = \frac{1}{2\pi^{\frac{1}{2}} \sqrt{\sigma_{c_i}^2}} \exp\left\{-\frac{1}{2} \left[(c_i - \mu_c) \frac{1}{\sigma_{c_i}^2} (c_i - \mu_c) \right]\right\}$$

em que, μ_c = média dos efeitos comuns de leitegada; $\sigma_{c_i}^2$ = componente de variância do efeito comum de leitegada.

Construção das distribuições “à priori” dos componentes de variância

A análise bayesiana exige atribuição de distribuições “à priori” para os efeitos do modelo.

Essas distribuições refletem a incerteza sobre os possíveis valores dos parâmetros antes dos dados serem coletados. Uma distribuição à priori não-informativa representa a ausência de informação sobre um possível valor, e por isso, é normalmente representada como uma constante. Em outras palavras, qualquer valor tem igual probabilidade de ocorrência. Na distribuição pouco informativa alguns valores são mais prováveis que outros, entretanto a variância ainda é grande, indicando que o intervalo de valores prováveis é ainda significativamente extenso. Na distribuição informativa a variância é pequena, sendo assim o intervalo de valores prováveis é mais curto e alguns poucos valores tem probabilidade de ocorrência alta.

Foram formulados três tipos de distribuições “à priori” para os componentes de variância com base nos valores de v_i e s_i^2 (não-informativa, pouco informativa e informativa).

Na Tabela 2 estão apresentados os valores das distribuições “à priori” para os modelos 1 e 2. As informações sobre o parâmetro de escala das distribuições “à priori” informativa - III, foi baseado nas estimativas dos componentes de variância de cada população (Landrace, Large White e Duroc), obtidas pelo método REML, por meio do programa MTDFREML (BOLDMAN et al., 1995). Para todos os casos, os valores iniciais usados para estimação dos componentes de variância foi 0, representando o desconhecimento prévio dos valores de variância. O parâmetro de forma foi definido com um valor que proporcionasse baixa variabilidade à distribuição, refletindo conhecimento acerca dos prováveis valores do componente de variância. A distribuição “à priori” pouco informativa - II foi construída de modo a refletir algum conhecimento acerca dos valores dos componentes de variância. Considerando o componente de variância genética aditiva, o parâmetro de escala foi tomado como um valor próximo à estimativa obtida pelo método REML. Os valores assumidos para os parâmetros de escala e forma da distribuição “à priori” não informativa - I, produz uma distribuição uniforme, indicando ignorância sobre os possíveis valores dos componentes de variância antes da coleta dos dados, uma vez que a probabilidade de ocorrência de qualquer valor do componente é a mesma. Procedimento semelhante foi adotado por Blasco et al. (1998).

TABELA 2 – Valores dos parâmetros de escala (S_i^2) e forma (ν_i) das distribuições “à priori” dos componentes de variância referentes aos modelo 1 e 2.

Prior	Variância					
	σ_a^2		σ_e^2		σ_c^2	
	S_a^2	ν_a	S_e^2	ν_e	S_c^2	ν_c
Modelo – 1						
I	0	-2	0	-2	-	-
II	4000,00	4	3500,00	5	-	-
III	3568,35	150	2984,19	55	-	-
Modelo – 2						
I	0	-2	0	-2	0	-2
II	3500,00	5	3869,86	9	506,78	9
III	2679,56	150	3013,32	120	673,24	150

I - Prior não-informativo; II - Prior pouco informativo; III- Prior informativo.

Estimação dos parâmetros por inferência bayesiana

As seqüências de Gibbs para os dois modelos foram geradas pelo programa MTGSAM (TASSELL & VLECK, 1996). A análise de convergência foi realizada por meio de múltiplas cadeias com diferentes especificações para os parâmetros comprimento total da seqüência, tamanho do período de descarte amostral (*burn-in*) e intervalo de amostragem (*thinning interval*). O programa Gibanal (KAAM, 1998) foi utilizado para análise das cadeias geradas e para representação gráfica das distribuições marginais à posteriori, a partir de arquivos contendo as amostras geradas pelo MTGSAM.

