



Florestas

DOCUMENTOS, 46

ISSN 1517-536X

**INFERÊNCIA BAYESIANA E SIMULAÇÃO
ESTOCÁSTICA (AMOSTRAGEM DE GIBBS) NA
ESTIMAÇÃO DE COMPONENTES DE VARIÂNCIA E
DE VALORES GENÉTICOS EM PLANTAS PERENES**

Marcos Deon Vilela de Resende

Colombo
2000



Exemplares desta publicação podem ser solicitados à:

Embrapa Florestas

Estrada da Ribeira km 111 - Caixa Postal 319

83411-000 - Colombo, PR Brasil

Fone: (0**41) 666-1313

Fax: (0**41) 666-1276

E-mail: sac@cnpf.embrapa.br

Tiragem: 300 exemplares

Comitê de Publicações:

Américo Pereira de Carvalho, Antônio Carlos de S. Medeiros, Edilson Batista de Oliveira, Erich Gomes Schaitza, Guiomar Moreira de Souza Braguinha (Secretaria Executiva), Honorino Roque Rodigheri, Jarbas Yukio Shimizu, José Alfredo Sturion, Moacir José Sales Medrado (Presidente), Patrícia Póvoa de Mattos, Rivail Salvador Lourenço, Sérgio Ahrens, Susete do Rocio C. Penteado.

Revisão gramatical: Elly Claire Jansson Lopes

Normalização: Lidia Woronkoff

RESENDE, M.D.V. de. Inferência Bayesiana e simulação estocástica (amostragem de Gibbs) na estimação de componentes de variância e de valores genéticos em plantas perenes. Colombo: Embrapa Florestas, 2000. 68p. (*Embrapa Florestas*. Documentos, 46).

ISSN 1517-536X

1. Genética. 2. Planta perene. 3. Inferência. I. Título. II. Série.

CDD 581.35

©Embrapa, 2000

Produção:

ÁREA DE COMUNICAÇÕES E NEGÓCIOS

Supervisor: Miguel Haliski

LAYOUT DA CAPA:

Cleide da S.N.F. de Oliveira

DIAGRAMAÇÃO

Marta de Fátima Vencato

IMPRESSÃO

Gráfica Radial - Fone: 333-9593

Dezembro/2000

Sumário

1	SINOPSE	5
2	INTRODUÇÃO	8
3	ESTIMAÇÃO E PREDIÇÃO BAYESIANA	9
3.1	CONCEITO DE PROBABILIDADE, PROBABILIDADE CONDICIONAL E TEOREMA DE BAYES	10
3.2	FUNÇÃO DE VEROSSIMILHANÇA	16
3.3	DISTRIBUIÇÕES A PRIORI	19
3.4	PRIORI NÃO INFORMATIVA	19
3.5	DETERMINAÇÃO SUBJETIVA DA PRIORI	21
3.6	DETERMINAÇÃO DA PRIORI POR FORMAS FUNCIONAIS	21
3.7	INFERÊNCIA BAYESIANA, INFERÊNCIA FREQUENTISTA E INFERÊNCIA VEROSSIMILHANÇA	22
4	FUNDAMENTOS BAYESIANOS DA PREDIÇÃO DE VALORES GENÉTICOS	24
5	ESTIMAÇÃO BAYESIANA DE ÍNDICES DE SELEÇÃO UNICARACTERÍSTICOS	25
6	RELAÇÃO ENTRE BLUP E ESTIMADORES BAYESIANOS	26
7	RELAÇÃO ENTRE ESTIMADORES DE MÁXIMA VEROSSIMILHANÇA E ESTIMADORES BAYESIANOS	29
8	ESTIMAÇÃO BAYESIANA DE ÍNDICES DE SELEÇÃO MULTICARACTERÍSTICOS ..	29
9	ESTIMAÇÃO BAYESIANA DE COMPONENTES DE VARIÂNCIA E RELAÇÃO COM ML E REML	30
10	ESTIMADORES MELHORADOS PARA A MÉDIA DE TRATAMENTOS	32
11	PROBLEMAS DA ABORDAGEM CLÁSSICA DE PREDIÇÃO DE VALORES GENÉTICOS E VANTAGENS DA ABORDAGEM BAYESIANA	34
12	IMPLEMENTAÇÃO PRÁTICA DA ANÁLISE BAYESIANA	41
13	DIAGNÓSTICO DE CONVERGÊNCIA	51
14	ESTIMAÇÃO BAYESIANA VIA MODELOS LINEARES ROBUSTOS	53
15	APLICAÇÃO A DADOS EXPERIMENTAIS	54
16	REFERÊNCIAS BIBLIOGRÁFICAS	62

INFERÊNCIA BAYESIANA E SIMULAÇÃO ESTOCÁSTICA (AMOSTRAGEM DE GIBBS) NA ESTIMAÇÃO DE COMPONENTES DE VARIÂNCIA E DE VALORES GENÉTICOS EM PLANTAS PERENES

Marcos Deon Vilela de Resende

1. Sinopse

A predição de valores genéticos e a estimação de componentes de variância são atividades essenciais no melhoramento de plantas perenes, sendo que o procedimento padrão de estimação/predição é o REML/BLUP (máxima verossimilhança restrita/melhor predição linear não viciada). Uma deficiência do procedimento REML/BLUP para a estimação/predição de componentes de variância/valores genéticos refere-se ao fato de que o método REML propicia intervalos de confiança apenas aproximados para os parâmetros genéticos, através do uso de aproximações e suposições de normalidade assintótica. Isto porque a distribuição e variância dos estimadores não são conhecidos e, assim, questões referentes à efetividade da seleção a ser praticada não podem ser respondidas com rigor.

A análise Bayesiana baseia-se no conhecimento da distribuição a posteriori dos parâmetros genéticos e possibilita a construção de intervalos de confiança (melhor definido como intervalo de probabilidade ou intervalo de confiança Bayesiano) exatos para as estimativas dos parâmetros genéticos. Assim, esta análise propicia uma descrição mais completa sobre a confiabilidade dos parâmetros genéticos do que o método REML (Gianola & Fernando, 1986; Resende, 1997; 1999).

A inferência estatística Bayesiana baseia-se na distribuição condicional do parâmetro (ϱ) dado o vetor de dados (y), ou seja na distribuição a posteriori do parâmetro dadas as observações fenotípicas, a qual equivale a:

$$f(\theta|y) = \frac{f(y|\theta) f(\theta)}{\int_{\mathcal{R}} f(y|\theta) f(\theta) d\theta}, \text{ em que:}$$

$f(y|\varrho)$ – função densidade de probabilidade da distribuição condicional de um vetor de observações (y) dado ϱ (denominada função de verossimilhança ou modelo para os dados).

$f(\eta)$ - função densidade de probabilidade da distribuição a priori, que é também a densidade marginal de η . Esta função denota o grau de conhecimento acumulado sobre η , antes da observação de y .

$\int_R f(y|\theta) f(\theta) d\theta$ - distribuição marginal ou preditiva de y com respeito a η , onde R é a amplitude da distribuição de η .

Verifica-se então que a distribuição a posteriori é proporcional a verossimilhança x priori, ou seja, a função de verossimilhança conecta a priori à posteriori, usando para isto os dados do experimento (observações). Dessa forma, a distribuição a posteriori contempla o grau de conhecimento prévio sobre o parâmetro (η) e também as informações adicionais propiciadas pelo experimento (y).

Para a situação em que tanto a distribuição a priori quanto a distribuição das informações são normais, ou seja, $\theta \sim N(u, r^2)$ e $(Y|\theta) \sim N(\theta, \sigma^2)$ com σ^2 conhecido, a distribuição a posteriori de η é também normal, ou seja $(\theta|Y=y) \sim N(u_1, r_1^2)$ onde:

$$\mu_1 = \frac{r^{-2}\mu + \sigma^{-2}y}{r^{-2} + \sigma^{-2}}; \quad r_1^{-2} = r^{-2} + \sigma^{-2}$$

Usando esta distribuição a posteriori, pode-se demonstrar que os fundamentos básicos da predição de valores genéticos são essencialmente de natureza Bayesiana, conforme apresentado inicialmente por Robertson (1955).

Os efeitos genéticos aditivos são definidos como desvios e uma população de efeitos genéticos apresenta zero como média e σ_a^2 como variância. A melhor predição do efeito genético de um indivíduo sem nenhuma informação, tomado aleatoriamente da população é a média populacional u , a qual pode ser tomada como o estimador a priori, cuja variância é σ_a^2 . Tomando uma informação y do indivíduo, um segundo estimador do efeito genético é o desvio $(y-u)$ fenotípico em relação à média populacional, o qual possui variância $\sigma_e^2 = \sigma_y^2 - \sigma_a^2$, em que σ_y^2 é a variância de y . Estes dois estimadores independentes podem ser combinados linearmente da melhor maneira possível, tomando as recíprocas das respectivas variâncias como pesos. Sob o enfoque Bayesiano, a esperança da distribuição a posteriori corresponde à média ponderada pela precisão, das médias da priori e da verossimilhança. Assim, tem-se:

$$\hat{a} = \left(\frac{\text{estimador a priori}}{\sigma_a^2} + \frac{\text{estimador informação observada}}{\sigma_y^2 - \sigma_a^2} \right) / \left(\frac{1}{\sigma_a^2} + \frac{1}{\sigma_y^2 - \sigma_a^2} \right) = \left(\frac{0}{\sigma_a^2} + \frac{Y - u}{\sigma_y^2 - \sigma_a^2} \right) / \left(\frac{1}{\sigma_a^2} + \frac{1}{\sigma_y^2 - \sigma_a^2} \right) = h^2 (y - u),$$

como no enfoque da inferência estatística clássica, onde h^2 é a herdabilidade. Este é também (considerando y como a média dos indivíduos selecionados) o estimador para o ganho genético para a seleção massal.

No contexto Bayesiano, as equações de modelo misto (denominadas equações de modelo misto de Robertson) são dadas por:

$$\begin{bmatrix} X'R^{-1}X + S^{-1} & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} E(\beta|y) \\ E(a|y) \end{bmatrix} = \begin{bmatrix} X'R^{-1}y + S^{-1}r_1 \\ Z'R^{-1}y + G^{-1}0 \end{bmatrix}$$

onde $r_1 = E(b)$ e $0 = E(a)$.

Tomando a informação a priori sobre os efeitos fixos (b) como não informativa (expressa como $S \rightarrow \infty$ e então $S^{-1} \rightarrow 0$), tem-se que esta equação resultante equivale às equações do modelo misto (EMM):

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}, \text{ em que:}$$

$$R = I\sigma_e^2; \quad G = A\sigma_a^2 \text{ e } S = \text{Var}(\beta).$$

O problema básico da implementação da análise Bayesiana refere-se à integração numérica (no espaço do parâmetro) da função densidade de probabilidade a posteriori. Tal integração, por métodos analíticos, é impossível na prática. Assim, tem sido utilizado o procedimento de simulação estocástica denominado amostragem de Gibbs (GS) para a viabilização da estimação Bayesiana (Sorensen, 1996).

Além das distribuições (normais) assumidas para os efeitos aleatórios (a) no modelo linear clássico e para a verossimilhança do vetor de observações (y), a abordagem Bayesiana requer atribuições para as distribuições a priori dos efeitos fixos e componentes de variância. De maneira geral, são atribuídas distribuições a priori não informativas ou uniformes para os efeitos fixos e componentes de variância, como forma de caracterizar um conhecimento a

priori vago sobre os referidos efeitos e componentes (Gianola & Fernando, 1986).

Utilizando-se distribuições a priori não informativas para os efeitos fixos e componentes de variância, tem-se que as modas das distribuições marginais a posteriori dos parâmetros genéticos correspondem às estimativas obtidas por REML e gera estimativas de componentes de variância pelo método VEIL (ou de verossimilhança integrada) apresentados por Gianola & Foulley (1990). A grande vantagem da análise Bayesiana, neste caso, refere-se à obtenção dos desvios padrões e intervalos de confiança exatos para os parâmetros genéticos bem como a obtenção de estimativas mais precisas. Resende (1999b) empregando as abordagens REML e GS obteve estimativas de $0,32 \pm 0,06$ e $0,34 \pm 0,005$, respectivamente, para a herdabilidade no sentido restrito para o caráter diâmetro em *Pinus caribaea* var. *hondurensis*, confirmando a grande precisão do método Bayesiano.

Quanto à estimação dos efeitos fixos (efeitos de blocos por exemplo) e a predição dos efeitos aleatórios (valores genéticos), tem-se que as médias das distribuições marginais a posteriori dos parâmetros de locação (efeitos fixos e aleatórios), dado os componentes de variância ou parâmetros de dispersão, equivalem às soluções das equações do modelo misto do BLUP desde que sejam atribuídas prioris não informativas para os efeitos fixos, prioris normais para os efeitos aleatórios e verossimilhança normal para o vetor de observações. Neste caso, a vantagem da abordagem Bayesiana refere-se à obtenção de valores genéticos preditos com menores desvios padrões.

Em resumo, a análise Bayesiana propicia estimativas mais precisas de componentes de variância, parâmetros genéticos, valores genéticos e ganhos genéticos. Adicionalmente, a estimação Bayesiana permite a análise exata de amostras de tamanho finito. Este último aspecto é muito importante, especialmente para programas de melhoramento baseados em conjuntos de dados desbalanceados, onde tal abordagem propicia uma elegante análise de amostra finita, a qual não pode ser obtida pela metodologia freqüentista de modelos mistos.

2. Introdução

A predição de variáveis aleatórias e a estimação de componentes de variância e efeitos fixos via metodologia de modelos mistos apresentam grande relevância na genética quantitativa aplicada ao melhoramento de plantas perenes. Embora a metodologia de modelos lineares mistos (MMLM) tratada pela abordagem da Inferência Estatística Freqüentista apresente várias propriedades desejáveis, publicações recentes têm mostrado uma superioridade da MMLM quando abordada do ponto de vista da Inferência Estatística Bayesiana.

Em 1986, o trabalho clássico de Gianola & Fernando (1986) propôs a abordagem Bayesiana como uma estratégia conceitual para resolução de problemas em melhoramento animal, a qual foi detalhada por Gianola et al. (1990). Devido às excelentes propriedades teóricas e práticas da análise Bayesiana, acredita-se que a mesma tornar-se-á rotineira em melhoramento de espécies perenes, transformando-se no procedimento padrão de estimação de componentes de variância e de predição de valores genéticos.

Os geneticistas quantitativos foram influenciados pelo renascimento Bayesiano que ocorreu em meados da década de 1960. Lindley & Smith (1972) apresentaram uma ligação entre modelos mistos e as abordagens Bayesianas. Dempfle (1977) e Ronningen (1971) pesquisaram as conexões entre BLUP e estimadores Bayesianos. Entretanto, somente na década de 1990, a utilização da Inferência Bayesiana em melhoramento animal tornou-se rotineira, destacando-se os trabalhos de Gianola & Foulley (1990); Gianola et al. (1990), Weigel & Gianola (1993); Wang et al. (1993, 1994a e b); Jensen et al. (1994); Sorensen et al. (1994); Varona et al. (1994); Varona (1994); Tassel et al. (1995); Rodriguez et al. (1996); Arendonk et al. (1996), Garcia-Cortez & Sorensen (1996) e Rekaya (1997).

Outros trabalhos, enfatizando a estimação Bayesiana de parâmetros genéticos, mas não no contexto dos modelos lineares mistos, foram desenvolvidos (Toro & Prunonosa, 1984; Saio & Hayashi, 1990) através do uso do espaço paramétrico como informações a priori. No Brasil, a análise Bayesiana vem sendo recomendada para avaliação genética em melhoramento animal (Magnabosco, 1997; Rosa, 1999) e florestal (Resende, 1997; 1999).

A estimação Bayesiana de variáveis aleatórias apresenta grande relevância em diversas áreas do conhecimento além da genética quantitativa, especialmente na biometria, econometria, engenharia, física e inteligência artificial, dentre outras. Segundo Sun et al. (1996), de maneira genérica, uma análise Bayesiana exata pode ser obtida para os modelos de componentes de variância baseados tradicionalmente na teoria da normalidade, permitindo, para qualquer parâmetro de interesse, uma detalhada inferência, para amostras de tamanho finito. Esta é uma das grandes vantagens da análise Bayesiana.

3. Estimação e predição Bayesiana

A origem da Inferência Estatística Bayesiana ocorreu com o lançamento da obra “An Essay Towards Solving a Problem in the Doctrine of Chances”, a título póstumo, de autoria do Reverendo Thomas Bayes. Esta obra foi publicada em 1763, por Richard Price, e apresenta o teorema denominado Teorema de Bayes (Murteira, 1990).

O teorema de Bayes fundamenta-se em probabilidades condicionais. Para enunciar tal teorema torna-se relevante apresentar alguns conceitos básicos de probabilidade, conceitos estes apresentados em vários livros de estatística, por exemplo em Mood et al. (1974).

3.1. Conceito de probabilidade, probabilidade condicional e Teorema de Bayes

O conceito de probabilidade é a base de toda a Estatística. A definição de probabilidade é aceita por praticamente todos os estatísticos. Entretanto, várias interpretações de probabilidade existem, destacando-se a Física ou Clássica, a Freqüentista, a Lógica e a Subjetiva.

(i) Probabilidade Freqüentista

Sob o enfoque freqüentista, as freqüências relativas de ocorrência dos resultados convergem para a probabilidade de ocorrência destes resultados quando o número n de repetições do experimento tende a ser grande. Por exemplo, a freqüência de um resultado A é dada por

$f_A = \frac{n_A}{n} \Rightarrow \lim_{n \rightarrow \infty} f_A = P[A]$, em que n_A é o número de vezes em que ocorreu o resultado A e $P[A]$ denota a probabilidade de A .

(ii) Probabilidade Física ou Clássica

A interpretação clássica de probabilidade baseia-se no conceito de resultados igualmente prováveis. Assim, a probabilidade de um resultado ou evento é dada pela razão $1/N$, em que N é o número de resultados possíveis.

(iii) Probabilidade Subjetiva

Segundo esta interpretação, a probabilidade que uma pessoa atribui a um possível resultado de algum processo representa seu próprio julgamento da chance de o resultado ser obtido. Assim, a probabilidade de um evento A é a medida do grau de confiança que a pessoa deposita em A .

(iv) Probabilidade Lógica

A interpretação lógica de probabilidade pode ser derivada em função do conhecimento científico que é de domínio público, ou seja, do conhecimento comum a todos.

A seguir são apresentadas algumas definições relevantes de Probabilidade.

Espaço de probabilidade: um espaço de probabilidade é a tripla $(\bar{W}, A, P[.])$, onde \bar{W} é o espaço amostral, A é uma coleção de eventos e $P[.]$ é uma função de probabilidade com domínio em A .

Probabilidade condicional: Se A e B são eventos em A , para um dado espaço de probabilidade $(\bar{w}, A, P[.])$, a probabilidade condicional de um evento A dado o evento B , indicado por $P[A|B]$ é definido por:

$$P[A|B] = \frac{P[A, B]}{P[B]} \text{ se } P[B] > 0,$$

$$P[A, B] = P[B] \cdot P[A|B] = P[A] \cdot P[B|A].$$

Desta forma, como $P[A|B]$ é uma probabilidade, valem para ela todas as propriedades de probabilidade. A prova de que a proporção $P[A, B]/P[B]$ é uma probabilidade vem da verificação dos axiomas:

- (i) $P[A|B] = P[A, B]/P[B] \geq 0$
- (ii) $P[\bar{w}|B] = P[\bar{w}, B]/P[B] = P[B]/P[B] = 1$
- (iii) Se $A_1, A_2, \dots, A_n, \dots$, é uma seqüência de eventos mutuamente exclusivos em A , então:

$$P\left[\bigcup_{i=1}^{\infty} A_i | B\right] = \sum_{i=1}^{\infty} P[A_i | B]$$

Regra da multiplicação: Para um dado espaço de probabilidade $(\bar{w}, A, P[.])$, se A_1, A_2, \dots, A_n são eventos em A , para o qual $[A_1, A_2, \dots, A_{n-1}] > 0$; então:

$$P[A_1, A_2, \dots, A_n] = P[A_1] \cdot P[A_2 | A_1] \cdot P[A_3 | A_1, A_2] \dots P[A_n | A_1, A_2, \dots, A_{n-1}]$$

Eventos independentes: Para um dado espaço de probabilidade $(\bar{w}, A, P[.])$, sejam A e B dois eventos em A . Os eventos A e B são definidos ser independentes se, e somente se, uma das condições for satisfeita:

- (i) $P[A, B] = P[A] \cdot P[B]$
- (ii) $P[A|B] = P[A]$ se $P[B] > 0$
- (iii) $P[B|A] = P[B]$ se $P[A] > 0$

Teorema de probabilidade total: Para um dado espaço de probabilidade $(\bar{w}, A, P[.])$, se B_1, B_2, \dots, B_n é uma coleção de eventos mutuamente disjuntos em A satisfazendo:

$\Omega = \bigcup_{j=1}^n B_j$ e $P[B_j] > 0$ para $j=1, 2, \dots, n$ então para todo $A \in \mathbf{A}$:

$$P[A] = \sum_{j=1}^n P[A | B_j] \cdot P[B_j] = P[A]$$

Considere o seguinte exercício: duas caixas possuem bolas vermelhas (V) e azuis (A). A caixa 1 (C_1) possui 60V e 40A enquanto a caixa 2 (C_2) possui 10V e 20A. Suponha que uma caixa seja tomada aleatoriamente e uma bola tomada aleatoriamente dentro desta caixa. Qual a probabilidade de que a bola seja vermelha?

