Avaliação da razão de verossimilhança generalizada em teste de hipóteses sobre o controle genético de características contínuas

Evaluation of the generalized likelihood ratio test of hypotheses on the genetic control of continuous traits

Renata Pires Gonçalves[1]; Eduardo Bearzoti[2]; Daniel Furtado Ferreira[2]

Resumo

Estudos de herança genética em plantas são realizados para caracterizar os efeitos genéticos e verificar a existência de um gene de efeito maior e/ou de genes de pequeno efeito ("poligenes"). Quando a característica de interesse é contínua, a verossimilhança é baseada em modelos de misturas de densidades normais. Uma vez que não há testes exatos evidentes para julgar a existência de um gene de efeito maior, a razão de verossimilhança generalizada é em geral utilizada, considerando a aproximação de qui-quadrado. Este trabalho objetivou avaliar esta estatística de teste através de simulação em computador. Dados foram simulados, considerando particularidades de genealogia típicas de tais estudos, e duas condições sob a hipótese de nulidade, ou seja, sem a presença de um gene de efeito maior e sem a presença de genes de pequeno efeito (poligenes), para avaliar o controle do erro tipo I. O poder do teste foi avaliado com ambos presentes. No processo de simulação, foram variados o tamanho de amostra e valores do coeficiente de herdabilidade. Resultados indicaram que, embora a distribuição empírica da razão de verossimilhança tenha se desviado significativamente da distribuição de qui-quadrado, houve controle do erro tipo I, considerando um nível de significância nominal de 5%. O poder é elevado para detectar poligenes e gene de efeito maior, em geral. O poder é baixo para detectar gene de efeito maior quando a proporção da variação genética explicada por este é baixa.
Palavras-chave: Misturas de normais, gene de efeito maior, simulação de dados.

Abstract

Studies of genetic control in plants are carried out to characterize genetic effects and detect the existence of a major gene and/or genes of minor effects (polygenes). If the trait of interest is continuous, the likelihood can be constructed based on a model with mixtures of normal densities. Once exact tests are not evident with such models, the likelihood ratio test is generally used, using the chi-square approximation. This work aimed at evaluating such test statistic using computer simulation. Data sets were simulated using generations typical in plant studies, under two conditions of null hypothesis, without a major gene, and without polygenes. The power of the test was evaluated with both types of genes present. Different sample sizes and values of heritability were considered. Results showed that, although the empirical densities of the test statistic departed significantly from a chi-square distribution, under null hypotheses, there was a reasonable control of type I error, with a significance level of 5%. The power of the test was generally high to detect polygenes and major genes. Power is low to detect a major gene only when it explains a low fraction of genetic variation.
Key words: Mixtures of normal, major gene, computer simulation.

[1] Prof. Adjunta, Colégio de Aplicação da UFV – Coluni, Viçosa-MG
[2] Prof. Adjunto, Dpto Ciências Exatas – UFLA, Lavras-MG

Recebido para publicação 19/04/04 Aprovado em 25/11/04

Introdução

Estudos de herança genética em plantas são realizados para caracterizar os efeitos genéticos associados a características econômicas de interesse. Tais estudos são particularmente úteis em características contínuas, para verificar a existência de um gene de efeito maior e/ou de genes de pequeno efeito ("poligenes").

Existindo um gene de efeito maior, é comum referir-se à herança como "monogênica". Em tais casos, é interessante estimar o valor genotípico médio de cada uma das três classes genotípicas, a saber, dos dois homozigotos e do heterozigoto.

A estimação de tais parâmetros, bem como o teste da hipótese de herança monogênica, têm em geral sido feitos utilizando dados observados em plantas das linhagens contrastantes "P_1" e "P_2", da geração "F_1" (cruzamento entre P_1 e P_2) e "F_2" (cruzamento entre plantas F_1), e dos retrocruzamentos "RC_1" (entre indivíduos F_1 e P_1) e "RC_2" (entre indivíduos F_1 e P_2). A dificuldade estatística inerente a estes estudos é a existência de misturas de distribuições (em geral normais), para as quais os estimadores não são evidentes.

Um método de estimação adequado para tanto é o da máxima verossimilhança, usado por Lynch e Walsh (1998) e Silva (2003) em problemas de misturas de distribuições em genética.

Silva (2003) descreve um modelo estatístico bastante parametrizado envolvendo misturas de densidades normais, o que dificulta a construção de testes exatos para julgar hipóteses de interesse. Assim, testes de razão de verossimilhanças podem ser empregados para tal, considerando a aproximação de qui-quadrado. Convém salientar que a construção de hipóteses a serem testadas depende do interesse do pesquisador. Assim sendo, pode-se exemplificar algumas hipóteses de nulidade correspondentes aos seguintes modelos:

(i) Contendo efeitos ambientais e poligênicos (sem gene de efeito maior).

(ii) Contendo efeitos ambientais e um gene de efeito maior (sem genes de pequeno efeito, ou seja, poligenes).