Para inferências acerca dos componentes de variâncias foram utilizadas cadeias de 150.000 iterações. Após o descarte das primeiras 1.000 iterações (período de burn-in) e a aplicação de um intervalo de amostragem com valor 10, obteve-se uma amostra de tamanho 14.900. O mesmo procedimento foi aplicado para as razões dos componentes de variância genético aditivo e de ambiente comum com a variância fenotípica, entretanto, o tamanho total da cadeia foi de 250.000 iterações com descarte de 2.000 amostras (burn-in). O intervalo de amostragem foi o mesmo utilizado para os componentes de variância, proporcionando uma amostra de tamanho 24.800 para inferências.

RESULTADOS E DISCUSSÃO

O procedimento apresentado se baseia na escolha de modelos por meio de comparações das estimativas à posteriori para diferentes tipos de informação à priori. Se a amostra é grande o suficiente, a função de verossimilhança sobrepõe a informação à priori, a qual passa a ter pouca importância nas inferências. Dessa forma não são esperadas grandes mudanças de comportamento das densidades das distribuições à posteriori, mesmo sob modelos mais parametrizados. Entretanto, se a amostra não contém informação suficiente, as estimativas sob modelos mais complexos são sensíveis à informação à priori nas decisões de análise que devem ser tomadas em diferentes momentos.

Nas Figuras 1, 2 e 3, estão as representações gráficas das distribuições à posteriori para os componentes do modelo 2, respectivamente, para as raças Duroc, Landrace e Large White. Para as três raças e para todos os parâmetros analisados, a distribuição à posteriori construída a partir de uma distribuição à priori informativa se desviou das duas outras distribuições. Esses resultados evidenciam que a informação incorporada às análises tem efeito sobre as estimativas obtidas. Distribuições à priori representando ausência ou pouca informação, quando utilizadas para inferências, geraram estimativas com valores próximos.

