

HERITABILITY OF POLYCHOTOMOUS CHARACTERS

DANIEL GIANOLA

*Department of Animal Science,
University of Illinois, Urbana 61801*

Manuscript received May 29, 1979
Revised copy received October 29, 1979

ABSTRACT

Characters with phenotypic expression consisting of a response in one of several mutually exclusive and exhaustive categories are considered. Formulae relating heritability in the discrete, outward scale to heritability in underlying normal and exponential scales are presented. For the normal case with two response categories, results reduce to the well-known formula for heritability of binary traits.

MOST applications of quantitative genetics theory to animal and plant breeding have been made with respect to characters showing a continuous distribution on a phenotypic scale. However, many traits such as tolerance to micro-nutrients and calving difficulty present a discrete distribution of phenotypes. On the basis that such traits can be expressed in terms of percentage incidence, WRIGHT (1920, 1926) developed the inverse probability transformation and used it to determine the relative amounts of genetic and environmental variation in digit numbers of guinea pigs (WRIGHT 1934a, 1934b). WRIGHT postulated that the dichotomy observed in guinea pigs (three-toed *vs* four-toed) was the result of a physiological threshold in a character affected by many Mendelian factors, each of which made a fairly constant contribution to variability in an underlying normal scale.

ROBERTSON, in an APPENDIX to a paper by DEMPSTER and LERNER (1950) derived a formula describing the relationship between heritability in a normally distributed underlying scale, where all genetic effects are additive, and heritability in an outward binary scale. The present paper presents general expressions relating heritability in underlying normal and exponential scales to heritability in an observed scale, where the expression of the character is a response in one of the several mutually exclusive and exhaustive categories.

THE MULTIPLE THRESHOLD MODEL

When an individual is subjected to a set of conditions defining a population in a statistical sense, a random variable, y , is assumed to arise in an underlying continuous scale. In this scale, there is a set of $m-1$ fixed thresholds defined by the vector $\mathbf{t}' = (t_1, t_2, \dots, t_{m-1})$ corresponding to m discontinuities in the observed scale. If $t_j < y < t_{j+1}$, for $j=0, \dots, m-1$ and with $t_0 = -\infty$ and $t_m = \infty$, then

the individual is scored as responding in the $j+1$ th category. The γ variable can be viewed as representing a linear combination of stochastically independent genetic and nongenetic factors, and this structure is envisaged as determining the probability of the character appearing. Let

$$\gamma_i = \mu + g_i + e_i \quad (1)$$

be the phenotype of the i th individual in an underlying continuous scale, g_i be its additive genetic value, and e_i be an environmental deviation. Further, let $E(\gamma_i) = \mu$, $E(g_i) = E(e_i) = 0$, and $\text{Cov}(g_i, e_i) = 0$ for all i . Then $\sigma^2 = \sigma_g^2 + \sigma_e^2$, with heritability defined as $h^2 = \sigma_g^2/\sigma^2$. The model in (1) can be standardized as

$$\gamma_i^* = \frac{\gamma_i - \mu}{\sigma} = \frac{g_i}{\sigma} + \frac{e_i}{\sigma} = g_i^* + e_i^* \quad (2)$$

with $E(\gamma_i^*) = 0$, $\text{Var}(g_i^*) = h^2$, $\text{Var}(e_i^*) = 1-h^2$ and $\text{Cov}(g_i^*, e_i^*) = 0$. In the standardized scale, the vector of thresholds becomes

$$\mathbf{t}^{**} = (t_1^*, \dots, t_{m-1}^*) = \left(\frac{t_1 - \mu}{\sigma}, \dots, \frac{t_{m-1} - \mu}{\sigma} \right).$$

