

Influência da Informação *a priori* na Avaliação Genética Animal Utilizando Dados Simulados¹

José Marques Carneiro Júnior², Giselle Mariano Lessa de Assis³, Ricardo Frederico Euclides⁴, Paulo Sávio Lopes⁴

RESUMO - Estudos de simulação foram conduzidos com o objetivo de verificar a influência da informação *a priori* nas avaliações genéticas dos animais. Foi simulado um genoma de 3.000 centimorgans, considerando-se uma única característica quantitativa, determinada por 800 locos, com dois alelos por loco, na qual a herdabilidade variou de 0,40 a 0,60. Foram simulados 1.500 machos e 1.500 fêmeas, que formaram a população-base. A partir da população-base, foram formados três tamanhos de populações: POP1 (100 animais), POP2 (300 animais) e POP3 (1.600 animais). Foram empregados três níveis de informação *a priori*: *priors* não-informativos (PNI), *priors* pouco informativos (PPI) e *priors* informativos (PI). Foram avaliadas também as conseqüências de se incluir nas análises valores de *priors* superestimados (com 50% de erro), considerados pouco informativos e informativos. Para verificação da influência de diferentes níveis de informação na avaliação genética, foram utilizadas a Porcentagem de Erro em relação ao valor verdadeiro dos componentes de variância, a Correlação de Spearman e o Quadrado Médio do Erro entre os valores genéticos reais e os preditos. As populações POP1 e POP2 apresentaram maiores porcentagens de erro e foram mais sensíveis às informações *a priori*. Maiores níveis de informação conduziram a melhores estimativas das variâncias, sendo que os QME apresentaram-se menores com o aumento do nível de informação, principalmente para as populações pequenas. Entretanto, as Correlações de Spearman permaneceram sempre inferiores a 0,70 para POP1 e POP2, indicando grandes alterações nas classificações dos animais. Conclui-se que existe uma perda ao se trabalhar com conjuntos de dados pequenos, mesmo quando informações *a priori* adicionais estão disponíveis.

Palavras-chave: análise Bayesiana, componentes de variância, informação *a priori*, parâmetros genéticos, simulação

Influence of Prior Information in Animal Genetic Evaluation Using Simulated Data

ABSTRACT - The influence of *a priori* information on animal genetic evaluation was studied using a simulated 3.000 centimorgans genome, for a single trait governed by 800 loci, with two alleles for locus, and heritability varying from 0.40 to 0.60. Three populations with different sizes were derived from the base-population constituted by 1.500 males and 1.500 females: POP1 (100 animals), POP2 (300 animals) and POP3 (1.600 animals). Three *a priori* information levels were considered: no-informative priors (PNI), slightly informative priors (PPI) and informative priors (PI). The effect of using overestimated priors (with 50% error), considering slightly informative and informative priors on genetic evaluation was also studied. The influence of different levels of prior information on genetic evaluation was evaluated by the following criteria: error percentage related to the true value of variance components, Spearman correlation and Mean Square Error between real and predicted genetic values. POP1 and POP2 presented larger Error Percentages and were more sensitive to prior information. Larger levels of information led to best variance estimates and QME became smaller by increasing the information level, mainly for small populations. However, Spearman correlation were lower than 0.70 for POP1 and POP2, indicating significant changes in ranking of animals. Therefore, genetic evaluation of small populations is problematic, even if additional prior information is available.

Key Words: *a priori* information, Bayesian analyses, genetic parameters, simulation, variance components

Introdução

A identificação e o acasalamento de animais geneticamente superiores é essencial para o progresso genético. Entretanto, para a realização da seleção, é necessário o conhecimento dos componentes de variância da população com a qual se está trabalhando. Segundo Kennedy (1981), essas variâncias

freqüentemente não são conhecidas, particularmente onde novas características ou populações provenientes de programas de melhoramento são avaliadas. Um recurso é primeiro estimar componentes de variância e, então, predizer os valores genéticos de interesse em um mesmo conjunto de dados.

Gianola & Foulley (1982) introduziram os métodos Bayesianos em melhoramento animal no contexto de

¹ Financiamento: CAPES.

² Zootecnista, Doutorando em Genética e Melhoramento - UFV, 36571-000, Viçosa, MG. Jmcj_jr@yahoo.com.br

³ Zootecnista, Doutorando em Genética e Melhoramento - UFV, 36571-000, Viçosa, MG. giselle@alunos.ufv.br

⁴ Professor do DZO/UFV, Departamento de Zootecnia, 36571-000, Viçosa, MG. rbaja@ufv.br; plopes@ufv.br

características de limiar e, posteriormente, Gianola & Fernando (1986) salientaram as possibilidades da exploração de técnicas Bayesianas nas avaliações genéticas. Embora os métodos Bayesianos sejam teoricamente poderosos, usualmente múltiplas integrais devem ser resolvidas, para obtenção de distribuições marginais *a posteriori* utilizadas para uma completa inferência Bayesiana. Essas integrais, com bastante frequência, são intratáveis analiticamente (Blasco, 2001), sendo necessário utilizar métodos de integração numérica, como os métodos de Monte Carlo Markov Chain (MCMC). Os métodos MCMC são uma família de processos iterativos que geram amostras de distribuições multivariadas e, entre eles, a Amostragem de Gibbs tem sido bastante utilizada para estimar distribuições marginais *a posteriori* a partir de densidades condicionais completas.

