

X Simpósio Brasileiro de Melhoramento Animal

Uberaba, MG – 18 a 23 de agosto de 2013

Critérios de convergência para algoritmos MCMC em modelos de regressão aleatória ajustados a dados de produção de leite de cabras da raça Alpina¹

Hinayah Rojas de Oliveira², Lais Costa Brito³, Geraldo Iria de Souza Júnior², Paula de Oliveira Barros², Robledo de Almeida Torres⁴, Marcelo Teixeira Rodrigues⁴

¹Trabalho financiado pela FAPEMIG

²Graduação em Zootecnia – UFV, Viçosa. E-mail: hinayah@hotmail.com

³Programa de Pós-Graduação em Zootecnia – UFV, Viçosa.

⁴Departamento de Zootecnia – UFV, Viçosa.

Resumo: Um dos complicadores na utilização da inferência bayesiana é a verificação do alcance da distribuição estacionária da cadeia de Markov, visto que algumas das metodologias desenvolvidas para este fim apresentam resultados contraditórios. Assim, objetivou-se avaliar se existem diferenças nos resultados dos métodos de Geweke (1992), Raftery-Lewis (1992b) e Heidelberg e Welch (1983) para os componentes de (co)variâncias dos coeficientes de polinômios ortogonais de Legendre. Os dados utilizados neste trabalho são de cabras da raça Alpina provenientes do Setor de Caprinocultura da Universidade Federal de Viçosa, Viçosa-MG. A edição dos dados foi realizada pelo software SAS 9.1 (2003) e a geração das cadeias foram realizadas nos programas *renumf90* e *gibbs3f90* (Misztal, 2012). Foi utilizado o pacote BOA do software R (2012) para verificação dos critérios de convergência. A variância genética-aditiva do coeficiente β_0 demorou mais tempo para atingir a estacionalidade da cadeia, visto que em todos os modelos este necessitou de um *burn-in* maior. Além disso, foi observado que após a realização do *burn-in* e do *thin* baseados no método de Raftery-Lewis, muitas das amostras não convergiram pelos demais métodos analisados, o que demonstra contradições entre os métodos. Dessa forma, é de extrema importância que o pesquisador se preocupe em utilizar estes métodos com cautela, considerando as diferenças de cada um.

Palavras-chave: distribuição estacionária, Geweke, Heidelberg e Welch, inferência bayesiana, Raftery-Lewis

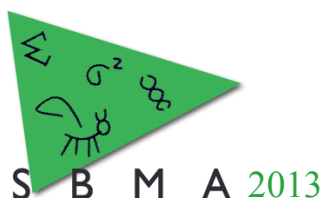
Convergence criteria for MCMC algorithms in random regression models adjusted to data on milk production in Alpine goats

Abstract: One of the complications to use bayesian inference is to verify when the stationary distribution chain is reached, consequence of contradictory results presented by methodologies developed to this purpose. The objective is to evaluate if exists differences among Geweke (1992), Raftery-Lewis (1992b) and Heidelberg e Welch (1983) methods to verify convergence of covariance coefficients of Legendre orthogonal polynomials chains. Data from Alpine goat herd of Universidade Federal de Viçosa, Viçosa - MG were used. Data edition and Markov chain generation were performed using SAS 9.1 (2003) and *gibbs3f90* software (Misztal, 2012). The convergence criteria were obtained using BOA package of software R (2012). The genetic additive variance of β_0 coefficient had bigger burn-in period and, consequently, is the component more complicated to reach the chain stationality. Besides, after realization of adequate burn-in period and thin based on Raftery-Lewis method, were observed non-convergence of many parameters by the other methods analyzed, which demonstrated incoherence between the methods. In this way, is extremely important that the researcher use these tools with prudence, considering the differences between each one.