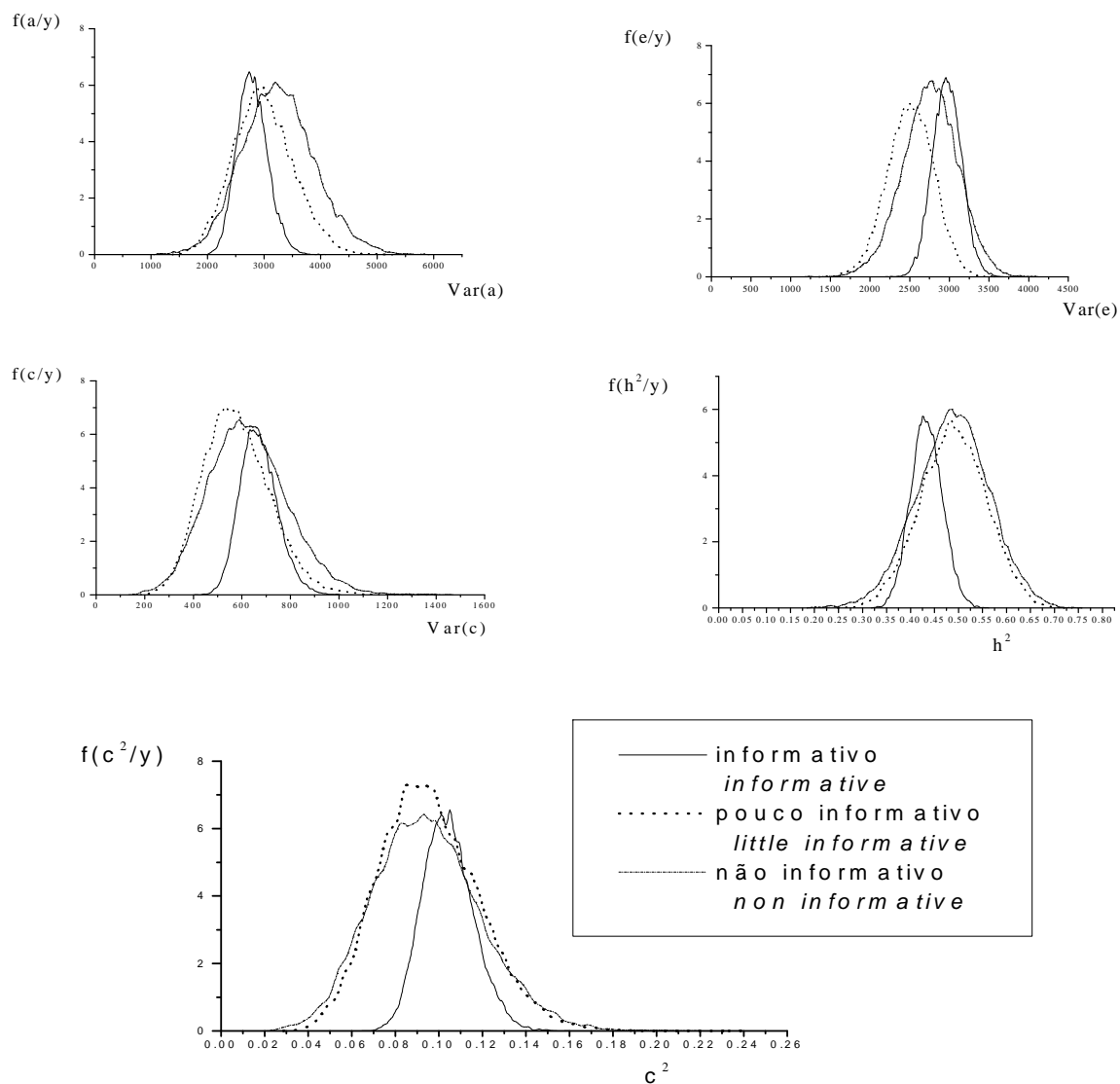


FIGURA 1 – Distribuições *à posteriori*, obtidas com base no modelo 2, para os componentes de variância genética aditiva, residual, ambiente comum, herdabilidade e razão da variância devido ao ambiente comum e variância fenotípica (c^2), de acordo com as distribuições “*à priori*” informativa, pouco informativa e não-informativa, para a raça Duroc.

As estimativas obtidas das distribuições *à posteriori* (médias) para os componentes de variância, herdabilidade e razão entre variância devido ao ambiente comum (c^2) e variância fenotípica (σ_p^2), de acordo com a raça, as distribuições “*à priori*” e o modelo, estão apresentadas na Tabela 3.

Ao analisar estas distribuições observa-se que são próximas para os três tipos de informação “*à priori*”

utilizada sob o modelo 1. As distribuições associadas às densidades *à priori* pouco informativas e não-informativas são muito semelhantes e, diferem relativamente pouco da distribuição associada à densidade *à priori* informativa. A maior diferença foi observada na raça Large White. O mesmo comportamento é observado para o modelo 2. Esses resultados corroboram as observações feitas com base nas Figuras 1, 2 e 3.

As distribuições à posteriori resultantes do uso de priors informativos, como esperado, apresentaram a menor variância, e a estimativa obtida foi a mais provável. Intervalos de alta densidade podem então ser construídos. Os valores contidos no intervalo são associados a altas probabilidades e, portanto, plausíveis de ocorrer.

Os intervalos de alta densidade definidos para a variância genética aditiva, variância residual e herdabilidade na raça Large White foram de 4600 – 5500, 4800 – 5400 e 0,46 – 0,52, respectivamente.

Para a raça Large White e para o modelo 1, as estimativas obtidas das distribuições associadas às densidades à priori pouco informativas e não-informativa para o componente de variância genética aditiva e herdabilidade estão fora do intervalo e são, portanto, valores menos significativos (Tabela 3). Entretanto, as estimativas obtidas para a variância residual estão próximas ao limite inferior do intervalo e, assim, são considerados valores plausíveis.

Apesar das diferenças entre estimativas obtidas das distribuições à priori não-informativa e pouco informativa com aquelas da distribuição à priori informativa, os resultados sugerem, por comparação entre os resultados de cada modelo, que para o modelo 1 a informação à priori tem pequeno efeito sobre as estimativas (Tabela 3).

Na raça Large White (modelo 2), as estimativas obtidas das distribuições à posteriori associadas às distribuições à priori pouco informativa e não-informativa são semelhantes, e tem seus valores mais próximos das estimativas obtidas da distribuição à posteriori obtida com a priori informativa quando comparadas as diferenças encontradas no modelo 1 (Tabela 3). Por exemplo, a diferença entre as estimativas de variância genética aditiva entre as distribuições obtidas a partir de priors não-informativa e informativa no modelo 1 foi 713,2, enquanto que no modelo 2 foi de 415,46.

