

X Simpósio Brasileiro de Melhoramento Animal

Uberaba, MG – 18 a 23 de agosto de 2013

Diagnóstico de Convergência de Múltiplas Cadeias de Markov via Monte Carlo usando polinômios ortogonais de Legendre para modelar a produção de leite de cabras da raça Alpina¹

Lais Costa Brito², Hinayah Rojas de Oliveira³, Yara Laureano da Cunha³, William Heleno Mariano³, Vinicius Silva Junqueira⁴, Paulo Sávio Lopes⁵

¹ Trabalho financiado pela FAPEMIG

² Programa de Pós-Graduação em Zootecnia – UFV, Viçosa. e-mail: laiscostabrito@gmail.com

³ Graduação – UFV, Viçosa. e-mail: hinayah@hotmail.com; yara.cunha@ufv.br; williamheleno@ig.com.br

⁴ Programa de Pós-graduação em Genética e Melhoramento. e-mail: junqueiravinicius@hotmail.com

⁵ Departamento de Zootecnia – UFV, Viçosa. e-mail: plopes@ufv.br

Resumo: Uma das principais críticas para quem adota métodos MCMC é determinar quando a cadeia atingiu a condição de equilíbrio para que assim, possam ser utilizadas amostras da mesma para obter características das distribuições estacionária de interesse. Dessa forma, objetivou-se com este trabalho verificar a estacionalidade das cadeias para as (co)variâncias dos coeficientes de polinômios ortogonais de Legendre utilizando múltiplas cadeias com diferentes pontos de partida. Foram utilizados 33844 dados provenientes de cabras da raça Alpina do Setor de Caprinocultura da Universidade Federal de Viçosa (Viçosa - MG). A edição dos dados foi realizada pelo programa SAS 9.1 (2003) e a geração das cadeias foram realizadas no programa gibbs3f90 (Miszta, 2012). Em todas as análises a variância genética aditiva do coeficiente foi o componente que necessitou de maior número de iterações para atingir estacionalidade, sendo que nos componentes de ambiente permanente este foi relativamente menor quando comparado aos componentes genético aditivo. Assim, o coeficiente do componente de variância genética aditiva do modelo de 3ª ordem para ambos componentes aleatórios (genético aditivo e ambiente permanente), foi o que apresentou maior número de iterações necessárias para que a cadeia atingisse o equilíbrio. Dessa forma, caso seja opção do pesquisador escolher um *burn-in* único para todos os componentes, esta escolha deve ser realizada sobre os componentes de variância genética aditiva.

Palavras-chave: regressão aleatória, coeficiente, convergência

Convergence diagnostic of multiple Markov Monte Carlo chains using Legendre orthogonal polynomials to model test-day milk yield of Alpine goats

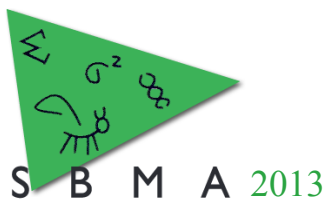
Abstract: One of the principal issues is to determine the moment when MCMC chain reach the equilibrium condition, to use their samples to obtain characteristics of the stationary distribution. The objective is to verify chain stationality of covariances coefficients of Legendre orthogonal polynomials using multiple chains with different initial values. Were used 33844 data from Alpine goat herd of Federal University of Viçosa, Viçosa-MG. Data edition and Markov chain generation were performed using SAS 9.1 (2003) and gibbs3f90 software (Miszta, 2012). All analysis demonstrates retard convergence to genetic additive variance of coefficient, and in permanent environment this coefficient had relatively more rapidly convergence. Then coefficient of genetic additive variance to the model of 3th order to both random components (additive and permanent environment) required more number of iterations to the chain reach the equilibrium. In this way, if necessary to the researcher choose a unique burn-in for all components, this choice should be made using genetic additive variance components.

Keywords: random regression, coefficient, convergence

Introdução

A amostragem de Gibbs é uma técnica iterativa de Monte Carlo utilizada para possibilitar inferências das distribuições condicionais completas *a posteriori*, considerando-as como as distribuições estacionárias de interesse. Usualmente, o descarte de algumas iterações no início da cadeia, denominadas de *burn-in*, são suficientes para assegurar que o algoritmo irá permitir excelentes inferências marginais.

Atualmente, vários métodos que objetivam estimar o período de *burn-in* estão disponíveis. O uso de múltiplas cadeias ou a obtenção de cadeias em paralelo, para obter a convergência com base na comparação dos componentes de variâncias é um destes. De forma geral, são usados desvios de números



X Simpósio Brasileiro de Melhoramento Animal

Uberaba, MG – 18 a 23 de agosto de 2013

aleatórios idênticos para múltiplas cadeias de Gibbs, de forma que os valores gerados serão os mesmos a cada iteração e as diferenças entre as cadeias paralelas serão decorrentes exclusivamente dos valores iniciais. Assim, com o aumento do número de iterações, o efeito destes valores iniciais sob a cadeia diminui e as cadeias paralelas coincidem, permitindo avaliar o quanto o algoritmo é sensível ao efeito dos valores iniciais na determinação do período de *burn-in*.

Dessa forma, objetivou-se com este trabalho verificar a estacionalidade das cadeias para os componentes de (co)variância dos coeficientes de polinômios ortogonais de Legendre utilizando análise visual de múltiplas cadeias de Gibbs.