Keywords: stationary distribution, Geweke, Heidelberg and Welch, bayesian inference, Raftery-Lewis

Introdução

A Inferência Bayesiana, técnica que vem sendo cada vez mais utilizada nos dias atuais, considera o uso de densidades de probabilidade para representar informações *a priori* sobre os parâmetros do modelo, que associada à distribuição conjunta dos dados dado os parâmetros, estima à distribuição *a posteriori*. Vale ressaltar que durante os anos de 1980 e 1990 as técnicas bayesianas foram abandonadas por requererem resolução de múltiplas integrais complicadas, mas após a descoberta da



X Simpósio Brasileiro de Melhoramento Animal

Uberaba, MG – 18 a 23 de agosto de 2013

técnica de Monte Carlo via Cadeias de Markov (MCMC), houve a possibilidade de solução para muitos problemas que não foram resolvidos no passado (Blasco, 2001). Dessa forma, a utilização de algoritmos que envolvem um processo estocástico de simulação tem a necessidade de verificar se a cadeia realmente atingiu o equilíbrio representando assim, a distribuição estacionária.

Existem várias metodologias utilizadas nos dias de hoje para avaliar se a cadeia alcançou sua distribuição estacionária, mas vale ressaltar que é de responsabilidade do analista verificar o número de amostras necessárias para fazer inferência estatística sobre os parâmetros da cadeia, sendo esta obtida pela avaliação conjunta do número de iterações da análise, da eliminação das primeiras iterações (*burn-in*) e do intervalo de amostragem (*thin*). Porém, um dos maiores problemas encontrados, é que muitos dos métodos apresentam resultados contraditórios.

Dessa forma, objetivou-se com este trabalho avaliar se há diferenças entre os critérios de Geweke (1992), Raftery-Lewis (1992b) e Heidelberg e Welch (1983) na verificação da convergência das cadeias geradas para os componentes de (co)variâncias dos coeficientes dos polinômios ortogonais de Legendre.

Material e Métodos

A base de dados é composta por informações de animais puros e mestiços da raça Alpina, provenientes do Setor de Caprinocultura da Universidade Federal de Viçosa, Viçosa - MG. As cabras são mantidas em baias coletivas sob o sistema de estabulação livre (*free stall*) e recebem alimentação à base de silagem de milho e feno como volumoso, além de mistura concentrada. Os controles leiteiros são efetuados semanalmente em ordenha mecânica, realizada duas vezes ao dia e consistem em anotar a produção de leite no dia do controle (PLDC), que representa a soma do leite coletado nos dois períodos. Foram utilizados 33844 registros de primeiras lactações coletadas no período de 1997 a 2013.

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. 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.

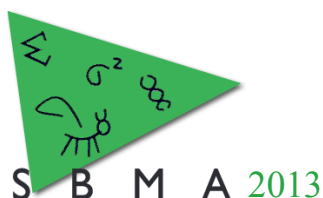
Adicionalmente considerou-se os efeitos aleatórios genético-aditivo, modelada por funções de covariância de 3ª a 4ª ordem, ambiente permanente (3ª a 4ª ordem) e homogeneidade de variância residual. Para recodificação do pedigree, geração das cadeias de Gibbs e organização das amostras dos componentes estimados foram utilizados os programas *renumf90*, *gibbs3f90* e *postgibbsf90*, respectivamente, utilizando as distribuições *a priori* preconizadas pelos referidos programas. Para verificação dos critérios de convergência foi utilizado o pacote *Bayesian Output Analysis* (BOA) do software R (2012). Após determinação do *thin* e do *burn-in* pelo método de Raftery-Lewis, foram realizados intervalos de amostragem de 100 e *burn-in* de 200, 4000 e 100000 amostras.

Resultados e Discussão

O coeficiente β_0 foi o que demorou mais tempo para atingir a estacionalidade da cadeia, visto que na maior parte dos modelos este necessitou de um *burn-in* maior, como pode ser observado na tabela 1.

Tabela 1: Modelo de média 4, variância genética aditiva e ambiente permanente de terceira ordem e heterogeneidade de variância residual (M4A3P3), analisado sem *burn-in* e sem *thin*.

