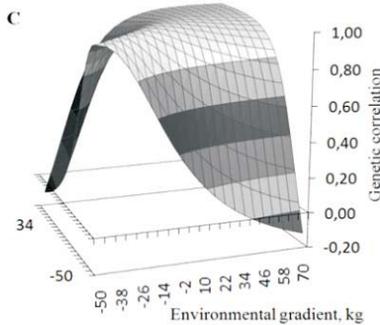


Aplicação da Inferência Bayesiana no Melhoramento Animal Usando o Programa Intergen Manual da Versão 1.2

$$P = G + E + G \times E$$



ISSN 1982-5390
Dezembro, 2010

*Empresa Brasileira de Pesquisa Agropecuária
Embrapa Pecuária Sul
Ministério da Agricultura, Pecuária e Abastecimento*

Documentos 111

Aplicação da Inferência Bayesiana no Melhoramento Animal Usando o Programa Intergen Manual da Versão 1.2

Fernando Flores Cardoso

Embrapa Pecuária Sul
Bagé, RS
2010

Exemplares desta publicação podem ser adquiridos na:

Embrapa Pecuária Sul

BR 153, km 603, Caixa Postal 242

96.401-970 - Bagé - RS

Fone/Fax: 55 53 3240-4650

<http://www.cppsul.embrapa.br>

sac@cppsul.embrapa.br

Comitê Local de Publicações

Presidente: Naylor Bastiani Perez

Secretária-Executiva: Graciela Olivella Oliveira

Membros: Daniel Portella Montardo, Eliara Quincozes, João Batista Beltrão Marques,

Magda Vieira Benavides, Naylor Bastiani Perez, Renata Wolf Suñé, Sergio Silveira Gonzaga

Supervisor editorial: Comitê Local de Publicações

Revisor de texto: Comitê Local de Publicações

Normalização bibliográfica: Graciela Olivella Oliveira

Tratamento de ilustrações: Roberto Cimirro Alves

Editoração eletrônica: Roberto Cimirro Alves

Foto(s) da capa e gráfico: Fernando Flores Cardoso

1ª edição online

Todos os direitos reservados.

A reprodução não-autorizada desta publicação, no todo ou em parte, constitui violação dos direitos autorais (Lei nº 9.610).

Dados internacionais de Catalogação na Publicação (CIP)

Embrapa Pecuária Sul

Cardoso, Fernando Flores

Aplicação da inferência bayesiana no melhoramento animal usando o programa Intergen : manual da versão 1.2 [recurso eletrônico] / Fernando Flores Cardoso.

-- Bagé: Embrapa Pecuária Sul, 2010.

(Documentos / Embrapa Pecuária Sul, ISSN 1982-5390 ; 111)

Sistema requerido: Adobe Acrobat Reader

Modo de acesso: <<http://cppsul.embrapa.br/unidade/publicacoes:list/258>>

Título da página Web (acesso em 30 dez. 2010)

1. Método estatístico. 2. Análise estatística. 3. Melhoramento genético animal.
I. Título. II. Série.

CDD 636.0821

Autores

Fernando Flores Cardoso

Médico Veterinário, Doutor (Ph.D.) em
Bioinformática - ênfase em Estatística Genômica,
Pesquisador da Embrapa Pecuária Sul,
Caixa Postal 242, BR 153 Km 603,
CEP 96401-970 - Bagé, RS – Brasil
fcardoso@cppsul.embrapa.br

Apresentação

Novas alternativas e meios de pesquisa e estudos sobre os sistemas produtivos, principalmente os de bovinos de corte são constantemente implementados e incorporados aos métodos e técnicas já existentes sobre o tema.

No contexto, a investigação científica é essencial para que possamos ter as bases para formular e orientar a melhor forma de condução de um dado sistema de criação de bovinos de corte, cujos indicadores de produção precisam ser continuamente acompanhados para que o sistema possa produzir seus melhores resultados, principalmente pelo ponto de vista do melhoramento genético animal.

Neste sentido, este trabalho visa apresentar a Aplicação da Inferência Bayesiana no Melhoramento Animal usando o Programa Intergen – Manual da Versão 1.2.. Este programa objetiva contribuir com o melhoramento genético animal nos rebanhos da Campanha do Rio Grande do Sul e do Brasil.

Assim, a Embrapa Pecuária Sul cumpre seu papel e contribui com a geração de conhecimentos capazes de gerar o desenvolvimento sustentável regional e nacional almejados.

Roberto Silveira Collares
Chefe-Geral

Sumário

Introdução.....	06
O que há de novo na Versão 1.2.....	07
Arquivo de parâmetros.....	07
Como rodar o programa	20
Como continuar uma cadeia já iniciada	20
Arquivos de saída do programa	21
Referências	24
Anexos	25

Aplicação da inferência bayesiana no melhoramento animal usando o Programa Intergen

Manual da Versão 1.2.

Fernando Flores Cardoso

Introdução

Este documento tem por objetivo descrever a funcionalidade e utilização da Versão 1.2 do programa Intergen, originalmente descrito em Cardoso (2008). O programa foi desenvolvido na Embrapa Pecuária Sul em linguagem Fortran 90/95 com capacidade de implementar modelos hierárquicos de Bayes e estimar seus parâmetros por meio de métodos Monte Carlo via Cadeias de Markov (SORENSEN; GIANOLA, 2002).

