Theor Appl Genet (1985) 71:451–454

Errors in genetic theory equations*

D. E. Rowe

Research Geneticist, USDA-ARS, College of Agriculture, Rm 323 A, University of Nevada, Reno, NV 89557, USA

Received June 29, 1985; Accepted July 25, 1985
Communicated by P. M. A. Tigerstedt

Summary. This study addresses the consequences of eliminating terms such as x^2 and x^3 from genetic equations when the variable x is known to be small. This paper indicates logically that to assign such terms a value of 0.0 requires knowing the magnitude of the coefficients for each of these terms as well as the magnitude of all other terms in a given expression. Since most genetic expressions of interest involve several unknowns, the elimination of these terms appears difficult to justify in most situations. The effects of the elimination of a single term from an expression in a classical plant breeding paper were investigated as a simple exemplifying case. In the example, the simplified equation for change in population mean with selection sometimes greatly overestimated the response to selection and in some cases also altered conclusions as to best procedure. Though simplified equations are usually much more tractable and interpretable, the bias which is introduced into the research results and the potential for propagation of such biases in subsequent studies indicates that no term can be uncritically ignored in a genetic equation. The obvious alternatives are (1) do not simplify by eliminating terms, (2) perform a complete error analysis, or (3) restrict the range of values for variables so that terms can be justifiably eliminated in the error analysis.

Key words: Selection response – Genetic Analysis – Genetic models

Introduction

Theoretical quantitative geneticists have, with some success, developed algebraic expressions predicting relative rates of improvement in population means with selection for various breeding methods under optimum and some interesting non-optimum selection conditions. Such research by quantitative geneticists has been taunted as their most important contribution to breeding. This conclusion is debatable, but this theoretical research has obvious and immediate applications to plant breeding.

To briefly outline such theoretical work, a parsimonious description of a genetic process is developed into an algebraic expression which the researcher manipulates to produce interpretable or meaningful equations, either stochastic or deterministic. If interpretable equations are not forthcoming, a numerical analysis may augment or replace analytical expressions and provide useful interpretation of the results.

The objective of this research is to scrutinize a small but potentially critical procedural point which occurs with some regularity in the development of response to selection equations and sporadically in other genetic theory research. Consider a model where x represents the change in gene frequency in response to selection or some other variable which is expected to be small. With an x much less than 1.0 and near 0.0, the square, cube, and higher exponentials of x are much closer to 0.0 and thus can be ignored because they are negligible. Then the equation is simplified by elimination of all terms involving x^2 and x^3.

Using this logic, expressions may be greatly reduced in complexity and those expressions which did not lend themselves to interpretation may be interpretable when expressed as linear in x. Also intractable equations may

* Cooperative investigations of the Alfalfa Production Research Unit, United States Department of Agriculture, Agricultural Research Service, and the Nevada Agricultural Experiment Station, Reno, NV

be sufficiently simplified that they can be incorporated into the next step of some modeled procedure and that step becomes solvable. The value of the elimination of the higher order terms of x usually increases the earlier it is used in model development.

The logic of this elimination of the higher powers of x is correct in the context presented which can be represented as the following expression:

$$y = x + x^2 = x\,(1 + x) \qquad (1)$$

where $y = x$ implies $(1 + x) = 1$ and that x is near 0.0. The difficulty with equation (1) is that it rarely, if ever, occurs in any useful genetic model. The common expression involving x in its simplest form is the following:

$$y = Ax + Bx^2 = x\,(A + Bx) \qquad (2)$$

where $y = Ax$ implies that $(A + Bx) = A$. To simplify this equation it is necessary that Bx approach 0.0 which is a certainty when both B and x are near zero. Another possibly sufficient condition is that x be near zero and that the ratio (A/B) be very large. In many modeling situations the complexity of a given expression or the existence of random variables of unknown magnitude, such as A and B, frustrates any determination that these requirements have been met. The real problem arises that the simplified predictive equations not only approximate natural processes but also approximate the original parsimonious interpretation of a genetic process. Thus a bias is introduced into an expression which may alter conclusions and be propogated in subsequent studies.

Error analysis

In the literature there is no systematic treatment of the bias or error in genetic equations introduced with the elimination of terms. Such analysis may be borrowed from that sometimes used in computer science and simulation studies (Chapter 3, Kennedy and Gentle 1980). There are the direct error analysis, which indicates the magnitude of error at each computational step, and the inverse error analysis, where bias is measured only in the final results. Even these analyses may not be informative about the potential for altering the conclusions of comparative studies which are often the explicit objectives of research on the responses to selection.