Material e Métodos

Os dados utilizados neste trabalho são provenientes de cabras da raça Alpina do Setor de Caprinocultura da Universidade Federal de Viçosa, Viçosa - MG. As cabras são mantidas sob o sistema de estabulação livre e recebem alimentação à base de silagem de milho e feno como volumoso, além de mistura concentrada, fornecida conforme suas exigências nutricionais. Os controles leiteiros são efetuados semanalmente em ordenha mecânica, realizada duas vezes ao dia.

A edição e organização dos dados foi realizada com o auxílio do software SAS 9.1 (*Statistical Analysis System*, versão 9.1), de forma que o arquivo de pedigree continha 5124 animais e 33844 registros de primeiras lactações coletadas no período de 1997 a 2013. Foram considerados efeitos fixos de ano e estação de parto, agrupamento genético, tipo de parto (até três filhotes), as covariáveis idade da cabra ao parto (de até 912 dias), considerando o efeito linear e quadrático para esta característica e as médias dos coeficientes da curva de lactação da população, de 4ª ordem.

Adicionalmente considerou-se os efeitos aleatórios genético aditivo (A), modelada por polinômios ortogonais de Legendre de 3ª e 4ª ordem, ambiente permanente (3ª e 4ª ordem - P) e homogeneidade de variância residual. Os modelos foram denominados de AXPX, em que X é a ordem dos polinômios avaliados. Para recodificação do pedigree e geração das cadeias de Gibbs foram utilizados os programas *renumf90* e *gibbs3f90*, respectivamente, utilizando as distribuições *a priori* preconizadas pelo último. Para avaliação do método de cadeia múltiplas, foram obtidas 3 (três) cadeias com diferentes pontos de partida (1, 0.1 e 0.01 para as variâncias) para cada modelo e o período de *burn-in* foi estimado quando a diferença máxima entre os valores gerados pelas cadeias em paralelo atingiam

Resultados e Discussão

Observou-se que dentre todas as cadeias obtidas referentes às estimativas de (co)variâncias dos coeficientes dos polinômios ortogonais de Legendre, o componente de variância genético aditivo relativo ao coeficiente (coeficiente linear) do modelo A3P3, foi o que apresentou maior número de iterações necessárias (7983) para que a cadeia atingisse o equilíbrio, conforme demonstrado na figura 1.

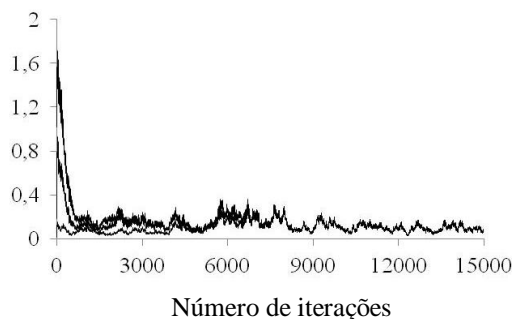
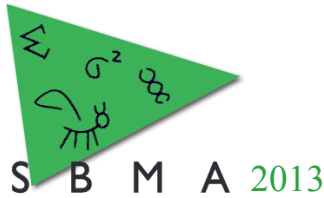


Figura 1: Gráfico de traço do componente de variância genético aditivo relativo ao coeficiente do modelo de 3ª ordem para o componente genético aditivo e ambiente permanente.

Quando as cadeias em paralelo são geradas para o componente de ambiente permanente, verifica-se maior estabilidade da mesma e menor número de iterações necessárias para que a cadeia



alcance o equilíbrio, quando comparada ao componente genético aditivo (Figura 2). De forma similar, o componente do modelo A4P3 foi o que mais demorou para atingir a estabilidade (somente a partir da iteração 2874).

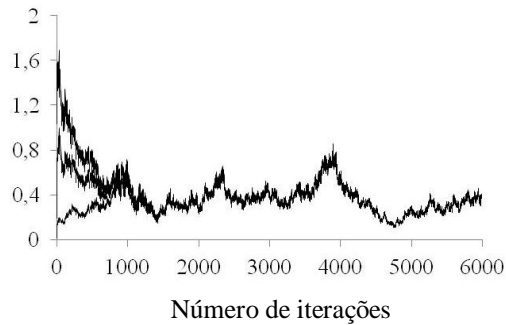


Figura 2: Gráfico de traço do componente de ambiente permanente relativo ao coeficiente do modelo de 4ª ordem para o componente genético aditivo e 3ª ordem para ambiente permanente.

Dentre todos os modelos testados, verificou-se ainda que o coeficiente do componente ambiente permanente do modelo A4P3 foi o que mais facilmente alcançou a estacionalidade, com exceção da variância residual, visto que a cadeia se apresentou bastante estável ao decorrer das iterações, necessitando de um *burn-in* relativamente baixo (21).

Conclusões

Em todas as análises verificou-se maior estabilidade da cadeia nos componentes de ambiente permanente quando comparado aos componentes genético-aditivo. Dentre todas as análises realizadas, o componente de variância genética-aditiva relativo ao coeficiente do modelo A3P3, foi o que apresentou maior número de iterações necessárias para que a cadeia atingisse o equilíbrio, enquanto o coeficiente do componente ambiente permanente do modelo A4P3 foi o que mais facilmente alcançou a estacionalidade. Dessa forma, caso seja opção do pesquisador padronizar suas análises escolhendo um *burn-in* único para todos os componentes, é importante que esta escolha seja realizada sobre os valores do componente de variância genética aditiva.

Agradecimentos

À Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG) pelo apoio e à equipe de trabalho do Grupo de Melhoramento de Caprinos da Universidade Federal de Viçosa.

Literatura citada

R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.

SAS Institute Inc. **Statistical Analysis System user's guide**. Version 9.1 ed. Cary: SAS Institute, USA, 2003.