If there are m possible response categories in the outward scale, for each g^* , there is a vector $\mathbf{G}' = (\theta_1, \theta_2, \dots, \theta_m)$, with $\sum_{j=1}^m \theta_j = 1$ corresponding to the distribution of responses in the observed scale, *i.e.*, θ_j is the probability of the response in the j th category, given that $g^* = k$, say. Since g^* and e^* are independent:

$$\begin{aligned} \theta_j &= \text{Prob}\{t_{j-1}^* < \gamma^* < t_j^* | g^* = k\} \\ &= \text{Prob}\{t_{j-1}^* - g^* < e^* < t_j^* - g^* | g^* = k\} \\ &= \int_{t_{j-1}^* - g^*}^{\infty} f(e^*) de^* - \int_{t_j^* - g^*}^{\infty} f(e^*) de^* \\ &= \gamma_{j-1} - \gamma_j, \end{aligned} \quad (3)$$

which is written in this form to facilitate comparison with ROBERTSON's developments (see DEMPSTER and LERNER 1950) and where $f(e^*)$ is the density function of e^* . Hence, $\mathbf{G}' = \gamma_0 - \gamma_1, \gamma_1 - \gamma_2, \dots, \gamma_{m-1}$, with $\gamma_0 = 1$ and $\gamma_m = 0$, is a function of g^* .

In most applications, the aggregate value of a genotype in the outward scale can be defined as:

$$w_i = \mathbf{a}'\mathbf{G}_i = \mathbf{a}'\mathbf{G}_i^A + \mathbf{a}'\mathbf{G}_i^R, \quad (4)$$

where $\mathbf{a}$ is an $m \times 1$ vector of scores or weights given to each of the possible response categories, and $\mathbf{G}_i^A$ and $\mathbf{G}_i^R$ are vectors of additive and nonadditive genetic effects, respectively. We now adopt the model of DEMPSTER and LERNER (1950) by letting

$$w_i = \alpha + \beta g_i^* + u_i, \quad (5)$$

where $\mathbf{a}'\mathbf{G}_i^A = \alpha + \beta g_i^*$ describes a linear relationship between the aggregate

additive genetic value in the outward scale and the additive genetic value in the underlying scale. From (4) and (5)

$$\text{Var}(\mathbf{a}'\mathbf{G}_i^A) = \beta^2 h^2 = \text{Cov}^2(\mathbf{a}'\mathbf{G}_i, g_i^*)/h^2. \quad (6)$$

Furthermore, $\text{Cov}(\mathbf{a}'\mathbf{G}_i, g_i^*) = \mathbf{a}'\text{Cov}(\mathbf{G}_i, g_i^*)$ and the j th element of $\text{Cov}(\mathbf{G}_i, g_i^*)$ is given by

$$\text{Cov}(\theta_j, g^*) = \int_{-\infty}^{\infty} \theta_j(g^*) g^* f(g^*) dg^*, \quad (7)$$

where θ_j is written as $\theta_j(g^*)$ to indicate its dependence on g^* and where $f(g^*)$ is the density function of g^* .

The phenotypic variance in the outward scale can be obtained by defining an $m \times 1$ vector of phenotypic response probabilities $\hat{\Pi}' = (\hat{\Pi}_1, \hat{\Pi}_2, \dots, \hat{\Pi}_m)$ with variance-covariance matrix with elements $\Pi_i(1 - \Pi_i)$, $i=1 \dots m$, and $-\Pi_i\Pi_j$ for $i \neq j$. If phenotypes are scored as $P = \mathbf{a}'\hat{\Pi}$, where $\mathbf{a}$ is as before, we have

$$\begin{aligned} \text{Var}(P) &= \mathbf{a}'\text{Var}(\hat{\Pi})\mathbf{a} \\ &= \sum_{i=1}^m a_i^2 \Pi_i(1 - \Pi_i) - 2 \sum_{i=1}^m \sum_{j=1}^m a_i a_j \Pi_i \Pi_j \end{aligned} \quad (8)$$

which in the case of two response categories and with $a_1=0$, $a_2=1$, reduces to $\Pi(1-\Pi)$, the well-known formula for the binomial distribution.