Example

The elimination of a single term from an equation for response to selection is used to demonstrate the potential effects on estimates and conclusions in comparative studies. Comstock et al. (1949) derived a response equation to compare a breeding procedure he introduced called reciprocal recurrent selection (RRS) to a test cross procedure (TX) for improvement of the interpopulation mean. The procedures compared are outlined below.

Assume there are two populations designated C and D in which plants are fertilized with pollen from a foreign population. The half-sib progenies of each plant are then grown and evaluated and the performance of these progenies are used as criteria for selection of best parents in a population. With RRS the foreign pollen placed on plants of population C is a representative sample from population D and that on plants of D comes from population C. This procedure is compared to the TX procedure where fertilization is effected in populations C and D using pollen from a third population designated T.

Comstock et al. (1949) developed equations for change in mean genotypic value with selection at a single locus with two genes (*B* and *b*) and diploid inheritance. The frequencies of gene *B* in populations C, D, and T are, respectively, p, r, and v. Symbols dp and dr indicate the changes in gene frequencies p and r, respectively, with selection. Population means are modeled with parameters a and u as indicated in Table 1. The change in population mean ($\Delta\bar{X}$) is as follows:

$$\Delta\bar{X} = [dp\,(1 + a) + dr\,(1 + a) - 2\,ar\,(dp) - 2\,ap\,(dr) - 2\,a\,(dp)\,(dr)]\,u \qquad (3)$$

Equation (3) was then simplified by assuming dp and dr would be small and that their crossproduct could be ignored to give the following simplified expression:

$$\Delta\bar{X} = [dp\,(1 + a) + dr\,(1 + a) - 2\,ar\,(dp) - 2\,ap\,(dr)]\,u \qquad (4)$$

Hereafter equations (3) and (4) are referred to, respectively, as the exact and approximate equations for change in population means with selection. The expression for change in gene frequency with selection in population C is:

$$dp = p\,(1 - p)\,(1 + a - 2\,at)\,sc \qquad (5)$$

where t is frequency of gene *B* in the pollen parent (either v or r in this case), c is the ratio $u/(4\,V_Y)^{\cdot 5}$, and s is the selection differential in multiples of the standard deviation of progeny means $(V_Y)^{\cdot 5}$. The dr was similarly calculated using the gene frequency r. When dp and dr were calculated the value of c was assumed to be 1.0 for computational simplicity.

Three types of genic action models are reported here in some detail: simple dominance and two types of overdominance. The RRS procedure was initially developed to exploit any performance advantage due to overdominance genic action.

The exact error, which is the deviation of the approximate prediction from the exact prediction expressed as a percentage of the exact prediction, is indicated in Fig. 1 for several genetic

Table 1. Genotypic values and assigned values for three types of modeled genic action at the digenic diploid locus

Genotype	Genotypic value	Modeled genic action		
		Simple dominance	Overdominance	
			Type 1	Type 2
BB	2u	1.0	0.5	0.1
Bb	u+au	1.0	1.0	1.0
bb	0	0.0	0.0	0.0