Tem-se:

$$P[C_1] = 0,5; \quad P[C_2] = 0,5$$

$$P[V \hat{\cap} C_1] = 0,60; \quad P[V \hat{\cap} C_2] = 10/30 = 0,3333$$

Aplicando o teorema da probabilidade total, tem-se:

$$\begin{aligned} P[V] &= P[V \cap C_1] + P[V \cap C_2] \\ &= P[V \hat{\cap} C_1] \cdot P[C_1] + P[V \hat{\cap} C_2] \cdot P[C_2] \\ &= 0,60 \cdot 0,5 + 0,3333 \cdot 0,5 \\ &= 0,4666 \end{aligned}$$

Teorema de Bayes: Para um dado espaço de probabilidade $(\omega, \mathbf{A}, P[\cdot])$, se B_1, B_2, \dots, B_n é uma coleção de eventos mutuamente disjuntos em \mathbf{A} satisfazendo:

$\Omega = \bigcup_{j=1}^n B_j$ e $P[B_j] > 0$ para $j=1, 2, \dots, n$ então para todo $A \in \mathbf{A}$:

$$P[B_k | A] = \frac{P[A | B_k] \cdot P[B_k]}{\sum_{j=1}^n P[A | B_j] \cdot P[B_j]}$$

Considere novamente o exemplo das bolas e caixas. Qual a probabilidade de ter sido tomada a caixa 1, sabendo-se que a bola amostrada foi a vermelha?

Aplicando-se o teorema de Bayes, tem-se:

$$P[C_1 | V] = \frac{P[V | C_1] \cdot P[C_1]}{P[V | C_1] \cdot P[C_1] + P[V | C_2] \cdot P[C_2]} = \frac{P[V | C_1] \cdot P[C_1]}{P[V]} = \frac{0,60 \cdot 0,5}{0,4666} = 0,6428$$

Assim, o Teorema de Bayes envolve os conceitos de Probabilidade Condicional e Probabilidade Total.

O Teorema de Bayes definido em termos de densidades de probabilidade, tem a seguinte formulação para a distribuição de uma variável aleatória contínua:

$$f(\theta|y) = \frac{f(y|\theta) f(\theta)}{\int_R f(y|\theta) f(\theta) d\theta} \quad (1), \text{ onde:}$$

α – vetor de parâmetros

y – vetor de dados ou de informações obtidas por amostragem

$f(\alpha|y)$ – distribuição condicional de α dado y , ou distribuição a posteriori (que é a base da estimação e predição Bayesiana).

$f(y|\alpha)$ – função densidade de probabilidade da distribuição condicional de uma observação (y) dado α (denominada função de verossimilhança ou modelo para os dados).

$f(\alpha)$ – função densidade de probabilidade da distribuição a priori, que é também a densidade marginal de α . Esta função denota o grau de conhecimento acumulado sobre α , antes da observação de y .

$f(y|\alpha)f(\alpha)$ – função densidade conjunta de y e α .

$f(y) = \int_R f(y, \theta) d\theta = \int_R f(y|\theta) f(\theta) d\theta = E_\theta [f(y|\theta)]$ - distribuição marginal ou preditiva de y com respeito a α , onde R é a amplitude da distribuição de α . E_α significa esperança com respeito à distribuição de α . (A integração da distribuição conjunta, no espaço paramétrico, produz a marginal de y).

Como $f(y)$ não é função de α (ou seja, $f(y)$ é constante para qualquer α), a forma usual da formulação de Bayes é: $f(\alpha|y) \propto f(y|\alpha) f(\alpha)$, onde \propto indica proporcionalidade.

A expressão (1) advém das expressões $f(\alpha, y) = f(y|\alpha) f(\alpha)$ e $f(\alpha, y) = f(\alpha|y) f(y)$, as quais são obtidas a partir do teorema da probabilidade condicional.

Verifica-se pela expressão (1) que um fator lógico que diferencia o enfoque Bayesiano da abordagem freqüentista refere-se ao tipo de informação utilizada. Na concepção Bayesiana, toda informação de que se dispõe é útil e deve ser utilizada. Por outro lado, a estatística clássica utiliza apenas observações de dados reais, desprezando-se as informações subjetivas (Gamerman & Migon, 1993).

A componente amostral ou experimental é comum aos modelos clássicos e Bayesianos, mas com interpretações diferentes. Embora os modelos Bayesianos passem por uma extensão dos modelos clássicos, existe uma

divergência fundamental entre os dois enfoques: no modelo clássico o parâmetro é um escalar ou um vetor desconhecido, porém fixo, ao passo que no modelo Bayesiano, o parâmetro é considerado como escalar ou vetor aleatório (não observável), pois para os Bayesianos tudo o que é desconhecido é incerto e, portanto, toda a incerteza deve ser quantificada em termos de probabilidade (Murteira, 1988; 1990). Em função disto, os modelos Bayesianos tratam formalmente a informação a priori, através da distribuição de probabilidade (subjéitiva ou lógica) a priori. As informações a priori e amostrais permitem a atualização periódica da distribuição de probabilidade a posteriori e, portanto, permite modificar e atualizar as estimativas dos parâmetros.

O teorema de Bayes é uma regra para atualização de probabilidades e pode ser utilizado para cálculo de probabilidades a posteriori em mais que um estágio. Considerando ainda o exemplo das duas caixas com bolas vermelhas (V) e azuis (A), qual a probabilidade de ter sido amostrada a caixa 1 (C_1), tendo-se amostradas 2 bolas vermelhas?

$$P[C_1|V_1 \cap V_2] = \frac{P[V_1 \cap V_2 | C_1].P[C_1]}{P[V_1 \cap V_2]} = \frac{P[V_2|C_1].P[C_1|V_1]}{P[V_2|C_1].P[C_1|V_1] + P[V_2|C_2].P[C_2|V_1]}$$

O raciocínio usado na expressão acima é de que, no cálculo da probabilidade a posteriori após a retirada da segunda bola vermelha V_2 , a probabilidade a priori a ser utilizada refere-se à probabilidade a posteriori no estágio anterior (após a retirada da primeira bola vermelha V_1). Assim, tem-se:

$$P[V_2 \hat{=} C_1] = P[V \hat{=} C_1] = 0,60;$$

$$P[C_1 \hat{=} V_1] = P[C_1 \hat{=} V] = 0,6428, \text{ conforme calculado anteriormente;}$$

$$P[V_2 \hat{=} C_2] = P[V \hat{=} C_2] = 0,3333;$$

$$P[C_2 \hat{=} V_1] = 1 - P[C_1 \hat{=} V] = 0,3572;$$

$$P[C_1|V_1 \cap V_2] = \frac{0,60 \cdot 0,6428}{0,60 \cdot 0,6428 + 0,3333 \cdot 0,3572} = 0,7641$$

A ordem com que as informações são processadas pelo Teorema de Bayes é irrelevante, sendo que a posteriori resultante é sempre a mesma. Também, as observações podem ser processadas em um único grupo, uma a uma, ou divididas em subgrupos que o resultado final permanece o mesmo (Gamerman, 1996).

A expressão (1) fornece a regra de atualização de probabilidades sobre α , partindo de $f(\alpha)$ e chegando a $f(\alpha|\frac{1}{2}y)$. Assim, a distribuição a posteriori é proporcional à verossimilhança X priori, ou seja, a função de verossimilhança conecta a priori à posteriori usando para isto os dados experimentais (amostrais).

Dessa forma, a distribuição a posteriori contempla o grau de conhecimento prévio sobre o parâmetro $f(\alpha)$ e também as informações adicionais propiciadas pelo experimento $f(y|\alpha)$. A distribuição $f(y)$ preditiva ou marginal de y , antes de se observar y é útil para verificar a adequação da priori através das predições que ela fornece para y . Após observar y é útil para testar o modelo como um todo (Gamerman & Migon, 1993).

Assumindo distribuição normal para a priori e para as observações, ou seja, $\theta \sim N(\mu, r^2)$ e $Y|\theta \sim N(\theta, \sigma^2)$ com σ^2 conhecido, a distribuição a posteriori de α é também normal, ou seja $\theta|Y=y \sim N(\mu_1, r_1^{-2})$, onde:

$$\mu_1 = \frac{r^{-2}\mu + \sigma^{-2}y}{r^{-2} + \sigma^{-2}}; \quad r_1^{-2} = r^{-2} + \sigma^{-2}$$

Assim, verifica-se que a precisão ou inverso da variância conota informação, ou seja, a relação $w = r^{-2}/(r^{-2} + \sigma^{-2})$, $w \in (0,1)$, mede a informação contida na priori em relação à informação total (priori + verossimilhança). Dessa forma, pode-se rescrever $\mu_1 = w\mu + (1-w)y$, de forma que a média da posteriori equivale à média ponderada pela certeza na priori e na verossimilhança. Verifica-se também que a precisão (r_1^{-2}) da posteriori equivale à soma das precisões da priori e da verossimilhança.

Em termos de estimação, enquanto para a estatística freqüentista possam existir vários estimadores para um determinado parâmetro, para a Estatística Bayesiana existe, em princípio, um único estimador, o qual conduz a estimativas que maximizam a função densidade de probabilidade a posteriori. Assim, inferências sobre α são realizadas a partir da densidade a posteriori através da expressão geral $p(\theta|y) = \int_R f(\theta|y) d\theta$ onde p denota probabilidade (Gianola & Fernando, 1986).

Ao nível do i -ésimo elemento do vetor α , a esperança condicional de α_i

dado y é
$$\frac{\int_R \theta_i f(y|\theta) f(\theta) d\theta}{\int_R f(y|\theta) f(\theta) d\theta}$$
, o qual é o usual estimador Bayesiano

de α_i .

O teorema de Bayes pode ser estendido para inferências sobre dados não observados (dados futuros, por exemplo) (Lindley, 1965; Gamerman & Migon, 1993). Sendo \tilde{y} um vetor de dados não observados, tem-se

$$f(\tilde{y}, \theta|y) = f(\tilde{y}|\theta, y) f(\theta|y), \quad (2)$$

que é a função densidade de probabilidade conjunta ou função preditiva. A função densidade de probabilidade marginal ou função preditiva de \tilde{y} é dada

$$\text{por } f(\tilde{y}|y) = \int_{R\theta} f(\tilde{y}, \theta|y) d\theta \quad (3)$$

$$= \int_{R\theta} f(\tilde{y}|\theta, y) f(\theta|y) d\theta$$

$$= \int_{R\theta} f(\tilde{y}|\theta) f(\theta|y) d\theta$$

$$= E_{\theta|y} [f(\tilde{y}|\theta)]$$

As expressões (2) e (3) representam a base para os problemas de predição Bayesiana. Do exposto, torna-se claro que a estimação trata da compreensão de um fenômeno (realidade) através de afirmações probabilísticas a respeito de quantidades não observáveis e a predição trata de se fazer afirmações probabilísticas acerca de quantidades a serem observadas no futuro.

3.2. Função de verossimilhança

A função de verossimilhança formaliza a contribuição dos dados amostrais para o conhecimento sobre \mathcal{Q} , visto que a mesma conecta a distribuição a priori à distribuição a posteriori.

A função de verossimilhança $[\ell(\theta; y)]$ de \mathcal{Q} é a função que associa a cada \mathcal{Q} , o valor $f(y|\mathcal{Q})$. Dessa forma $\theta \rightarrow \ell(\theta; y) = f(y|\theta)$. A função $\ell(\theta; y)$ associa (para um valor fixo de y) a probabilidade de ser observado y a cada valor de \mathcal{Q} . Assim, quanto maior o valor de ℓ maiores são as probabilidades atribuídas pelo particular valor de \mathcal{Q} considerado, ao valor fixado de y . Ao fixar um valor de y e variar os valores de \mathcal{Q} observa-se a plausibilidade ou verossimilhança de cada um dos valores de \mathcal{Q} . É interessante observar que: $\int_{R} f(y|\theta) d\theta = 1$ mas $\int_{\theta} \ell(\theta; y) d\theta = K \neq 1$, ou seja a integral da função densidade de probabilidade equivale a 1, mas a função de verossimilhança não integra 1 (Gamerman & Migon, 1993).

É importante ressaltar que os termos probabilidade e verossimilhança são conceitos diferentes. No cômputo da verossimilhança, fixa-se a amostra ou conjunto de dados (y) e varia-se o parâmetro \mathcal{Q} , procurando encontrar o

parâmetro verossímil ou plausível com o conjunto de dados. Por outro lado, no cálculo de uma probabilidade utiliza-se de uma distribuição com parâmetro α conhecido e calcula-se a probabilidade de se observar um determinado valor $y = y_0$. Em outras palavras, o problema da probabilidade é prever a chance de ocorrer y sabendo α e o problema da verossimilhança é fazer afirmações sobre α com base no valor observado y .

No cálculo de probabilidades fixa-se α (conhecido) e varia-se y ao passo que no cálculo da verossimilhança fixa-se y e varia-se α .

Como exemplo, considere a distribuição $y \sim \text{Binomial}(2, \alpha)$, em que:

$2 = n =$ número de repetições do experimento;

$\alpha =$ parâmetro da binomial = probabilidade de sucesso;

$y = 0, 1, 2 =$ espaço amostral (y é o número de sucessos);

$\alpha \in (0, 1) =$ espaço paramétrico ou domínio de variação do parâmetro α .

A função de probabilidade equivale a $p(y|\alpha) = \binom{2}{y} \alpha^y (1 - \alpha)^{2-y}$, a qual

equivale à própria função de verossimilhança, $\ell(\alpha; y)$, quando se observa $y = y_0$. Considerando os diferentes elementos do espaço amostral, tem-se as seguintes funções de verossimilhança:

$$(i) \quad \text{se } y = 0, \text{ então } \ell(\alpha; y = 0) = \binom{2}{0} \alpha^0 (1 - \alpha)^2 = (1 - \alpha)^2;$$

$$(ii) \quad \text{se } y = 1, \text{ então } \ell(\alpha; y = 1) = \binom{2}{1} \alpha^1 (1 - \alpha)^1 = 2\alpha(1 - \alpha);$$

$$(iii) \quad \text{se } y = 2, \text{ então } \ell(\alpha; y = 2) = \binom{2}{2} \alpha^2 (1 - \alpha)^0 = \alpha^2.$$

Tendo-se observado $y = 0$, variando-se α no intervalo de 0 a 1 na função de verossimilhança $\ell(\alpha; y = 0) = (1 - \alpha)^2$, verifica-se que o ponto de máximo desta função ocorre quando $\alpha = 0$. Por outro lado, tendo-se observado $y = 1$, variando-se α no intervalo de 0 a 1 na função $\ell(\alpha; y = 1) = 2\alpha(1 - \alpha)$, verifica-se que o ponto de máximo ocorre quando $\alpha = \frac{1}{2}$. Finalmente, tendo-se observado $y = 2$, verifica-se que o máximo de $\ell(\alpha; y = 2) = \alpha^2$ equivale a 1. Estas funções de verossimilhança encontram-se plotadas na Figura 1.

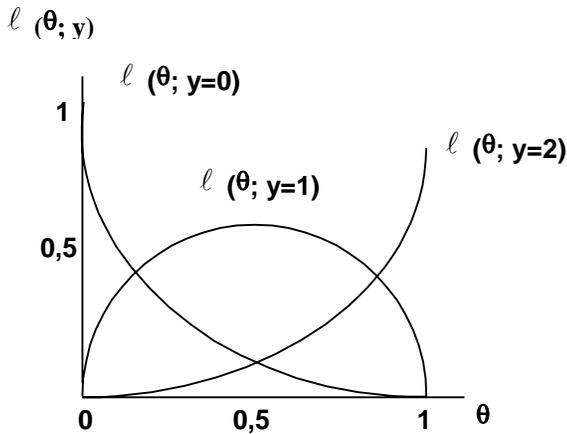


Figura 1. Funções de verossimilhança para diferentes valores observados $y = 0, 1, 2$.

Em resumo, se for observado:

$y = 0 \Rightarrow \hat{\theta} = 0$ maximiza $\ell(\theta; 0)$;

$y = 1 \Rightarrow \hat{\theta} = \frac{1}{2}$ maximiza $\ell(\theta; 1)$;

$y = 2 \Rightarrow \hat{\theta} = 1$ maximiza $\ell(\theta; 2)$.

Genericamente, uma função de verossimilhança pode ser construída conforme descrito a seguir. Sendo, y_1, y_2, \dots, y_n uma amostra aleatória independente e identicamente distribuída, com função de probabilidade ou função densidade de probabilidade dada por $f(y_i|\theta)$, em que θ é o parâmetro da função, a função de verossimilhança de $\underline{y} = (y_1, y_2, \dots, y_n)$ é dada por $\ell(\theta; \underline{y}) = f(\underline{y}|\theta) = f(y_1, y_2, \dots, y_n|\theta)$, função essa que associa a cada θ , um valor de $f(y_i|\theta)$.

A função conjunta $\ell(\theta; \underline{y}) = f(\underline{y}|\theta) = f(y_1, y_2, \dots, y_n|\theta)$ fatora nas marginais de forma que equivale ao produto $\ell(\theta; \underline{y}) = f(y_1|\theta) \dots f(y_n|\theta) =$

$\prod_{i=1}^n f(y_i|\theta)$. Como exemplo, considere a distribuição normal $N(\theta, \sigma^2)$. Neste

caso, a função de verossimilhança desta normal equívale a $\ell(\underline{\theta}; \underline{y}) = f(\underline{y} | \underline{\theta})$

$$= \prod_{i=1}^n f_N(y_i; \theta, \sigma^2) = \prod_{i=1}^n \frac{1}{[2\pi\sigma^2]^{1/2}} \exp\left\{-\frac{1}{2} \frac{(y_i - \theta)^2}{\sigma^2}\right\} \alpha \exp\left\{-\frac{n}{2\sigma^2} (\bar{y} - \theta)^2\right\},$$

onde \bar{y} é a média aritmética dos y_i .

A função de verossimilhança é a base do Princípio da Verossimilhança, o qual postula que toda a informação contida na amostra ou experimento encontra-se representada nesta função. Existem diferentes verossimilhanças que podem ser empregadas na estimação paramétrica tais quais verossimilhança incondicional marginal, verossimilhança condicional e verossimilhança parcial. Gianola et al. (1989) apresentam detalhes sobre a estimação envolvendo estes diferentes conceitos.

3.3. Distribuições a priori

A distribuição a priori constitui-se no único elemento novo na análise Bayesiana em relação à análise clássica. A determinação desta distribuição é subjetiva, entretanto dados experimentais podem ser utilizados nesta etapa. A premissa desta distribuição é que ela represente o estado atual de conhecimento sobre θ , antes de serem analisados os resultados experimentais. Zellner (1971) distingue dois tipos de informação a priori, uma baseada em dados passados e outra completamente subjetiva (não baseada em dados). Na presença de dados experimentais prévios ou quando baseada em fundamentos teóricos, a priori deve ser obtida de uma maneira científica (por formas funcionais).

A distribuição a priori tem importância maior quando a quantidade (n) de dados experimentais atuais é pequena. Caso contrário a priori tende a ser dominada (pois recebe menor peso) pela verossimilhança, ou seja, com $n \rightarrow \infty$, as abordagens frequentista e Bayesiana são coincidentes. Em outras palavras, o peso da verossimilhança é maior do que o peso da priori, se a variância dos dados observados é menor que a variância da priori. É importante ressaltar, entretanto, que atribuir probabilidade zero a priori, significa atribuir zero também a posteriori pois, neste caso, sendo a posteriori = priori x verossimilhança, não há verossimilhança que faça a posteriori diferente de zero.