Estas densidades condicionais completas correspondem à distribuição de um ou de vários parâmetros especificados condicionais a todos os outros parâmetros do modelo. A Amostragem de Gibbs pode ser considerada uma implementação prática, pois, o conhecimento das distribuições condicionais é suficiente para determinar uma distribuição conjunta (Wang et al., 1994).

Em muitas situações, informações adicionais estão disponíveis sobre o valor do parâmetro desconhecido. Caso esta informação possa ser utilizada para se construir uma distribuição *a priori* para o parâmetro de interesse, os métodos Bayesianos podem ser empregados para se estimar o valor do parâmetro desconhecido. Como os componentes de variância são variáveis aleatórias na inferência Bayesiana, torna-se necessário assumir a forma de suas distribuições. Usualmente, as distribuições *a priori* admitidas para os componentes de variância são as distribuições de qui-quadrado invertidas escalonadas independentes. A função densidade de probabilidade de uma variável aleatória X que possui distribuição de qui-quadrado invertida escalonada é a seguinte:

$$f_x(x) = \left[\Gamma\left(\frac{\nu}{2}\right) 2^{\nu/2} \right]^{-1} (\nu s^2)^{\nu/2} x^{-\left(\frac{\nu}{2}+1\right)} \exp\left(-\frac{\nu s^2}{2x}\right) \propto x^{-\left(\frac{\nu}{2}+1\right)} \exp\left(-\frac{\nu s^2}{2x}\right)$$

em que ν representa o parâmetro da distribuição que corresponde ao “grau de confiança” do componente de variância e s^2 , o parâmetro da distribuição que pode ser interpretado como um valor *a priori* do componente de variância.

De acordo com os parâmetros dessa distribuição (ν, s^2) , diferentes curvas podem ser obtidas. Assumindo valores de $\nu = -2$ e $s^2 = 0$, por exemplo, tem-se uma distribuição uniforme, indicando que a probabilidade de ocorrência de qualquer valor no intervalo é a mesma. Assim, diferentes níveis de informação *a priori* podem ser considerados conforme os valores estipulados para esses hiperparâmetros. Pode-se, portanto, assumir desde *priors* não-informativos até *priors* informativos que informam que o valor utilizado, *a priori*, se aproxima do verdadeiro valor do parâmetro. Quanto maior a falta de informação *a priori*, maior a variância da distribuição, ou seja, maior o grau de incerteza sobre o verdadeiro valor do parâmetro. Ao contrário, tem-se uma pequena variância da distribuição quando o *prior* é informativo.

A distribuição *a priori* de um parâmetro é uma função de probabilidade ou função densidade de probabilidade expressando o grau de confiança sobre o valor do parâmetro, antes de se observar uma amostra da variável aleatória, cuja função de distribuição depende do parâmetro (Larson, 1982).

Wang et al. (1993) avaliaram o efeito de quatro níveis de informação *a priori* em dados sob seleção e ilustraram como informações externas podem contribuir com as informações geradas pelos dados. Os autores observaram que as estimativas foram mais precisas quando maiores níveis de informação *a priori* foram utilizados, demonstrando que a análise Bayesiana pode ser empregada para incorporar informações externas em uma situação prática de melhoramento animal.

Rodrigues et al. (1996), trabalhando com dados reais de suínos Landrace, utilizaram dois níveis de informação *a priori*: *priors* não-informativos (*flat priors*), refletindo pouco conhecimento sobre os componentes de variância e *priors* informativos. Os autores concluíram que a informação externa na inferência não foi importante. Para as análises em que foi dado maior peso para as informações externas, os resultados foram basicamente iguais àqueles obtidos com *flat priors*.

Weigel & Gianola (1992), trabalhando com dados simulados, verificaram a utilização de um método Bayesiano para estimação de componentes de variância dentro de rebanho. O método combina as estimativas obtidas entre os rebanhos com as informações geradas pelos dados dentro dos rebanhos. Nesse estudo, foram simulados diferentes tamanhos

de rebanhos e diversos pesos para as estimativas obtidas entre os rebanhos. Os autores encontraram que a acurácia da estimativa da variância de touro foi maior quando ambas as informações foram utilizadas: dentro e entre rebanhos. Entretanto, o aumento verificado na acurácia da variância residual com a incorporação da informação *a priori* foi mínimo. Os autores concluíram que a utilização da informação *a priori* pode permitir a predição de valores genéticos menos influenciados pelos problemas decorrentes da heterogeneidade de variâncias.

Assis (2005) avaliou a influência da informação *a priori* em dados simulados sob seleção utilizando diferentes níveis de informação *a priori*, em dois tamanhos de população sob seleção. A autora concluiu que populações de menor tamanho são mais influenciadas pela informação *a priori*. Contudo, não encontrou diferenças para a classificação dos indivíduos.