O Intergen está disponível na página eletrônica da Embrapa Pecuária Sul (EMBRAPA PECUÁRIA SUL, 2010). Os modelos possíveis de serem ajustados no Intergen contemplam situações complexas abordadas em estudos de genética quantitativa, utilizando dados de desempenho de animais pecuários, incluindo, entre outros:

- Modelo de incerteza de paternidade, na presença de acasalamentos com reprodutores múltiplos;
- Modelo multirracial, para populações cruzadas com múltiplas composições raciais;
- Modelo de normas de reação, estudos de interação genótipo-ambiente;
- Modelo robusto, para conjuntos de registros com dados extremos;
- Modelo estrutural de variâncias, para ajustar heterogeneidade de variância residual.

O software é baseado na família de programas BLUPF90 desenvolvidas por Misztal et al. (2002) e utiliza bibliotecas para geração de números aleatórios e para operações com matrizes esparsas (MISZTAL, 2009) para implementar as metodologias dos modelos mistos (HENDERSON, 1984) e MCMC.

O programa INTERGEN, apesar de permitir modelagem complexa, possui uma interface bastante simples, sendo acionado através de uma janela do Command Prompt do DOS ou terminal Linux e é controlado por um arquivo de parâmetros ("parameterfile"), o qual contém as informações sobre os arquivos de dados e de pedigree, sobre os efeitos no modelo e sobre a cadeia MCMC a ser implementada.

O que há de novo na Versão 1.2.

Um novo algoritmo de Metropolis-Hastings foi desenvolvido e implementado para permitir ajustar modelos de normas de reação com heterogeneidade de variância residual. Isto permite estimar conjuntamente a variação residual por meio de uma função exponencial do gradiente ambiental e os valores da covariável ambiental desconhecida que é utilizada para caracterizar esse gradiente de ambiente.

Arquivo de parâmetros

Uma vez acionado, o programa INTERGEN solicita ao usuário a especificação de um arquivo de parâmetros, que deve ser preparado previamente em texto não formatado tipo ASCII.

O arquivo de parâmetros apresenta um formato geral, descrito com detalhes abaixo. As seções são definidas por cabeçalhos em letras maiúsculas em negrito, que devem aparecer exatamente como descritas e numa única linha do arquivo ASCII – embora aqui possam aparecer em duas linhas por limitação de tamanho de linha no presente documento. Na linha subsequente, os campos que são definidos pelo usuário são

listados e campos opcionais aparecem entre parênteses.

Comentários podem ser incluídos no arquivo de parâmetros, iniciando a linha de comentário com #.

Título de descrição da análise em formato livre

MCMC_CHAIN: TOTAL_CYCLES BURN_IN THINNING_INTERVAL

número total de ciclos MCMC, período de aquecimento, intervalo para salvar amostras

Exemplo

400000 100000 100

SEED

número inteiro

Detalhes

Semente para iniciar o processo de geração de variáveis aleatórias, o que permite repetir exatamente a mesma cadeia.

Exemplo

123

RESTART: Y/N? [CYCLE_TO_RESTART]

y/n, (ciclo para continuar a cadeia)

Detalhes

Informa se análise é continuidade de cadeia já em andamento?

y = continuar cadeia anterior (deve informar o ciclo para reiniciar), n = iniciar nova.

Exemplo

y 10001

DATAFILE NAME N_RECORDS

nome do arquivo de dados, número de linhas de dados

Detalhes

Os arquivos de dados são arquivos texto não formatado (ASCII), com dados para todos os efeitos no modelo em colunas, com pelo menos um espaço em branco entre colunas, e uma linha para cada registro.

No caso de modelo animal, deve conter a identificação (ID) do animal ao qual o registro pertence, de acordo com a codificação no arquivo de pedigree.

Todos os efeitos classificatórios devem ser recodificados de 1 ao número de classes. Não existe uma ordem necessária dos efeitos no arquivo, entretanto, por questão de organização, recomendamos o seguinte padrão: primeira coluna com ID do animal, a seguir ID da mãe se houver efeito materno, depois listar os demais efeitos classificatórios, seguidos das covariáveis e finalmente das variáveis respostas. Dados perdidos devem ser codificados com 0.

Exemplo

manimal.dat 22124

NUMBER_OF_TRAITS

número de características respostas na análise

Exemplo

2

NUMBER_OF_EFFECTS

número de efeitos no modelo

Exemplo

5

OBSERVATION(S)

coluna(s) no arquivo de dados correspondentes às variáveis resposta

Detalhes

Deve informar tantas colunas quantas variáveis respostas houver, uma coluna por variável.

Exemplo

- 12 13 → duas características, com valores nas colunas 12 e 13 do arquivo de dados.
 10 → uma característica, com valores na coluna 10 do arquivo de dados.

WEIGHT(S)

(coluna(s) onde estão as ponderações para as variáveis respostas)

Detalhes

Uma coluna por variável resposta; deixe em branco para análise sem ponderação.

Exemplo

- 11 → ponderações na coluna 11 do arquivo de dados.
 → sem ponderação.

EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS TYPE_OF_EFFECT SAVE_SAMPLES? [EFFECT NESTED]

Posição(ões) no arquivo de dados - uma coluna por variável resposta, número de níveis, tipo de efeito, salvar amostras? y/n, (posição(ões) na qual o efeito está aninhado - uma coluna por variável resposta)

Detalhes

Seção dos efeitos no modelo com uma linha para cada efeito.

Os tipos de efeito possíveis no Intergeren são:

- **cross** = classificatório,
- **cov** = covariável,
- **unknowncov** = efeito classificatório a ser usado como covariável para normas de reação via regressão aleatória,
- **rnorm** = normas de reação lineares via regressão aleatória,
- **ram** = efeito de modelo animal reduzido com reprodutores múltiplos.

A opção salvar amostras? y = sim deve ser usada com cautela, especialmente para efeitos com muitos níveis, para evitar a geração de arquivos muito grandes e que o algoritmo gaste muito tempo escrevendo em disco rígido.