3.4. Priori não informativa

Existem situações em que o conhecimento sobre determinado fenômeno é vago ou inexistente. Nestes casos, a distribuição a priori é dita vaga, difusa

ou não informativa, não precisa ou de variância alta, significando que a densidade a priori reflete ignorância (Jeffreys, 1961).

Assim, a priori não informativa poderia ser representada por $f(\varrho)$ constante; $-\infty < \varrho < \infty$ (4)

Uma priori deste tipo poderia ser especificada através de uma distribuição com variância tendendo ao infinito (Gianola & Fernando, 1986) e neste caso, maior ênfase ou peso serão dados à verossimilhança (dados experimentais ou amostrais).

Entretanto, para inferência sobre uma variância s^2 , prioris da forma mencionada tornam-se informativas (Jeffreys, 1961) e uma priori não informativa poderia ser representada por $f(\sigma^2) \propto \frac{1}{\sigma^2}$; $0 < \sigma^2 < \infty$.

Para prioris não informativas, supuseram-se inicialmente distribuições uniformes como representantes de situações onde não se dispõe de informação inicial, ou seja, $f(\varrho)$ a constante (ϱ variando na reta), fato que implica em não favorecer qualquer valor particular de ϱ (Bayes, 1763).

Gamerman & Migon (1993) apresentam algumas dificuldades inerentes a esta escolha:

(i) $f(\varrho)$ é imprópria, ou seja, a integral sobre todos os possíveis valores de ϱ não converge: $\int f(\theta) d\theta \rightarrow \infty$.

(ii) se $\phi = \phi(\theta)$ é transformação 1 a 1 (biunívoca) de ϱ e se ϱ tem distribuição uniforme então pelo teorema de transformações de

variáveis, a densidade de ϕ será: $f(\phi) = f(\theta(\phi)) \left| \frac{d\theta}{d\phi} \right| \propto \left| \frac{d\theta}{d\phi} \right|$.

Assim, o raciocínio que conduz à especificação $f(\varrho)$ a constante, deveria levar também a $f(\phi)$ a constante, o que não é verdade. O ideal seria estabelecer uma regra que fosse invariante e que $f(\phi)$ não fosse imprópria.

Entretanto, o interesse principal reside na distribuição a posteriori e como esta é, em geral, própria (pois (4) é absorvida na constante de integração associada com a posteriori) mesmo quando a priori não o é, a eventual impropriedade das distribuições a priori não é importante (Gamerman & Migon, 1993). A classe de prioris não informativas proposta por Jeffreys (1961) é invariante, mas eventualmente imprópria.

3.5. Determinação subjetiva da priori

Sendo α uma quantidade desconhecida e considerando os seus possíveis valores, a probabilidade a priori de cada valor possível de α pode ser avaliada diretamente através de instrumentos auxiliares (loterias, roletas) quando α for discreta. No caso de distribuições contínuas, têm sido sugeridas as abordagens do histograma, da função distribuição e da verossimilhança relativa (Gamerman & Migon, 1993).

3.6. Determinação da priori por formas funcionais

De algum conhecimento sobre α pode-se descrever sua densidade por uma determinada forma funcional, ou eventualmente, pode-se definir uma família paramétrica de densidades. Adicionalmente deve-se verificar se a densidade escolhida está representando bem a informação disponível. Gamerman & Migon (1993) citam que, por exemplo, poder-se-iam fazer as seguintes considerações sobre α :

- (i) α se distribui simetricamente em relação à moda;
- (ii) a densidade de α decai rapidamente quando se afasta da moda, implicando uma variância pequena;
- (iii) intervalos muito afastados da moda têm probabilidade desprezível.

Tais considerações poderiam caracterizar aproximadamente a família da distribuição normal com seus hiperparâmetros, ou seja, parâmetros que auxiliam a especificação da priori.

Uma abordagem sistemática para a determinação de distribuições a priori pode ser enunciada tal qual o caso das distribuições conjugadas. A propriedade de preservação da classe de distribuições quando a distribuição do parâmetro é atualizada define conjugação.

No tópico 3.1., enunciou-se que se a distribuição dos dados é $(y|\alpha) \sim N(\alpha, s^2)$ e a distribuição a priori é $\alpha \sim N(u, r^2)$, a distribuição a posteriori de α é $(\alpha|y) \sim N(m_1, r_1^2)$. Assim, tem-se a priori e a posteriori com distribuição normal, ou seja, priori normal conjugada a posteriori normal. Este tipo de priori, segundo Gamerman & Migon (1993) torna a análise mais tratável, permitindo explorar a natureza seqüencial do método Bayesiano.

Sendo $F = \{f(y|\theta), \theta \in \theta\}$ uma família de distribuições amostrais, uma classe de distribuição Y é conjugada a F se:

$$\forall f \in F \text{ e } f(\theta) \in \Psi \text{ então } f(\theta|y) \in \Psi$$

Quanto à definição de distribuições conjugadas, algumas ressalvas devem ser feitas (Gamerman & Migon, 1993):

- (i) a classe Υ pode ser muito ampla, por exemplo, quando se tem $\Upsilon = \{\text{todas as distribuições}\}$ e F qualquer. Neste caso, a definição não tem valor prático;
- (ii) a classe Υ pode ser muito restrita, por exemplo, quando $\Upsilon = \{P: P(\alpha = \alpha_0) = 1\}$. Isto implica que qualquer amostra não dará informação alguma, pois tem-se a certeza de que $\alpha = \alpha_0$ e, portanto, esta certeza será mantida na posteriori. Assim, a definição também não terá valor prático. Dessa forma, recomenda-se então a atribuição de probabilidade positiva a todo e qualquer valor possível de α por mais improvável que ele seja.

Em conclusão, a classe Υ deve ser ampla o suficiente para proporcionar a escolha de uma distribuição a priori conveniente e, ao mesmo tempo deve ser restrita o suficiente para que a definição de distribuições conjugadas seja útil.

Como principais famílias conjugadas podem ser citadas:

- (i) família de distribuições Beta conjugada à Binomial;
- (ii) família de distribuições Normais conjugada à Normal;
- (iii) família de distribuições Gama conjugada à Poisson;
- (iv) família de distribuições Gama conjugada à Exponencial;
- (v) Normal-Gama conjugada à Normal.

3.7. Inferência Bayesiana, Inferência freqüentista e Inferência verossimilhança

A inferência estatística tem evoluído muito nas últimas décadas, principalmente em função do aumento dos recursos computacionais. Desde o início do desenvolvimento da estatística até recentemente, houve grande supremacia da **inferência freqüentista**, a qual, em geral, tem sido creditada como Fisheriana e se apóia, principalmente, na consistência assintótica através da qual, um modelo estimado converge para o verdadeiro modelo. A propriedade de se incorporar a informação a priori na análise, sob o enfoque Bayesiano, sempre gerou muita controvérsia, em determinada fase do desenvolvimento da estatística (Efron, 1978; Lindley, 1978; Smith, 1984). Atualmente, este tipo de discussão parece não ser mais relevante, com as diferenças entre as

duas abordagens bem compreendidas e as virtudes de cada uma delas utilizadas quando mais conveniente (Gamerman, 1996).

Atualmente, uma das marcantes características da inferência estatística moderna é o papel central da função de verossimilhança, a qual foi introduzida por Fisher (1922). Segundo Lindsey (1999), poucos estatísticos modernos percebem que Fisher nunca foi um freqüentista para inferência (embora ele tenha geralmente usado uma interpretação freqüentista de probabilidade), sendo muito mais próximo aos Bayesianos, tais quais Jeffreys, do que da escola de Neyman-Pearson (os verdadeiros freqüentistas, que introduziram a teoria dos testes de hipóteses entre 1936 e 1938). Segundo tal autor, a inferência estatística no sentido Fisheriano enfatiza a obtenção do máximo de informação, dado o conjunto de observações, sem incorporação de conhecimento e informação a priori, exceto aquelas necessárias para a construção de modelos (o princípio da verossimilhança baseia-se na suposição de que o modelo ou a função são verdadeiros e que somente os valores corretos dos parâmetros necessitam ser determinados). Portanto, a inferência Fisheriana refere-se essencialmente à **inferência verossimilhança**.

O uso da função de verossimilhança como um meio direto de fazer inferência, conforme sugerido por Fisher, é um enfoque recente em estatística (Edwards, 1972; Bardorff-Nielsen, 1976; Gomes, 1981; Pereira, 1997; Lindsey, 1999), caracterizando a **inferência verossimilhança**.

O artigo de Pereira (1997) caracteriza muito bem os pontos fortes da **inferência verossimilhança**, tais quais: (i) a função de verossimilhança fornece uma medida exata da incerteza, resumindo toda a informação que os dados fornecem sobre um parâmetro desconhecido θ ; (ii) a verossimilhança é uma medida relativa, de forma que não é possível obter uma medida de plausibilidade absoluta e por isto, somente as razões de verossimilhança são relevantes; (iii)

a função de verossimilhança relativa, dada por $R(\theta; y) = \frac{\theta^y (1-\theta)^{n-y} n^n}{y^n (n-y)^{n-y}}$ pode

ser utilizada para obtenção de níveis de significância aproximados, através dos testes da razão de verossimilhança; (iv) existe uma relação entre a função de verossimilhança e os métodos convencionais de teoria de grandes amostras, ou seja, além de fornecer uma medida exata de plausibilidade, a função de verossimilhança dá uma indicação de aplicabilidade de resultados assintóticos em casos particulares; (v) os métodos de verossimilhança são exatos tanto para pequenas quanto para grandes amostras, não dependendo de propriedades assintóticas das várias distribuições amostrais.

Assim, em **inferência Bayesiana**, certos métodos que assumem distribuições a priori não informativas, são essencialmente de **inferência verossimilhança**, tais como o método VEIL (ou da verossimilhança integrada de

Gianola & Foulley, 1990) de estimação de componentes de variância, os quais mantêm a propriedade de conduzir a análise exata de amostra de tamanho finito.

4. Fundamentos Bayesianos da predição de valores genéticos

Os fundamentos básicos da predição de valores genéticos são essencialmente de natureza Bayesiana, conforme verificado inicialmente por Robertson (1955) e Dempfle (1977) e mostrado a seguir, considerando a distribuição a priori, normal, conjugada à posteriori normal.

Os efeitos genéticos aditivos são definidos como desvios e uma população de efeitos genéticos apresenta zero como média e σ_a^2 como variância. A melhor predição do efeito genético de um indivíduo sem nenhuma informação, tomado aleatoriamente da população é a média populacional μ , a qual pode ser tomada como o estimador a priori, cuja variância é σ_a^2 . Tomando uma informação y do indivíduo, um segundo (dado observado) estimador do efeito genético é o desvio $(y - \mu)$ fenotípico em relação à média populacional, o qual possui variância $\sigma_e^2 = \sigma_y^2 - \sigma_a^2$. Estes dois estimadores independentes podem ser combinados linearmente da melhor maneira possível, tomando as recíprocas das respectivas variâncias como pesos. Sob o enfoque Bayesiano, a esperança da distribuição a posteriori corresponde à média ponderada pela precisão, das médias da priori e da verossimilhança (tópico 3.1.). Assim, tem-se:

$$\hat{a} = \left(\frac{\text{estimador a priori}}{\sigma_a^2} + \frac{\text{estimador informação observada}}{\sigma_y^2 - \sigma_a^2} \right) / \left(\frac{1}{\sigma_a^2} + \frac{1}{\sigma_y^2 - \sigma_a^2} \right) = \left(\frac{0}{\sigma_a^2} + \frac{Y - \mu}{\sigma_y^2 - \sigma_a^2} \right) / \left(\frac{1}{\sigma_a^2} + \frac{1}{\sigma_y^2 - \sigma_a^2} \right) = h^2 (y - \mu),$$

como no enfoque freqüentista, onde h^2 é a herdabilidade. Este é também (considerando y como a média dos indivíduos selecionados) o estimador para o ganho genético para a seleção massal, conforme apresentado e derivado da mesma forma por Lerner (1958).

No melhoramento, duas fontes independentes de informação têm importante aplicação prática, mesmo quando se tem apenas o estimador a priori. Por exemplo, quando se deseja selecionar 30 indivíduos para reprodução, mas a partir de um teste de apenas 40 indivíduos tomados aleatoriamente da população. Neste caso, o procedimento mais eficiente é selecionar os 20 melhores (que provavelmente estarão acima da média populacional) identificados

com base no teste e tomar os outros 10 (não testados) aleatoriamente da população. Espera-se que, dentre os 10, em torno de 5 estejam acima da média populacional. Robertson (1955) demonstrou claramente a utilidade de duas fontes de informação independentes para derivar várias fórmulas de predição.

5. Estimação Bayesiana de índices de seleção unicaracterísticos

A predição de valores genéticos de genitores com base em testes de progênies é derivada a seguir, com base em Gianola & Fernando (1986), visando elucidar a relação entre índice de seleção e estimadores Bayesianos.

Considere uma amostra de 5 genitores não aparentados, tomados aleatoriamente de uma população de genitores e suas capacidades de transmissão (metade de seus valores genéticos) representados pelo vetor $\underline{g} = [S_1, \dots, S_5]$, distribuídos a priori como $N \sim (g, \sigma_g^2)$. Logicamente, neste caso, o interesse reside nas inferências sobre \underline{g} , dados os resultados do teste de progênies. O modelo para a média da progênie do genitor i é: $\bar{y}_i = \mu + S_i + \bar{e}_i$; $i=1, \dots, S$, onde μ é uma constante conhecida e $\bar{e}_i \sim N(0, \sigma_e^2/n_i)$ é a média dos resíduos associados com as observações da progênie n_i do genitor i . Uma vez que μ é assumido como conhecido, pode-se escrever: $w_i = \bar{y}_i - \mu = S_i + \bar{e}_i$.

$$E(w_i|S_i) = S_i \text{ e } Var(w_i|S_i) = \sigma_e^2/n_i.$$

Uma vez que os genitores não são aparentados, a densidade a priori é dada por:

$$f(\theta) = \prod_{i=1}^S \frac{1}{\sqrt{2\pi\sigma_g^2}} \exp\left\{-\frac{(S_i - g)^2}{2\sigma_g^2}\right\}$$

A função de verossimilhança, com $Cov(\bar{e}_i, \bar{e}_j) = 0, i \neq j$, é

$$f(w_1, \dots, w_S | \theta) =$$

$$\prod_{i=1}^S \sqrt{\frac{n_i}{2\pi\sigma_e^2}} \exp\left\{-\frac{n_i(w_i - S_i)^2}{2\sigma_e^2}\right\}.$$

Com a suposição de independência

$$f(S_i|w_1, \dots, w_S) = f(S_i|w_i); \quad i = 1, \dots, S$$

A densidade a posteriori da capacidade de transmissão do i -ésimo genitor é então:

$$f(S_i|w_i) \propto \exp \left[-\frac{1}{2} \left(\frac{\sigma_s^2 + \sigma_e^2/n_i}{\sigma_s^2 \sigma_e^2/n_i} \right) X \left(S_i - \frac{w_i \sigma_s^2 + g \sigma_e^2/n_i}{\sigma_s^2 + \sigma_e^2/n_i} \right)^2 \right].$$

Este resultado equívale a uma densidade normal com média:

$$E(S_i|w_i) = \frac{w_i \sigma_s^2 + g \sigma_e^2/n_i}{\sigma_s^2 + \sigma_e^2/n_i} = \frac{g (\sigma_s^2)^{-1} + w_i (\sigma_e^2/n_i)^{-1}}{(\sigma_s^2)^{-1} + (\sigma_e^2/n_i)^{-1}}.$$

Observa-se que a média a posteriori corresponde à média ponderada da média a priori (g) e da média da amostra (w_i), com os pesos sendo as recíprocas de suas variâncias.

Denominando $a = (\sigma_s^2)^{-1} / [(\sigma_s^2)^{-1} + (\sigma_e^2/n_i)^{-1}]$, tem-se:

$$\begin{aligned} E(S_i|w_i) &= ag + (1-a) w_i \\ &= g + (1-a)(w_i - g) = g + (1-a)(\bar{y}_i - \mu - g), \end{aligned}$$

que é a média condicional de uma distribuição normal bivariada.

Sendo $(1-a) = n_i / (n_i + \frac{\sigma_e^2}{\sigma_s^2})$, e sendo $g = 0$, neste caso tem-se:

$$E(S_i|w_i) = \frac{n_i}{n_i + \frac{\sigma_e^2}{\sigma_s^2}} (\bar{y}_i - \mu) = \frac{n_i}{n_i + (4-h^2)/h^2} (\bar{Y}_i - \mu),$$

que equívale ao clássico índice de seleção para avaliação da capacidade de transmissão de um touro via teste de progênies (Vleck et al., 1987; Vleck, 1993).

6. Relação entre BLUP e estimadores Bayesianos

Duas derivações da metodologia BLUP sob o enfoque Bayesiano são apresentadas a seguir, uma delas com base em Ronningen (1971) e Dempfle (1977, 1989) e outra com base em Robertson (1955).

Para ambas as derivações tem-se o seguinte modelo e suposições, assumindo G e R conhecidos:

$$y = Xb + Za + e = Wt + e; \quad W = [X \ Z]; \quad t' = [b' \ a'];$$

$$E(a) = 0; \quad E(e) = 0; \quad E(y) = Xb;$$

$$\text{Var}(a) = G; \quad \text{var}(e) = R; \quad \text{cov}(a, e) = 0; \quad \text{var}(y) = ZGZ' + R.$$

A primeira derivação (Dempfle, 1977) usa o teorema de Bayes e assume-se que adicionalmente $t \sim N(r, M)$; $e \sim N(0, R)$; $y \sim N(Xb, WMW' + R)$, onde:

$$E(t) = E \begin{pmatrix} \beta \\ a \end{pmatrix} = r = \begin{pmatrix} r_1 \\ 0 \end{pmatrix} \quad e \quad \text{var}(t) = M = \begin{pmatrix} S & 0 \\ 0 & G \end{pmatrix}$$

Dado $t, y \sim N(Wt, R)$, e dado $r, t \sim N(r, M)$. A verossimilhança de y dado t é

$L(t; y) = k_1 \exp[-(y - Wt)'R^{-1}(y - Wt)/2] = f(y|t)$, e a densidade da distribuição a priori de t é

$$f(t) = k_2 \exp[-(t - r)'M^{-1}(t - r)/2].$$

Assim, a distribuição a posteriori de t é proporcional a

$$f(t|y) \propto L(t; y) f(t) \propto k_3 \exp\{-(y - Wt)'R^{-1}(y - Wt) + (t - r)'M^{-1}(t - r)/2\}.$$

Esta equação pode ser expressa também em termos da densidade conjunta a posteriori dada por:

$$\begin{aligned} f(a, b) \propto & \exp\{-\frac{1}{2}(y - Xb - Za)'R^{-1}(y - Xb - Za)\} \\ & \times \exp\{-\frac{1}{2}(b - r_1)'S^{-1}(b - r_1)\} \\ & \times \exp\{-\frac{1}{2}(a - E(a))'G^{-1}(a - E(a))\}. \end{aligned}$$

Uma vez que esta distribuição é simétrica e unimodal (normal), a moda, a mediana e a média são idênticas e uma grande classe de funções de perda comum (função de perda quadrática, função de perda absoluta ou função de perda uniforme) conduz ao mesmo estimador. Determinando a moda, obtém-se o vetor médio da distribuição conjunta a posteriori, por maximização e não integração. Diferenciando a expressão com respeito a t (b e a) e igualando a zero, obtém-se:

$$[W'R^{-1}W + M^{-1}]\hat{t} = W'R^{-1}y + M^{-1}r, \quad \text{onde } \hat{t} = E[t|y].$$

Este sistema é equivalente a:

$$\begin{bmatrix} X'R^{-1}X + S^{-1} & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} E(\beta|y) \\ E(a|y) \end{bmatrix} = \begin{bmatrix} X'R^{-1}y + S^{-1}r_1 \\ Z'R^{-1}y + G^{-1}0 \end{bmatrix} \quad (5)$$

onde $r_1 = E(b)$ e $0 = E(a)$.

Tomando a distribuição a priori sobre os efeitos fixos como não informativa (expressa como $S \rightarrow \infty$ e então $S^{-1} \rightarrow 0$), tem-se que esta equação resultante equívale às equações do modelo misto (EMM):

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

A equação para \hat{t} , pode ser derivada também pela maximização de $f(y, t)$ para variações em t , sendo o estimador, neste caso, denominado máximo a posteriori (MAP) (Henderson, 1982a, 1984).