Nota-se que a influência da informação *a priori* sobre os resultados, na área de melhoramento animal, ainda é contraditória e pouco estudada. Além disso, são escassos os trabalhos que demonstram o efeito de se assumir *priors* inadequados na avaliação genética dos animais.

Assim, os objetivos neste trabalho foram verificar a influência da informação *a priori* na avaliação genética animal, utilizando inferência Bayesiana, em populações simuladas com diferentes tamanhos, e avaliar o efeito de se assumirem valores inadequados de *priors* nos componentes de variâncias genética e residual e na predição dos valores genéticos.

Material e Métodos

Os dados utilizados neste trabalho foram simulados por meio do sistema computacional GENESYS (Euclides, 1996). O sistema foi desenvolvido em linguagem FORTRAN e permite a simulação de genomas de certa complexidade para estudos comparativos de métodos de seleção, testes de pressuposições, avaliação de novas metodologias de seleção, entre outros, dispensando, portanto, animais e laboratórios.

Foi simulado um genoma de 3.000 centimorgans de comprimento, considerando-se uma única característica quantitativa, determinada por 800 locos com dois alelos por loco.

Os locos quantitativos foram simulados segundo a distribuição normal e estavam distribuídos em 20

pares de cromossomos de tamanhos aleatórios. Os locos não possuíam desvios de dominância nem interações epistáticas entre si. O genoma não possuía cromossomo sexual e as frequências alélicas iniciais eram iguais em ambos os sexos.

A distribuição matemática empregada para a simulação das frequências iniciais dos locos quantitativos nas populações foi a uniforme, com média de 0,50. Foi utilizado um efeito sistemático de rebanho, com seis níveis, e os efeitos de ambiente não-controláveis foram simulados conforme a distribuição normal. O processo de simulação foi repetido 100 vezes.

Segundo a estrutura genômica proposta, foram simulados 1.500 machos e 1.500 fêmeas, que formaram a população-base.

As populações geradas foram submetidas a duas gerações de seleção ao acaso, sendo simulados três diferentes tamanhos de população: POP1 (100 animais), POP2 (300 animais) e POP3 (1.600 animais), nos quais a herdabilidade foi de 0,42, 0,43 e 0,54 respectivamente. Foi simulado também um efeito “fixo” de rebanho com seis subclasses.

A tendência dos parâmetros simulados foi próxima à de uma situação real para a característica peso ao sobreano, em rebanhos da raça nelore, como descrito por Balieiro (2001).

Os dados foram submetidos à análise para estimação dos parâmetros genéticos, utilizando-se a inferência Bayesiana, por meio do programa MTGSAM (*Multiple Trait Gibbs Sampling in Animal Models*), descritos por Van Tassel & Van Vleck (1995). Os parâmetros genéticos também foram estimados pelo método clássico REML, descrito por Boldman et al. (1995), por meio do programa MTDFREML (*Estimativa pelo Método da Máxima Verossimilhança Restrita adotando-se o procedimento Livre de Derivadas*).

Foi empregado o seguinte modelo animal, descrito sob a forma matricial:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{g} + \mathbf{e}$$

em que \mathbf{y} = vetor das observações da característica medida nos animais; \mathbf{X} = matriz de incidência de efeitos “fixos”; $\boldsymbol{\beta}$ = vetor de efeitos “fixos” de rebanho; \mathbf{Z} = matriz de incidência dos efeitos aleatórios; \mathbf{g} = vetor de efeitos aleatórios (valores genéticos dos animais); \mathbf{e} = vetor de erros aleatórios.

Foram utilizados três níveis de informação *a priori*: *priors* não-informativos (PNI); *priors* pouco infor-

mativos (PPI) e *priors* informativos (PI). No primeiro conjunto de *priors*, foram assumidos valores de $\nu = -2$ e $s^2 = 0$, que indicou vago conhecimento sobre o componente de variância. No segundo conjunto de *priors*, foram assumidos valores de $\nu = 5$ para a variância genética aditiva e $\nu = 15$ para a variância ambiental, sendo s^2 igual ao valor verdadeiro do componente de variância a ser estimado, indicando *priors* pouco informativos sobre o componente de variância. No terceiro conjunto de *priors*, foram assumidos valores de $\nu = 60$ para a variância genética aditiva, $\nu = 100$ para a variância ambiental e s^2 igual ao valor verdadeiro do componente de variância a ser estimado, indicando *priors* informativos. Como o hiperparâmetro ν representa o “grau de confiança”, quanto maior o seu valor, menor a variância da distribuição *a priori* do valor indicado pelo hiperparâmetro s^2 .

Com o objetivo de verificar o efeito de se assumir valores inadequados de *priors*, para ambos os componentes de variâncias genética e residual, utilizou-se um valor de *prior* (s^2) com 50% de erro de superestimação, baseando-se nos valores verdadeiros dos componentes de variâncias conhecidos no processo de simulação. Estes valores adulterados foram adicionados às análises como pouco informativos e informativos, como descrito anteriormente.