Por se tratar de uma distribuição aproximada, estudos de simulação são interessantes para obter conhecimento acerca de sua qualidade (quanto a níveis de significância reais e poder do teste) sob diferentes circunstâncias, em termos de tamanho da amostra, número de genes e herdabilidade.

Este trabalho objetivou avaliar a qualidade do teste da razão de verossimilhanças em termos do controle do erro tipo I, bem como avaliar o poder do teste, sob diferentes condições de tamanho de amostra e magnitude de parâmetros genéticos.

Metodologia

O teste da razão de verossimilhanças estudado neste trabalho é baseado em modelos probabilísticos proposto por Silva (2003) (Tabela 1) para estudos de herança genética utilizando dados das gerações P_1, P_2, F_1, F_2, RC_1 e RC_2 , anteriormente citados.

Para cada modelo descrito anteriormente a descrição dos parâmetros são definidos como: σ^2 variância ambiental; m uma constante de referência; Ao efeito aditivo do gene de efeito maior; Do efeito de dominância do gene de efeito maior; [a] é o afastamento provocado pelos poligenes da média dos pais homozigotos da linhagem com maior expressão do caráter (componente de média poligênico aditivo); [d] é o afastamento devido aos poligenes da média dos pais heterozigotos (componente de média poligênico de dominância); V_A é a variância poligênica aditiva; V_D a variância poligênica de dominância e S_{AD} o componente da variação relativa aos produtos dos efeitos poligênicos aditivos pelos efeitos poligênicos de dominância (MATHER; JINKS, 1984).

Tabela 1. Modelos genéticos e seus respectivos parâmetros, adaptado de Silva (2003)

MODELO	Herança	Gene Maior	Poligenes	PARÂMETROS
1	Maior+Poligenes	dom. e adit.	dom. e adit.	μ , A, D, [a], [d], V_A, V_D, S_{AD}, σ^2
2	Maior+Poligenes	dom. e adit.	adit.	μ , A, D, [a], V_A, σ^2
3	Maior+Poligenes	adit.	dom. e adit.	μ , A, [a], [d], V_A, V_D, S_{AD}, σ^2
4	Maior+Poligenes	adit.	adit.	μ , A, [a], V_A, σ^2
5	Só Poligenes	—	dom. e adit.	μ , [a], [d], V_A, V_D, S_{AD}, σ^2
6	Só Poligenes	—	adit.	μ , [a], V_A, σ^2
7	Só Gene Maior	dom. e adit.	—	μ , A, D, σ^2
8	Só Gene Maior	adit.	—	μ , A, σ^2
9	Nenhum	—	—	μ , σ^2

Os efeitos genéticos (tanto maior quanto poligênicos) apresentam dominância, ou seja, um desvio entre o heterozigoto e a média dos homozigotos. Estatisticamente, a existência de dominância implica em parâmetros adicionais no modelo. Sob normalidade, tem-se uma única densidade em cada uma das gerações P_1, P_2 e F_1, enquanto que nos retrocruzamentos e geração F_2 ocorrem misturas de duas e três densidades normais, respectivamente. Representando a densidade normal por $\varphi(\cdot)$, tem-se as seguintes densidades para cada uma das gerações acima:

$$\begin{array}{ll} P_1\colon\ \varphi(\mu_1,\sigma_1^2) & RC_1\colon\ \frac{1}{2}\varphi(\mu_{41},\sigma_4^2)+\frac{1}{2}\varphi(\mu_{42},\sigma_4^2) \\ P_2\colon\ \varphi(\mu_2,\sigma_2^2) & RC_2\colon\ \frac{1}{2}\varphi(\mu_{51},\sigma_5^2)+\frac{1}{2}\varphi(\mu_{52},\sigma_5^2) \\ F_1\colon\ \varphi(\mu_3,\sigma_3^2) & F_2\colon\ \frac{1}{4}\varphi(\mu_{61},\sigma_6^2)+\frac{1}{2}\varphi(\mu_{62},\sigma_6^2)+\frac{1}{4}\varphi(\mu_{63},\sigma_6^2) \end{array} \quad (1)$$

As médias e variâncias das densidades acima, por sua vez, podem ser expressas como:

$$\mu_1 = \mu - [a] - A$$
$$\mu_3 = \mu + [d] + D$$
$$\mu_{42} = \mu - \frac{[a]}{2} + \frac{[d]}{2} + D$$
$$\mu_{52} = \mu + \frac{[a]}{2} + \frac{[d]}{2} + D$$
$$\mu_{62} = \mu + \frac{[d]}{2} + D$$

$$\mu_2 = \mu + [a] + A$$
$$\mu_{41} = \mu - \frac{[a]}{2} + \frac{[d]}{2} - A$$
$$\mu_{51} = \mu + \frac{[a]}{2} + \frac{[d]}{2} + A$$
$$\mu_{61} = \mu + \frac{[d]}{2} - A$$
$$\mu_{63} = \mu + \frac{[d]}{2} + A$$

$$\sigma_1^2 = \sigma_2^2 = \sigma_3^3 = \sigma^2$$
$$\sigma_4^2 = \sigma^2 + \frac{V_A}{2} + V_D - S_{AD}$$
$$\sigma_5^2 = \sigma^2 + \frac{V_A}{2} + V_D + S_{AD}$$
$$\sigma_6^2 = \sigma^2 + V_A + V_D$$

Dados foram simulados para as seis gerações, conforme as densidades em (1), e sob H_0, em duas situações: (A) Sem a presença de um gene de efeito maior e (B) Sem a presença de genes de pequeno efeito, ou seja, poligenes.