A segunda derivação baseia-se em Robertson (1955) e fundamenta-se na combinação de dois estimadores (fontes de informação) independentes:

(i) o estimador BLUE ou de mínimos quadrados generalizados de t

$$\hat{t}_1 = (W'R^{-1}W)^{-1} W'R^{-1}y \quad e$$

(ii) o estimador a priori de t

$$\hat{t}_2 = E(t) = r.$$

Estes estimadores apresentam as matrizes de variância-covariância

$$\text{var}(\hat{t}_1) = (W'R^{-1}W)^{-1} M \quad e \quad \text{var}(\hat{t}_2) = 0,$$

com variâncias do erro de estimação

$$\text{var}(\hat{t}_1 - t) = (W'R^{-1}W)^{-1} \quad e \quad \text{var}(\hat{t}_2 - t) = M.$$

Combinando-se os dois estimadores obtém-se:

$$\hat{t}_c = [\text{var}(\hat{t}_1 - t)^{-1} + \text{var}(\hat{t}_2 - t)^{-1}]^{-1} \{ \text{var}(\hat{t}_1 - t)^{-1} \hat{t}_1 + \text{var}(\hat{t}_2 - t)^{-1} \hat{t}_2 \}, \quad \text{que equívale a} \\ [W'R^{-1}W + M^{-1}] \hat{t}_c = W'R^{-1}y + M^{-1}r.$$

Fazendo as mesmas suposições sobre os efeitos fixos, esta expressão eqüivale às equações de modelo misto. As equações de modelo misto apresentadas em (5) são denominadas equações de modelo misto de Robertson.

7. Relação entre estimadores de máxima verossimilhança e estimadores Bayesianos

Assumindo como uniforme a distribuição a priori dos parâmetros a serem estimados e maximizando (obtendo a moda) a distribuição a posteriori, o estimador resultante é eqüivalente ao de máxima verossimilhança – ML (Henderson, 1984; Gianola & Fernando, 1986). De fato maximizando $f(a, b)$ (tópico anterior) (mas considerando uma priori não informativa para b) com respeito a a e b obtém-se um estimador denominado de máxima verossimilhança, por Henderson et al. (1959), embora $f(a, b)$ (tópico anterior) não seja uma função de verossimilhança e sim uma densidade a posteriori. Mesmo assim, pode ser obtido a partir das EMM que

$$E(\beta | y) = (X'V^{-1}X)^{-1} X'V^{-1}y = \hat{\beta} \text{ e}$$

$$E(a|y) = GZ'V^{-1} [y - X\beta] = \hat{a}$$

$\hat{\beta}$ é um estimador GLS e também ML de b e \hat{a} é um estimador ML de $E(a|b, y)$, eqüivalendo à média da distribuição condicional na qual b é fixado.

8. Estimação Bayesiana de índices de seleção multicaracterísticos

Com a técnica dos índices de seleção multicaracterísticos deseja-se prever um vetor não observável $T = A\alpha_1$ (denominado agregado genotípico) em que A e α_1 referem-se a uma matriz conhecida, contendo os pesos econômicos relativos dos diferentes caracteres e um vetor de valores genéticos, respectivamente.

O vetor de dados é dado por:

$$y = \alpha_2 + Z\alpha_1 + e, \text{ onde:}$$

α_2 - vetor de constantes especificando a locação da distribuição marginal de y

Z - matriz de incidência para α_1

Para este modelo tem-se:

$$\begin{bmatrix} \theta_1 \\ e \\ y \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ \theta_2 \end{bmatrix}, \begin{bmatrix} G & 0 & GZ' \\ 0 & R & R \\ ZG & R & V \end{bmatrix} \right), \text{ onde } V = ZGZ' + R \text{ e } G \text{ e } R \text{ são conhecidas.}$$

A densidade a posteriori de α_1 é $f(\alpha_1 | y)$ a $f(y | \alpha_1) \cdot f(\alpha_1)$, a qual tem distribuição normal pois $f(y | \alpha_1)$ e $f(\alpha_1)$ também tem distribuição normal. Esta posteriori equívale a:

$$f(\alpha_1 | y) \propto \exp \left\{ -\frac{1}{2} (y - \alpha_2 - Z\alpha_1)' R^{-1} (y - \alpha_2 - Z\alpha_1) \right\} \times \\ \times \exp \left\{ -\frac{1}{2} (\alpha_1 - 0)' G^{-1} (\alpha_1 - 0) \right\}, \text{ pois } E(\alpha_1) = 0$$

Uma vez que esta distribuição é simétrica e unimodal, a média e a moda são a mesma, de forma que a distribuição pode ser diferenciada com respeito a α_1 e igualada a zero para se obter a média a posteriori. Fazendo-se isto obtém-se

$$Z'R^{-1}Z + G^{-1} E(\alpha_1 | y) = [Z'R^{-1}(y - \alpha_2) + G^{-1}0]$$

Resolvendo-se esta equação para $E(\alpha_1 | y)$ obtém-se

$$E(\alpha_1 | y) = (Z'R^{-1}Z + G^{-1})^{-1} [Z'R^{-1}(y - \alpha_2) + G^{-1}0]$$

que equívale à equação para o BLP ou índice de seleção associado a α_1 . Dessa forma, necessita-se ainda incluir as informações econômicas para se obter T.

Uma vez que T é linear em α_1 , a posteriori T é também normal com vetor médio $E(T | y) = A E(\alpha_1 | y)$.

Obtém-se, assim, o estimador Bayesiano de T, bastando para isto, encontrar a média da distribuição a posteriori de $\alpha_1 | y$ e posteriormente incorporar as informações econômicas (A).

9. Estimação Bayesiana de componentes de variância e relação com ML e REML

No contexto dos modelos lineares mistos, os valores genéticos (α_1) são preditos simultaneamente à estimação dos efeitos fixos (α_2) e dos componentes de variância (α_3). Na abordagem Bayesiana, a avaliação genética pode ser obtida, de maneira geral, pela construção da densidade a posteriori $f(\alpha_1, \alpha_2, \alpha_3 | y)$ e, se necessário, pela integração de $f(\alpha_1, \alpha_2, \alpha_3 | y)$ em relação a α_2 e α_3 . Estes (α_2 e α_3) são denominados parâmetros de "nuisance" e por isso devem ser integrados fora, exceto α_2 em alguns casos, onde o mesmo constitui-se em uma parte integrante da função de mérito total (neste caso, a função de mérito depende da combinação linear de α_1 e α_2).

A obtenção de α_1 requer a integração ou o conhecimento de α_2 e α_3 . Henderson (1973) propôs o método BLUP para situações em que α_3 é conhecido e α_2 não o é. Para situações em que α_3 não é conhecido, este autor sugeriu que o procedimento de máxima verossimilhança (ML) propiciaria estimativas razoáveis. Conforme Gianola & Fernando (1986), argumentos Bayesianos, que não requerem normalidade e linearidade permitem validar a intuição de Henderson.

A distribuição de α_1 , α_2 , e α_3 , dado y é proporcional a

$$f(\alpha_1, \alpha_2, \alpha_3 | y) \propto f(y | \alpha_1, \alpha_2, \alpha_3) \cdot f(\alpha_1, \alpha_2, \alpha_3)$$

Concentrando o interesse em α_1 (o vetor de valores genéticos), deve-se integrar α_2 e α_3 através de

$$f(\theta_1 | y) = \int_{R_{\theta_2}} \int_{R_{\theta_3}} f(\theta_1 | \theta_2, \theta_3, y) \cdot f(\theta_2, \theta_3 | y) \cdot d\theta_2 d\theta_3$$

Tomando a distribuição conjunta a posteriori de forma que a maioria da densidade esteja na moda $(\hat{\theta}_2, \hat{\theta}_3)$, tem-se:

$$f(\theta_1 | y) \doteq f(\theta_1 | \theta_2 = \hat{\theta}_2, \theta_3 = \hat{\theta}_3, y).$$

Usando prioris não informativas para α_2 e α_3 , tem-se que $\hat{\theta}_2$ e $\hat{\theta}_3$ são precisamente estimadores ML de α_2 e α_3 , pois neste caso $f(\alpha_2, \alpha_3 | y) \propto f(y | \alpha_2, \alpha_3)$, ou seja a densidade de α_2 e α_3 dado y é proporcional à função de verossimilhança, de forma que a moda da posteriori conjunta corresponde ao máximo da função de verossimilhança, produzindo estimadores ML.

Uma abordagem alternativa para inferência sobre α_1 consiste em obter

$$f(\theta_1, \theta_2 | y) \doteq f(\theta_1, \theta_2 | \theta_3 = \hat{\theta}_3, y), \text{ onde } \hat{\theta}_3 \text{ refere-se à moda da}$$

densidade marginal de α_3 , dado y . Para obtenção de $\hat{\theta}_3$ deve-se integrar α_2 em $f(\theta_2, \theta_3 | y) \propto f(y | \theta_2, \theta_3)$ e então maximizar $f(\alpha_3 | y)$. Usando-se uma

priori não informativa para α_3 , sob normalidade $\hat{\theta}_3$ é um estimador de máxima verossimilhança restrita (REML) para α_3 (Harville, 1977). Assim, se o interesse reside na inferência conjunta para α_1 e α_2 basta usar

$$f(\theta_1, \theta_2 | y) \doteq f(\theta_1, \theta_2 | \theta_3 = \hat{\theta}_3, y), \text{ que sob normalidade é equivalente a solução das equações de modelo misto com } \alpha_3 \text{ substituído pelas estimativas REML de } \alpha_3 \text{ (desde que se tenha usado prioris não informativas para } \alpha_2 \text{ e } \alpha_3).$$

Inferências sobre componentes de variância devem ser baseadas em $f(\alpha_3 | y)$ a $f(y | \alpha_3) \cdot f(\alpha_3)$, onde α_3 contém variâncias e, portanto, $f(\alpha_3 | y)$ é definida na amplitude $(0, \infty)$ para cada um dos elementos de α_3 , de forma que nunca surgem problemas de estimativas negativas de componentes de variância (Box & Tiao, 1973). $f(\alpha_3 | y)$ é obtida integrando-se α_1 em $f(\alpha_1, \alpha_2, \alpha_3 | y)$ produzindo $f(\alpha_2, \alpha_3 | y)$ e integrando-se α_2 nesta última. Neste caso, $f(\alpha_2, \alpha_3 | y)$ conduz aos estimadores ML de α_2 e α_3 e $f(\alpha_3 | y)$ conduz a um estimador REML de α_3 . Segundo Gianola & Fernando (1986), isto (eliminação das influências de α_2 ou dos efeitos fixos) mostra precisamente porque REML deve ser preferido em relação a ML, ou seja, estes argumentos são mais fortes do que os apresentados por Patterson & Thompson (1971), que enfatizaram a propriedade de vício do ML.

10. Estimadores melhorados para a média de tratamentos

A análise Bayesiana pode ser útil também no contexto da análise de experimentos de variados tipos e objetivos na pesquisa agropecuária.

Considerando o delineamento experimental de blocos ao acaso com t tratamentos, b blocos e n plantas por parcela (plantas obtidas aleatoriamente de uma população), tem-se o modelo matemático:

$Y_{ij} = \mu + t_i + b_j + e_{ij}$, onde μ , t_i , b_j e e_{ij} referem-se aos efeitos da média geral, de tratamentos, de blocos e do resíduo, respectivamente. Nas análises clássicas, os efeitos t_i e e_{ij} podem ser assumidos como aleatórios e b_j como fixo, visto que estas suposições não alteram os resultados da análise de variância e das comparações de médias, conforme Steel & Torrie (1980).

Aplicando o mesmo princípio (regressão) da predição de valores genéticos para a predição da média de tratamentos, tem-se que a fração de $(\bar{Y}_i - \bar{Y}_{...})$ (efeito de tratamento) que realmente determina a média do tratamento é $R^2(\bar{Y}_i - \bar{Y}_{...})$, em que:

$R^2 = \sigma_t^2 / (\sigma_e^2 / b + \sigma_t^2) = 1 - 1/F$, onde F é o valor do teste F de Snedecor da análise de variância.

Assim, a determinação (R^2) associada ao efeito de tratamento pode ser generalizada para efeitos fixos ou aleatórios de tratamentos, pois depende apenas de F que é idêntico para ambos os modelos. A média (valor esperado em uma nova repetição do ensaio ou em uma aplicação do tratamento ao nível operacional) verdadeira do tratamento (M) é, então, melhor predita, fazendo-se $M = (1 - 1/F)(\bar{Y}_{i..} - \bar{Y}_{...}) + \bar{Y}_{...} = (1 - 1/F) \bar{Y}_{i..} + (1/F) \bar{Y}_{...}$. Em outras palavras, a

melhor predição da média é dada pela fração $(1-1/F)$ da média aritmética do tratamento mais a fração $(1/F)$ da média geral. Esta expressão para a média, além de ser mais coerente com a prática, considera adequadamente a precisão experimental na estimação da média de tratamento, pois quanto mais preciso ($> F$) o experimento, menor é a correção $(-1/F)$ aplicada sobre as médias de tratamento. Neste caso (agora supondo fixos os efeitos de tratamentos), níveis de significância diferentes (por exemplo 5% e 1%) conduzirão a médias ajustadas diferentes, embora na metodologia convencional conduzissem a inferência e médias idênticas.

Do ponto de vista da estatística freqüentista, a média amostral é o melhor estimador não viciado (BLUE) da média populacional. O estimador apresentado, que apresenta um regressor (no caso $1-1/F$) da média amostral do tratamento em relação à média geral é viciado (Gianola & Fernando, 1986). Entretanto, existem estimadores e preditores viciados que propiciam menor erro quadrático médio do que os estimadores BLUE (Henderson, 1984), sendo assim vantajosos em relação aos estimadores não viciados (Efron, 1975).

Os três principais procedimentos de estimação caracterizam os estimadores: melhor estimador linear não viciado (BLUE); estimador de mínimo erro quadrático médio e estimador Bayesiano. Esses três estimadores são similares e idênticos sob certas circunstâncias (Henderson, 1984).

A utilização e reconhecimento das vantagens dos estimadores viciados iniciou-se com o trabalho de Stein (1955) que constituiu um verdadeiro paradoxo na Estatística. Stein demonstrou que a média aritmética é estimador não admissível, isto é, que existem estimadores que propiciam menor erro quadrático médio ou menor risco que a média aritmética. Neste contexto, James & Stein (1961), apresentaram um estimador melhorado para a média populacional, o qual é dado por $M^* = k (\bar{Y}_{i...} - \bar{Y}_{...}) + \bar{Y}_{...}$, onde k é um fator regressor da média amostral de determinado tratamento sobre a média geral. Constata-se assim, a similaridade entre M e M^* .

O fator k apresentado por James & Stein é dado por

$$K = 1 - \frac{(t-3) \sigma^2}{\sum (\bar{Y}_{i...} - \bar{Y}_{...})^2}, \text{ onde } t \text{ refere-se ao número de tratamentos e } \sigma^2$$

refere-se à variância residual. Notando a similaridade entre $(t-3)$ e o número de graus de liberdade $(t-1)$ para tratamentos, verifica-se também a grande similaridade entre k e $(1-1/F)$. Stein (1955) demonstrou que a média aritmética é um estimador admissível apenas quando existem apenas uma ou duas médias a serem estimadas, de forma que a correção somente torna-se necessária quando três ou mais tratamentos forem considerados, conforme intuitivamente

esperado. O estimador melhorado apresentado por James & Stein (1961), não necessita de qualquer suposição referente aos efeitos fixos ou aleatórios ou distribuições das médias a serem estimadas (Efron & Morris, 1977). Os estimadores de mínimo erro quadrático médio têm recebido mais atenção recentemente (Bibby & Toutenburg, 1977).

O procedimento de estimação Bayesiana é mais antigo (1763) do que o método de Stein e também minimiza o erro quadrático esperado, de forma que o estimador de James-Stein é muito similar ao estimador de Bayes, tornando-se idênticos para grande número de tratamentos (Efron & Morris, 1977). Em inferência Bayesiana não existe qualquer distinção entre efeitos fixos ou aleatórios, sendo que os parâmetros a serem estimados são considerados variáveis aleatórias (Gianola & Fernando, 1986) que devem ser estimadas considerando as incertezas associadas a elas. No caso de inferências sobre médias populacionais de tratamentos, sob o enfoque Bayesiano, Box & Tiao (1973) apresentam como regressor exatamente a quantidade $(1 - 1/F)$.

Com base no exposto, recomenda-se, para inferências (com menor risco) práticas na área de pesquisa agropecuária, a apresentação das médias de tratamento como $M = (1-1/F) \bar{Y}_{i..} + (1/F)\bar{Y}_{...}$, onde F , $\bar{Y}_{i..}$ e $\bar{Y}_{...}$ referem-se ao valor de F para tratamentos na análise de variância, média de tratamentos e média geral, respectivamente.

11. Problemas da abordagem clássica de predição de valores genéticos e vantagens da abordagem Bayesiana

No procedimento de modelos mistos para avaliação genética, torna-se necessária a especificação de propriedades das distribuições dos vetores \underline{a} e \underline{e} , principalmente em termos de seus primeiro e segundo momentos. Para derivação do BLUP não é necessário assumir uma forma de distribuição, devendo-se assumir apenas que:

$$E(\underline{a}) = \underline{0}; \quad E(\underline{e}) = \underline{0}; \quad \text{Var}(\underline{a}) = \underline{G}; \quad \text{Var}(\underline{e}) = \underline{R}; \quad \text{COV}(\underline{a}, \underline{e}) = \underline{0}$$

Devido a isto tem-se:

$$E(\underline{y}) = \underline{X}\underline{b} \quad \text{Var}(\underline{y}) = \underline{Z}\underline{G}\underline{Z}' + \underline{R} = \underline{V}$$

Estas pressuposições são válidas em ausência de seleção. Em presença de seleção, $E(\underline{a})$ e $\text{COV}(\underline{a}, \underline{e})$ podem ser diferentes de zero bem como \underline{G} e \underline{R} podem ser alteradas em relação a seus valores originais para uma população não selecionada e não endógama.

Como propriedades desejáveis do BLUP, podem ser ressaltadas (Henderson, 1982a; Ronningen & Vleck, 1985):

- (i) é um preditor não viciado
- (ii) minimiza a variância do erro de predição dentro da classe das predições lineares não viciadas
- (iii) minimiza o quadrado médio dos erros de predição dentro da classe das predições lineares invariantes a translação (não afetadas por mudanças nos efeitos fixos)
- (iv) maximiza a correlação entre \hat{a}_i e a_i dentro da classe das predições lineares não viciadas.

Assumindo distribuição normal multivariada para a e e , propriedades adicionais interessantes são asseguradas (Henderson, 1982a).

- (i) se os indivíduos candidatos à seleção apresentam a mesma quantidade de informação, a seleção maximiza a média esperada dos valores genéticos dos indivíduos selecionados $[E(a|y)]$. $(E(a|y))$ foi denominado "*Best Predictor*" por Henderson, 1973).
- (ii) \hat{a} é estimador ML de $E(a|y)$
- (iii) \hat{a} é estimador MAP (máximo a posteriori) de a
- (iv) $\hat{\beta}$ e \hat{a} são estimativas Bayesianas, tomando-se como função de perda os desvios absolutos ou quadráticos e como estimativas a priori $E(a) = 0$ e $\text{Var}(a) = S$, com $S^{-1} \geq 0$.
- (v) As soluções BLUP \hat{a} e \hat{e} , podem ser usadas em algoritmos eficientes para estimação de componentes de variância por ML e REML, os quais permitem ao BLUP, controlar vícios devidos a certos tipos de seleção.

Com relação aos efeitos da seleção na predição de valores genéticos, Henderson (1975, 1982b) mostrou (assumindo normalidade) que se os componentes de variância da população não selecionada são conhecidos, as equações de modelo misto ignorando a seleção ainda conduzem a predição BLUP sob seleção, desde que o critério de seleção seja função linear das observações e de translação invariante (ou seja, que o critério de seleção não envolva os efeitos fixos). Entretanto, predições envolvendo efeitos fixos baseadas em um modelo que ignora a seleção, não são BLUP sob seleção. Fries & Schenckel (1993) abordam com detalhes esta questão.