A mesma análise feita para a raça Large White, no modelo 1 é feita na raça Duroc, no modelo 2. As seguintes regiões de alta densidade são definidas para os componentes de variância genética aditiva, componente de variância residual, herdabilidade, componente de ambiente comum de leitegada e c^2 , as quais são 2500 – 3000; 2750 – 3250; 0,40 – 0,47; 600 – 750 e 0,09 – 0,11, respectivamente.

Para todos os intervalos existem estimativas muito próximas dos limites ou fora desses, e são consideradas menos plausíveis (Tabela 3).

A comparação do deslocamento das curvas pelo distanciamento ou aproximação, fornece evidências da força da função de verossimilhança sob os modelos considerados.

TABELA 3 – Estimativas (médias) dos componentes de variância, herdabilidade (h^2) e razão entre variância devido ao ambiente comum e variância fenotípica (c^2), de acordo com as raças e distribuições “à priori”.

	Duroc			Landrace			Large White		
	I	II	III	I	II	III	I	II	III
Modelo 1									
σ_a^2	3905,02	3858,95	3695,88	4712,35	4580,61	4298,05	5847,06	5833,55	5133,86
σ_e^2	2831,90	2847,49	2928,57	5135,71	5200,14	5354,46	4877,50	4878,17	5231,49
h^2	0,57	0,57	0,56	0,47	0,46	0,44	0,54	0,54	0,49
Modelo 2									
σ_a^2	3253,11	2984,17	2790,12	2870,22	2784,45	2064,17	4226,15	4152,06	3810,69
σ_e^2	2758,09	2509,29	2971,26	5442,88	5485,69	5831,47	4774,13	4809,06	4973,58
σ_c^2	626,45	579,20	669,87	1081,25	1063,58	1179,84	1431,60	1420,87	1458,27
h^2	0,48	0,48	0,43	0,30	0,29	0,22	0,40	0,39	0,37
c^2	0,09	0,09	0,10	0,11	0,11	0,12	0,13	0,13	0,14

I - Prior não-informativo; II- Prior pouco informativo; III- Prior informativo; σ_a^2 - variância genética aditiva; σ_e^2 - variância residual; σ_c^2 - variância de ambiente comum. - variância genética aditiva; - variância residual; - variância comum; h^2 - herdabilidade; c^2 - razão entre variância devido ao ambiente comum e variância fenotípica.

Ao se comparar as distribuições obtidas com os diferentes *prioris* pelo o modelo 2, observa-se que as distribuições referentes aos componentes de variância e as herdabilidade para a raça Duroc (Tabela 3), apresentaram maior distanciamento quando comparadas às do modelo 1. Considerando o componente de variância genética aditiva, a diferença entre os resultados das distribuições *à priori* não-informativa e informativa obtidos sob o modelo 1 foi de 209,14 e sob o modelo 2 foi de 462,99. O mesmo comportamento é observado para a raça Landrace, embora

em menor proporção. A diferença entre os componentes de variância genética para essa raça sob o modelo 1 foi de 414,13 e sob o modelo 2 foi de 806,05. Na raça Large White, as distribuições apresentaram-se bem próximas àquelas observadas para o modelo 1.

Esses resultados indicam que a função de verossimilhança sofre influência da informação *à priori* sob o modelo 2, principalmente para a raça Duroc, em que se observa maiores variações nas predições de acordo com as diferentes informações *à priori*.