A opção [EFFECT NESTED] é utilizada para especificar efeitos de regressão aleatória e de normas de reação com covariáveis desconhecidas, onde,

- *posição no arquivo de dados* = a posição da covariável ou polinômio ortogonal (uma coluna por variável resposta);
- *número de níveis* = número de níveis do efeito aninhado (p.ex. efeito de animal),
- *tipo de efeito* = **cov** para modelo de regressão aleatória e **rnorm** para modelo de normas de reação com covariáveis desconhecidas;
- *salvar amostras?* (**y** = *sim*, **n** = *não*);
- *posição na qual o efeito está aninhado* = por. ex. a posição do efeito animal no arquivo de dados (uma coluna por variável resposta).

Exemplo

- 4 1000 cov n 1 → os valores do polinômio na coluna 4 do arquivo de dados, 1000 níveis para o efeito aninhado, tipo *cov*, não salvando amostras (que é a opção recomendada neste caso) e efeito aninhado (p.ex., animal) na coluna 1.
- 1 1520 cross n
- 2 1520 cross n
- 3 1 cov n
- 4 50 cross n → Modelo univariado com quatro efeitos, três classificatórios e uma covariável.

Modelos diferentes para cada característica podem ser especificados. O valor zero (0) para a posição no arquivo de dados de um efeito é usado para codificar a ausência de tal efeito para uma característica em particular.

Exemplo (MISZTAL, 2007)

- 1 0 1000 cross n
- 0 7 2000 cross n → Variável 1 tem grupo de contemporâneos (GC) na posição 1 com 1000 níveis e variável 2 tem GC na posição 7 com 2000 níveis.
- 1 7 2000 cross n → Mesmo exemplo acima, em uma linha, onde o número de níveis é máximo entre os dois.

No caso de modelo animal reduzido (tipo de efeito = **ram**) esse(s) efeito(s) deve(m) ser o(s) primeiro(s) a aparecer(em) na lista da seção EFFECTS.

RANDOM_RESIDUAL: TYPE PRIOR_DEGREES_OF_BELIEF

tipo de distribuição residual, grau de confiança nos valores a priori

Detalhes

Os tipos de pressuposições possíveis são seis sucintamente, temos:

- **homogeneous** = Normal homoscedástico: Distribuição normal (gaussiana) com v variância homogênea $e \sim N(0, \sigma^2_e)$ – que é a pressuposição usual;

IMPORTANTE: Esse é a única opção para análises multivariadas. As seguintes são válidas somente para análises univariadas.

- **structural** = Normal heteroscedástico: Distribuição normal com variância heterogênea $e_{(i)} \sim N(0, \sigma^2_{e(i)})$, onde $\sigma^2_{e(i)}$ é variância específica dos erros $e_{(i)}$ de uma combinação de subclasses i definida por um modelo estrutural;
- **student_t** = Student t homoscedástico: Distribuição Student t com variância homogênea $e_{(i)} \sim N(0, \sigma^2_e/w_{(i)}) \rightarrow y_{(i)} \sim t_{(v)}(0, V^2_e)$ - modelo robusto;
- **struct_student_t** = Student t heteroscedástico: Distribuição Student t com variância heterogênea $e_{(i)} \sim N(0, \sigma^2_e/w_{(i)}) \rightarrow y_{(i)} \sim t_{(v)}(0, V^2_{e(i)})$ - modelo estrutural robusto;
- **slash** = Slash homoscedástico: Distribuição Slash com variância homogênea $e_{(i)} \sim N(0, \sigma^2_e/w_{(i)}) \rightarrow y_{(i)} \sim \text{Slash}_{(v)}(0, V^2_e)$ - modelo robusto;
- **struct_slash** = Slash heteroscedástico: Distribuição Slash com variância heterogênea $e_{(i)} \sim N(0, \sigma^2_{e(i)}/w_{(i)}) \rightarrow y_{(i)} \sim \text{Slash}_{(v)}(0, V^2_{e(i)})$ - modelo estrutural robusto.

Para os graus de confiança nos valores a priori da(s) (co)variância(s) dos efeitos aleatórios, quanto maior o valor especificado, maior será a influência dos valores a priori nos resultados da análise. Especificando-se o número de variáveis na análise é equivalente a uma priori difusa pouco informativa e, portanto, a inferência é basicamente a partir da informação dos dados.

Exemplo

homogeneous 1 → resíduos normais homocedásticos com 1 grau de confiança no valor a priori da variância residual.

**METROPOLIS_STEP_OF_STRUCTURAL_EFFECTS:
ROUNDS_WITHIN_CYCLE_TUNING_SKIP**

(opcional)

número de passos de Metropolis, intervalo de sintonia

Detalhes

Esta seção opcional é necessária somente se houver efeito estrutural nos resíduos, tipos *structural*, *struct_student_t* ou *struct_slash*.

Nela é especificado o número de passos do algoritmo de Metropolis-Hastings (MH) dentro de cada ciclo da Cadeia e intervalo de ciclos para ajustar a variância da distribuição de propostas no algoritmo de MH durante o período de aquecimento (opções avançadas do modelo estrutural).

Exemplo

5 10

NUMBER_OF_STRUCTURAL_EFFECTS (opcional)

número de efeitos para a variância residual

Detalhes

Esta seção opcional é necessária somente se houver efeito estrutural nos resíduos, tipos *structural*, *struct_student_t* ou *struct_slash*.

Nela é definido quantos efeitos do modelo também afetam a variância residual.