Torna-se claro então, que BLUP sob um modelo de seleção requer, pelo menos, o conhecimento das variâncias antes da seleção. Estimativas de componentes de variância pelos métodos ML e REML são livres de algumas formas de vício devidos à seleção (Thompson, 1979; 1982; Gianola et al., 1989).

Assumir distribuição normal multivariada de y , a e e , implica assumir que os caracteres sejam determinados por um infinito número de locos não ligados de pequenos efeitos aditivos (modelo infinitesimal) (Mrode, 1996). Com o modelo infinitesimal, alterações na variância genética devidas à seleção, tais como desequilíbrio gamético (covariância negativa entre frequências alélicas em diferentes locos) ou mudanças devidas a endogamia e deriva genética são consideradas nas equações de modelo misto através da inclusão da matriz de parentesco completa (Sorensen & Kennedy, 1983). Alterações na variância devidas a cruzamentos preferenciais, também são consideradas da mesma forma. Considerando o modelo infinitesimal, os componentes de variância podem ser estimados de forma não viciada por REML, desde que os dados utilizados incluam informações sobre as quais a seleção foi baseada.

Dentre as propriedades desejáveis do BLUP, duas estão associadas à propriedade “Best”: minimização do quadrado médio do erro de predição e maximização da correlação entre os valores genéticos verdadeiros e os preditos. A terceira refere-se à propriedade de não vício [$E(a|\hat{a}) = \hat{a}$]. Estas três propriedades são interessantes do ponto de vista estatístico, entretanto, somente a propriedade de maximização da correlação entre a e \hat{a} apresenta uma conexão direta com o problema do melhoramento por seleção. Uma vez que o objetivo das funções dos dados é ordenar os candidatos à seleção e praticar a seleção de forma a se maximizar o ganho genético com seleção, as propriedades de mínima variância e normalidade assintótica talvez sejam irrelevantes (Gianola & Fernando, 1986; Gianola, 1995). Esta assertiva pode ser confirmada considerando que com dados desbalanceados, independentemente da distribuição, o ordenamento dos candidatos com base em $E(a|y)$ e a seleção daqueles com os maiores valores, maximiza a média dos indivíduos selecionados, conforme demonstrado por Bulmer (1980); Goffinet (1983) e Fernando & Gianola (1986).

A seguir serão relatados alguns pontos desfavoráveis da abordagem clássica e vantagens da inferência Bayesiana para estimação. As propriedades desejáveis dos estimadores pontuais e intervalares na estatística clássica são baseadas em hipotéticas repetições do experimento em um número infinito de vezes e não considera apenas os dados disponíveis. Isto significa que, na teoria de amostragem, η é uma quantidade conhecida fixa e o intervalo de confiança é aleatório. Assim, sob hipotética repetição de y , espera-se que o intervalo contenha η em uma certa proporção das amostras (repetições). Entretanto,

tendo-se os dados observados (uma repetição do vetor y), α_1 estará dentro ou fora do intervalo. Em análise Bayesiana, inferências são realizadas diretamente da distribuição a posteriori, baseando-se apenas nos dados disponíveis e na distribuição a priori. A partir da posteriori, estimativas pontuais podem ser obtidas para resumir as características de tal distribuição e também estimativas intervalares e inferências probabilísticas podem ser realizadas. Por exemplo, no caso da seleção de genitores com base em testes de progênie, pode-se fazer inferências probabilísticas sobre a capacidade de transmissão do genitor

o dado w_i , fazendo-se $Z = \frac{S_i - E(S_i | w_i)}{\sqrt{Var(S_i | w_i)}} \sim N(0,1)$ e computar para quaisquer

valores fixados q_1 e q_2 , a probabilidade $P\{q_1 < S_i < q_2 | w_i\} = F(z_2) - F(z_1)$, onde F é a função distribuição acumulada da normal padrão.

Os procedimentos freqüentistas de estimação, ML e REML, apresentam propriedades desejáveis que são bem definidas somente para amostras muito grandes, ou seja, somente apresentam justificativas assintóticas. Entretanto, a maioria dos programas de melhoramento de espécies perenes baseia-se em amostras de tamanho finito (pequeno). O teorema de Bayes propicia soluções precisas para o problema de amostras de tamanho finito, pois para cada conjunto de dados, pequeno ou grande, existe uma distribuição a posteriori exata para realização de inferências. Existe também uma teoria assintótica Bayesiana, de forma que com grandes amostras obtém-se resultados similares aos obtidos com o método ML ou REML.

Em muitos casos, a distribuição dos dados (y) não depende exclusivamente do grupo de parâmetros de interesse (α_1 , por exemplo), mas depende também de outros parâmetros incidentais denominados parâmetros de "nuisance" (α_2 , por exemplo). Neste caso, a densidade conjunta a posteriori possui a forma $f(\alpha_1, \alpha_2 | y) \propto f(y | \alpha_1, \alpha_2) f(\alpha_1, \alpha_2)$, onde $f(\alpha_1, \alpha_2)$ é a densidade conjunta a priori de α_1 e α_2 . Uma vez que o interesse concentra-se em inferências sobre α_1 na ausência de α_2 , os parâmetros "nuisance" podem ser eliminados integrando-se a densidade conjunta a posteriori com respeito a α_2 . Assim, tem-se:

$$\begin{aligned} f(\theta | y) &= \int_{R_{\theta_2}} f(\theta_1, \theta_2 | y) d\theta_2 \\ &= \int_{R_{\theta_2}} f(\theta_1 | \theta_2, y) f(\theta_2, y) d\theta_2 \end{aligned}$$

onde R_{θ_2} indica o espaço de α_2 (Zellner, 1971).

Como exemplo considere a questão da predição do valor genético de um indivíduo (α_i) a partir de uma média fenotípica (y_i). A priori, $\alpha_i \sim N.I.I.D. (0, h^2)$, $E(y_i | \alpha_i) = \alpha_i$ e $\text{Var}(y_i | \alpha_i) = 1 - h^2$, onde h^2 é a herdabilidade. Neste caso, a densidade condicional $f(\alpha_i | h^2 = K, y_i)$ é normal com média $E(\alpha_i | h^2 = K, y_i) = K y_i$.

Uma vez que existe apenas um dado (ponto), pode-se expressar $f(h^2 | y_i)$ a $f(h^2)$, onde $f(h^2)$ e $f(h^2 | y_i)$ são as densidades a priori e marginal a posteriori da herdabilidade, respectivamente. Suponha que a priori h^2 pode assumir três valores (0; 0,10 e 0,20) com probabilidades $P[h^2 = 0] = 0,10$; $P[h^2 = 0,10] = 0,40$; $P[h^2 = 0,20] = 0,50$. A herdabilidade neste exemplo é um parâmetro de “nuisance” uma vez que o interesse reside na inferência marginal sobre o valor genético. Assim, h^2 deve ser eliminado usando

$f(\theta_1 | y) = \int_{R_{\theta_2}} f(\theta_1 | \theta_2, y) f(\theta_2 | y) d\theta_2$ da seguinte forma:

$$\begin{aligned} f(\alpha_i | y_i) &= f(\alpha_i | h^2 = 0, y_i) \times 0,10 \\ &+ f(\alpha_i | h^2 = 0,10, y_i) \times 0,40 \\ &+ f(\alpha_i | h^2 = 0,20, y_i) \times 0,50 \\ &= 0,10 E(\alpha_i | h^2 = 0, y_i) && \text{(usando o fato de que a} \\ &+ 0,40 E(\alpha_i | h^2 = 0,10, y_i) && \text{densidade condicional} \\ &+ 0,50 E(\alpha_i | h^2 = 0,20, y_i) && f(\theta_1 | h^2 = K, y_i) \text{ é normal} \\ &= 0,14 y_i && \text{com média} \\ & && E(\theta_1 | h^2 = K, y_i) = K y_i. \end{aligned}$$

O valor 0,14 corresponde à média a priori da distribuição dos possíveis valores de herdabilidade. A expressão $f(\theta_1 | y) = 0,14 y_i$ ilustra que a herdabilidade foi efetivamente integrada (eliminada), pois esta expressão não depende mais de h^2 (Gianola & Fernando, 1986).

Segundo Box & Tiao (1973) é muito importante examinar cuidadosamente $f(\alpha_2 | y)$. Se $f(\alpha_2 | y)$ é aguda, isto é, quando a maioria da densidade está próxima de sua moda ($\hat{\theta}_2$, diga-se), então $f(\theta_1 | y) \doteq f(\theta_1 | \theta_2) = \hat{\theta}_2, y$ e a moda ou estimativas do tipo máxima verossimilhança (ML é modal pois refere-se a um ponto de máximo) podem ser usados como verdadeiros valores de $f(\alpha_1 | \alpha_2, y)$. Como mostrado, isto é equivalente a realizar inferências com base na distribuição marginal de α_1 (valores genéticos) dado y (dados). Quanto às estimativas de componentes de variância, Gianola & Foulley (1990) mostraram que estimativas obtidas por REML são marginais apenas com respeito aos efeitos fixos mas são condicionais a outros parâmetros “nuisance” do modelo. A análise Bayesiana propicia uma marginalização adicional que é particularmente interessante para modelos com grande número de componentes de variância.

Quanto à questão da predição de valores genéticos sob modelo de seleção, em análise Bayesiana, as inferências sobre valores genéticos e parâmetros genéticos devem ser realizadas com base na densidade a posteriori construída como se a seleção e cruzamentos preferenciais não tivessem ocorrido. Os resultados obtidos dessa forma equivalem aos obtidos pela abordagem clássica, assumindo normalidade multivariada e são válidos para qualquer distribuição. Estes resultados são demonstrados a seguir com base em Gianola & Fernando (1986).

Considere α_1 , um vetor desconhecido de valores genéticos, e α_2 (efeitos fixos) e α_3 (componentes de variância), vetores desconhecidos de parâmetros de "nuisance" e as decisões de seleção e cruzamento baseadas em dados disponíveis em y_0 (vetor $n_0 \times 1$). Dependendo de y_0 , dados adicionais y_i ($i = 1, \dots, s$) são obtidos a partir de decisões de seleção e cruzamentos baseadas em y_0 . Em ausência de seleção e cruzamentos preferenciais $y_0, y_1, \dots, y_s, \alpha_1, \alpha_2$ e α_3 têm distribuição conjunta com densidade $f(y_0, y_1, \dots, y_s, \alpha_1, \alpha_2, \alpha_3)$.

A densidade marginal de $f(y_0, y_i, \alpha_1, \alpha_2, \alpha_3)$ é obtida após apropriada integração e as inferências devem ser baseadas em $f(\alpha_1, \alpha_2, \alpha_3 | y_0, y_i)$, com posterior integração (eliminação) de α_2 e α_3 .

Sendo R_i , a discreta e disjunta partição do espaço n_0 -dimensional, a densidade conjunta de $y_0, y_i, \alpha_1, \alpha_2, \alpha_3$ é definida se $y_0 \in R_i$. Assim, pode-se escrever:

$$f(y_0, y_i, \alpha_1, \alpha_2, \alpha_3 | y_0 \in R_i) = \frac{f(y_0, y_i, \theta_1, \theta_2, \theta_3)}{P[y_0 \in R_i]}$$

Usando probabilidade condicional, tem-se que:

$$f(y_0, y_i | y_0 \in R_i) = \frac{f(y_0, y_i)}{P[y_0 \in R_i]}$$

onde $f(y_0, y_i)$ é a densidade conjunta de y_0 e y_i , considerando que a seleção não tenha ocorrido. Combinando estas duas últimas expressões obtém-se:

$$\begin{aligned} f(\alpha_1, \alpha_2, \alpha_3 | y_0, y_i, y_0 \in R_i) &= \frac{f(y_0, y_i, \theta_1, \theta_2, \theta_3) / P[y_0 \in R_i]}{f(y_0, y_i) / P[y_0 \in R_i]} \\ &= f(\alpha_1, \alpha_2, \alpha_3 | y_0, y_i) \end{aligned}$$

Este resultado prova que, se as decisões de seleção e cruzamento forem baseadas em y_0 e, se y_0 corresponde aos dados disponíveis para análise, as inferências sobre valores genéticos, efeitos fixos e componentes de variância devem ser realizados a partir da densidade a posteriori considerando que a seleção não tenha ocorrido, concordando com os resultados de Henderson (1975, 1982b) e Thompson (1982) que são dependentes da suposição de normalidade. Assim, a distribuição condicional, dado o conjunto completo de dados, é a mesma distribuição em ausência ou presença de seleção e cruzamentos preferenciais e, para obter este resultado, a abordagem Bayesiana não assumiu linearidade, normalidade ou invariância a translação.

Em resumo, podem-se relatar as principais vantagens dos métodos Bayesianos, conforme sumarização realizada por Resende (1997):

- rotineiramente componentes de variância são estimados pelo método REML e utilizados na predição de valores genéticos e cálculo do ganho genético. Entretanto, através deste procedimento, a distribuição e variância dos estimadores não são conhecidos e assim questões referentes à efetividade da seleção não podem ser respondidas com rigor. O método Bayesiano de estimação de componentes de variância, implementado via simulação estocástica como por exemplo através do procedimento denominado amostragem de Gibbs (GS), propicia uma descrição mais completa sobre a confiabilidade dos parâmetros genéticos do que o método REML. A abordagem Bayesiana (conhecimento da distribuição a posteriori dos parâmetros) possibilita a construção de intervalos de confiança (ou de probabilidade) exatos para as estimativas dos parâmetros genéticos, estando a confiabilidade (precisão) dos parâmetros associada à quantidade e precisão das informações utilizadas. Dessa forma, o método GS propicia estimativas mais precisas de componentes de variância, parâmetros genéticos, valores genéticos e ganhos genéticos. O método REML propicia apenas intervalos de confiança aproximados para os parâmetros genéticos, através do uso de aproximações e suposições de normalidade (argumentos assintóticos). De maneira geral, o método GS apresenta estimativas com menores desvios padrões que o método REML, especialmente para valores baixos de herdabilidade. A diferença de precisão entre os dois métodos reduz-se para conjuntos de dados extremamente grandes e, neste caso, a tendência é obterem-se resultados similares.
- permite a utilização de informações a priori, a eliminação de parâmetros de "nuisance", a integrada estimação – predição – decisão e a análise exata de amostras de tamanho finito (pois os dados são fixados na distribuição a posteriori). Este último aspecto é muito importante, especialmente para programas de melhoramento baseados em pequenos conjuntos de dados desbalanceados, onde a análise Bayesiana propicia uma elegante análise de amostra finita, a qual não pode ser obtida pela metodologia padrão de

modelos mistos. O teorema de Bayes propicia soluções precisas para o problema de amostras de tamanho finito, pois para cada conjunto de dados, pequeno ou grande, existe uma distribuição a posteriori exata para a realização de inferências. Existe também, uma teoria assintótica Bayesiana, de forma que, com grandes amostras obtêm-se resultados similares aos obtidos com o método REML. Quanto aos parâmetros de “nuisance”, verifica-se que as estimativas de componentes de variância obtidas por REML são marginais apenas com respeito aos efeitos fixos mas são condicionais a outros parâmetros “nuisance” do modelo. A análise Bayesiana permite uma marginalização adicional particularmente interessante para modelos com grande número de componentes de variância.

- a principal característica da análise Bayesiana refere-se à riqueza da informação propiciada por esta abordagem, possibilitando a obtenção de estimativas pontuais e os intervalos de probabilidade para as distribuições a posteriori dos parâmetros. Assim, inferências sobre parâmetros genéticos, valores genéticos e efeito de grupos genéticos podem ser realizados com segurança.
- sob modelo de seleção, as inferências sobre valores genéticos e parâmetros genéticos podem ser realizados com base na densidade a posteriori, construída como se seleção e cruzamentos preferenciais não tivessem ocorrido, pois a distribuição destes parâmetros é a mesma com ou sem seleção e cruzamentos preferenciais (desde que seja usada a matriz de parentesco completa e também os dados originais que orientaram as decisões de seleção).
- a análise Bayesiana é capaz de resolver o problema dos valores genéticos ordenados, os quais apresentam distribuição desconhecida e cujas propriedades da distribuição normal não se aplicam.
- a abordagem Bayesiana apresenta, também, vantagens computacionais pois não exige solução para as equações de modelo misto e, portanto, pode ser praticada em microcomputadores exigindo pouca memória (similar aos requerimentos de processos iterativos), embora o tempo de processamento seja muito maior.

12. Implementação prática da análise Bayesiana

Os resultados de interesse gerados pela análise Bayesiana são, em geral, as distribuições marginais a posteriori dos parâmetros genéticos, efeito de grupo genético e valores genéticos. Assim, inferências baseadas na média, mediana, moda e desvios padrões destas distribuições são realizadas na prática.

O problema básico da implementação da análise Bayesiana refere-se à integração numérica. A integração (no espaço do parâmetro) da função densidade de probabilidade a posteriori, por exemplo:

$$E [g(\theta)|y] = \int_{R_{\theta}} g(\theta) p(\theta|y) d\theta, \text{ onde:}$$

$g(\theta) = \theta$, para obtenção da média a posteriori e

$g(\theta) = (\theta - \mu)^2$, $\mu = E(\theta|y)$, para obtenção da variância a posteriori ou risco de Bayes, pode ser realizada através dos métodos (Gamerman, 1997): (i) analítico para aproximação de integral; (ii) automáticos ou de quadratura; (iii) simulação estocástica para obtenção de distribuições a posteriori. Os métodos analíticos para aproximação de integrais, tais como a Discrepância de Kulback Liebler, a Aproximação Assintótica à Normal e o Método de Laplace são funcionais até determinada dimensão do problema e estão associados a resultados assintóticos e à normalidade. O método de quadratura é puramente matemático e vem sendo substituído vantajosamente por métodos computacionalmente intensivos, os quais são essencialmente estatísticos, tais como os métodos de simulação estocástica. Dentre os métodos de simulação estocástica destacam-se: (i) Monte Carlo; (ii) Monte-Carlo com Função de Importância; (iii) Reamostragem ou Bootstrap Bayesiano; (iv) Monte-Carlo – Cadeias de Markov.

Dentre estas três grandes classes de algoritmos para aproximar as integrais, os métodos de Monte-Carlo são largamente indicados e utilizados para integração multivariada. Os métodos de Monte-Carlo referem-se a processos de aproximação de valores esperados (integrais com respeito a uma distribuição de probabilidade) por meio de amostras, podendo ser referidos também como um caso especial de simulação de um processo estocástico.

Em genética quantitativa, para implementação prática da análise Bayesiana, uma das maiores dificuldades técnicas é a marginalização. A obtenção de distribuições marginais por processos analíticos é praticamente impossível (Sorensen, 1996). Assim, a obtenção da distribuição marginal a posteriori (marginalização da distribuição conjunta a posteriori) tem sido obtida pelo método da amostragem de Gibbs (GS) através da amostragem e atualização das distribuições condicionais.

O método da amostragem de Gibbs pertence à classe de métodos, denominada Monte-Carlo – Cadeias de Markov, a qual é sustentada em propriedades das Cadeias de Markov. O nome Gibbs advém da distribuição de Gibbs, que é muito utilizada na área de Física Estatística ou Mecânica Estatística. O amostrador de Gibbs explorando as distribuições condicionais completas através de algoritmo iterativo foi proposto inicialmente por Geman & Geman

(1984) para aplicações na área de processamento de imagens. Entretanto, somente em 1990, este trabalho foi divulgado para toda a comunidade da área de estatística através de Gelfand & Smith (1990) que publicaram em periódico da área de estatística, trabalho comparando o amostrador de Gibbs com outros processos de simulação estocástica. A partir de então, houve uma revolução na inferência estatística com uma impressionante realização de trabalhos baseados nesta técnica (Gamerman, 1996).

Para ilustrar a aplicação da técnica da amostragem de Gibbs na avaliação genética de espécies perenes será considerado o modelo individual univariado, conforme Sorensen (1996).

Modelo

$y = Xb + Za + e$, onde:

y = vetor de dados, de ordem n ;

b = vetor de efeitos fixos, de ordem p ;

a = vetor de valores genéticos aditivos, de ordem q ;

e = vetor de erros, de ordem n ;

X, Z = matrizes de incidência que associam b e a aos dados (y).

Assume-se inicialmente que a distribuição condicional dos dados, b , a e σ_e^2 é normal multivariada:

$y|\beta, a, \sigma_e^2 \sim N(X\beta + Za, I\sigma_e^2)$, onde I é a matriz identidade e σ_e^2 a variância residual.

Assumindo o modelo quantitativo infinitesimal, tem-se que a distribuição de a é também normal multivariada:

$a|A, \sigma_a^2 \sim N(O, A\sigma_a^2)$, onde A é a matriz de parentesco genético aditivo e σ_a^2 é a variância genética aditiva na população base.