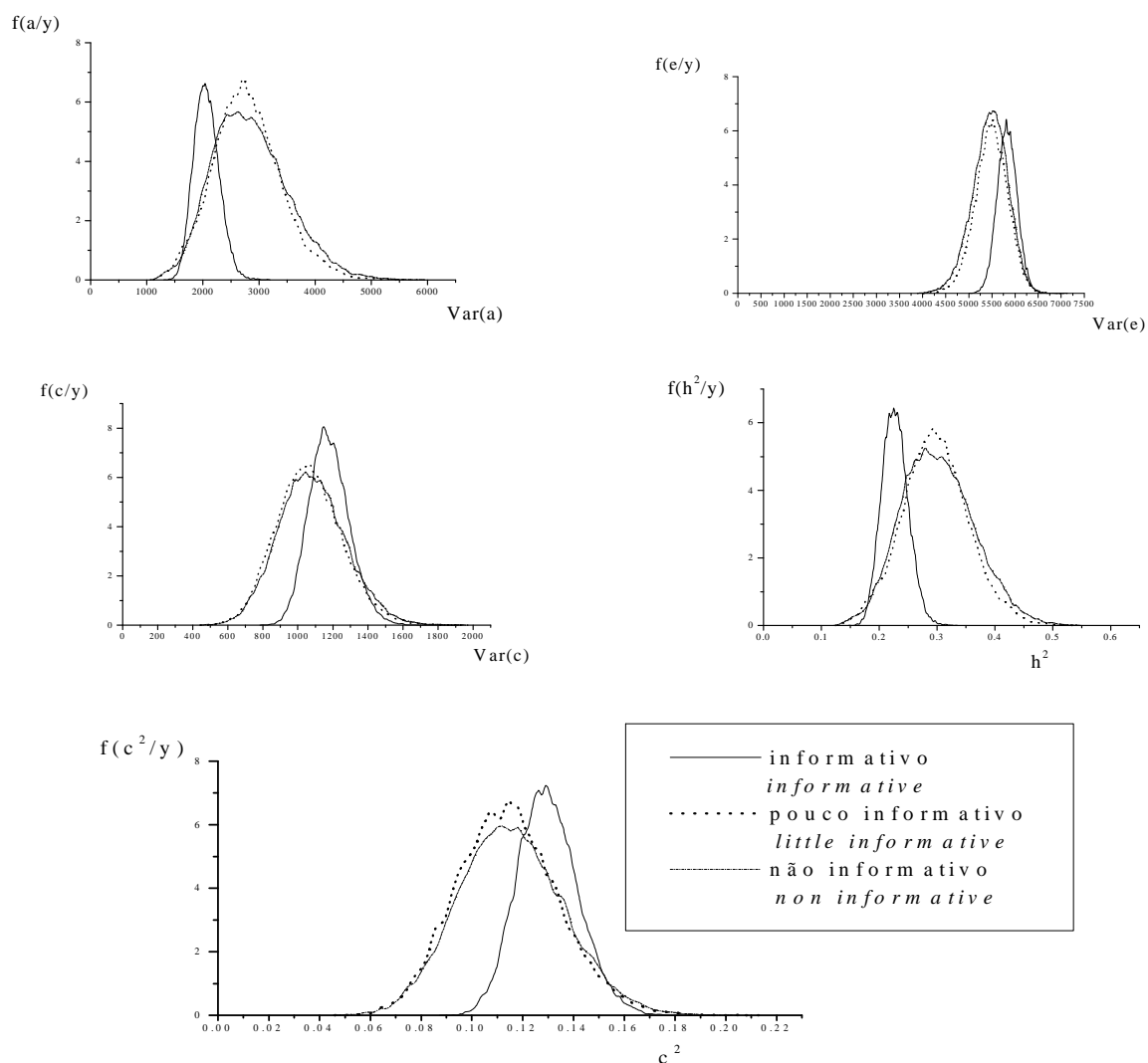


FIGURA 2 – Distribuições *à posteriori*, obtidas com base no modelo 2, para os componentes de variância genética aditiva, residual, ambiente comum, herdabilidade e razão da variância devido ao ambiente comum e variância fenotípica (c^2), de acordo com as distribuições “*à priori*” informativa, pouco informativa e não-informativa, para a raça Landrace.

A técnica sugerida neste estudo auxilia a escolha de um modelo operacional, dada a amostra e a quantidade de informação antes da coleta dos dados. Assim, após a definição de alguns modelos para descrição do fenômeno, utiliza-se o conjunto de dados nas análises com diferentes distribuições *à priori*. Se para um modelo as estimativas não se alteram com os diferentes níveis de informação *à priori*, significa que as estimativas são fortemente baseadas nos dados (via função de verossimilhança) e o conjunto de dados é adequado para análises sob esse modelo. Se as estimativas obtidas variam

consideravelmente sob diferentes *prioris*, então pode-se concluir que a função de verossimilhança é dependente da contribuição *à priori* para a produção das estimativas. Nesses casos, alguns problemas de pequeno número de informações nas subclasses podem surgir e gerar estimativas viesadas. Ao menos que uma distribuição *à priori* informativa esteja disponível, deve-se optar pela utilização de modelos menos parametrizados. Dessa forma, o procedimento auxilia na escolha do modelo que fornece estimativas mais adequadas à quantidade de informações disponível.