Exemplo

4

**STRUCTURAL_EFFECTS: LINE_FROM_EFFECTS_SECTION
SAVE_SAMPLES? (opcional)**

linha da seção EFFECTS, salvar amostras? (y/n)

Detalhes

Esta seção opcional é necessária somente se houver efeito estrutural nos resíduos, tipos *structural*, *struct_student_t* ou *struct_slash*, devendo ter uma linha para cada efeito. A linha da seção EFFECTS, se refere ao número de ordem da linha dessa seção de efeitos no modelo linear onde se especificou o efeito que vai afetar também a dispersão dos resíduos. Pode-se salvar amostras desses fatores escalares usando *y* (= sim).

Exemplo

4 y → usar efeito da quarta linha da seção EFFECTS: e salvar amostras.
 6 y
 7 y
 9 n → efeito da nona linha da seção EFFECTS: e não salvar amostras

RESIDUAL_PRIOR_(CO)VARIANCES

valores a priori para a matriz de covariância residual

Exemplo

200 10
 10 100 → Para um modelo bivariado variâncias a priori na diagonal e covariâncias fora dela.
 200 0
 0 100 → Covariância é assumida nula e mantida em zero (0) durante a análise.

RANDOM_GROUP (opcional)

linha(s) do(s) efeito(s) aleatórios na seção EFFECTS:

Detalhes

Esta seção opcional é necessária quando existem efeitos aleatórios no modelo. Pode especificar mais de uma linha, em caso de efeitos aleatórios correlacionados, por ex., direto e materno. Efeitos correlacionados devem necessariamente ser consecutivos na seção *EFFECTS*.

Outras seções descritas abaixo são necessárias para complementar a descrição dos efeitos aleatórios. A seção *RANDOM_GROUP* e as subsequentes são repetidas para cada efeito ou grupo de efeitos aleatórios.

Exemplo

1 2 → efeitos na primeira e segunda linha da seção EFFECTS são aleatórios e correlacionados.

RANDOM_TYPE PRIOR_DEGREES_OF_BELIEF (opcional)

tipo aleatório, grau de confiança nos valores a priori

Detalhes

Esta seção opcional necessariamente acompanha a seção *RANDOM_GROUP*.

O tipo de efeito aleatório pode ser uma das seguintes possibilidades:

- **diagonal** = efeito não correlacionado,
- **add_sire** = efeito aditivo de touro,
- **add_animal** = efeito aditivo de animal,
- **add_an_ms** = efeito aditivo animal com reprodutores múltiplos e grupos genéticos,
- **add_an_mb** = efeito aditivo animal multirracial com grupos genéticos,
- **diag_mb** = efeito não correlacionado multirracial.

Exemplo

add_animal 2 → efeito de modelo animal com 2 graus de confiança no valor a priori.

PEDIGREEFILE: NAME N_ANIMAL N_GENETIC_GROUPS [N_BREEDS] (opcional)

(nome do arquivo de pedigree, número de animais), (número de grupos genéticos), (número de raças)

Detalhes

Esta seção opcional necessariamente acompanha a seção *RANDOM_GROUP*.

Os diferentes tipos de efeitos aleatórios requerem diferentes arquivos de pedigree.

De forma geral, o arquivo de pedigree deve ter uma linha por animal com pelo menos três colunas.

Todos os animais devem constar no arquivo de pedigree identificados de 1 ao número total de animais no arquivo, incluindo os animais base sem pais conhecidos.

Não há a necessidade de os pais terem número menor que os filhos, exceto para o

modelo animal reduzido (tipo de efeito = **ram** e tipo aleatório = **add_an_ms**). Quando pai, mãe ou ambos forem desconhecidos devem ser identificados com 0. Outras colunas serão necessárias dependendo do tipo de efeito aleatório associado com o pedigree em questão, conforme descrito a seguir:

• **diagonal ou diag_mb:**

Não requer arquivo de pedigree, pois os níveis não são correlacionados dentro de variável. Ex: grupos contemporâneos aleatórios e efeito de ambiente permanente da vaca.

• **add_sire:**

número do animal, número do pai, número do avô materno

• **add_animal:**

número do animal, número do pai, número da mãe

• **add_an_ms:**

número do animal, número da mãe, número do pai (ou -1 se for animal com incerteza de paternidade - reprodutores múltiplos), 1/d, indicador se o animal tem progênie no arquivo (1=sim, 0=não).

Aqui, d = proporção da variância genética que é atribuída à segregação mendeliana para o animal em questão.

• **add_an_mb:**

número do animal, número do pai, número da mãe, proporção da raça 1,..., proporção da raça n.

Grupos genéticos (fixos) estimados conjuntamente com os efeitos aditivos dos animais nos tipos **add_an_ms** e **add_an_mb** são especificados atribuindo-se a eles números maiores que o número de animais no arquivo de pedigree. Por exemplo, se há 100 animais no total e desejamos criar dois grupos genéticos, os animais são associados a eles através de pais fictícios com números 101 e 102, dependendo do grupo ao qual pertencem.

Exemplo

mbreed1.ped 100 2 2 → no arquivo *mbreed1.ped* existem 100 animais com dois grupos genéticos de uma população formada por duas raças.

Uma possível linha desse arquivo seria:

23 101 101 0.625 0.375 → para um animal número 23, com pais desconhecidos,

mas associado ao grupo genético 101, é 5/8 da raça 1
e 3/8 da raça 2.

**MULTIPLE_SIRES: MAX_N_FOR_MCMC [FILE: NAME & DIMENSION]
[DIRICHLET_PRIORS] (opcional)**

número máximo de reprodutores múltiplos (RM) para inferência MCMC, (nome do
arquivo de RM, dimensão do arquivo), (hiperparâmetros da distribuição a priori
Dirichlet)

Detalhes

Seção opcional, acompanha a seção *RANDOM_GROUP* somente se o tipo de efeito
aleatório = **add_an_ms** ou **add_an_mb**, que define inferência na incerteza de
paternidade quando há RM.