Os parâmetros de interesse para inferências são: β , a , σ_a^2 e σ_e^2 . Para conduzir a análise Bayesiana torna-se necessário especificar as distribuições a priori para β , σ_a^2 e σ_e^2 (a distribuição de a já foi especificada).

Como priori para b pode-se assumir $p(b)$ a constante, que especifica aproximadamente a noção de conhecimento a priori vago para b . Esta

distribuição a priori é imprópria, mas pode-se tornar própria, desde que se especifiquem os limites superior e inferior para $p(b)$.

As distribuições a priori dos componentes de variância (σ_e^2 e σ_a^2) poderiam ser uniforme da forma $p(\sigma_i^2) \propto \text{constante}$, $0 \leq \sigma_i^2 < \sigma_{i \max}^2$ ($i = e, a$), onde, de acordo com o conhecimento acumulado sobre o caráter, $\sigma_{i \max}^2$ seria o valor máximo que σ_i^2 poderia assumir, a priori. Alternativamente, poderia ser especificada uma priori mais informativa para os componentes de variância, assumindo uma distribuição qui-quadrado escalonada invertida, da forma:

$$p(\sigma_i^2 | v_i, S_i^2) \propto (\sigma_i^2)^{-(v_i/2)+1} \exp\left[-\frac{v_i S_i^2}{2 \sigma_i^2}\right] \quad (i = e, a), \text{ onde } v \text{ são os graus}$$

de liberdade da distribuição qui-quadrado e S_i^2 , o valor inicial da variância. Esta distribuição reduz-se a uma distribuição uniforme imprópria se $v_i = -2$ e $S_i^2 = 0$.

Definidas estas distribuições, pode-se agora escrever a distribuição conjunta a posteriori dos parâmetros do modelo.

$$p(\beta, a, \sigma_a^2, \sigma_e^2 | y) \propto p(\beta, a, \sigma_a^2, \sigma_e^2) p(y | \beta, a, \sigma_a^2, \sigma_e^2) \\ = p(\beta) p(a | \sigma_a^2) p(\sigma_a^2) p(\sigma_e^2) p(y | \beta, a, \sigma_a^2, \sigma_e^2), \text{ onde}$$

omitiu-se o condicionamento nos hiperparâmetros (parâmetros que auxiliam na especificação da priori) e na conhecida matriz de parentesco A .

Considerando a distribuição a priori dos componentes de variância como uma qui-quadrado escalonada invertida, tem-se que a distribuição conjunta a posteriori pode ser reescrita:

$$p(\beta, a, \sigma_a^2, \sigma_e^2 | y) \propto \sigma_e^{2-\left(\frac{n+v_e}{2}+1\right)} \exp\left[-\frac{(y - X\beta - Za)'(y - X\beta - Za) + v_e S_e^2}{2 \sigma_e^2}\right] \\ \sigma_a^{2-\left(\frac{q+v_a}{2}+1\right)} \exp\left[-\frac{(a' A^{-1} a + v_a S_a^2)}{2 \sigma_a^2}\right]$$

Desejando assumir distribuição a priori uniforme para σ_a^2 e σ_e^2 , basta fazer $v_i = -2$ e $S_i^2 = 0$ ($i = a, e$) na expressão acima.

Para implementação do GS, deve-se derivar todas as distribuições condicionais a posteriori a partir da distribuição conjunta a posteriori apresentada acima.

Denominando-se $X\beta + Za = W\theta$, onde $W = [X \ Z]$ e $\theta' = [\beta' \ a']$, tem-se que a matriz dos coeficientes das equações de modelo misto é dada por

$$C = W' W + S, \text{ onde } \Sigma = \begin{bmatrix} 0 & 0 \\ 0 & a^{-1}\sigma_e^2/\sigma_a^2 \end{bmatrix}. \text{ A distribuição condicional a}$$

posteriori de θ é:

$\theta \mid \sigma_a^2, \sigma_e^2, y \sim N(\hat{\theta}, C^{-1}\sigma_e^2)$, em que $\hat{\theta}$ é dado por $C\hat{\theta} = W'y$, ou seja, pelas equações de modelo misto.

Para derivar a distribuição condicional a posteriori para b_i (o i -ésimo elemento do vetor b) devem-se fazer as partições:

$$\theta' = (\theta'_i, \theta'_{-i}); \quad W' = (W'_i \ W'_{-i}); \quad C = \begin{bmatrix} C_{i,i} & C_{i,-i} \\ C_{-i,i} & C_{-i,-i} \end{bmatrix} = \left\{ \begin{array}{ccc} X'_i X_i & X'_i X_{-i} & X'_i Z \\ X'_{-i} X_i & X'_{-i} X_{-i} & X'_{-i} Z \\ Z' X_i & Z' X_{-i} & Z' Z + A^{-1}\alpha \end{array} \right\}$$

onde: $\alpha = \sigma_e^2/\sigma_a^2$, β_{-i} é o vetor de efeitos fixos com o i -ésimo elemento excluído, de ordem $p-1$, X_i é o i -ésimo vetor coluna da matriz X e X_{-i} é a matriz X com X_i excluído.

Assim, tem-se: $\theta_i \mid \theta_{-i}, \sigma_a^2, \sigma_e^2, y \sim N(\hat{\theta}_i, C_{i,i}^{-1}\sigma_e^2)$, em que $\hat{\theta}_i$ é dado por $C_{i,i}\hat{\theta}_i = (W'_i \ y - C_{i,-i}\theta_{-i})$.

Tem-se também:

$W'_i y = X'_i y$; $(X'_i \ X_i) \hat{\beta}_i = X'_i y - X'_i X_{-i} \beta_{-i} - X'_i Za$ e finalmente pode-se escrever a distribuição condicional a posteriori de b_i :

$$\beta_i \mid \beta_{-i}, a, \sigma_a^2, \sigma_e^2, y \sim N(\hat{\beta}_i, (X'_i X_i)^{-1}\sigma_e^2), \quad (3)$$

em que: $\hat{\beta}_i = (X'_i X_i)^{-1} X'_i (y - X_{-i} \beta_{-i} - Za)$

Para derivar a distribuição condicional a posteriori de a_i , deve-se proceder de maneira similar. Assim, tem-se:

$$A^{-1} = \begin{bmatrix} A_{ii}^{-1} & A_{i,-i}^{-1} \\ A_{-i,i}^{-1} & A_{-i,-i}^{-1} \end{bmatrix}; \quad C = \begin{bmatrix} X'X & X'z_i & X'Z_{-i} \\ z_i'X & z_i'z_i + A_{ii}^{-1}\alpha & z_i'Z_{-i} + A_{i,-i}^{-1}\alpha \\ Z_{-i}'X & Z_{-i}'z_i + A_{-i,i}^{-1}\alpha & Z_{-i}'Z_{-i} + A_{-i,-i}^{-1}\alpha \end{bmatrix}$$

$$W_i' y = z_i' y; \quad (z_i' z_i + A_{ii}^{-1}\alpha) \hat{a}_i = z_i' y - z_i' X\beta - (z_i' Z_{-i} + A_{i,-i}^{-1}\alpha) a_{-i} \\ = z_i' y - z_i' X\beta - A_{i,-i}^{-1}\alpha a_{-i}, \quad \text{pois } z_i' Z_{-i} = 0$$

Finalmente, a distribuição condicional a posteriori de a_i é:

$$a_i | \beta, a_{-i}, \sigma_a^2, \sigma_e^2, y \sim N(\hat{a}_i, (z_i' z_i + A_{ii}^{-1}\alpha)^{-1} \sigma_e^2) \quad (4)$$

A obtenção das distribuições condicionais a posteriori completas dos componentes de variância será descrita a seguir. Para σ_a^2 , deve-se tomar da distribuição conjunta a posteriori somente os termos que envolvem σ_a^2 . Assim:

$$p(\sigma_a^2 | \beta, a, \sigma_e^2, y) \propto (\sigma_a^2)^{-\left(\frac{q+v_a+1}{2}\right)} \exp\left[\frac{-a'A^{-1}a + v_a S_a^2}{2 \sigma_a^2}\right] \\ = (\sigma_a^2)^{-\left(\frac{\tilde{v}_a+1}{2}\right)} \exp\left[\frac{-\tilde{v}_a \tilde{S}_a^2}{2 \sigma_a^2}\right], \quad \text{em que}$$

$$\tilde{S}_a^2 = (a'A^{-1}a + v_a S_a^2) / \tilde{v}_a \quad \text{e} \quad \tilde{v}_a = q + v_a$$

A expressão de $p(\sigma_a^2 | \beta, a, \sigma_e^2, y)$ é proporcional à distribuição qui-quadrado (χ^2) escalonada invertida, com \tilde{v}_a graus de liberdade e parâmetro de escala $a'A^{-1}a + v_a S_a^2$.

Assim:

$$\sigma_a^2 | \beta, a, \sigma_e^2, y \sim \tilde{v}_a \tilde{S}_a^2 \chi_{\tilde{v}_a}^2 \quad (5)$$

e para amostrar desta distribuição deve-se inicialmente trabalhar com um qui-quadrado com $\tilde{v}_a = q + v_a$ graus de liberdade, inverter este número e posteriormente multiplicá-lo por $a' A^{-1} a + v_a S_a^2$.

De maneira similar tem-se para σ_e^2 :

$$\begin{aligned} p(\sigma_e^2 | \beta, a, \sigma_a^2, y) &\propto (\sigma_e^2)^{-\left(\frac{n+v_e}{2}+1\right)} \exp \left[\frac{(y - X\beta - Za)'(y - X\beta - Za) + v_e S_e^2}{2 \sigma_e^2} \right] \\ &= (\sigma_e^2)^{-\left(\frac{\tilde{v}_e}{2}+1\right)} \exp \left[\frac{-\tilde{v}_e \tilde{S}_e^2}{2 \sigma_e^2} \right], \text{ onde} \end{aligned}$$

$$\tilde{v}_e = n + v_e \quad \tilde{S}_e^2 = ((y - X\beta - Za)'(y - X\beta - Za) + v_e S_e^2) / \tilde{v}_e.$$

A expressão de $p(\sigma_e^2 | \beta, a, \sigma_a^2, y)$ é proporcional à distribuição qui-quadrado escalonada invertida, com $\tilde{v}_e = n + v_e$ graus de liberdade e parâmetro de escala $(y - X\beta - Za)'(y - X\beta - Za) + v_e S_e^2$.

Assim, tem-se:

$$\sigma_e^2 | \beta, a, \sigma_a^2, y \sim \tilde{v}_e \tilde{S}_e^2 \chi_{\tilde{v}_e}^{-2} \quad (6)$$

A implementação da amostragem de Gibbs consiste em amostrar sucessivamente de (3), (4), (5) e (6).

Considere o exemplo:

Indivíduo	Bloco	Pai	Mãe	Y_i
1	1	-	-	-
2	2	-	-	-
3	1	1	-	Y_3
4	2	1	2	Y_4
5	1	3	4	Y_5
6	2	1	4	Y_6
7	1	5	6	-

Considerando o modelo individual tem-se $\beta = (\beta_1, \beta_2)'$ para os blocos 1 e 2 e $a = (a_1, \dots, a_7)'$ para os valores genéticos aditivos dos sete indivíduos. Assumindo distribuições a priori uniforme para os componentes de variância, tais que $v_i = -2$, $S_i^2 = 0$, ($i = a, e$) e valores iniciais $\sigma_a^2 = 5,0$ e $\sigma_e^2 = 5,0$, deve-se trabalhar com as equações de modelo misto atualizando o valor $\alpha = \sigma_e^2 / \sigma_a^2$, a cada iteração. A matriz dos coeficientes das equações de modelo misto em função de α equivale a:

$$C = \alpha \begin{bmatrix} \beta_1 & \beta_2 & a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 \\ 2,00 & 0,00 & 0,00 & 0,00 & 1,00 & 0,00 & 1,00 & 0,00 & 0,00 \\ 0,00 & 2,00 & 0,00 & 0,00 & 0,00 & 1,00 & 0,00 & 1,00 & 0,00 \\ 0,00 & 0,00 & 2,33\alpha & 0,50\alpha & -0,66\alpha & -0,50\alpha & 0,00\alpha & -1,00\alpha & 0,00\alpha \\ 0,00 & 0,00 & 0,50\alpha & 1,50\alpha & 0,00\alpha & -1,00\alpha & 0,00\alpha & 0,00\alpha & 0,00\alpha \\ 1,00 & 0,00 & -0,66\alpha & 0,00\alpha & 1+2,83\alpha & 0,50\alpha & -1,00\alpha & 0,00\alpha & 0,00\alpha \\ 0,00 & 1,00 & -0,50\alpha & -1,00\alpha & 0,50\alpha & 1+3,00\alpha & -1,00\alpha & -1,00\alpha & 0,00\alpha \\ 1,00 & 0,00 & 0,00\alpha & 0,00\alpha & -1,00\alpha & -1,00\alpha & 1+2,62\alpha & 0,62\alpha & -1,23\alpha \\ 0,00 & 1,00 & -1,00\alpha & 0,00\alpha & 0,00\alpha & -1,00\alpha & 0,62\alpha & 1+2,62\alpha & -1,23\alpha \\ 0,00 & 0,00 & 0,00\alpha & 0,00\alpha & 0,00\alpha & 0,00\alpha & -1,23\alpha & -1,23\alpha & 2,46\alpha \end{bmatrix}$$

Os parâmetros de locação são $\theta' = (\beta_1, \beta_2, a_1, \dots, a_7)$ e assumindo que os valores iniciais são zero, tem-se $\theta^{(0)} = 0$ e pode-se iniciar a amostragem com uma cadeia de comprimento igual a n ($i = 1, \dots, n$).

A primeira amostragem para $\beta_1^{(1)}$ é obtida de **(3)** ou seja obtém-se $\beta_1^{(1)}$ de $\beta_1 | \beta_2, a, \sigma_a^2, \sigma_e^2, y \sim N(\hat{\beta}_1, (2)^{-1} \sigma_e^2)$. Dados os valores iniciais,

a média da distribuição normal é $\hat{\beta}_1 = (Y_3 + Y_5)/2$ e a variância é $(2)^{-1}\sigma_e^2$, ou seja, 5/2.

$\beta_2^{(1)}$ por sua vez, é retirado de uma distribuição normal com média $\hat{\beta}_2 = (Y_4 + Y_6)/2$ e variância 5/2.

$a_1^{(1)}$ por sua vez, é retirado de $a_1|\beta, a_{-1}, \sigma_a^2, \sigma_e^2, y \sim N(\hat{a}_1, (2,33\alpha)^{-1}\sigma_e^2)$, que provém de **(4)**, onde $\hat{a}_1 = 0/2,33\alpha = 0$, $\sigma_e^2 = 5$ e $\alpha = 1$.

$a_2^{(1)}$ é retirado de uma distribuição normal $N(\hat{a}_2, (1,50\alpha)^{-1}\sigma_e^2)$, onde $\hat{a}_2 = (0 - 0,50\alpha a_1^{(1)})/(1,50\alpha)$, $\sigma_e^2 = 5$ e $\alpha = 1$, onde 0,50 provém de C.

Para o indivíduo número 6, o valor genético na primeira iteração, $a_6^{(1)}$, é retirado de:

$N(\hat{a}_6, (1 + 2,62\alpha)^{-1}\sigma_e^2)$, onde

$$\hat{a}_6 = \frac{Y_6 - \beta_2^{(1)} - (-1,00\alpha a_1^{(1)} - 1,00\alpha a_4^{(1)} + 0,62\alpha a_5^{(1)})}{1 + 2,62\alpha}, \sigma_e^2 = 5 \text{ e } \alpha = 1$$

Para o indivíduo número 7, o valor genético na primeira iteração, $a_7^{(1)}$, é retirado de:

$N(\hat{a}_7, (2,46\alpha)^{-1}\sigma_e^2)$, onde

$$\hat{a}_7 = \frac{0 - (-1,23\alpha a_5^{(1)} - 1,23\alpha a_6^{(1)})}{2,46\alpha}, \sigma_e^2 = 5 \text{ e } \alpha = 1$$

Tendo-se amostrados todos os parâmetros de locação do modelo, deve-se computar:

$$SS_e^{(1)} = (y - X\beta^{(1)} - Za^{(1)})'(y - X\beta^{(1)} - Za^{(1)})$$

$$SS_a^{(1)} = (a^{(1)})'A^{-1}a^{(1)}$$

A primeira iteração do amostrador é completada, retirando-se os componentes de variância, usando $SS_a^{(1)}$ e $SS_e^{(1)}$:

$$\sigma_a^2 | \beta, a, \sigma_e^2, y \sim SS_a^{(1)} \chi_{q-2}^{-2} \quad (7)$$

$$\sigma_e^2 | \beta, a, \sigma_a^2, y \sim SS_e^{(1)} \chi_{n-2}^{-2} \quad (8)$$

A segunda iteração inicia-se através de atualizações das equações de modelo misto com $\alpha = \sigma_e^2 / \sigma_a^2$, onde σ_e^2 e σ_a^2 são os valores amostrados de (7) e (8) e prossegue com a amostragem de $\theta^{(2)}$ da mesma maneira realizada para $\theta^{(1)}$ e com o cômputo de $SS_e^{(2)}$ e $SS_a^{(2)}$ da mesma maneira realizada para $SS_e^{(1)}$ e $SS_a^{(1)}$, e assim sucessivamente.

Em termos mais simples, o algoritmo GS pode ser apresentado de forma resumida:

1. Fornecer os valores iniciais dos parâmetros de locação e dispersão do modelo. Estes valores iniciais podem ser calculados através de procedimentos padrões tais como a estimação de componentes de variância por REML ou quadrados mínimos. Assumindo a média geral \bar{y} como único efeito fixo, pode-se calcular \bar{y} como a média aritmética das observações e $a_i = h^2(y_i - \bar{y})$. Devem ser fornecidos os valores iniciais para \bar{y}_i , a_i , σ_e^2 , σ_a^2 e $\alpha = \sigma_e^2 / \sigma_a^2$.

2. Gerar valores para os efeitos fixos. Sendo o único efeito fixo, a média geral, tem-se:

$$\hat{\bar{y}} = \bar{y} + \text{rnd } \sigma_e / (n)^{1/2}$$

3. Gerar valores para os efeitos aleatórios:

$$\hat{a} = a_i + \text{rnd } [(1 - r_{\hat{a}}^2) \sigma_a^2]^{1/2}, \text{ onde } r_{\hat{a}} \text{ é a acurácia dada por } r_{\hat{a}} = (1 - PEV_i / \sigma_a^2)^{1/2}, \text{ onde } PEV_i \text{ é o iésimo elemento da inversa da matriz dos coeficientes das EMM multiplicado por } \sigma_e^2.$$

4. Calcular a soma de quadrados do resíduo (SSE) e a variância residual σ_e^2 . Assumindo que a distribuição a priori para a variância residual é a inversa de uma qui-quadrado, tem-se:

$$SSE = \sum (y_i - \hat{y} - \hat{a}_i)^2$$

$$\sigma_e^2 = \frac{SSE}{X_n^2}$$

5. Gerar um valor para a variância dos efeitos aleatórios de valores genéticos.

$$\sigma_a^2 = \frac{\hat{a}' A^{-1} \hat{a}}{X_q^2}$$

6. Calcular o novo valor do parâmetro

$$\hat{\alpha} = \frac{\sigma_e^2}{\sigma_a^2}$$

7. Repetir os passos de (2) a (6) até que se obtenha a convergência da cadeia.

13. Diagnóstico de convergência

Este tema foi apresentado em detalhes por Gamerman (1996) e aqui serão apresentados alguns aspectos práticos principais, com base na referida literatura.

As cadeias de Markov estão inseridas no contexto da teoria dos processos estocásticos, teoria esta, definida como a parte dinâmica da teoria de probabilidades, onde se estuda uma coleção de variáveis aleatórias, com respeito a sua interdependência e comportamento limite. Para a inferência Bayesiana, é de maior relevância o estudo do comportamento assintótico da cadeia, quando o número de iterações tende a ∞ , uma vez que a inferência deve ser baseada na distribuição (a posteriori) estacionária, ou seja, em equilíbrio.

À medida em que o número de iterações aumenta, a cadeia se aproxima da condição de equilíbrio. Dessa forma, é necessário assumir a convergência em uma determinada iteração cuja distribuição esteja próxima da distribuição

em equilíbrio (atingido teoricamente quando $n \gg \frac{1}{\epsilon}$), ou seja, após um número suficientemente grande de iterações.