Para cada uma das situações acima citadas, foi avaliada a qualidade do teste da razão de verossimilhanças em termos do controle do erro tipo I, o poder do mesmo foi avaliado com ambos (poligenes e gene de efeito maior) presentes. Preliminarmente, simulou-se amostras de tamanho 50, 50, 50, 150, 150 e 200 observações para as gerações P_1, P_2, F_1, RC_1, RC_2 e F_2 respectivamente, que consistem em tamanhos de amostras usuais no melhoramento genético; de maneira análoga, também foram simuladas 25, 25, 25, 75, 75 e 100 observações para as gerações acima citadas, correspondendo à metade daqueles tamanhos usuais. Admitiu-se, sem perda de generalidade, que $\mu = 1000$ e s^2=1000. Além disso, estabeleceu-se valores de herdabilidade iguais a 0,2 e 0,7. A obtenção dos demais parâmetros é descrita abaixo.

Comparação entre Modelos 1 e 5 sob H_0

Na primeira situação (A) atribuiu-se valor igual a zero aos parâmetros **A** e **D**. Optou-se, neste estudo, por admitir dominância completa, ou seja, efeitos aditivos iguais aos de dominância. Assim, para o poligene i tem-se $a_i = d_i$. Seja a variância aditiva V_A expressa por $V_A = \sum_{i=1}^{g} \frac{a_i^2}{2}$. Neste estudo, admitiu-se arbitrariamente um número de poligenes g igual a 100 e assim $a_1 = a_2 = ... = a_{100} = a$. Assim, tem-se que: $V_A = \frac{100}{2}a^2$ e portanto $a = \frac{\sqrt{2V_A}}{10}$. A variância de dominância, representada por V_D, é expressa por $V_D = \sum_{i=1}^{g} \frac{d_i^2}{4}$. Com as pressuposições acima, e substituindo-se d por a, tem-se que $V_D = \frac{100}{4}a^2$, logo $V_A = 2V_D$.

Definindo-se h_P^2 como o coeficiente de herdabilidade dos poligenes, sua expressão é dada por $h_P^2 = \frac{V_A + V_D}{\frac{A^2}{2} + \frac{D^2}{4} + V_A + V_D + \sigma^2}$, ou seja, a fração da variância total que é explicada pelos poligenes. Substituindo-se $V_A = 2V_D$, A = D = 0 e $\sigma^2 = 1000$ na expressão de h_P^2 tem-se: ,

$$h_P^2 = \frac{2V_D + V_D}{2V_D + V_D + 1000} \Rightarrow 1000h_P^2 + 3V_Dh_P^2 = 3V_D \Rightarrow (3h_P^2 - 1)V_D = -1000h_P^2$$

logo, as variâncias de dominância e aditiva são dadas respectivamente por $V_D = \frac{\sigma^2 h_P^2}{3(1 - h_P^2)}$ e $V_A = \frac{2\sigma^2 h_P^2}{3(1 - h_P^2)}$.

A expressão de [a] é dada por $[a] = \sum_{i=1}^{g} a_i$; logo $[a] = 100a \Rightarrow [a] = 100\frac{\sqrt{2V_A}}{10} \Rightarrow [a] = 10\sqrt{2V_A}$. Sabendo-se que [d] é a soma dos efeitos de dominância dos poligenes e que [a] = [d] (dominância completa), tem-se que: $[d] = \sum_{i=1}^{g} d_i$, logo $[d] = 10\sqrt{2V_A}$.

Como [a] = [d], tem-se que o componente da variação relativa aos produtos dos efeitos poligênicos aditivos pelos efeitos poligênicos de dominância S_{AD} é dado como sendo o oposto da variância aditiva, ou seja, $S_{AD} = -V_A$.

Considerou-se um valor pequeno e um valor alto para h_P^2, conforme detalhado a seguir.

• Valores dos parâmetros sob $h_P^2 = 0,2$

Para $h_P^2 = 0,2$ tem-se que as variâncias aditiva e de dominância, o componente da variação relativa aos produtos dos efeitos poligênicos aditivos pelos efeitos poligênicos de dominância e a soma dos efeitos aditivos e de dominância dos poligenes valem, respectivamente: $V_A = 166,67$;

$V_D = 83,335$; $S_{AD} = -166,67$ e [a] = [d] = 182,576.