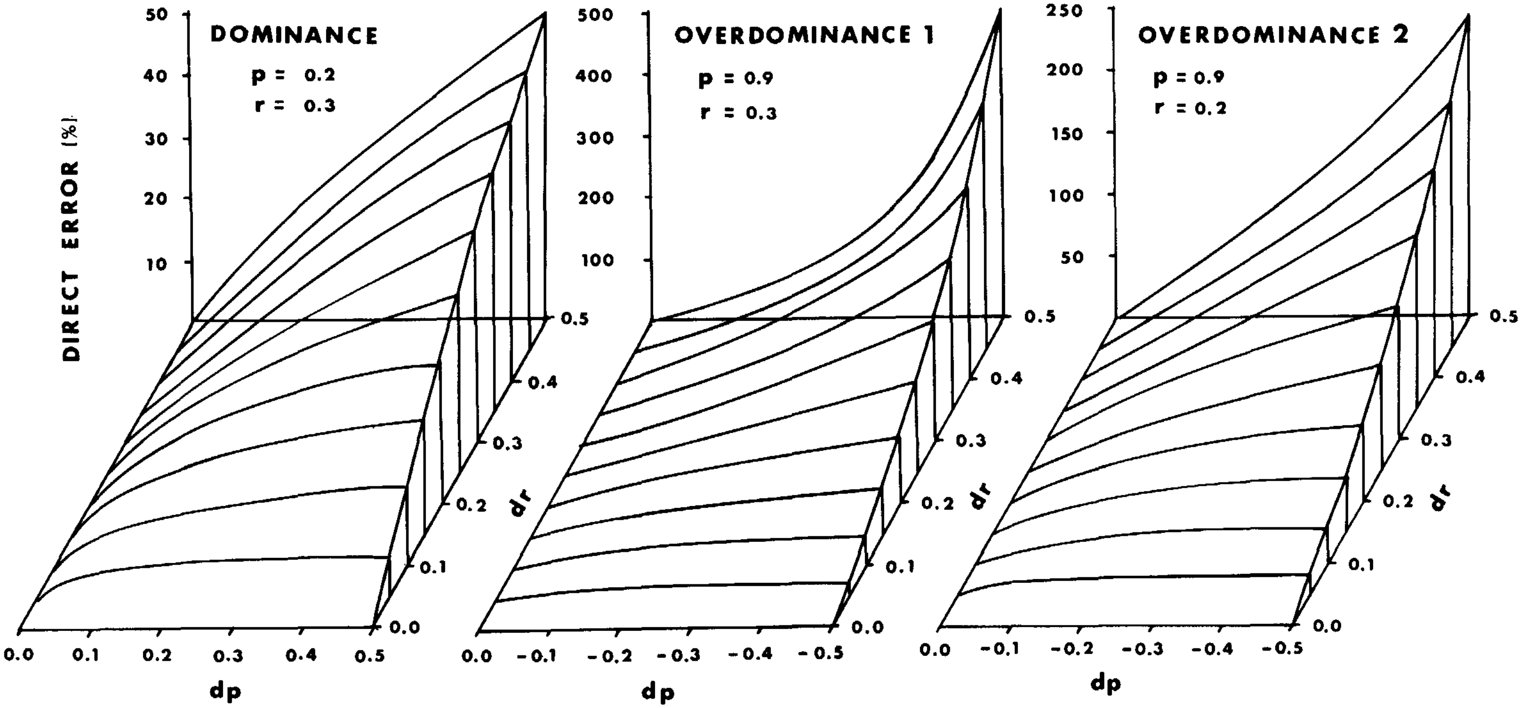

Fig. 1. The direct error in the estimates of the means of populations with simple dominance (dominance) and overdominance types 1 and 2 genic action (overdominance 1 and overdominance 2, respectively) with indicated gene frequencies p and r, and changes in gene frequencies, dp and dr

p = 0.1

r =	.1	.2	.3	.4	.5	.6	.7	.8	.9
0.0	•	•	•	•	•	+			
0.1		•	•	+	+				
0.2									
0.3									
0.4									
v = 0.5									
0.6									
0.7									
0.8									
0.9									
1.0									

0.2

r =	.1	.2	.3	.4	.5	.6	.7	.8	.9
0.0	•	•	•	•	+	+			
0.1	•	•	•	•	+	+			
0.2			•	+	+				
0.3									
0.4									
v = 0.5									
0.6									
0.7									
0.8									
0.9									
1.0									

0.3

r =	.1	.2	.3	.4	.5	.6	.7	.8	.9
0.0	•	•	•	•	+	+			
0.1	•	•	•	•	+	+			
0.2		•	•	•	+				
0.3				•	+				
0.4									
v = 0.5									
0.6									
0.7									
0.8									
0.9									
1.0									

0.4

r =	.1	.2	.3	.4	.5	.6	.7	.8	.9
0.0	•	•	•	+	+	+			
0.1	+	•	•	•	+	+			
0.2		+	•	•	+				
0.3			•	•	+				
0.4					+				
v = 0.5									
0.6									
0.7									
0.8									
0.9									
1.0								+	+

0.5

r =	.1	.2	.3	.4	.5	.6	.7	.8	.9
0.0	•	+	+	+	+				
0.1	+	+	+	+	+				
0.2		+	+	+	+				
0.3			+	+	+				
0.4				+	+				
v = 0.5									
0.6									
0.7						+			
0.8						+	+		
0.9						+	+	+	
1.0						+	+	+	+

0.6

r =	.1	.2	.3	.4	.5	.6	.7	.8	.9
0.0	+	+	+	+					
0.1		+	+	+					
0.2									
0.3									
0.4									
v = 0.5									
0.6									
0.7					+	•	+		
0.8					+	+	•	+	
0.9					+	+	•	•	+
1.0					+	+	+	•	•