A forma básica de se obter uma amostra de tamanho m da posteriori é produzir m cadeias independentes (geradas a partir de m valores iniciais diferentes) e, após a convergência, retirar os valores da última iteração de cada cadeia. Outra opção consiste em retirar m amostras da mesma cadeia após a convergência, visto que se estará amostrando da distribuição a posteriori em equilíbrio. Neste último caso, é importante relatar que as amostras sucessivas não são independentes, de forma que torna-se necessário descartar várias iterações entre cada duas amostras a serem salvas. Como o processo é Markoviano, a dependência diminui com o aumento da distância entre iterações, obtendo-se, assim, independência entre as amostras salvas.

Considerando a segunda opção, no contexto do diagnóstico da convergência, tornam-se relevantes as quantidades: M = número de iterações pré-convergência ou período de descarte ou período de aquecimento da cadeia; N = número de iterações após a convergência; K = número de iterações entre amostras sucessivas ou intervalo entre amostras. O tamanho total da cadeia é dado por $T = M + N$.

O valor de K pode ser determinado calculando as autocorrelações na série de valores gerados e verificando a partir de qual ponto pode-se considerar as autocorrelações como nulas. Uma vez que o valor de K é muito menor que o de M , os métodos baseados em uma única cadeia (mais longa) são preferidos computacionalmente.

Outra forma de análise de convergência refere-se à estimação do erro de Monte Carlo, conforme apresentado no tópico 15. Em resumo, para a aplicação prática, algumas recomendações são apresentadas a seguir.

Devido ao fato de que valores aleatórios são utilizados inicialmente como realização do conjunto de parâmetros, é necessário um período de descarte de amostras até que as amostras de GS possam ser consideradas como provenientes da distribuição conjunta a posteriori, ou seja da distribuição em equilíbrio estacionário. Em geral, tem sido utilizado o esquema tradicional de cadeia longa (única) de Gibbs, onde o processo de reamostragem é contínuo. Assim, de maneira geral, um grande (da ordem de 10.000 a 1.000.000) número de ciclos tem sido utilizado, sendo descartadas as primeiras amostras (da ordem de poucos milhares) e amostras de cada parâmetro são salvas a cada pequeno (da ordem de 50 a 100) número de iterações. O intervalo entre amostras salvas é necessário como forma de obtenção de amostras independentes, visto que amostras sucessivas apresentam correlação serial. O número total de amostras salvas é utilizado para cômputo das estimativas pontuais e intervalares de interesse. O software MTGSAM (Tassell & Vleck, 1995) tem sido utilizado com sucesso para a implementação da análise Bayesiana.

14. Estimação Bayesiana via modelos lineares robustos

No tópico 12 foram assumidas verossimilhanças normais para os dados observados. Entretanto, a distribuição Gaussiana usualmente assumida nos modelos de genética quantitativa nem sempre se ajusta aos dados observados. Neste caso, a presença de dados atípicos (outliers) podem distorcer os resultados e inferências obtidos com base na suposição de normalidade, visto que a distribuição normal não é robusta a observações discrepantes, conforme relato de Rogers & Tukey (1972).

Visando contornar este problema, distribuições simétricas de caudas longas (tais como a t de Student, a Slash e a normal contaminada) tem sido aplicadas recentemente (Strandén, 1996; Rosa, 1999) visando a estimação robusta. Estas distribuições mencionadas pertencem ao grupo de distribuições denominado normal/independente. Conforme relatado por Rosa (1999) a distribuição normal/independente pode ser definida como a distribuição de um vetor k dimensional dado por $y = m + e | (w)^{1/2}$, onde m é o vetor de locação de dimensão k , w é uma variável aleatória positiva com densidade $p(w | v)$, onde v é um parâmetro de robustez, e e é um vetor aleatório k dimensional, normalmente distribuído com média 0 e matriz de covariância $\hat{\Sigma}$ não singular.

Neste caso, a distribuição de y , condicional a w , é normal com média m e matriz de covariância $\hat{\Sigma} | w$. A distribuição de y é dada por:

$$p(y | \mu, \Sigma, v) = \int p(y | \mu, \Sigma, w) p(w | v) dw$$

Dessa forma, a distribuição de w define a distribuição de y . As distribuições normal/independente têm algumas propriedades comuns, tais quais (Lange & Sinsheimer, 1993): (i) se $y \sim p(y | \mu, \hat{\Sigma}, v)$, ou seja se y tem distribuição normal/independente, então $E[y | \mu, \hat{\Sigma}, v] = m$ e $\text{Var}[y | \mu, \hat{\Sigma}, v] = E[w^{-1} | v] \hat{\Sigma}$, desde que existam o primeiro e segundo momentos; (ii) $p(y | \mu, \hat{\Sigma}, v)$ é simétrica e unimodal, de forma que a moda e a mediana são iguais à média m .

Dentre as distribuições robustas mencionadas, a t de Student tem um único parâmetro de robustez, que são os graus de liberdade da distribuição. Quanto menor o número de graus de liberdade, mais pesadas são as caudas das distribuições t. Na distribuição Slash, a robustez também é definida por um único parâmetro. Por outro lado, a distribuição normal contaminada apresenta dois parâmetros de robustez, sendo que um fornece a proporção de contaminantes e o outro define como a variância das duas (contaminantes e não contaminantes) populações difere (Rosa, 1999). Verifica-se que estas distribuições robustas são modificações da distribuição normal (Gaussiana), modificações estas em função dos parâmetros de robustez. Assim, a distribuição Gaussiana é um caso particular de distribuição normal/independente

eqüivalendo às distribuições t e Slash com parâmetro v infinito ou à distribuição normal contaminada quando a proporção de contaminantes é nula.

A implementação da análise Bayesiana usando distribuições normal/independente ao invés da suposição Gaussiana pode ser realizada via amostrador de Gibbs, de maneira similar à descrita no tópico 12.

Em melhoramento animal, a presença de dados discrepantes pode ser devida ao tratamento preferencial aos melhores animais, conforme constatado por Strandén (1996). Neste caso, os modelos robustos tendem a fornecer resultados mais plausíveis e devem ser adotados.

15. Aplicação a dados experimentais

A seguir são apresentados resultados obtidos por Resende (1999), relativos à avaliação genética de um caráter quantitativo em uma espécie vegetal perene.

Foram utilizados dados de quatro experimentos (testes de progênies de irmãos germanos) desbalanceados de *Pinus caribaea* var. *hondurensis* com um total de 89 progênies, sendo 65 diferentes e algumas comuns aos experimentos. As 65 progênies foram obtidas de 76 genitores e no total foram avaliados 3.886 indivíduos. Os experimentos foram implantados pela Champion Papel e Celulose S.A., no município de Santana no Estado do Amapá, no delineamento em blocos casualizados com nove repetições e cinco plantas por parcela. Considerou-se a variável diâmetro avaliada aos nove anos e meio e o modelo aditivo infinitesimal.

O modelo linear univariado eqüivale a:

$$y = Xb + Za + Wc + e, \text{ em que:}$$

y , b , a , c e e = vetores de dados, de efeitos fixos (blocos), de valores genéticos, de efeitos de ambiente comum (parcela) e erros aleatórios, respectivamente.

X , Z e W = matrizes de incidência para b , a e c , respectivamente.

As seguintes distribuições foram assumidas:

$$y|\beta, a, c, \sigma_e^2 \sim N (X\beta + Za + Wc, I\sigma_e^2)$$

$$a|A, \sigma_a^2 \sim N (0, A\sigma_a^2)$$

$$c|\sigma_c^2 \sim N (0, I\sigma_c^2)$$

$$e|\sigma_e^2 \sim N (0, I\sigma_e^2)$$

$COV(a, c')=0$; $COV(a, e')=0$; $COV(c, e')=0$, em que:

A = matriz de parentesco genético aditivo.

I = matriz identidade de ordem apropriada.

σ_a^2 , σ_c^2 e σ_e^2 = variâncias genética aditiva, de ambiente comum da parcela e residual, respectivamente.

Sob este modelo, a herdabilidade (h^2) e a correlação devida ao ambiente comum da parcela (c^2) são, respectivamente:

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_c^2 + \sigma_e^2}; \quad c^2 = \frac{\sigma_c^2}{\sigma_a^2 + \sigma_c^2 + \sigma_e^2}$$

A análise Bayesiana deste modelo linear mais geral (incluindo o efeito de parcela) é apresentada a seguir:

Distribuição conjunta a posteriori dos parâmetros do modelo

$$p(\beta, a, c, \sigma_a^2, \sigma_c^2, \sigma_e^2|y) \propto p(\beta, a, c, \sigma_a^2, \sigma_c^2, \sigma_e^2) p(y|\beta, a, c, \sigma_a^2, \sigma_c^2, \sigma_e^2) \\ = p(\beta) p(a|\sigma_a^2) p(c|\sigma_c^2) p(\sigma_a^2) p(\sigma_c^2) p(\sigma_e^2) p(y|\beta, a, c, \sigma_a^2, \sigma_c^2, \sigma_e^2)$$

Assumindo a distribuição a priori dos componentes da variância como uma qui-quadrado escalonada invertida, tem-se que:

Distribuição conjunta a posteriori

$$p(\beta, a, c, \sigma_a^2, \sigma_c^2, \sigma_e^2 | y) \propto \sigma_e^{2-\left(\frac{n+v_e}{2}+1\right)} \exp\left[(y - X\beta - Za - Wc)'(y - X\beta - Za - Wc) + v_e S_e^2\right]$$

$$\sigma_a^{2-\left(\frac{q+v_a}{2}+1\right)} \exp\left[-\frac{(a' A^{-1} a + v_a S_a^2)}{2 \sigma_a^2}\right]$$

$$\sigma_c^{2-\left(\frac{p+v_c}{2}+1\right)} \exp\left[-\frac{(c' c + v_c S_c^2)}{2 \sigma_c^2}\right]$$

Incorporando na matriz C (item 12) as submatrizes relacionando W com Z e X, obtém-se as distribuições condicionais a posteriori para cada parâmetro:

Distribuição condicional a posteriori de b_i

$$\beta_i | \beta_{-i}, a, c, \sigma_a^2, \sigma_c^2, \sigma_e^2, y \sim N(\hat{\beta}_i, (X_i' X_i)^{-1} \sigma_e^2),$$

$$\text{em que: } \hat{\beta}_i = (X_i' X_i)^{-1} X_i' (y - X_{-i} \beta_{-i} - Za - Wc)$$

Distribuição condicional a posteriori de a_i

$$a_i | \beta, a_{-i}, c, \sigma_a^2, \sigma_c^2, \sigma_e^2, y \sim N(\hat{a}_i, (Z_i' Z_i + A_{i,i}^{-1} \alpha)^{-1} \sigma_e^2)$$

Distribuição condicional a posteriori de c_i

$$c_i | \beta, a, c_{-i}, \sigma_a^2, \sigma_c^2, \sigma_e^2, y \sim N(\hat{c}_i, (W_i' W_i + \alpha^*)^{-1} \sigma_e^2)$$

Distribuição condicional a posteriori de σ_a^2

$$\sigma_a^2 | \beta, a, c, \sigma_c^2, \sigma_e^2, y \sim \tilde{v}_a \tilde{S}_a^2 \chi_{\tilde{v}_a}^2$$

Distribuição condicional a posteriori de σ_c^2

$$\sigma_c^2 | \beta, a, c, \sigma_a^2, \sigma_e^2, y \sim \tilde{\nu}_c \tilde{S}_c^2 \chi_{\tilde{\nu}_c}^2$$

Distribuição condicional a posteriori de σ_e^2

$$\sigma_e^2 | \beta, a, c, \sigma_a^2, \sigma_c^2, y \sim \tilde{\nu}_e \tilde{S}_e^2 \chi_{\tilde{\nu}_e}^2 \text{ em que } \alpha^* = \frac{\sigma_e^2}{\sigma_c^2} \text{ e os demais parâmetros}$$

são definidos analogamente ao item 12.

Além das distribuições assumidas para os efeitos aleatórios no modelo linear freqüentista e para a verossimilhança do vetor de observações, a abordagem Bayesiana requer atribuições para as distribuições a priori dos efeitos fixos e componentes de variância. Uma distribuição a priori não informativa ou uniforme foi atribuída a b , refletindo conhecimento a priori vago sobre os efeitos fixos, de forma que $p(b) \propto K$ (constante). Para os componentes de variância, distribuições c^2 inversas foram assumidas como priori, conforme especificado no tópico anterior. Considerou-se adicionalmente $u_i = -2$ e $S_i^2 = 0$, de forma que a distribuição se tornasse uniforme, e, portanto, não informativa.

Para a inferência Bayesiana sobre os parâmetros de interesse empregou-se a técnica da amostragem de Gibbs. O principal aspecto deste procedimento refere-se ao fato das inferências basearem-se na distribuição marginal a posteriori dos parâmetros, sendo que a marginalização da distribuição conjunta a posteriori é obtida via o amostrador de Gibbs, através de amostragens e atualizações das distribuições condicionais. Em resumo, a abordagem Bayesiana baseia-se na construção da distribuição marginal a posteriori de um parâmetro de interesse, tratando-o como uma variável aleatória e aplicando cálculo de probabilidades. Este procedimento implica problemas multidimensionais, uma vez que todos os outros parâmetros do modelo devem ser integrados (eliminados), fato que raramente é possível usando os métodos numéricos padrões. O procedimento iterativo da amostragem de Gibbs refere-se a uma técnica de integração estocástica que cria uma cadeia de Markov, que é uma distribuição (conjunta a posteriori) estacionária associada à distribuição a posteriori de interesse. Tomando-se amostras, iterativamente, das distribuições condicionais a posteriori, com contínua atualização, obtém-se a distribuição conjunta a posteriori em equilíbrio e, após um número de iterações suficientemente grande, a última amostra desta seqüência e qualquer amostra subsequente é uma amostra da distribuição conjunta a posteriori. Este resultado implica que cada

coordenada do vetor de amostras retiradas, $\theta^n = [\beta^n \ a^n \ c^n \ \sigma_a^{2(n)} \ \sigma_c^{2(n)} \ \sigma_e^{2(n)}]$, é uma amostra da distribuição marginal a posteriori apropriada.

Para a análise Bayesiana empregou-se o software MTGSAM (Tassell & Vleck, 1995) e para a análise clássica empregou-se o software MTDFREML (Boldman et al., 1995). Foram avaliados vários tamanhos de cadeia de Markov e intervalos entre amostragens.

Para análise de convergência do processo de estimação foram consideradas cadeias de tamanhos 10.000, 15.000 e 20.000 ciclos, intervalos de 50 e 100 entre amostras dos parâmetros de interesse e período de descarte de 2.000 ciclos. Alguns resultados referentes aos parâmetros genéticos são apresentados na Tabela 1.

Tabela 1. Resultados comparativos entre algumas estratégias de amostragem de Gibbs, empregadas para estimação (médias) dos parâmetros genéticos herdabilidade (h^2) e correlação de ambiente comum (c^2).

Tamanho da cadeia	Intervalo entre amostras	h^2	c^2
10.000	100	0,34	0,032
20.000	100	0,34	0,032
15.000	50	0,38	0,029
20.000	50	0,36	0,030

Os resultados apresentados na Tabela 1, juntamente com outros não apresentados, permitiram concluir que a convergência do processo de amostragem ocorreu com intervalos de 100 entre amostras e com tamanho de cadeia de 10.000 ciclos (visto que a cadeia de 10.000 ciclos encontra-se dentro da cadeia de 20.000 ciclos e que ambas conduziram a resultados idênticos).

Outra forma de análise de convergência refere-se à estimação do erro de Monte Carlo, que é uma estatística associada ao erro de estimação de determinado parâmetro devido ao número de amostras utilizadas na cadeia de Gibbs, sendo que este erro é inversamente proporcional ao tamanho da cadeia (Tassel & Vleck, 1996). Este erro pode ser calculado através da variância dos parâmetros amostrados sucessivamente a cada intervalo dividida pelo número de amostras salvas, sendo que a raiz quadrada deste erro fornece uma aproximação para o desvio padrão do erro associado ao comprimento da cadeia.

Considerando a cadeia de tamanho 20.000 e intervalo de 100 entre amostras, o desvio padrão do erro associado ao tamanho da cadeia foi de 0,0004 para a herdabilidade (Tabela 2), resultado que revela que os 18.000 ciclos, após o período de descarte, foram suficientes para obtenção de precisas estimativas das médias a posteriori. Dessa forma, optou-se por apresentar os resultados referentes ao tamanho de cadeia de 20.000 ciclos e com amostras salvas a cada 100 ciclos. O tempo de processamento para esta análise foi de cerca de 10 horas em um computador Pentium 100 Mhz, com 16 MB de memória.

Na Tabela 2 são apresentadas as estimativas de parâmetros genéticos pelos procedimentos REML e GS.

Tabela 2. Estimativas descritivas das distribuições marginais a posteriori da herdabilidade (h^2), correlação devida ao ambiente comum (c^2) e variância fenotípica (σ_y^2) obtidas pelo procedimento da amostragem de Gibbs (GS), bem como estimativas destes parâmetros obtidas pelo método da máxima verossimilhança restrita (REML). Caráter diâmetro em *Pinus caribaea* var. *hondurensis*.

Parâmetro	GS				REML	
	Média \pm s*	Moda	Mediana	s**	Intervalo de Confiança (95%)	Estimativa
h^2	0,34 \pm 0,005	0,34	0,34	0,0004	0,33—0,35	0,32 \pm 0,06
c^2	0,032 \pm 0,003	0,032	0,032	0,0002	0,026—0,038	0,031 \pm 0,01
σ_y^2	8,2441 \pm 0,02	-	-	-	-	8,0624

s* = desvio padrão; s** = desvio padrão Monte Carlo.

Os resultados pontuais obtidos pelas abordagens freqüentista e Bayesiana foram similares (Tabela 2), conforme esperado. Uma vez que foram utilizadas distribuições a priori não informativas na análise Bayesiana, as modas das distribuições marginais a posteriori dos parâmetros genéticos foram similares às correspondentes estimativas REML. Do ponto de vista Bayesiano, as estimativas obtidas por REML correspondem às modas das distribuições conjuntas a posteriori dos componentes de variância, dada a utilização de prioris uniformes para os efeitos fixos e componentes de variância. A técnica GS possibilitou estimativas de componentes de variância pelo método VEIL (ou da verossimilhança integrada) apresentado por Gianola & Foulley (1990).

A grande vantagem da análise Bayesiana, neste caso, refere-se à obtenção dos desvios padrões e intervalos de confiança exatos para os parâmetros genéticos e valores genéticos preditos. Verificou-se que a técnica GS propiciou estimativas muito mais precisas que aquelas obtidas por REML.

Estimativas com precisões similares empregando o procedimento GS foram obtidas por Varona et al. (1994), para caracteres quantitativos em gado de leite e por Soria et al. (1998) em *Eucalyptus globulus*. Na Tabela 3 são apresentados os valores genéticos preditos dos melhores genitores, obtidos pelos procedimentos GS e BLUP, bem como os ganhos genéticos associados.

Tabela 3. Estimativas descritivas das distribuições marginais a posteriori dos valores genéticos preditos dos 10 melhores genitores, obtidos pelo procedimento da amostragem de Gibbs (GS), bem como valores genéticos preditos pelo procedimento BLUP. Caráter diâmetro em *Pinus caribaea* var. *hondurensis*.

Genitor	Média ± Desvio Padrão	Ordem (GS)	BLUP ± SEP*	Acurácia (BLUP)	Ordem (BLUP)
130	2,70 ± 0,03	1	2,61 ± 0,80	0,86	1
49	2,23 ± 0,07	2	2,15 ± 0,94	0,81	3
104	2,19 ± 0,06	3	2,26 ± 0,86	0,84	2
73	2,07 ± 0,07	4	1,96 ± 0,99	0,78	5
114	2,03 ± 0,09	5	1,96 ± 1,06	0,75	6
56	1,99 ± 0,07	6	1,89 ± 0,96	0,80	7
100	1,97 ± 0,03	7	1,97 ± 0,78	0,87	4
92	1,90 ± 0,03	8	1,82 ± 0,79	0,87	8
76	1,80 ± 0,05	9	1,81 ± 0,96	0,80	9
125	1,79 ± 0,03	10	1,69 ± 0,65	0,91	10

	GS	BLUP
Ganho genético (cm)	2,066 ± 0,04	(1,99 — 2,14)**
Ganho genético (%)	13,93 ± 0,27	(13,40 — 14,46)**

* desvio padrão do erro de predição.