• Valores dos parâmetros sob $h_P^2 = 0,7$

Para $h_P^2 = 0,7$ tem-se que as variâncias aditiva e de dominância, o componente da variação relativa aos produtos dos efeitos poligênicos aditivos pelos efeitos poligênicos de dominância e a soma dos

efeitos aditivos e de dominância dos poligenes valem, respectivamente: $V_A = 1555,56$;

$V_D = 777,78$; $S_{AD} = -1555,56$ e [a] = [d] = 557,774.

Comparação entre Modelos 1 e 7 sob H_0

Na situação (B) atribuiu-se valor igual a zero para os parâmetros [a], [d], V_A, V_D e S_{AD}.

Definindo-se h^2_M como o coeficiente de herdabilidade do gene de efeito maior, sua expressão é dada por $h^2_M = \dfrac{\frac{A^2}{2} + \frac{D^2}{4}}{\frac{A^2}{2} + \frac{D^2}{4} + V_A + V_D + \sigma^2}$, com variância genética de referência em $F_2 = \frac{A^2}{2} + \frac{D^2}{4}$. Supondo-se que o efeito aditivo e o efeito de dominância do gene de efeito maior são iguais e substituindo A = D , $V_A = 2V_D = 0$ e $\sigma^2 = 1000$ tem-se:

$$h^2_M = \frac{\frac{A^2}{2} + \frac{A^2}{4}}{\frac{A^2}{2} + \frac{A^2}{4} + \sigma^2} \Rightarrow$$

$$\Rightarrow h^2_M = \frac{\frac{3}{4}A^2}{\frac{3}{4}A^2 + 1000} \Rightarrow$$

$$\Rightarrow \frac{3}{4}A^2 h^2_M + 1000h^2_M = \frac{3}{4}A^2 \Rightarrow$$

$$\Rightarrow \frac{3}{4}A^2(1 - h^2_M) = 1000h^2_M \Rightarrow$$

$$\Rightarrow A^2 = \frac{4}{3}\cdot 1000 \cdot \frac{h^2_M}{1 - h^2_M}.$$

Dessa forma tem-se que o efeito aditivo do gene de efeito maior é dado por $A = 2\sqrt{\frac{1000h^2_M}{3(1 - h^2_M)}}$ e como A = D tem-se que $D = 2\sqrt{\frac{1000h^2_M}{3(1 - h^2_M)}}$.

Da mesma maneira, dois valores de h^2_M foram considerados, 0,2 e 0,7.

• Valores dos parâmetros sob $h^2_M = 0,2$

Para $h^2_M = 0,2$ tem-se que os efeitos aditivo e de dominância do gene de efeito maior são dados por: A = D = 18,257.

• Valores dos parâmetros sob $h^2_M = 0,7$

Para $h^2_M = 0,7$ tem-se que os efeitos aditivo e de dominância do gene de efeito maior são dados por: A = D = 55,777.

Cálculo dos parâmetros para avaliação do poder do teste

Considerou-se a partição $h^2 = h^2_M + h^2_P$. Analogamente ao que foi descrito anteriormente, tem-se:

$$h^2_M = \frac{\frac{A^2}{2} + \frac{D^2}{4}}{\frac{A^2}{2} + \frac{D^2}{4} + V_A + V_D + \sigma^2}$$

$$h^2_P = \frac{V_A + V_D}{\frac{A^2}{2} + \frac{D^2}{4} + V_A + V_D + \sigma^2}$$

substituindo-se A = D e $V_A = 2V_D$ em h^2_M e h^2_P, obtém-se:

$$h^2_M = \frac{\frac{3}{4}A^2}{\frac{3}{4}A^2 + 3V_D + \sigma^2} \text{ e } h^2_P = \frac{3V_D}{\frac{3}{4}A^2 + 3V_D + \sigma^2}.$$

Da mesma forma, substituindo-se $V_D = \frac{\sigma^2 h^2}{3(1 - h^2)}$ em h^2_P tem-se:

$$h_P^2 = \frac{3V_D}{\frac{3}{4}A^2 + 3V_D + \sigma^2} \Rightarrow$$

$$\Rightarrow \frac{3}{4}A^2h_P^2 + 3V_Dh_P^2 + \sigma^2h_P^2 = 3V_D \Rightarrow$$

$$\Rightarrow 3(1 - h_P^2)V_D = \frac{3}{4}A^2h_P^2 + \sigma^2h_P^2 \Rightarrow$$

$$\Rightarrow V_D = \frac{\frac{3}{4}A^2h_P^2 + \sigma^2h_P^2}{3(1 - h_P^2)}.$$

Substituindo-se $V_D = \frac{\frac{3}{4}A^2h_P^2 + \sigma^2h_P^2}{3(1 - h_P^2)}$ em h_M^2, tem-se que:

$$h_M^2 = \frac{\frac{3}{4}A^2(1 - h_P^2)}{\frac{3}{4}A^2 + 3V_D + \sigma^2} \Rightarrow$$

$$\Rightarrow \frac{3}{4}A^2h_M^2 + \sigma^2h_M^2 = \frac{3}{4}A^2(1 - h_P^2) \Rightarrow$$

$$\Rightarrow \frac{3}{4}A^2\left[h_M^2 - (1 - h_P^2)\right] = -\sigma^2h_M^2 \Rightarrow$$

$$\Rightarrow A^2 = \frac{\sigma^2h_M^2}{\frac{3}{4}\left[(1 - h_P^2) - h_M^2\right]}.$$