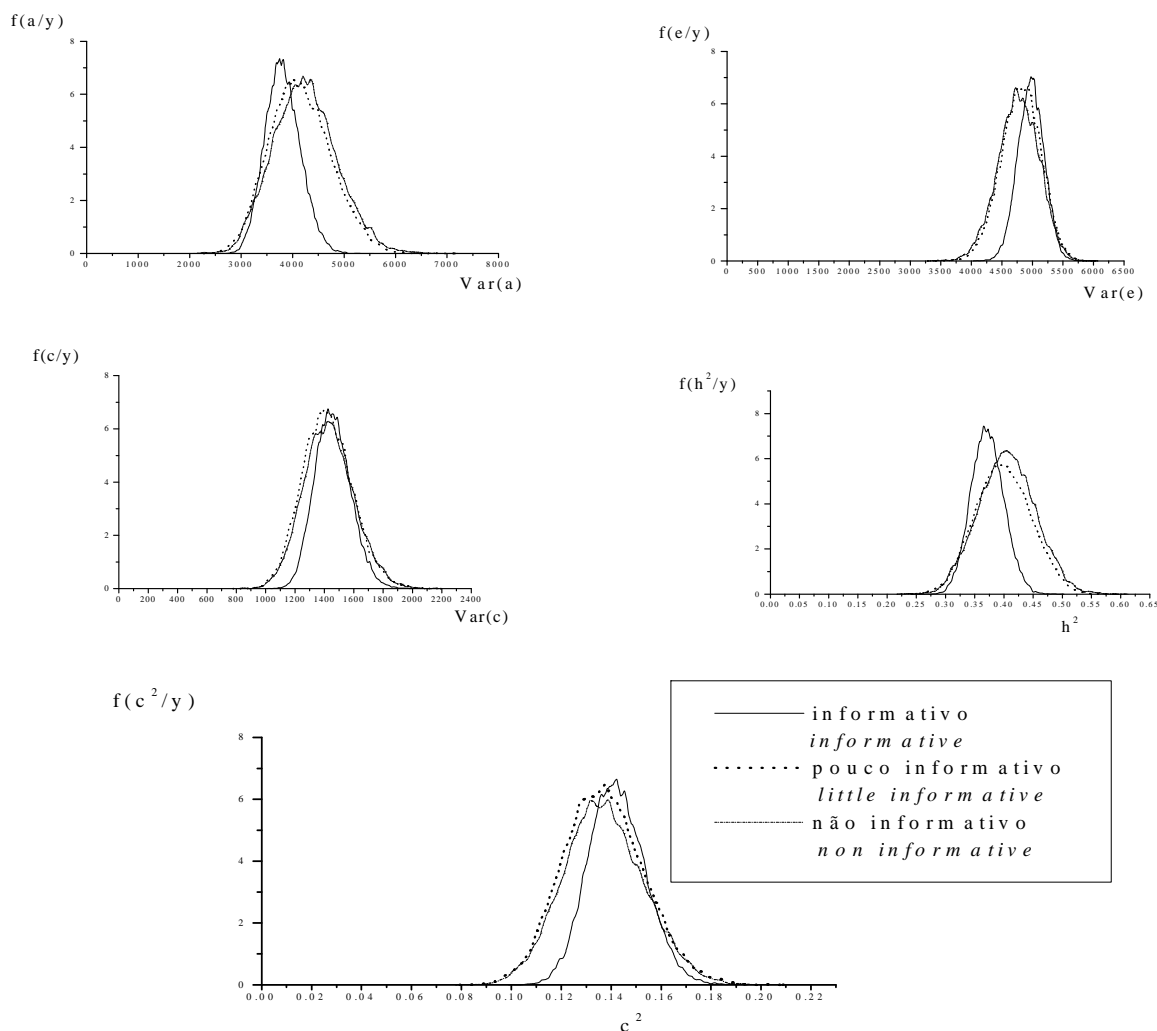


FIGURA 3 – Distribuições, obtidas com base no modelo 2, para os componentes de variância genética aditiva, residual, ambiente comum, herdabilidade e razão da variância devido ao ambiente comum e variância fenotípica (c^2), de acordo com as distribuições “*à priori*” informativa, pouco informativa e não-informativa, para a raça Large White.