O *número máximo de possíveis pais para fazer inferência* define o tipo de inferência
com respeito às probabilidades de paternidade. Deve-se usar 0 para análise sem RM
com os tipos aleatórios **add_an_ms** e **add_an_mb**, e usar 1 para utilizar a matriz de
parentesco médio, que também será usada quando o tamanho do grupo de RM
exceder o número máximo de possíveis pais para fazer inferência MCMC.

Se a análise incluir animais com RM deve-se informar *o nome e o número de linhas
do arquivo de RM*. Nesse arquivo são indicados quais os possíveis pais para cada
indivíduo associado a um grupo RM por -1 na coluna do número do pai no arquivo
de pedigree. O arquivo de RM tem a seguinte estrutura para cada animal:

número de possíveis pais

identificação do pai 1

probabilidade a priori do pai 1

...

identificação do pai n

probabilidade a priori do pai n

Por exemplo, para um animal com 3 pais possíveis (4, 5 ou 6) com iguais
probabilidades de paternidade a priori, temos,

3

4

0.3333333333

5

0.3333333333

6

0.3333333333

O programa INTERGEN cada vez que encontra um animal com incerteza de paternidade no arquivo de pedigree, busca no arquivo de RM quais os possíveis pais. Portanto, nesse arquivo de RM os possíveis pais de cada animal devem aparecer na mesma ordem que os animais com paternidade incerta são relacionados no arquivo de pedigree e a estrutura descrita acima deve ser repetida para cada um desses animais. Também a identificação dos pais deve ser consistente com a do arquivo de pedigree.

Finalmente, nesta seção o *valor para hiperparâmetros da distribuição a priori Dirichlet das probabilidades de paternidade* define as priores para as probabilidades de paternidade. Se for igual a 0 não usa a distribuição Dirichlet, se o valor for maior que 0, usa esse valor para todos os parâmetros alfa da distribuição Dirichlet em todos os grupos de RM, e se for menor que 0, usa valores definidos pelo usuário no arquivo de RM para especificar os parâmetros alfas.

Exemplo

10 animodel.ms 252 -1 → Calcula probabilidades posteriores de paternidade para grupos de até 10 RM e para grupos com 11 ou mais pais candidatos usa o parentesco médio. O arquivo de RM *animodel.ms* tem 252 linhas e os valores de alfas da priori Dirichlet serão as probabilidades especificadas nesse arquivo.

METROPOLIS_STEP_OF_MULTIBREED_(CO)VARIANCES: ROUNDS_WITHIN_CYCLE TUNING_SKIP (opcional)

número de passos de Metropolis, intervalo de sintonia

Detalhes

Seção opcional, acompanha a seção *RANDOM_GROUP* somente se o tipo de efeito aleatório for multirracial, isto é, *RANDOM_TYPE* = **add_an_mb** ou **diag_mb**. Define o número de passos do algoritmo de Metropolis-Hastings (MH) dentro de cada ciclo da Cadeia e intervalo de ciclos para ajustar a variância da distribuição de propostas no algoritmo de MH durante o período de aquecimento (opções avançadas do modelo multirracial).

Exemplo

2 20

→ Faz dois ciclos de MH dentro de cada ciclo da cadeia MCMC e ajusta a variância da distribuição de propostas baseada na taxa de aceitação média a cada 20 ciclos durante a primeira metade do período de aquecimento.

(CO)VARIANCES (opcional)

valores a priori para a matriz de covariância dos efeitos aleatórios

Detalhes

Seção opcional que necessariamente acompanha a seção *RANDOM_GROUP*. A estrutura da matriz de (co)variância (*G*) permite múltiplos efeitos aleatórios e múltiplas características. Considerando um exemplo com dois efeitos (*a* e *m*) e duas características (1 e 2), temos a seguinte matriz simétrica, onde tem-se blocos diagonais de cada efeito com variâncias e covariâncias entres as diferentes características e blocos fora da diagonal com as covariâncias entre os diferentes efeitos e/ou características:

$$G = \begin{bmatrix} \sigma_{a_1}^2 & \sigma_{a_1a_2} & \sigma_{a_1m_1} & \sigma_{a_1m_2} \\ \sigma_{a_1a_2} & \sigma_{a_2}^2 & \sigma_{a_2m_1} & \sigma_{a_2m_2} \\ \sigma_{a_1m_1} & \sigma_{a_2m_1} & \sigma_{m_1}^2 & \sigma_{m_1m_2} \\ \sigma_{a_1m_2} & \sigma_{a_2m_2} & \sigma_{m_1m_2} & \sigma_{m_2}^2 \end{bmatrix}$$

Embora essas matrizes sejam simétricas, todos seus os elementos devem ser especificados na seção *(CO)VARIANCES*.

Exemplo

200 10 10 0
 10 100 0 15
 10 0 50 2
 0 15 2 300

→ Note que $\sigma_{a_1m_1} = \sigma_{a_2m_1} = 0$, isto é, a covariância é assumida nula entre os efeitos *a* e *m* nas diferentes características e mantida em zero (0) durante a análise.

Como rodar o programa

Para rodar o programa, primeiro deve-se colocar os arquivos de parâmetros, de dados e de pedigree (e de reprodutores múltiplos, se for o caso) e o executável *intergen1.2.exe* no mesmo diretório. Depois abrir uma janela de command prompt do DOS (ou terminal do Linux), ir até o diretório onde foram colocados os arquivos e digitar: *intergen1.2* (ou *./intergen1.2*) <enter>, o programa então irá solicitar o nome do arquivo de parâmetros, através da seguinte pergunta:

name of parameter file?

deve-se então digitar o *nome do arquivo de parâmetros* <enter>.