** intervalo com 95% de confiança (probabilidade).

*** cálculo considerando uma acurácia média de 0,88.

As diferenças observadas nos valores genéticos preditos pelos procedimentos GS e BLUP/REML foram pequenas, conduzindo a uma ligeira alteração na ordem dos melhores pelos dois procedimentos (Tabela 3). Do ponto de vista Bayesiano, uma vez que nenhuma distinção existe entre efeitos fixos e aleatórios, as soluções das equações de modelo misto do BLUP correspondem às médias das distribuições marginais a posteriori dos parâmetros de locação (efeitos fixos e aleatórios), dados os componentes de variância, ou parâmetros de dispersão. As condições para que haja esta correspondência entre BLUP e GS para os parâmetros de locação são: atribuição de prioris uniformes (não informativas) para os efeitos fixos, prioris normais para os efeitos aleatórios e verossimilhança normal para o vetor de observações. Estas premissas foram atendidas neste trabalho, fato que explica os resultados obtidos.

Segundo Wang et al. (1994), uma deficiência do procedimento BLUP/REML é que os erros de estimação dos componentes de dispersão não são considerados por ocasião da predição de valores genéticos. A abordagem Bayesiana e a teoria de probabilidade a ela associada impõem que as inferências

devem ser baseadas nas distribuições marginais a posteriori dos parâmetros de interesse, de forma que toda a incerteza sobre os parâmetros é considerada completamente. Isto é conseguido a partir da distribuição conjunta a posteriori de todos os parâmetros. Desta distribuição, a distribuição marginal a posteriori do valor genético de um indivíduo é obtida por sucessivas integrações (eliminações) de todos os parâmetros de “nuisance” do modelo, tais como efeitos fixos, outros efeitos aleatórios (ambiente comum) e componentes de variância e parâmetros associados (h^2 e c^2 no caso).

O aspecto ressaltado por Wang et al. (1994) é muito relevante considerando, principalmente, que para que as propriedades desejáveis do BLUP sejam asseguradas, é necessário o conhecimento exato dos componentes de variância. Dessa forma, a análise Bayesiana, que permite estimativas quase exatas dos componentes de variância, tende a conduzir também a uma maior aproximação entre ganhos genéticos preditos e realizados com seleção. Este último aspecto pode, em parte, ser visualizado através dos desvios padrões dos ganhos genéticos apresentados na Tabela 3, sendo que o procedimento GS propiciou uma inferência muito mais segura (menor desvio padrão ou risco) sobre o ganho genético a ser capitalizado com a seleção dos 10 melhores genitores.

Os parâmetros genéticos iniciais utilizados foram obtidos pelo procedimento REML e as distribuições a priori foram tomadas como não informativas. Se as distribuições a priori para estes parâmetros fossem tomadas como informativas, menores desvios padrões para os parâmetros seriam obtidos. Quanto às estimativas pontuais (médias a posteriori), tem sido verificado que os valores obtidos são similares quando se consideram prioris muito informativas e não informativas (Tassell & Vleck, 1996).

As seguintes conclusões podem ser relatadas: (i) a análise Bayesiana propicia resultados adicionais àqueles obtidos pela abordagem freqüentista, destacando-se os intervalos de confiança Bayesianos para as estimativas de parâmetros genéticos, valores genéticos e ganhos genéticos; (ii) as estimativas dos parâmetros genéticos, valores genéticos e ganhos genéticos pelo procedimento da amostragem de Gibbs tendem a ser mais precisas do que pelo procedimento REML/BLUP; (iii) a análise Bayesiana é uma técnica elegante e flexível que permite a simultânea estimação dos parâmetros genéticos, efeitos fixos e valores genéticos de maneira precisa, mesmo para amostras de tamanho finito.

É importante relatar que o uso da abordagem Bayesiana em genética quantitativa não se restringe aos modelos lineares Gaussianos mas, engloba também os modelos lineares generalizados (Tempelman, 1998), os modelos não lineares e os modelos robustos (Gianola, 2000). De maneira geral, além da riqueza de informações propiciada, a análise Bayesiana oferece maior

flexibilidade quanto aos modelos de análise, visto que permite o ajuste de modelos de grande complexidade.

16. Referências Bibliográficas

- ARENDRONK, J.A.M. van ; ROSMEVLEN, C. van; JANSSE, L.L.G.; KNOL, E.F. Estimation of direct and maternal genetic (co)variances for survival within litters of piglets. **Livestock Production Science**, v.46, p.163-171, 1996.
- BARDORFF-NIELSEN, O. Plausibility inference. **Journal of the Statistical Society, Series B**, v.38, p.103-131, 1976.
- BAYES, T. An essay towards solving a problem in the doctrine of chances. **Philos. Trans. R. Soc. London**, v.53, 370-418, 1763.
- BIBBY, J.; TOUTENBURG, H. **Prediction and improved estimation in linear models**. New York: Wiley and Sons, 1977.
- BOLDMAN, K.G.; KRIESE, L.A.; VLECK, L.D. van.; KACHMAN, S.D. **A manual for use of MTDFREEM**: a set of programs to obtain estimates of variances and covariances. Washington: ARS / USDA, 1993. 120p.
- BOX, G.E.P.; TIAO, G.C. **Bayesian inference in statistical analysis**. Reading: Addison-Wesley Publ.Co., 1973.
- BULMER, M.G. **The mathematical theory of quantitative genetics**. Charedon Press, Oxford, 1980.
- DEMPFLE, L. Relation entre BLUP (Best linear unbiased prediction) et estimateurs bayesiens. **Ann. Genet. Sel. Anim.**, v.9, p.27-32, 1977.
- DEMPFLE, L. Estimation of breeding values. In: HILL, W.G.; MACKAY, T.F.C., eds. **Evolution and animal breeding**. Wallingford: CAB International, , 1989, p.181-188.
- EDWARDS, A.W.F. **Likelihood**. Cambridge: Cambridge University Press, 1972.
- EFRON, B. Biased versus unbiased estimation. **Advances in Mathematics**, v.16, p.259-277, 1975.
- EFRON, B. Why isn't everyone a Bayesian? **American Statistician**, v.40, p.11, 1986.
- EFRON, B.; MORRIS, C. Stein's paradox in statistics. **Scientific American**, v.236, n.5, p.119-127, 1977.
- FISHER, R.A. On the mathematical foundations of theoretical statistics. **Phil. Trans. R. Soc. Lond. Serie A**, n.222, p.309-368, 1922.

- FRIES, L.A.; SCHENKEL, F.S. Estimation and prediction under a selection model. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 30., 1993, Rio de Janeiro. **Anais dos Simpósios. Rio de Janeiro: Sociedade Brasileira de Zootecnia**, 1993. p.1-20.
- GAMERMAN, D. **Simulação estocástica via cadeias de Markov**. Caxambú: Associação Brasileira de Estatística, 1996. 196p.
- GAMERMAN, D. **Markov Chain Monte Carlo**: stochastic simulation for bayesian inference. Boca Raton: CRC Press, 1997.
- GAMERMAN, D.; MIGON, H.S. **Inferência Estatística**: uma abordagem integrada. Rio de Janeiro: Instituto de Matemática / UFRJ, 1993. 207p.
- GARCIA-CORTES, L.A.; SORENSEN, D. On a multivariate implementation of the Gibbs samples. **Genetique, Selection, Evolution**, v.28, p.121-126, 1996.
- GELFAND, A.E.; SMITH, A.F.M. Sampling-based approaches to calculating marginal densities. **Journal of the American Statistical Association**, v.85, p.398-409, 1990.
- GEMAN, S.; GEMAN, D. Stochastic relaxation, gibbs distribution and the Bayesian restoration of imagens. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.6, p.721-741, 1984.
- GIANOLA, D. Statistics in animal breeding. **Journal of the American Statistical Association**, v.95, p.296-299, 2000.
- GIANOLA, D. Inferences about best, worst and ordered genetic values in a small population with unknown variance. In: CONGRESSO NACIONAL DE GENÉTICA, 41., 1995, Caxambú. **Anais**. Ribeirão Preto: Sociedade Brasileira de Genética, 1995. Palestra.
- GIANOLA, D.; FERNANDO, R.L. Bayesian methods in animal breeding theory. **Journal of Animal Science**, v.63, p.217-244, 1986.
- GIANOLA, D.; FERNANDO, R.L.; IM, S.; FOULLEY, J.L. Likelihood estimation of quantitative genetic parameters when selection occurs: models and problems. **Genome, Ottawa**, v.31, p.768-777, 1989.
- GIANOLA, D.; FOULLEY, J.L. Variance estimation from integrated likelihood (VEIL). **Genetics Selection Evolution**, v.22, p.403-417, 1990.
- GIANOLA, D.; IM, S.; MACEDO, F.W. A framework for prediction of breeding value. In: GIANOLA, D.; HAMMOND, K., ed. **Advances in statistical methods for genetic improvement of livestock**. Berlin: Springer Verlag, 1990. p.210-238.

- GOFFINET, B. Selection on selected records. **Genetique, Selection, Evolution**, v.15, p.91, 1983.
- GOMES, M.I. Verossimilhança e inferencia estatística. In: COLÓQUIO DE ESTATÍSTICA E INVESTIGAÇÃO OPERACIONAL, 2., 1981. **Actas**. Rio de Janeiro, 1981. p.230-246.
- HARVILLE, D.A. Maximum likelihood approaches to variance component estimation and to related problems. **Journal of the American Statistics Association**, Washington, v.72, p.320-328, 1977.
- HENDERSON, C.R. **Aplications of linear models in animal breeding**. Guelph: University of Guelph, 1984. 462p.
- HENDERSON, C.R. Avaliação de vacas e touros. In: SIMPÓSIO BRASILEIRO DE MELHORAMENTO GENÉTICO DE BOVINOS, 1, 1982. **Anais**. Coronel Pacheco: EMBRAPA-CNPGL, 1982a. p.137-168.
- HENDERSON, C.R. Best linear unbiased estimation and prediction under a selection model. **Biometrics**, v.31, p.423-449, 1975.
- HENDERSON, C.R. Best linear unbiased prediction in populations that have undergone selection. In: BARTON, R.A.; SMITH, W.C., ed. **Proceedings of the World Congress on Sheep and Beef Cattle Breeding**. Palmerston North: Dunmore Press 1982b, v.1, p.191-200.
- HENDERSON, C.R. Sire evaluation and genetic trends. In: Animal Breeding and Genetics Symposium in Honor of J. Lush, 1973, Champaign. **Proceedings**. Champaign: American Society of Animal Science, 1973. 1973, p.10-41.
- HENDERSON, C.R.; KEMPTHORNE, O.; SEARLE, S.R.; VON KROSIGH, C.M. The estimation of environmental and genetic trends from records subject to culling. **Biometrics**, v.15, p.192, 1959.
- JAMES, W.; STEIN, C. Estimation with quadratic loss. **Proceedings of the Fourth Barkeley Symposium on Mathematical Statistics and Probability**, v.1, p.361-379, 1961.
- JEFFREYS, H. **Theory of Probability**. Oxford: Clarendon Press, 1961.
- JENSEN, J.; WANG, C.S. ; SORENSEN, D.A.; GIANOLA, D. Bayesian inference on variance and covariance components for traits influenced by maternal and *direct genetic effects using the Gibbs sampler*. **Acta Agriculturae Scandinavica**, v.44, p.192-201, 1994.
- LANGE, K.; SINSHEIMER, J.S. Normal/independent distributions and their applications in robust regression. **Journal of the American Statistical Association**, v.2, p.175-198, 1993.

- LERNER, I.M. **The genetic basis of selection**. New York: John Wiley & Sons, 1957. 298p.
- LINDLEY, D.W. **Introduction to probability and statistics**. Cambridge: Cambridge University Press, 1965.
- LINDLEY, D.V. The Bayesian approach. **Scandinavian Journal of Statistics**, v.5, p.1-26, 1978.
- LINDLEY, D.V.; SMITH, A.F.M. Bayes estimates for the linear model. **Journal of the Royal Statistical Society, Series B**, v.34, p.1-41, 1972.
- LINDSEY, J.K. Some statistical heresies. *Journal of the Royal Statistical Society, Series B*, v.61, p. 1- 26 , 1999.
- MAGNABOSCO, C.V. **Estimativas de parâmetros genéticos em características de crescimento de animais** da raça Nelore usando os métodos REML e Amostragem de Gibbs. Ribeirão Preto: USP / FMRP. 1997. 83p. Tese Doutorado.
- MOOD, A.M.; GRAYBILL, F.A.; BOES, D.C. **Introduction to the theory of statistics**. McGraw-Hill, 1974.
- MRODE, R.A. **Linear models for the prediction of animal breeding values**. Wallingford: CAB International, 1996, 187p.
- MURTEIRA, B.J.F. **Estatística: inferência e decisão**. Lisboa: Imprensa Nacional, Casa da Moeda, 1988.
- MURTEIRA, B.J.F. **Probabilidade e estatística: inferência estatística**. 2.ed. Lisboa: McGraw-Hill, 1990. v.2.
- PATTERSON, H.D.; THOMPSON, R. Recovery of inter-block information when block sizes are unequal. **Biometrika**, London, v.58, p.545-554, 1971.
- PEREIRA, B.B. Inferência verossimilhança. **Boletim da Associação Brasileira de Estatística**, v.38, p.31-42, 1997.
- REKAYA, R. **Análisis Bayesiano de datos de producción en los días de control para la selección de caracteres lecheros**. Madrid: Universidad Politécnica de Madrid. 1997. Doctoral Tesis.
- RESENDE, M.D.V. de. Avanços da genética biométrica florestal. In: BANDEL, G.; VELLO, N.A.; MIRANDA FILHO, J.B., eds. In: ENCONTRO SOBRE TEMAS DE GENÉTICA E MELHORAMENTO: Genética Biométrica Vegetal, 14, 1997, Piracicaba. Anais... Piracicaba: ESALQ, 1997. p.20-46.
- RESENDE, M.D.V. de. **Predição de valores genéticos, componentes de variância, delineamentos de cruzamento e estrutura de populações no melhoramento florestal**. Curitiba: Universidade Federal do Paraná, 1999. 434p. Tese Doutorado.

- ROBERTSON, A. Prediction equations in quantitative genetics. **Biometrics**, v.11, p.95-98, 1955.
- RODRIGUEZ, M.C.; TORO, M.; SILIO, L. Selection on lean growth in a nucleus of Landrace pigs: an analysis using Gibbs sampling. **Animal Science**, v.63, p.243-253, 1996.
- ROGERS, W.H.; TUKEY, J.W. Understanding some long-tailed distributions. **Statistica Neerlandica**, v.26, p.211-226, 1972.
- RONNINGEN, K. Some properties of the selection index derived by "Henderson's mixed model method". **Z. Tierz Zuchtungsbiol**, v.88, p.186, 1971.
- RONNINGEN, K.; VLECK, L.D. van. Selection index theory with practical applications. In: CHAPMAN, A.B. (ed.). **World animal science A4: general and quantitative genetics**. Oxford: Elsevier Science Publ., 1985, p.187-225.
- ROSA, G.J.M. Robust mixed linear models in quantitative genetics: Bayesian analysis via Gibbs sampling. In: LOPES, P.S.; EUCLYDES, R.F.; TORRES, R.A.; GUIMARÃES, S.E.F., eds. **International Symposium on Animal Breeding and Genetics**. Viçosa: Universidade Federal de Viçosa, 1999. p.133-159.
- SAIO, K.; HAYASHI, T. Bayesian estimation of genetic parameters. **Japanese Journal of Breeding**, v.40, p.63-75, 1990.
- SMITH, A.F.M. Bayesian statistics. Present position and potencial developments: some personal views. **Journal of the Royal Statistical Society, Séries A**, v.147, p.245-259, 1984.
- SORENSEN, D.A. **Gibbs Sampling in Quantitative Genetics**. Denmark: Danish Institute of Animal Science, Department of Breeding and Genetics. 1996. 186p. (Intern Report, 82).
- SORENSEN, D.A.; KENNEDY, B.W. The use of the relationship matrix to account for genetic drift variance in the analysis of genetic experiments. **Theoretical and Applied Genetics**, v.66, p.217-220, 1983.
- SORENSEN, D.A.; WANG, C.S.; JENSEN, J.; GIANOLA, D. Bayesian analysis of genetic change due to selection using Gibbs sampling. **Genetics, Selection, Evolution**, v.26, p.333-360, 1994.
- SORIA, F.; BASURCO, F.; TOVAL, G.; SILIO, L.; RODRIGUEZ, M.C.; TORO, M. Na application of Bayesian techniques to the genetic evaluation of growth traits in *Eucalyptus globulus*. **Canadian Journal of Forestry Research**, v.28, p.1286-1204, 1998.
- STEEL, R.G.D.; TORRIE, J.H. **Principles and procedures of statistics**. 2.ed. New York: McGraw-Hill, 1980. 633p.

- STEIN, C. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. **Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability**, v.1, p.197-206, 1955.
- STRANDEN, I.J. **Robust mixed effects linear models with t distributions and application to dairy cattle breeding**. Madison: University of Wisconsin, 1996. Ph.D. Thesis.
- SUN, L. et al. Bayesian methods for variance component models. **J. Am. Stat. Assoc.**, v.91, n.434, p.743-52, 1996.
- TASSELL, C.P. van; CASELLA, G.; POLLAK, E.J. Effects of selection on estimates of variance components using Gibbs sampling and restricted maximum likelihood. **Journal of Dairy Science**, v.78, p.678-692, 1995.
- TASSELL, C.P. van ; VLECK, L.D. van. **A manual for use of MTGSAM: a set of FORTRAN programs to apply Gibbs sampling to animal models for variance component estimation**. Washington: USDA / ARS, 1995. 82p.
- TASSELL, C.P. van; VLECK, L.D. van. Multiple-trait Gibbs sampler for animal models: flexible programs for Bayesian and likelihood-based covariance component inference. **Journal of Animal Science**, v.74, p.2586-2597, 1996.
- TEMPELMAN, R.J. Generalized linear mixed models in dairy cattle breeding. **J. Dairy Sci.**, v.81, p.1428-1444, 1998.
- THOMPSON, R. Recent developments in the estimation of variance components and their application to the estimation of genetic parameters. In: BARTON, R.A.; SMITH, W.C., ed. **Proceedings of the World Congress on Sheep and Beef Cattle Breeding**. Palmerston North, Dunmore Press, 1982. v.1, p.217-224.
- THOMPSON, R. Sire evaluation. **Biometrics**, v.35, p.339-353, 1979.
- TORO, M.A.; PRUNONOSA, D.V. The use of prior information in the estimation of heritability by parent-offspring regression. **Genetique, Selection, Evolution**, v.16, p.177-184, 1984.
- VARONA, L. **Aplicaciones del muestreo de Gibbs en modelos de Genética Cuantitativa: analisis de un caso de heterogeneidad de varianzas**. Local: Universidad de Zaragoza, 1994. PhD. Thesis.
- VARONA, L.; MORENO, C.; GARCIA-CORTES, L.A.; ALTARRIBA, J. Estimación multicarácter de componentes de varianza y covarianza en vacuno lechero mediante muestreo de Gibbs. **Revista Portuguesa de Zootecnia**, v.1, p.185-195, 1994.
- VLECK, L.D. van. **Selection Index and Introduction to Mixed Model Methods**. Boca Raton: CRC Press, 1993. 512p.

- VLECK, L.D. van; POLLAK, E.J.; OLTENACU, E.A.B. **Genetics for the animal sciences**. New York: W.H. Freeman, 1987. 391p.
- WANG, C.S.; GIANOLA, D.; SORENSEN, D.A.; JENSEN, J.; CHRISTENSEN, A.; RUTLEDGE, J.J. Response to selection for litter size in Danish Landrace pigs: a Bayesian analysis. **Theor. Appl. Genet.**, v.88, p.220, 1994a.
- WANG, C.S.; RUTLEDGE, J.J.; GIANOLA, D. Bayesian analysis of mixed linear models via Gibbs sampling with an application to litter size in Iberian pigs. **Genetics Selection Evolution**. v.26, p.91, 1994b.
- 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, 1993.
- WEIGEL, K.A.; GIANOLA, D. A computationally simple bayesian method for estimation of heterogeneous within-herd phenotypic variances. **Journal of Dairy Science**, v.76, p.1455-1465, 1993.
- ZELLNER, A. **An introduction to Bayesian inference in econometrics**. New York: J. Wiley and Sons, 1971.