Para o vetor de efeitos “fixos”, considerou-se uma distribuição uniforme, que indicou vago conhecimento sobre esse vetor. Para os vetores de efeitos aleatórios, considerou-se distribuição normal.

Na amostragem de Gibbs, o tamanho da cadeia foi de 1.000.000 amostras para as populações POP1 e POP2 e 200.000 para a população POP3. As amostras geradas nos primeiros ciclos foram descartadas, por não pertencerem à distribuição estacionária, sendo este o período de “burn in” de 100.000 iterações para as populações POP1 e POP2 e de 10.000 iterações para a população POP3. O intervalo de retirada de amostras (*thinning interval*) foi de 50 para as populações POP1 e POP2 e de 20 para a população POP3. O critério de convergência e as análises das cadeias de Gibbs foram baseados na inspeção visual dos gráficos do valor de cada amostra e no programa GIBANAL (Van Kaam, 1998).

A acurácia das estimativas foi avaliada por meio da comparação entre os componentes de variância estimados e os verdadeiros que são conhecidos para dados simulados. A partir das porcentagens de erro entre os componentes de variâncias verdadeiros e os

estimados, foi possível verificar o efeito dos diferentes níveis de informação *a priori* em populações com diferentes tamanhos. Quanto mais os valores se aproximaram de zero, mais acurada foi a estimativa. Calculou-se a Porcentagem de Erro por meio da seguinte expressão:

$$PE = \frac{(\hat{\sigma}_i^2 - \sigma_i^2)}{\sigma_i^2} 100, \quad i = a, e$$

em que *PE* é a porcentagem de erro na estimação do componente de variância; $\hat{\sigma}_i^2$, o componente de variância estimado; σ_i^2 , o componente de variância verdadeiro; e $i = a, e$, as variâncias genéticas aditivas e ambientais, respectivamente.

Verificou-se também o efeito do nível de informação *a priori* na predição dos valores genéticos, por meio da correlação de Spearman (CS) e pelo quadrado médio do erro (QME). As correlações de Spearman entre valores genéticos preditos e verdadeiros foram obtidas pelo pacote estatístico *Statistical Analysis System - SAS* (Littell et al., 1991). Os valores genéticos preditos foram comparados com os valores genéticos verdadeiros, por meio do Quadrado Médio do Erro (QME), como descrito por Harvey (1979), citado por Schenkel (1998):

$$QME = \sum_{i=1}^n \frac{1}{n} (\hat{g}_i - g_i)^2$$

em que n é o número de animais e \hat{g}_i refere-se ao valor genético predito do $i^{\text{ésimo}}$ animal e ao valor genético verdadeiro do $i^{\text{ésimo}}$ animal. Quanto mais próximo de zero, maior a acurácia na predição do valor genético.

A vantagem da simulação neste processo de comparação de metodologias é que as variâncias e os valores genéticos verdadeiros são conhecidos.

Resultados e Discussão

Na Tabela 1 são apresentadas as estimativas da média e da moda *a posteriori* dos componentes de variâncias genéticas aditiva e ambiental, obtidos para as populações POP1 (100 indivíduos), POP2 (300 indivíduos) e POP3 (1.600 indivíduos), sob diferentes níveis de informação *a priori* e os valores verdadeiros dos componentes de variância.

Tabela 1 - Estimativas dos componentes de variâncias genéticas aditiva ($\hat{\sigma}_a^2$) e ambiental ($\hat{\sigma}_e^2$), obtidos pela metodologia Bayesiana, sob diferentes *priors*, e valores verdadeiros (VV) para populações com três diferentes tamanhos

Table 1 - Additive ($\hat{\sigma}_a^2$) and environmental ($\hat{\sigma}_e^2$) genetic variance estimates, obtained by Bayesian methodology under different *priors*, and true values (VV) for populations with three different sizes

População <i>Population</i>		$\hat{\sigma}_a^2$			W	$\hat{\sigma}_e^2$			W
		PNI	PI	I		NI	PI	I	
POP1	Média	26,21	127,92	174,46	179,01	383,20	301,68	260,75	251,02
	Moda	0,46	84,37	163,57		408,91	290,24	254,38	
POP2	Média	72,91	113,21	176,52	189,32	361,83	329,24	281,30	251,46
	Moda	37,10	87,50	168,34		358,50	340,75	274,67	
POP3	Média	284,83	279,83	294,23	303,98	288,22	289,93	277,32	257,65
	Moda	266,41	277,56	286,21		302,99	296,59	286,80	

Média (*Mean*); Moda (*Mode*).

NI= Não-informativo; PI = Pouco informativo; I = Informativo; VV = Valores verdadeiros.

NI = *No-informative priors*; (PI) = *Slightly informative priors*; (I) = *Informative priors*; VV = *True values*.

De modo geral, pode ser observado que os resultados que mais se afastaram dos valores reais foram aqueles obtidos com as populações POP1, sob *priors* não-informativos. Para o componente de variância ambiental, aparentemente os resultados foram mais consistentes que para a variância genética aditiva.