Essa resposta pode ser incluída em um arquivo texto (digamos *run*), p.ex., o arquivo *run* (ASCII) tem o seguinte conteúdo:

```
parameterfile
```

Assim podemos chamar o programa da seguinte forma:

```
C:\...caminho...\intergen1.2 < run > out <enter>
```

Neste caso, o INTERGEN obtém o nome do arquivo de parâmetros do arquivo *run* e toda a saída de tela é colocada no arquivo *out*.

IMPORTANTE: Muitos “crashes” do programa podem estar associados a erros no número de níveis dos efeitos do modelo e na especificação das posições dos efeitos. Muitas vezes esses erros vêm associados com a palavra “hash” ou “ACCESS VIOLATION”.

Como continuar uma cadeia já iniciada

Em inferência MCMC, é comum necessitar alongar uma cadeia já finalizada para aumentar o conteúdo de informação ou por ter uma cadeia interrompida por falta de energia ou pela necessidade de desligar o computador. Nestes casos, o INTERGEN através da seção RESTART:

Y/N? [CYCLE_TO_RESTART] oferece a oportunidade de continuar a cadeia por mais ciclos, sem ter que passar por novo período de aquecimento ou perder os ciclos de uma cadeia interrompida.

Para tal, o programa, cada vez que grava os arquivos de saída, grava também a última amostra de todos os parâmetros nos arquivos *solutions*, *varcomp*, *structural_r* e *robustness_w*.

Para continuar/reiniciar a cadeia, o usuário deve observar na primeira linha de um dos arquivos acima qual foi o último ciclo gravado, por exemplo, no *varcomp*:

Variance components after: 44000 rounds and burn-in of 10000.

Neste caso o último ciclo foi o 44.000 e, portanto, a cadeia deve ser reiniciada no 44.001, alterando a seção correspondente no arquivo de parâmetros da seguinte forma:

RESTART: Y/N? [CYCLE_TO_RESTART]

y 44001

IMPORTANTE: Existe um passo não automático muito importante de ser completado pelo usuário. Nos arquivos *solutionsam*, *varcompsam*, e *loglike_rnd* as amostras são salvas com maior frequência que nos outros acima e, portanto, o usuário deverá abrir manualmente com um editor de textos esses arquivos, ir até o final do arquivo e apagar todas as amostras salvas após o último ciclo gravado nos arquivos *solutions*, *varcomp*, *structural_r* e *robustness_w* (44.000 no exemplo acima), salvando os arquivos com os nomes originais sem colocar nenhuma extensão.

Arquivos de saída do programa

Os seguintes arquivos de saída são gerados pelo programa INTERGEN:

Solutions: Média e desvio padrão a posteriori dos efeitos no modelo após

o período de aquecimento e última amostra para continuar cadeia.

Solutionsam: Amostras dos efeitos no modelo na seção EFFECTS onde opção SAVE_SAMPLES = y.

Varcomp: Média e desvio padrão a posteriori dos componentes de variância após o período de aquecimento e última amostra para recomeçar cadeia. Os efeitos dos RANDOM_GROUP aparecem primeiro e na mesma ordem que são especificados no arquivo *parameterfile* e por últimos aparecem os componentes de (co)variância residual.

Varcompsam: Amostras dos componentes de variância e, se for o caso, efeitos do modelo estrutural heteroscedástico. Os componentes de variância aparecem na mesma ordem do **varcomp** e dessas matrizes, por serem simétricas, são listados somente elementos da diagonal inferior, da seguinte forma: v11, v21, v22, v13, v23, v33... Os fatores multiplicativos dos efeitos estruturais para variância residual aparecem após a variância residual e na mesma ordem que são especificados no *parameterfile*.

structural_r: Média e desvio padrão a posteriori dos efeitos estruturais na variância residual após o período de aquecimento e última amostra para recomeçar cadeia, listados na mesma ordem que são especificados no *parameterfile*.

robustness_w: Média e desvio padrão a posteriori das variáveis de ponderação do modelo robusto após o período de aquecimento e última amostra para recomeçar cadeia, para cada registro, na mesma ordem do arquivo de dados.

Loglike_rnd: O logaritmo da distribuição condicional de geração dos dados (primeiro estágio do modelo hierárquico) para critérios de escolha de modelo e medidas de ajuste.

Loglike_obs: Informações para critérios de escolha de modelo e medidas de ajuste.

Mh_dbeliefchg: Sintonia do fator escalar multiplicativo da variância das distribuições proposta do algoritmo de MH (se for o caso).

Mh_acceptance: Taxa de aceitação de valores propostos no algoritmo de MH (se for o caso).

Msiresppr: Média e desvio padrão a posteriori das probabilidades de paternidade após o período de aquecimento e última amostra para recomeçar cadeia (se for o caso). Este arquivo ainda salva a proporção de vezes que cada pai candidato foi escolhido com pai verdadeiro no ciclo MCMC.

Msirespprsam: Amostras das probabilidades de paternidade para os possíveis pais do primeiro animal que aparece com incerteza de paternidade no arquivo de pedigree.

Referências

CARDOSO, F. F. **Manual de utilização do programa INTERGEN – versão 1.0 em estudos de genética quantitativa animal**. Bagé: Embrapa Pecuária Sul, 2008. 74 p. (Embrapa Pecuária Sul. Documentos, 74).

EMBRAPA PECUÁRIA SUL. **Software INTERGEN 1.0**. Disponível em: <<http://www.cppsul.embrapa.br/unidade/servicos/intergen>>. Acesso em: 30 dez. 2010.

HENDERSON, C. R. **Applications of linear models in animal breeding**. Guelph: University of Guelph, 1984. 462 p.