Dessa forma tem-se que o efeito aditivo do gene de efeito maior é dado por:

$$A = 2\sqrt{\frac{\sigma^2h_M^2}{3\left[1 - (h_P^2 + h_M^2)\right]}}.$$

e substituindo-se

$$A = 2\sqrt{\frac{\sigma^2h_M^2}{3\left[1 - (h_P^2 + h_M^2)\right]}}$$

em V_D, obtém-se que a variância dos efeitos poligênicos de dominância é dado por:

$$V_D = \frac{\frac{\sigma^2h_M^2h_P^2}{\left[1 - (h_P^2 + h_M^2)\right]} + \sigma^2h_P^2}{3(1 - h_P^2)}.$$

Considerou-se combinações de valores 0,2 e 0,7 para h_M^2 e h_P^2, o que determinou valores específicos para os parâmetros, conforme detalhado a seguir.

• Valores dos parâmetros sob $h_P^2 = 0,2$ e $h_M^2 = 0,2$

Para $h_P^2 = 0,2$ e $h_M^2 = 0,2$ tem-se que os efeitos aditivos e de dominância do gene de efeito maior, as variâncias aditiva e de dominância, o componente da variação relativa aos produtos dos efeitos poligênicos aditivos pelos efeitos poligênicos de dominância e a soma dos efeitos aditivos e de dominância dos poligenes valem, respectivamente: A = D = 21,082 ; $V_A = 222,22$; $V_D = 111,11$; $S_{AD} = -222,22$ e [a] = [d] = 210,817.

• Valores dos parâmetros sob $h_P^2 = 0,2$ e $h_M^2 = 0,7$

Para $h_P^2 = 0,2$ e $h_M^2 = 0,7$ tem-se que os efeitos aditivos e de dominância do gene de efeito maior, as variâncias aditiva e de dominância, o componente da variação relativa aos produtos dos efeitos poligênicos aditivos pelos efeitos poligênicos de dominância e a soma dos efeitos aditivos e de dominância dos poligenes valem, respectivamente: A = D = 96,609; $V_A = 1333,334$; $V_D = 666,667$; $S_{AD} = -1333,334$ e [a] = [d] = 516,398.

• Valores dos parâmetros sob $h_P^2 = 0,7$ e $h_M^2 = 0,2$

Para $h_P^2 = 0,7$ e $h_M^2 = 0,2$ tem-se que os efeitos aditivos e de dominância do gene de efeito maior, as variâncias aditiva e de dominância, o componente da variação relativa aos produtos dos efeitos poligênicos aditivos pelos efeitos poligênicos de dominância e a soma dos efeitos aditivos e de dominância dos

poligenes valem, respectivamente:A = D = 51,639; V_A = 4666,666; V_D = 2333,333; S_{AD} = –4666,666 e [a] = [d] = 966,092.

Descrição do programa de simulação

Especificados valores para os parâmetros, os dados eram simulados de maneira usual, ou seja, sorteando-se valores entre 0 e 1, os quais eram usados na inversa da função distribuição de probabilidade normal. No caso de o dado pertencer a uma geração contendo um mistura de distribuições (um retrocruzamento ou F_2), então a simulação era feita em duas etapas. Primeiramente, sorteava-se de qual componente da mistura o indivíduo pertencia, conforme probabilidades apropriadas, obedecendo à proporção 1:1 nos retrocruzamentos e 1:2:1 em F_2. Em seguida, sorteava-se uma realização de variável aleatória normal, correspondente ao componente em questão, pelas técnicas usuais de simulação.

O ajustamento aos modelos foi feito da mesma maneira que Silva (2003), alternando o método numérico de Quase-Newton com o de Powell.

As rotinas computacionais necessárias foram escritas em linguagem Delphi, versão 6.

Cada conjunto de dados propiciou um valor de razão de verossimilhanças, dado por Mood, Graybill e Boes, (1974):

$$LR = -2 \ln \frac{L(M_0)}{L(M_1)}$$

sendo $L(M_j)$ a verossimilhança máxima sob o modelo M_j , com j = 0 ou 1.