Na Tabela 2 são apresentadas as estimativas da média e da moda *a posteriori* dos componentes de variância genética aditiva e ambiental, obtidos para populações POP1, POP2 e POP3, a partir de *priors* com 50% de erro de superestimação, tanto para a variância genética aditiva quanto para a variância ambiental.

Foram obtidos diferentes valores para a média e a moda quando *priors* contendo erros foram utilizados. Maiores diferenças ocorreram para as populações POP1 e POP2 no componente de variância genético aditivo. Para a população POP1, a estimativa da média obtida para a variância genética aditiva com *prior* adequado foi de 174,46, próximo ao valor real 179,01 (Tabela 1). Entretanto, quando *priors* com erros foram considerados, esta estimativa mudou para 224,00 (Tabela 2), acompanhando o valor do *prior* com erro. Tendência semelhante foi verificada para a variância ambiental.

Constam na Figura 1 as Porcentagens de Erro das estimativas dos componentes de variâncias genéticas aditiva e ambiental para as populações com diferentes tamanhos e sob três níveis de informação. Para

comparação, foram utilizados os valores das médias *a posteriori*. De modo geral, constatou-se que a POP1 apresentou estimativas menos acuradas, observando-se que a introdução de maiores níveis de informação resultou em menores Porcentagens de Erro, tanto para a variância genética aditiva quanto para a variância ambiental, a qual foi menos influenciada pela informação *a priori*.

Os componentes de variância genética aditiva foram subestimados para todas as populações em estudo (Figura 1). Sob *priors* não-informativos, as Porcentagens de Erro para a variância genética aditiva foram de -85; -25 e -3%, para as populações POP1, POP2 e POP3, respectivamente, indicando que maiores problemas de estimação ocorrem para populações pequenas. Conforme maiores níveis de informação foram introduzidos, menores Porcentagens de Erro foram observadas, alcançando apenas -3% de erro no nível mais informativo. Entretanto, com o aumento do tamanho da população, a importância da informação *a priori* decresceu. Para a população POP3 as Porcentagens de Erro foram de -6; -8 e -3% para os níveis PNI, PPI e PI, respectivamente, indicando que conjuntos maiores de dados são menos sensíveis à informação *a priori*. Esses resultados estão de acordo com aqueles reportados por Gianola & Fernando (1986), Wang et al. (1994) e Blasco et al. (1998), em que, à medida que o tamanho da amostra aumenta, a

Tabela 2 - Estimativas dos componentes de variâncias genéticas aditiva ($\hat{\sigma}_a^2$) e ambiental ($\hat{\sigma}_e^2$), obtidos pela metodologia Bayesiana, com *priors* contendo 50% de erro considerados pouco-informativos e informativos, e valores verdadeiros (entre parênteses), para populações com três diferentes tamanhos

Table 2 - Addictive ($\hat{\sigma}_a^2$) and environmental ($\hat{\sigma}_e^2$) genetic variance estimates, obtained by Bayesian methodology with 50% error priors, considered slightly informative and informative, and true values (in parentheses) for populations with three different sizes

População		$\hat{\sigma}_a^2$		Prior	$\hat{\sigma}_e^2$		Prior
Population		PPI	PI		PPI	PI	
POP1	Média	139,20	224,00	268,52(179,01)	318,29	331,28	376,53(251,02)
	Moda	91,89	211,52		311,65	318,13	
POP2	Média	124,53	207,66	283,98(189,32)	333,27	315,80	377,19(251,46)
	Moda	111,53	203,35		326,73	318,72	
POP3	Média	293,61	318,30	455,97(303,98)	283,92	281,30	386,48(257,65)
	Moda	299,39	318,10		288,76	281,43	

Média (Mean); Moda (Mode).

NI= Não-informativo; PI = Pouco informativo; I = Informativo; VV = Valores verdadeiros.

NI = No-informative priors; (PI) = Slightly informative priors; (I) = Informative priors; VV = True values.