MISZTAL, I. **BLUPF90 family of programs**. Disponível em: <<http://nce.ads.uga.edu/~ignacy/newprograms.html>>. Acesso em: mar. 2007.

MISZTAL, I. **Ignacy Misztal**. Disponível em: <<http://nce.ads.uga.edu/~ignacy/>>. Acesso em: 30 dez. 2010.

MISZTAL, I.; TSURUTA, S.; STRABEL, T.; AUVRAY, B.; DRUET, T.; LEE, D. H. BLUPF90 and related programs (BGF90). In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 7, 2002, Montpellier. **Proceedings...** Montpellier: INRA, 2002. 1 CD-ROM. Communication no 28-07.

SORENSEN, D. A.; GIANOLA, D. **Likelihood, Bayesian and MCMC methods in quantitative genetics**. New York: Springer-Verlag, 2002. 740 p.

Anexos

A seguir é apresentada uma série de arquivos de parâmetros para rodar determinados modelos, freqüentemente utilizados em análises de genética quantitativa baseada em métodos bayesianos. Os exemplos incluindo arquivos de dados, pedigree e parâmetros, estão disponíveis na página da Embrapa Pecuária Sul (EMBRAPA PECUÁRIA SUL, 2010).

A.1. Modelo animal com prioris próprias

#Exemplo 1 - Modelo animal

MCMC_CHAIN: TOTAL_CYCLES BURN_IN THINNING_INTERVAL

11000 1000 10

SEED

2

RESTART: Y/N? [CYCLE_TO_RESTART]

n

DATAFILE NAME N_RECORDS

animodel.dat 240

NUMBER_OF_TRAITS

1

NUMBER_OF_EFFECTS

2

OBSERVATION(S)

4

WEIGHT(S)

EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS

TYPE_OF_EFFECT SAVE_SAMPLES? [EFFECT NESTED]

1 300 cross n #efeito_do_animal

#Pratica - Modelo animal

3 2 cross n #sexo

RANDOM_RESIDUAL: TYPE PRIOR_DEGREES_OF_BELIEF

homogeneous 1

RESIDUAL_PRIOR_(CO)VARIANCES

150

RANDOM_GROUP

1

RANDOM_TYPE PRIOR_DEGREES_OF_BELIEF

add_animal 1

PEDIGREEFILE: NAME N_ANIMAL N_GENETIC_GROUPS [N_BREEDS]

animodel.ped 300 0

(CO)VARIANCES

100

A.2. Modelo animal usando uma matriz do numerador do parentesco médio de Henderson.

```
# Exemplo 2 - Modelo animal com matriz de parentesco medio MCMC_CHAIN:
TOTAL_CYCLES BURN_IN THINNING_INTERVAL
11000 1000 10
SEED
2
RESTART: Y/N? [CYCLE_TO_RESTART]
n
DATAFILE NAME N_RECORDS
animodel.dat 240
NUMBER_OF_TRAITS
1
NUMBER_OF_EFFECTS
2
OBSERVATION(S)
4
WEIGHT(S)

EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS
TYPE_OF_EFFECT SAVE_SAMPLES? [EFFECT NESTED]
1 300 cross n #efeito_do_animal
3 2 cross n #sexo
RANDOM_RESIDUAL: TYPE PRIOR_DEGREES_OF_BELIEF
homogeneous 10000
RESIDUAL_PRIOR_(CO)VARIANCES
150
RANDOM_GROUP
1
RANDOM_TYPE PRIOR_DEGREES_OF_BELIEF
add_an_ms 1
PEDIGREEFILE: NAME N_ANIMAL N_GENETIC_GROUPS [N_BREEDS]
animodelrm.peda 300 0
MULTIPLE_SIRES: MAX_N_FOR_MCMC [FILE: NAME & DIMENSION]
[DIRICHLET_PRIORS]
1 animodelrm.ms 252
(CO)VARIANCES
100
```

A.3. Modelo animal com grupos genéticos de Westell.

```
# Exemplo 3 - Modelo animal com grupos genéticos MCMC_CHAIN:
TOTAL_CYCLES BURN_IN THINNING_INTERVAL
11000 1000 10
SEED
2
RESTART: Y/N? [CYCLE_TO_RESTART]
n
DATAFILE NAME N_RECORDS
animodel.dat 240
NUMBER_OF_TRAITS
1
NUMBER_OF_EFFECTS
2
OBSERVATION(S)
4
WEIGHT(S)

EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS
TYPE_OF_EFFECT SAVE_SAMPLES? [EFFECT NESTED]
1 301 cross n #efeito_do_animal
3 2 cross n #sexo
RANDOM_RESIDUAL: TYPE PRIOR_DEGREES_OF_BELIEF
homogeneous 10000
RESIDUAL_PRIOR_(CO)VARIANCES
150
RANDOM_GROUP
1
RANDOM_TYPE PRIOR_DEGREES_OF_BELIEF
add_an_ms 1
PEDIGREEFILE: NAME N_ANIMAL N_GENETIC_GROUPS [N_BREEDS]
animodelrm.pedg 300 1
MULTIPLE_SIRES: MAX_N_FOR_MCMC [FILE: NAME & DIMENSION]
[DIRICHLET_PRIORS]
0
(CO)VARIANCES
100
```

A.4. Modelo de incerteza de paternidade para filhos de acasalamentos com reprodutores múltiplos.