Resultados e Discussões

Avaliação sob H_0, diferentes tamanhos de amostra e h^2

As tabelas 3 e 4, a seguir, ilustram os resultados obtidos para controle do erro tipo I, convencionando-se por "n" o conjunto de tamanhos de amostra iguais a 50, 50, 50, 200, 150 e 150 para as gerações P_1, P_2, F_1, F_2, RC_1 e RC_2, respectivamente, e coeficientes de herdabilidade iguais a 0,2 e 0,7. Na Tabela 2 obtém-se os resultados com amostra n, enquanto que na tabela 3 estão apresentados os resultados referentes à metade deste tamanho da amostra, ou seja, n/2, para os mesmos coeficientes de herdabilidade.

Tabela 2. Resultados sob H_0, com amostra de tamanho n e diferentes valores do coeficiente de herdabilidade h^2.

Modelos sob	Hipótese de Nulidade	h^2	Amostras Válidas	Taxa de Rejeição Empírica	Intervalo de Confiança	Kolmogorov Smirnov
1 versus 5	Só Poligenes	0,2	129	0,00775194	[0,00019624, 0,04243306]	0,7760165**
1 versus 5	Só poligenes	0,7	54	0,03703704	[0,00451736, 0,12747179]	0,5798769**
1 versus 7	Só gene maior	0,2	651	0,00000000	[0,00000000, 0,00459117]	0,3278636**
1 versus 7	Só gene maior	0,7	766	0,02088773	[0,01198498, 0,03369879]	0,2567153**

Tabela 3. Resultados sob H_0, com amostra de tamanho n/2 e diferentes valores do coeficiente de herdabilidade h^2.

Modelos sob Comparação	Hipótese de Nulidade	h^2	Amostras Válidas	Taxa de Rejeição Empírica	Intervalo de Confiança	Kolmogorov Smirnov
1 versus 5	Só poligenes	0,2	233	0,00429185	[0,00010865, 0,02367915]	0,7745634**
1 versus 5	Só poligenes	0,7	99	0,01010101	[0,00025570, 0,05499635]	0,6261338**
1 versus 7	Só gene maior	0,2	766	0,01436031	[0,00718964, 0,02554845]	0,3415536**
1 versus 7	Só gene maior	0,7	878	0,02505695	[0,01576805, 0,03769155]	0,241019**

Embora tenha-se simulado 1000 amostras em cada configuração, nem sempre todas eram aproveitadas em função de o critério de convergência não ser satisfeito, quando eventualmente o limite máximo de 1000 iterações por amostra era atingido. Isto refletiu em diferentes números de amostras válidas (Tabelas 2 e 3). Cada conjunto de valores de LR era submetido ao teste de Kolmogorov-Smirnov, para testar sua aderência à distribuição assintótica esperada de qui-quadrado com 2 graus de liberdade na situação 1 versus 5 e com 5 graus de liberdade na situação 1 versus 7. Além disso, a taxa empírica de rejeição de H_0 foi calculada, obtendo-se um intervalo de confiança exato para esta, baseado na distribuição binomial, para confrontar com o valor nominal de significância utilizado (5%).

Os resultados da simulação mostraram que os valores de razão de verossimilhanças apresentaram uma taxa de rejeição empírica inferior ao valor nominal do nível de significância (0,05). Pode-se dizer, contudo, que os valores com $h^2 = 0,7$, na comparação entre os modelos 1 e 5, foram estatisticamente iguais ao nível nominal, a julgar pelos intervalos de confiança.

As amostras de tamanho n apresentaram maior taxa de rejeição de H_0 (Tabela 2) em relação às amostras de tamanho n/2 (Tabela 3), à exceção da comparação 1 versus 7, sob $h^2 = 0,2$. No entanto, deve-se notar que o número de amostras válidas neste caso foi pequeno em relação aos demais da mesma configuração, e assim a taxa real de rejeição de H_0 foi estimada com menor precisão.

De qualquer maneira, a tendência geral de a taxa de rejeição de H_0 se aproximar do nível de significância nominal, com o aumento do tamanho da amostra, foi coerente com o esperado, por se tratar de uma distribuição assintótica de qui-quadrado. É possível que, com amostras ainda maiores, tais taxas se aproximassem ainda mais.

O coeficiente de herdabilidade apresentou uma tendência geral de apresentar uma menor taxa de rejeição de H_0, quando aquele era igual a 0,2, e uma maior taxa, mais próxima ao valor nominal, com $h^2 = 0,7$ (Tabelas 2 e 3).

Estes resultados em conjunto mostram que condições de menor precisão (menor tamanho de amostra e menor h^2) distanciam as taxas de rejeição de H_0 frente ao valor nominal de significância. No entanto, este distanciamento consistentemente se apresentou no sentido de reduzir a taxa de rejeição de H_0. Assim, pode-se dizer que, em condições menos favoráveis quanto à precisão, continua havendo um controle satisfatório do erro tipo I, ainda que à custa de um maior conservadorismo do teste.