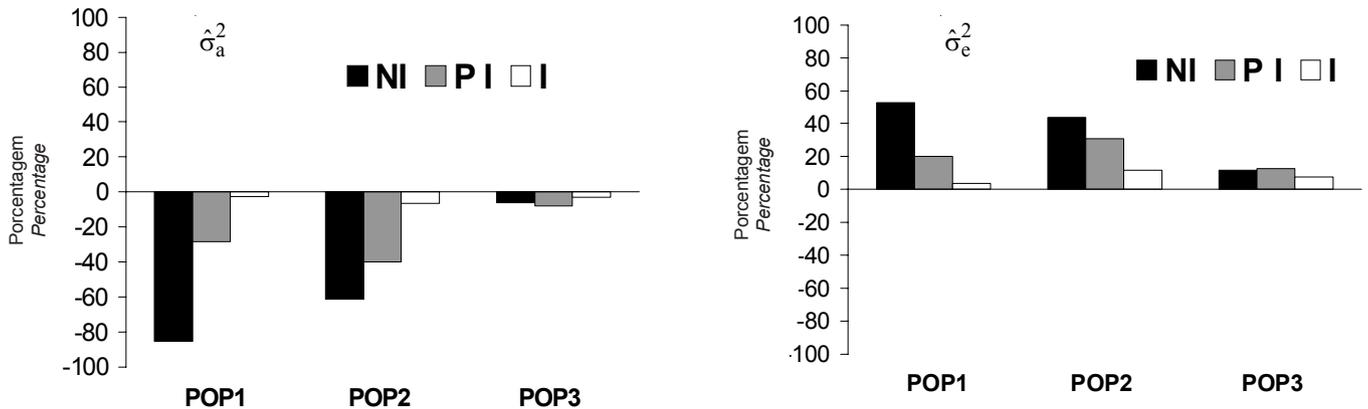


Figura 1 - Porcentagens de erros observados das estimativas dos componentes de variâncias genéticas aditiva (à esquerda) e ambiental (à direita), obtidos pela metodologia Bayesiana para POP1 (100 animais), POP2 (300 animais) e POP3 (1.600 animais), sob três níveis de informação *a priori*: *prior* não-informativo (PNI); *prior* pouco informativo (PPI) e *prior* informativo (PI).

Figure 1 - Error percentages of estimated additive genetic (on left side) and environmental (on right side) variance components, obtained by Bayesian methodology, in POP1 (100 animals), POP2 (300 animals) and POP3 (1.600 animals), under three levels of prior information: no-informative prior (PNI); slightly informative prior (PPI) and informative prior (PI).

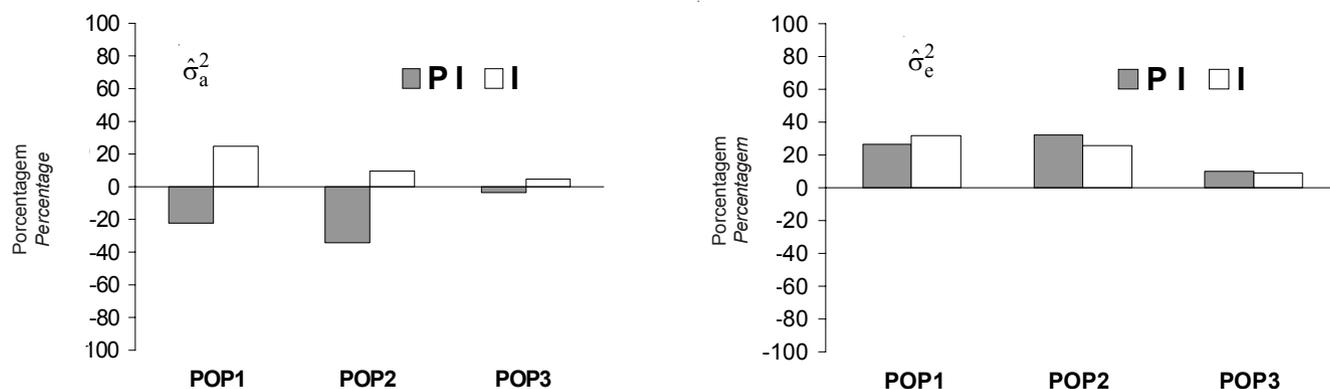


Figura 2 - Porcentagens de erros observados das estimativas dos componentes de variâncias genéticas aditiva (à esquerda) e ambiental (à direita), obtidos pela metodologia Bayesiana para POP1 (100 animais), POP2 (300 animais) e POP3 (1.600 animais), quando *priors* contendo 50% de erro foram adicionados como pouco informativos (PPI) e informativos (PI).

Figure 2 - Error percentages of estimated additive genetic (on left side) and environmental (on right side) variance components, obtained by Bayesian methodology, in POP1 (100 animals), POP2 (300 animals) and POP3 (1.600 animals), using 50% error priors with two informative levels: slightly informative (PPI) and informative (PI).

distribuição *a posteriori* se assemelha à distribuição representada pela função de verossimilhança. Portanto, o conhecimento *a priori* tende a perder a importância relativa com o aumento do volume dos dados das populações. Segundo Van Tassel et al. (1995), a utilização da informação *a priori* pode contribuir para a estimação dos componentes de variância, principalmente naquelas situações em que o conjunto de dados é pequeno e alguma informação adicional a respeito dos seus parâmetros está disponível.

Na Figura 2 são apresentadas as Porcentagens de Erro na estimação dos componentes de variância genética aditiva e ambiental quando foram utilizadas informações *a priori* pouco informativa e informativa, com 50% de erro. A população de maior tamanho (POP3) foi menos influenciada pelos erros nas informações *a priori*. Para a população pequena (POP1), quando maior confiabilidade foi atribuída ao valor *a priori*, a estimativa da variância genética aditiva apresentou erro próximo de 30% (Figura 2), confirmando as observações anteriores, em que a população de menor tamanho foi a mais influenciada pela informação *a priori*.

Como a simulação realizada foi em nível genético, os verdadeiros valores genéticos dos animais foram conhecidos, o que permitiu a comparação da classificação dos animais com base nos valores genéticos verdadeiros e os preditos. Os resultados das Correlações de Spearman (CS) e dos Quadrados Médios dos Erros (QME) são apresentados na Tabela 3.