Os resultados obtidos com o enfoque bayesiano, para a raça Large White, sugerem que as estimativas obtidas sob o modelo 2, não sofrem alterações consideráveis da informação *à priori*. Entretanto, a amostra das raças Landrace e Duroc, não são suficientes para produzir funções de verossimilhança fortes, e assim, as estimativas obtidas sob o modelo 2 sofrem alterações quando diferentes tipos de informação *à priori* são utilizadas.

Como citado anteriormente, se o conjunto de dados não é grande (como o da raça Duroc) a utilização de um modelo mais parametrizado (modelo 2) pode ser problemático, uma vez que algumas subclasses conterão uma quantidade pequena de dados e conseqüentemente contribuirão para uma verossimilhança fraca, e portanto, susceptível a informação *à priori*.

Se o melhorista sente dificuldade na formulação da distribuição *à priori* (normalmente é o caso na prática), então, o modelo 1 (mais simples), deve ser usado para o processo de estimação, uma vez que fornecerá resultados mais consistentes do que o modelo 2 (mais complexo), utilizando representações imprecisas do conhecimento anterior à coleta de dados. Com o passar do tempo, o programa de melhoramento irá possuir maior volume de dados e, assim, aumentará a possibilidade de construção de uma distribuição *à priori* com mais confiança.

CONCLUSÕES

Uma vez que a quantidade e qualidade dos dados são adequadas, a *informação à priori* tem pouca influência nas estimativas obtidas e assim, inferências podem ser realizadas utilizando modelos mais parametrizados. Por outro lado, quando a *informação à priori* é importante na obtenção dos valores estimados, a escolha do modelo deve ser feita de maneira criteriosa, em que a possibilidade de representação fidedigna do estado de conhecimento anterior à coleta dos dados, deve ser um dos fatores importantes a ser considerado na análise. Finalmente, se a representação matemática do conhecimento *à priori* não é

clara e/ou fácil, deve escolher modelos menos parametrizados (mais simples) para as inferências, uma vez que esses têm a possibilidade de serem menos sensíveis a esse tipo de informação. Assim, para a raça Duroc e Landrace, o modelo 1 deve ser utilizado e para a raça Large White as análises podem ser conduzidas com base no modelo 2, mais parametrizado.

REFERÊNCIAS BIBLIOGRÁFICAS

BLASCO, A.; SORENSEN, D.; BIDANEL, J. P. Bayesian inference of genetic parameters and selection response for litter size components in pigs. **Genetics**, [S.l.], v. 149, n. 1, p. 301-306, May 1998.

BOLDMAN, K. G.; KRIESE, L. A.; VLECK, L. D. van. **A manual for use of MTDFREML**: set programs to obtain estimates of variances and covariances [DRAFT]. Lincoln: Agricultural Research Service, 1995. 115 p.

GIANOLA, D.; FERNANDO, R. L. Bayesian methods in animal breeding theory. **Journal Animal Science**, Amsterdam, v. 63, n. 1, p. 217-244, Ago. 1986.

KAAM, J. B. C. H. M. van. "GIBANAL": analyzing program for Markov Chain Monte Carlo sequences. Version 2.10. Wageningen: Agricultural University, 1998.

TASSELL, C. P. van; VLECK, L. D. van. Multiple-trait Gibbs sampler for animal models: flexible programs for bayesian and likelihood-based (co)variance component inference. **Journal Animal Science**, Amsterdam, v. 74, n. 11, p. 2586-2597, Nov. 1996.

WANG, C. S.; GIANOLA, D.; SORENSEN, D. A.; JENSEN, J.; CHRISTENSEN, A.; RUTLEDGE, J. J. Response to selection for litter size in Danish Landrace pigs: a Bayesian analysis. **Theory Applied Genetics**, [S.l.], v. 88, n. 2, p. 220-230, Mar./Apr. 1994.