Embora as densidades sob H_0 tenham sido estatisticamente diferentes das distribuições esperadas de qui-quadrado, pelo teste Kolmogorov-Smirnov (Tabelas 2 e 3), seria interessante confrontar, ainda que visualmente, as densidades empíricas da estatística LR com estas distribuições esperadas. Tais densidades estão apresentadas nas Figuras 1 e 2, representando apenas parte das situações, a título de ilustração. De fato, pode-se observar claras

semelhanças com a distribuição de qui-quadrado. Há uma evidente assimetria à direita nos casos apresentados. Além disso, esta assimetria é menos acentuada na comparação entre os modelos 1 e 7, com 5 graus de liberdade, em relação à comparação entre os modelos 1 e 5 (2 graus de liberdade). Isto foi coerente com o comportamento de uma qui-quadrado, a qual tende à simetria, à medida que o número de graus de liberdade tende ao infinito. As características apresentadas nas Figuras 1 e 2 são válidas para outras situações (dados não apresentados).

Assim, não obstante a significância do teste de Kolmogorov-Smirnov, as Figuras 1 e 2, apenas como caráter ilustrativo, apontam para uma consistência do critério de teste utilizado, ainda que aproximada (assintótico).

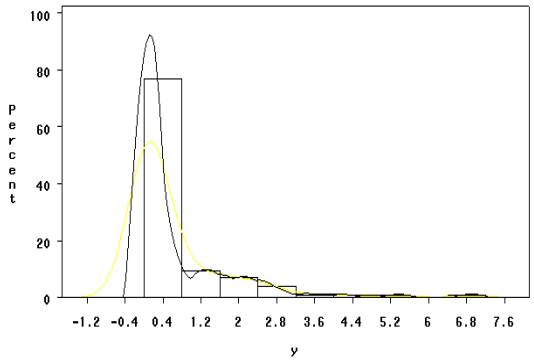

Figura 1. Densidade da razão de verossimilhanças estimada via simulação de dados, utilizando um estimador "Kernel". Modelo 1 versus 5, amostra n e $h^2 = 0,2$.

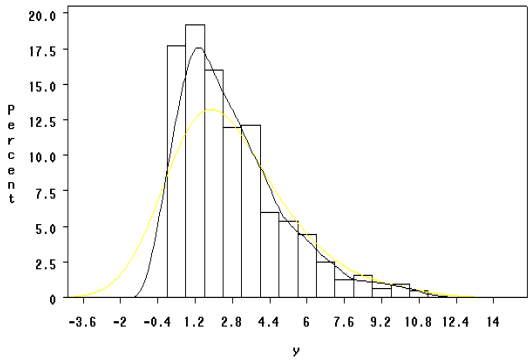

Figura 2. Densidade da razão de verossimilhanças estimada via simulação de dados, utilizando um estimador "Kernel". Modelo 1 versus 7, amostra n e $h^2 = 0,7$.

Avaliação do poder do teste sob diferentes tamanhos de amostra e h^2

As tabelas 4 e 5 ilustram os resultados obtidos para avaliação do poder do teste. Na Tabela 4, foi utilizado tamanho de amostra n, ou seja, 50, 50, 50, 200, 150 e 150 observações para as gerações P_1, P_2, F_1, F_2, RC_1 e RC_2, respectivamente, e uma combinação dos coeficientes de herdabilidade h^2_M e h^2_P, iguais a 0,2 e 0,7. Na tabela 5 estão apresentados os resultados referentes à metade do tamanho da amostra acima, ou seja, tamanho n/2, para as mesmas combinações entre os coeficientes de herdabilidade.

Tabela 4. Avaliação do poder do teste para amostra de tamanho n.

Modelos sob Comparação	Hipótese de Nulidade	h^2_P	h^2_M	Amostras Válidas	Taxa de Rejeição Empírica	Intervalo de Confiança
1 versus 5	Só poligenes	0,2	0,2	626	0,04632588	[0,03124122, 0,06585622]
1 versus 5	Só poligenes	0,2	0,7	415	1,00000000	[0,99280736, 1,00000000]
1 versus 5	Só poligenes	0,7	0,2	160	0,11875000	[0,07302985, 0,17921524]
1 versus 7	Só gene maior	0,2	0,2	761	1,00000000	[0,99607116, 1,00000000]
1 versus 7	Só gene maior	0,2	0,7	518	1,00000000	[0,99423342, 1,00000000]
1 versus 7	Só gene maior	0,7	0,2	174	1,00000000	[0,98293050, 1,00000000]

Tabela 5. Avaliação do poder do teste para amostra de tamanho n/2.