Independentemente do nível de informação utilizado, a população POP3 apresentou os melhores resultados. A Correlação de Spearman para esta população, para os três níveis de informação *a priori*, foi de 0,75 e o valor de QME foi próximo de 125, comprovando que o aumento no nível de informação *a priori* não alterou os resultados. Por outro lado, para a população POP1, o QME foi de 168, 127 e 118 para os níveis PNI, PPI e PI, respectivamente, indicando que valores genéticos mais acurados foram obtidos quando maiores níveis de informação *a priori* foram utilizados. Porém, a mais acurada estimação dos componentes de variância e a melhor predição dos valores genéticos, decorrentes da introdução de *priors* informativos não alteraram a classificação dos animais. Para esta situação, foram encontrados valores de Correlação de Spearman próximos de 0,50, indicando que, em populações de tamanho pequeno, a escolha dos melhores indivíduos pode ser comprometida mesmo quando se utiliza *priors* informativos, considerando-se, inclusive, o valor verdadeiro do parâmetro considerado na análise.

Para a população POP2, quando maiores níveis de informação foram utilizados, verificaram-se diminuição no QME e aumento nas Correlações de Spearman. Contudo, mesmo com incremento de informação *a priori*, a Correlação de Spearman não ultrapassou 0,70. Segundo Crews & Franke (1998), correlações de ordem inferiores a 0,70 podem resultar em alterações na classificação dos animais, comprometendo a seleção e o

Tabela 3 - Correlações de Spearman (CS) e quadrados médios dos erros (QME) entre os valores genéticos verdadeiros e preditos, obtidos pela metodologia Bayesiana, em populações com diferentes tamanhos e diferentes níveis de informação *a priori*

Table 3 - Spearman correlations (SC) and mean square errors (MSE) among true and predicted breeding values, obtained by Bayesian methodology in populations with different sizes and different levels of prior information

População Population	PNI		PPI		PI	
	CS SC	QME MSE	CS SC	QME MSE	CS SC	QME MSE
POP1	0,48	167,63	0,50	126,82	0,50	118,11
POP2	0,61	137,45	0,64	122,98	0,66	110,35
POP3	0,75	125,27	0,75	125,45	0,75	124,59

NI= Não-informativo; PI = Pouco informativo; I = Informativo; VV = Valores verdadeiros.

NI = No-informative priors; (PI) = Slightly informative priors; (I) = Informative priors; VV = True values.

progresso genético. Portanto, apesar de melhores resultados serem obtidos com maiores níveis de informação, não são tão acurados quanto os obtidos com grandes conjuntos de dados. Pode-se inferir que existe perda ao se trabalhar com conjuntos de dados pequenos, mesmo quando informações adicionais estão disponíveis.

Os resultados obtidos pelo método tradicional de análise REML-BLUP, tanto para a Correlação de Spearman quanto para os valores de QME, foram semelhantes aos obtidos pelo método Bayesiano quando *priors* não-informativos foram utilizados. Entretanto, a predição dos valores genéticos foi sensivelmente mais acurada quando maiores níveis de informação *a priori* foram adicionados, apesar de não ter alterado a classificação dos indivíduos.

Na Tabela 4 são apresentadas as Correlações de Spearman e os QME, para as três populações em análise, quando *priors* contendo 50% de erro foram adicionados como pouco informativos e informativos.

Tanto a CS quanto o QME não foram influenciados pelos erros inseridos nos *priors*. Segundo Annala et al (1995), o viés causado pelo uso de estimativas de componentes de variância não influi na classificação dos animais.

Para efeito de comparação, foi empregado também o método tradicional de análise REML – BLUP, sendo que, para a população POP1, a convergência não foi atingida. Os resultados para as populações POP2 e POP3 foram semelhantes aos das estimativas das modas obtidas pelo método Bayesiano.

Conclusões

Populações de tamanho pequeno apresentam maiores problemas na estimação de componentes de variância e na predição dos valores genéticos.

Tabela 4 - Correlações de Spearman (CS) e quadrados médios dos erros (QME) entre os valores genéticos reais e preditos, obtidos pela metodologia Bayesiana para POP1 (100 animais), POP2 (300 animais) e POP3 (1.600 animais); quando *priors* contendo 50% de erro foram adicionados como pouco informativos (PPI) e informativos (PI)

Table 4 - Spearman correlations (CS) and mean square errors (QME) among real and predicted genetic values, obtained by Bayesian methodology in POP1 (100 animals), POP2 (300 animals) and POP3 (1.600 animals), using 50% error priors with two informative levels: slightly informative (PPI) and informative (PI)

População Population	PPI		PI	
	CS	QME	CS	QME
POP1	0,50	126,22	0,50	118,03
POP2	0,64	120,94	0,66	110,12
POP3	0,75	124,89	0,75	124,31

NI= Não-informativo; PI = Pouco informativo; I = Informativo; VV = Valores verdadeiros.

NI = No-informative priors; (PI) = Slightly informative priors; (I) = Informative priors; VV = True values.

A inclusão de informação *a priori* proporciona melhores resultados para estimação dos componentes de variância, sobretudo em populações pequenas. À medida que o tamanho da amostra aumenta, a importância da informação *a priori* decresce.