Exemplo 4 - Modelo de incerteza de paternidade

MCMC_CHAIN: TOTAL_CYCLES BURN_IN THINNING_INTERVAL

11000 1000 10

SEED

2

RESTART: Y/N? [CYCLE_TO_RESTART]

n

DATAFILE NAME N_RECORDS

animodel.dat 240

NUMBER_OF_TRAITS

1

NUMBER_OF_EFFECTS

2

OBSERVATION(S)

4

WEIGHT(S)

EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS

TYPE_OF_EFFECT SAVE_SAMPLES? [EFFECT NESTED]

1 300 cross n #efeito_do_animal

3 2 cross n #sexo

RANDOM_RESIDUAL: TYPE PRIOR_DEGREES_OF_BELIEF

homogeneous 10000

RESIDUAL_PRIOR_(CO)VARIANCES

150

RANDOM_GROUP

1

RANDOM_TYPE PRIOR_DEGREES_OF_BELIEF

add_an_ms 1

PEDIGREEFILE: NAME N_ANIMAL N_GENETIC_GROUPS [N_BREEDS]

animodelrm.ped 300 0

MULTIPLE_SIRES: MAX_N_FOR_MCMC [FILE: NAME & DIMENSION]

[DIRICHLET_PRIORS]

10 animodelrm.ms 252 -1

(CO)VARIANCES

100

A.5. Modelo estrutural para variâncias residuais, com heterogeneidade de variância e robustez a dados extremos.

Exemplo 5 - Modelo animal heteroscedastico robusto

MCMC_CHAIN: TOTAL_CYCLES BURN_IN THINNING_INTERVAL

11000 1000 10

SEED

2

RESTART: Y/N? [CYCLE_TO_RESTART]

n

DATAFILE NAME N_RECORDS

animodelh.dat 240

NUMBER_OF_TRAITS

1

NUMBER_OF_EFFECTS

2

OBSERVATION(S)

4

WEIGHT(S)

EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS

TYPE_OF_EFFECT SAVE_SAMPLES? [EFFECT NESTED]

1 300 cross n #efeito_do_animal

3 2 cross n #sexo

RANDOM_RESIDUAL: TYPE PRIOR_DEGREES_OF_BELIEF

struct_student_t 1

METROPOLIS_STEP_OF_STRUCTURAL_EFFECTS:

ROUNDS_WITHIN_CYCLE TUNING_SKIP

1 10

NUMBER_OF_STRUCTURAL_EFFECTS

1

STRUCTURAL_EFFECTS: LINE_FROM_EFFECTS_SECTION

SAVE_SAMPLES?

2 y

RESIDUAL_PRIOR_(CO)VARIANCES

150

RANDOM_GROUP

1

RANDOM_TYPE PRIOR_DEGREES_OF_BELIEF

add_animal 1

PEDIGREEFILE: NAME N_ANIMAL N_GENETIC_GROUPS [N_BREEDS]

animodelh.ped 300 0

(CO)VARIANCES

100

A.6. Modelo animal multirracial.

Exemplo 6 - Modelo multirracial

MCMC_CHAIN: TOTAL_CYCLES BURN_IN THINNING_INTERVAL

4000 2000 20

SEED

123

RESTART: Y/N? [CYCLE_TO_RESTART]

n

DATAFILE NAME N_RECORDS

mbreed1.dat 4000

NUMBER_OF_TRAITS

1

NUMBER_OF_EFFECTS

3

OBSERVATION(S)

4

WEIGHT(S)

EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS

TYPE_OF_EFFECT [EFFECT NESTED]

1 4000 cross n

2 1 cov n

3 1 cov n

RANDOM_RESIDUAL: TYPE PRIOR_DEGREES_OF_BELIEF

homogeneous 1

RESIDUAL_PRIOR_(CO)VARIANCES

100

RANDOM_GROUP

1

RANDOM_TYPE PRIOR_DEGREES_OF_BELIEF

add_an_mb 3

PEDIGREEFILE: NAME N_ANIMAL N_GENETIC_GROUPS [N_BREEDS]

mbreed1.ped 4000 0 2

MULTIPLE_SIRES: MAXN_FULL [FILE: NAME DIMENSION]

[ALPHA_PRIORS]

0

METROPOLIS_STEP_OF_MULTIBREED_(CO)VARIANCES:

ROUNDS_WITHIN_CYCLE TUNING_SKIP

5 10

(CO)VARIANCES

100 20

20 50

A.7. Modelo de normas de reação.

Exemplo 7 - Modelo normas de reação

```
MCMC_CHAIN: TOTAL_CYCLES BURN_IN THINNING_INTERVAL
20000 5000 30
SEED
123
RESTART: Y/N? [CYCLE_TO_RESTART]
n
DATAFILE NAME N_RECORDS
rnorm_ex1.dat 1000
NUMBER_OF_TRAITS
1
NUMBER_OF_EFFECTS
4
OBSERVATION(S)
4
WEIGHT(S)
EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS
TYPE_OF_EFFECT [EFFECT NESTED]
2 1 cov n
3 50 unknowcov n
2 1520 cov n 1
3 1520 rnorm n 1
RANDOM_RESIDUAL: TYPE PRIOR_DEGREES_OF_BELIEF
homogeneous 1
RESIDUAL_PRIOR_(CO)VARIANCES
100
RANDOM_GROUP
2
RANDOM_TYPE PRIOR_DEGREES_OF_BELIEF
diagonal 1
PEDIGREEFILE: NAME N_ANIMAL N_GENETIC_GROUPS [N_BREEDS]
(CO)VARIANCES
50
RANDOM_GROUP
3 4
RANDOM_TYPE PRIOR_DEGREES_OF_BELIEF
add_animal 2
PEDIGREEFILE: NAME N_ANIMAL N_GENETIC_GROUPS [N_BREEDS]
rnorm_ex1.ped 1520 0
(CO)VARIANCES
20 6
6 2
```


Embrapa

Pecuária Sul

CGPE 9143

Ministério da
Agricultura, Pecuária
e Abastecimento