Modelos sob comparação	Hipótese de Nulidade	h^2_P	h^2_M	Amostras Válidas	Taxa de Rejeição Empírica	Intervalo de
1 versus 5	Só poligenes	0,2	0,2	614	0,04234528	[0,02784459, 0,06143191]
1 versus 5	Só poligenes	0,2	0,7	409	1,00000000	[0,99270223, 1,00000000]
1 versus 5	Só poligenes	0,7	0,2	159	0,09433962	[0,05376619, 0,15081892]
1 versus 7	Só gene maior	0,2	0,2	840	1,00000000	[0,99644000, 1,00000000]
1 versus 7	Só gene maior	0,2	0,7	632	1,00000000	[0,99527113, 1,00000000]
1 versus 7	Só gene maior	0,7	0,2	199	1,00000000	[0,98505881, 1,00000000]

Percebe-se que também aqui houve uma variação quanto ao número de amostras válidas, em função de nem sempre o critério de convergência ter sido satisfeito, com o número máximo de simulações estipulado (1000).

Nas Tabelas 4 e 5 percebe-se uma tendência geral de alto poder para a estatística de teste. A única exceção verificada foi na detecção do gene de efeito maior, em ambos os tamanhos amostrais, quando a herdabilidade referente àquele era baixa (0,2). A detecção dos poligenes foi significativa em todas as amostras simuladas, em qualquer combinação das herdabilidades h^2_M e h^2_P.

O baixo poder para detectar o gene de efeito maior com $h^2_M = 0,2$ não é realmente um aspecto muito desfavorável, uma vez que, sendo h^2_M baixa, tal gene tende a assumir um caráter mais de poligene do que propriamente de efeito maior.

Desta maneira, os resultados em conjunto permitem depreender que a estatística de teste, ainda que utilizando um critério de decisão aproximado (assintótico), mostrou-se bastante satisfatória, no sentido que promoveu um controle eficaz do erro tipo I, ainda que conservador, sem com isso acarretar grande prejuízo ao poder estatístico, nas condições avaliadas. Este aspecto torna-se mais significativo por ter sido verificado em amostras de tamanho inferior (n/2) aos geralmente praticados pelos melhoristas de plantas.

A qualidade observada do critério de decisão assintótico é ainda mais expressiva quando se leva em conta o tempo computacional relativamente longo para se obter convergência, o que dificulta a implementação de métodos potencialmente mais adequados, como métodos de reamostragem (bootstrap) ou um enfoque Bayesiano utilizando o amostrador de Gibbs. No entanto, seria interessante estudar a qualidade de tais métodos, futuramente.

Realmente, o aspecto computacional foi um dos fatores mais limitantes neste estudo, uma vez que cada configuração demandava um tempo computacional entre 50 e 168 horas, em geral. Isto limitou o número de configurações estudadas. Seria interessante, contudo, em estudos futuros, avaliar valores mais abrangentes de herdabilidade e de tamanhos de amostra. Outro fator que poderia ser estudado é o controle do erro tipo I e o poder de teste para identificar a existência de interações alélicas de dominância, seja nos poligenes, seja no gene de efeito maior.

É importante ressaltar que o baixo número de amostras válidas observado em diversas situações não significa que existiria uma alta probabilidade de não-convergência em uma situação real, conforme pode parecer ‘a primeira vista. Este baixo número

foi conseqüência de o número máximo de iterações estipulado ser limitado (1000), caso contrário o tempo computacional seria muito elevado, por serem muitas as amostras simuladas. Em uma situação real, por se tratar de um único conjunto de dados, é perfeitamente viável estipular um número máximo bem maior de iterações, aumentando sobremaneira a probabilidade de convergência. Isto foi feito em algumas condições e tal comportamento pôde ser verificado (dados não apresentados).

Conclusões

Os resultados da simulação mostraram que a densidade de razão de verossimilhanças, sob H_0, desvia-se significativamente da distribuição assintótica de qui-quadrado.

Apesar da conclusão (i), pôde-se verificar que o teste da razão de verossimilhança propiciou um controle adequado do erro tipo I, sendo considerado rigoroso.

No controle do erro tipo I, em todos os casos, a taxa de rejeição empírica de H_0 esteve abaixo de 0,05.

Em geral coeficientes de herdabilidade mais baixos ($h^2 = 0,2$) tenderam a abaixar a taxa de rejeição de H_0.

A redução do tamanho da amostra reduziu a taxa de rejeição de H_0.

A detecção de poligenes é feita com poder elevado, em todas as situações estudadas.

O gene de efeito maior e detectado com alto poder, a não ser quando a proporção da variação genética por ele explicada é baixa.

O teste da razão de verossimilhança utilizando o critério de decisão assintótico é satisfatório para uso em estudos de herança.

Referências

LYNCH, M.; WALSH, B. *Genetics and analysis of quantitative traits*. Sunderland: Sinauer Associates, 1998.
.

MATHER, K.; JINKS, J. L. *Introdução à genética biométrica***.** Ribeirão Preto: Sociedade Brasileira de Genética, 1984.

MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. *Introduction to the theory of statistics*. 3.ed. Tókio: McGraw-Hill Kogakusha, 1974.

SILVA, W. P. *Estimador de máxima verossimilhança em mistura de normais: uma aplicação à genética*. 2003. Dissertação (Mestrado) – Universidade Federal de Lavras, Lavras, 2003.