0.7

r =	.1	.2	.3	.4	.5	.6	.7	.8	.9
0.0									
0.1									
0.2									
0.3									
0.4									
v = 0.5									
0.6									
0.7							+		
0.8					+	•	•	•	
0.9					+	•	•	•	•
1.0					+	+	•	•	•

0.8

r =	.1	.2	.3	.4	.5	.6	.7	.8	.9
0.0									
0.1									
0.2									
0.3									
0.4									
v = 0.5									
0.6									
0.7									
0.8						+	•		
0.9					+	•	•	•	•
1.0				+	+	•	•	•	•

0.9

r =	.1	.2	.3	.4	.5	.6	.7	.8	.9
0.0									
0.1									
0.2									
0.3									
0.4									
v = 0.5									
0.6									
0.7									
0.8									
0.9					+	•	•		
1.0			+	+	•	•	•	•	

Fig. 2. A comparison of RRS and TX procedures at gene frequencies p, r, and v where superiority of RRS and TX are indicated by a blank and a dot, respectively. The plus sign indicates superiority of RRS with only the exact equation

situations. This figure indicates the error for increasing or decreasing values of dp and dr of 0.0 to 0.5 independent of selection procedure. Thus this figure indicates the overestimates of $\Delta\bar{X}$ by equation (4) but is not informative about any comparison of RRS and TX procedures.

The comparison of procedures was made by calculating expected gains with selection with approximate and exact equations and then determining if conclusions made with the approximate equation were inaccurate. The comparisons were made at frequency increments of 0.1 (Fig. 2).

With complete and incomplete dominance (not shown) the conclusions as to best procedure were the same for the approximate and exact equations. But with overdominance the approximate equation indicated incorrectly the inferiority of the RRS procedure for about 10% of the comparison with overdominance type 1 and 12% of the comparisons with overdominance type 2 as shown in Fig. 2. With reference to Fig. 2, 57% of the times that the approximate equation indicated the superiority of the TX procedure it was in error.

Discussion

The mathematical models of genetic events are expected to approximate the results in any given, real data situation. The models developed incorporate simplifying assumptions first to remove "noise" from the event by eliminating factors which are considered extraneous or unimportant and second to make calculations tractable and interpretable. Usually the simplifying assumptions are indicated in a paper and, though their precise effects on the final conclusions may not be obvious, it is understood that results should be considered within the framework of the simplifying assumptions. The limitations of theoretical results are usually acknowledged in the better papers and have been the topic of other papers (Robertson 1963; Kempthorne 1977).

In contrast the consequences of elimination of a term from a genetic equation because it appeared negligible are not always acknowledged and have received little critical attention. This paper, using the simplest possible example of equation simplification through term elimination, has demonstrated the potential for developing genetic equations which only approximate the genetic process which was to be modeled. The potential also exists for development, in the extreme situation, of equations whose results are the antithesis of the process modeled.

There are several solutions to this problem with equation simplification the most obvious of which is to not eliminate any terms from the genetic equation. This approach may result in intractable equations and certainly will result in more complex expressions which may or may not be interpretable analytically.

An alternative is to perform an error analysis, as might be found in a computer simulation paper, on the simplified equations. Equation complexity may frustrate any such error analysis and the analysis may greatly exceed the effort expended on development of the original equation. Even with such analysis, decision criteria will have to be developed on how much error is tolerable.

The third alternative is to acknowledge what terms have been eliminated and restrict the values of the other variables forcing some terms to zero. A limited error analysis might be necessary to verify that eliminated terms are negligible. The interpretation of the results would then be made within the context of restricted values of the variables. Probably, simplified equations which alter qualitative results should not be used, but simplified equations that result in quantitative changes in the results without altering conclusions may be used in some situations.

None of these alternatives is as easy as ignoring this procedural problem. But this author suggests the potential for developing and propogating misinformation in genetic theory should deter theoreticians from uncritical elimination of any term in a genetic equation.

References

Comstock RE, Robinson HF, Harvey PH (1949) A breeding procedure designed to make maximum use of both general and specific combining ability. Agron J 41:360–367

Kempthorne O (1977) Status of quantitative genetic theory. In: Pollok E, Kempthorne O, Bailey TB Jr (eds) Proc Int Conf Quant Genet. Iowa State Press, Ames IA, pp 719–760

Kennedy WJ Jr, Gentle JE (1980) Statistical computing. Marcel Dekker, New York Basel

Robinson A (1963) Discussion: some comments on quantitative genetic theories. In: Hanson WD, Robinson HF (eds) Statistical genetics and plant breeding. Natl Acad Sci Natl Res Council 1982, pp 108–115