A predição dos valores genéticos é menos influenciada pela informação *a priori* que a estimação dos componentes de variância, de modo que a escolha dos melhores indivíduos é mais prejudicada em populações com pequeno número de animais, mesmo quando maiores níveis de informação são disponíveis.

Agradecimento

À CAPES, pela concessão da bolsa de estudos ao primeiro e segundo autores durante o programa de doutoramento.

Literatura Citada

- ANNALA, M.; SANCHES-PALMA, A.; MUÑOZ-SERRANO, A. et al. Simulation analysis with BLUP methodology of different data structures in goat selection schemes in Spain. **Small Ruminant Research**, v.17, n.1, p.51-55, 1995.
- ASSIS, G.M.L. **Efeito do número de genes na avaliação genética utilizando dados simulados**. Viçosa, MG: Universidade Federal de Viçosa, 2005. 150p. Tese (Doutorado em Genética e Melhoramento) - Universidade Federal de Viçosa, 2005.
- BALIEIRO, J.C.C. **Heterogeneidade de variância na avaliação genética de bovinos da raça Nelore**. Viçosa, MG: Universidade Federal de Viçosa, 2001. 88 p. Tese (Doutorado em Genética e Melhoramento) - Universidade Federal de Viçosa, 2001.
- BLASCO, A.; SORENSEN, D.; BIDANEL J.P. Bayesian Inference of Genetic Parameters and Selection Response for Litter Size Components in Pigs. **Genetics**, v.149, p.301-306, 1998.
- BOLDMAN, K.G.; KRIESE, L.A.; Van VLECK, L.D. et al. **A manual for use of MTDFREML: a set of programs to obtain estimates of variances and covariances (DRAFT)**. Lincoln: Department of Agriculture/Agriculture Research Service, 1995. 120p.
- CREWS JR., D.H.; FRANKE, D.E. Heterogeneity of variances for carcass traits by percentage brahman inheritance. **Journal of Animal Science**, v.76, n.7, p.1803-1809, 1998.
- EUCLYDES, R.F. **Uso do sistema para simulação Genesys na avaliação de métodos de seleção clássicos e associados a marcadores moleculares**. Viçosa, MG: Universidade Federal de Viçosa, 1996. 150p. Tese (Doutorado em Genética e Melhoramento) - Universidade Federal de Viçosa, 1996.
- GIANOLA, D.; FOLLEY, J.L. Non linear prediction of latent genetic liability with binary expression: an empirical Bayes approach. In: WORLD CONGRESS GENETIC APPLIED LIVESTOCK PRODUCTION, 5., 1982, Madrid. **Proceedings...** Madrid: 1982. v.7, p.293-303.
- GIANOLA, D.; FERNANDO, L. Bayesian methods in animal breeding theory. **Journal of Animal Science**, v.63, p.217- 244, 1986.
- KENNEDY, B.W. Variance component estimation and prediction of breeding values. **Canadian Journal of Genetics and Cytology**, v.23, n.4, p.565-578, 1981.
- LARSON, H.J. **Introduction to probability theory and statistical inference**. 3.ed. Singapore: John Wiley & Sons, 1982. 637p.
- LITTELL, R.C.; FREUND, R.J.; SPECTOR, P.C. **SAS[®] System for linear models**. 3.ed. Cary: 1991. 329p.
- RODRIGUES, M.C.; TORO, M.; SILIÓ, L. Selection on lean growth in a nucleus of Landrace pigs: an analysis using Gibbs sampling. **Journal of Animal Science**, v.63, p.243-253, 1996.
- SCHENKEL, F.S. **Studies on effects of parental selection on estimation of genetic parameters and breeding values of metric traits**. Ghelph: University of Ghelph, 1998. 191p. (Ph.d. Thesis) - University of Ghelph, 1998.
- Van KAAM, J.B.C.H.M. **"GIBANAL" – Analyzing program for Markov Chain Monte Carlo Sequences (Version 2.10)**. Wageningen: Department of Animal Sciences, Agricultural University, 1998.
- Van TASSELL, C.P.; Van VLECK, L.D. **A manual for use of MTGSAM. A set of Fortran programs to apply Gibbs sampling to animal models for variance components estimation**. Lincoln: U.S. Department of Agriculture, Agricultural Research Service, 1995. 85p. (Manual)
- WANG, C.S.; GIANOLA, D.; SORENSEN, D.A. et al. Response to selection for litter size in Danish Landrace pigs: a Bayesian analysis. **Theory Applied Genetics**, v.88, p.220-230, 1994.
- WANG, C.S.; RUTLEDGE, J.J.; GIANOLA, D. Marginal inferences about variance components in a mixed linear model using Gibbs sampling. **Genetics Selection Evolution**, v.25, p.41-62, 1993.
- WEIGEL, D.A.; GIANOLA, D. Estimation of heterogeneous within-herd variance components using empirical Bayes methods: a simulation study. **Journal of Dairy Science**, v.75, p.2824-2833, 1992.

Recebido em: 22/12/04

Aceito em: 14/06/05