M4A3P3	Genética Aditiva		Ambiente Permanente		Resíduo	
	BURN-IN	THIN	BURN-IN	THIN	BURN-IN	THIN
β_0	2670	445	537	179	4	2
$\beta_0\beta_1$	714	238	297	99	-	-
$\beta_0\beta_2$	1008	252	414	138	-	-
β_1	429	143	339	113	-	-
$\beta_1\beta_2$	435	145	348	116	-	-
β_2	1032	258	363	121	-	-



Ainda na mesma tabela, pode-se observar que o coeficiente β_1 foi aquele que convergiu mais facilmente, visto que na maior parte dos modelos necessitou de um *burn-in* e de um *thin* de menores valores. Além disso, a necessidade de *burn-in* foi quase sempre menor nas cadeias de ambiente permanente quando comparado às cadeias do efeito aleatório genético aditivo, isto porque o segundo tende a demorar mais para atingir a estacionalidade pelo fato de necessitar da matriz de parentesco dos animais. Um dos fatores que contribuem para aumentar a necessidade de *thin* é a maior autocorrelação entre as amostras dos coeficientes, que como pode ser verificado, foi superior no coeficiente β_0 .

Foi observado que após a realização do *burn-in* e do *thin* adequados ao método de Raftery-Lewis, muitas das amostras não convergiram pelos demais métodos analisados, o que demonstra contradições entre os métodos, conforme pode ser observado na tabela 2. Vale ressaltar que mesmo quando extrapolado o valor de *burn-in* exigido nas análises para 100000, não foi observado diferença entre o comportamento dos coeficientes de (co)variância dos polinômios ortogonais de Legendre.

Tabela 2: Comparação entre os resultados dos métodos de Geweke (Gew), Raftery-Lewis (R e L) e Heidelberg e Welch (H e W), para o modelo M4A3P3, analisado com *burn-in* de 4000 e com *thin* de 100.

M4A3P3	Genética Aditiva			Ambiente Permanente			Resíduo		
	R e L	Gew	H e W	R e L	Gew	H e W	R e L	Gew	H e W
β_0	Não	Sim	Sim	Sim	Sim	Sim	Sim	Não	Sim
$\beta_0\beta_1$	Sim	Sim	Não	Sim	Sim	Sim	-	-	-
$\beta_0\beta_2$	Sim	Não	Sim	Sim	Sim	Sim	-	-	-
β_1	Sim	Não	Sim	Sim	Não	Sim	-	-	-
$\beta_1\beta_2$	Sim	Não	Sim	Sim	Não	Sim	-	-	-
β_2	Não	Não	Sim	Sim	Não	Sim	-	-	-

Sim: Quando a cadeia atingiu a convergência; Não: Quando a cadeia não atingiu a convergência.

Conclusões

Dentre todos os coeficientes de (co)variância dos polinômios ortogonais de Legendre o β_0 foi aquele que apresentou maior dificuldade para atingir a estacionalidade. Além disso, os resultados apresentados demonstram algumas discordâncias entre os métodos de Geweke, Raftery-Lewis e Heidelberg e Welch. Dessa forma, é de extrema importância que o pesquisador se preocupe em utilizar estes métodos com cautela, levando em consideração a qualidade do banco de dados, as diferenças de cada método e a complexidade da análise empregada.

Literatura citada

- BLASCO, A. The Bayesian controversy in animal breeding. **Journal of Animal Science**, v. 79, p. 2023-2046, 2001.
- GEWEKE, J. Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. In: **Bayesian Statistics 4**, Oxford: Oxford University Press, 1992. p. 169-193.
- HEIDELBERGER, P. AND WELCH, P.D. Simulation Run Length Control in the Presence of an Initial Transient. **Operations Research** 31:1109-1144. 1983.
- RAFTERY, A.E. AND LEWIS, S. How Many Iterations in the Gibbs Sampler? In: **Bayesian Statistics 4**, Oxford: Oxford University Press, 1992b. p. 763-773.