THE NORMAL CASE

If g^* and e^* are normally distributed, ROBERTSON's results (see DEMPSTER and LERNER 1950) can be directly extended (using equations 3 and 7) to obtain:

$$\begin{aligned} \text{Cov}(\theta_j, g^*) &= \int_{-\infty}^{\infty} \gamma_{j-1} g^* f(g^*) dg^* - \int_{-\infty}^{\infty} \gamma_j g^* f(g^*) dg^* \\ &= h^2(z_{j-1} - z_j), \end{aligned} \quad (9)$$

where z_{j-1} and z_j are the ordinates of a standard normal density function at points t_{j-1} and t_j corresponding to the thresholds between categories $j-1$ and j , and j and $j+1$, respectively. From equations 6 and 9

$$\text{Var}(\mathbf{a}'\mathbf{G}_i^A) = h^2 \left[\sum_{i=1}^{m-1} z_i (a_{i+1} - a_i) \right]^2, \quad (10)$$

which in the case of two response categories and one threshold and when $|a_2 - a_1| = 1$, becomes $z^2 h^2$, which is the expression obtained by ROBERTSON (see DEMPSTER and LERNER 1950) for the additive genetic variance in the outward scale. The heritability in the outward scale is

$$h_0^2 = h^2 \left[\sum_{i=1}^{m-1} z_i (a_{i+1} - a_i) \right]^2 / \left[\sum_{i=1}^m a_i^2 \Pi_i(1 - \Pi_i) - 2 \sum_{i=1}^m \sum_{j=1}^m a_i a_j \Pi_i \Pi_j \right], \quad (11)$$

which for two response categories becomes

$$h_0^2 = \frac{h^2 z^2}{\Pi(1 - \Pi)} \quad (12)$$

identical to the expression derived by ROBERTSON. Note that while the additive genetic variance in the outward scale with two categories of response in general depends on $\mathbf{a}$, the heritability is invariant to the scoring procedure.

THE EXPONENTIAL CASE

Let the environmental and genetic components have density functions:

$$f(e^*) = \frac{1}{\beta_1} e^{-e^*/\beta_1}, \quad e^* > 0$$

$$f(g^*) = \frac{1}{\beta_2} e^{-g^*/\beta_2}, \quad g^* > 0$$

in which case it is possible to show that γ^* is not exponentially distributed and has density function:

$$f(\gamma^*) = \frac{1}{\beta_1 - \beta_2} [e^{-\gamma^*/\beta_1} - e^{-\gamma^*/\beta_2}], \quad \gamma^* > 0.$$

From equation (3)

$$\begin{aligned} \theta_j &= \int_{t^*_{j-1}-g^*}^{t^*_j-g^*} f(e^*) de^* = \\ &= e^{-(t^*_{j-1}-g^*)/\beta_1} - e^{-(t^*_j-g^*)/\beta_1}. \end{aligned} \quad (13)$$

Likewise, from equations (3) and (7)

$$\begin{aligned} \text{Cov}(\theta_j, g^*) &= \int_0^\infty (\gamma_{j-1} - \gamma_j) g^* f(g^*) dg^* \\ &= \frac{1}{\beta_2} [e^{-t^*_{j-1}/\beta_1} - e^{-t^*_j/\beta_1}] \int_0^\infty g^* e^{-\alpha g^*} dg^*, \end{aligned}$$

where $\alpha = \frac{1}{\beta_2} - \frac{1}{\beta_1}$. From the result $\int_0^\infty g^* e^{-\alpha g^*} dg^* = 1/\alpha^2$, $\alpha \neq 0$, then

$$\begin{aligned} \text{Cov}(\theta_j, g^*) &= \frac{1}{\beta_2 \alpha^2} [e^{-t^*_{j-1}/\beta_1} - e^{-t^*_j/\beta_1}] \\ &= \frac{1}{\beta_2 \alpha^2} [m_{j-1} - m_j]. \end{aligned} \quad (14)$$

From equations (6) and (14), we then have

$$\text{Var}(\mathbf{a}'\mathbf{G}^A) = \frac{1}{(h\beta_2\alpha^2)^2} [a_1 + \sum_{i=1}^{m-1} (a_{i+1} - a_i) m_i]^2. \quad (15)$$

The heritability in the outward scale is

$$\begin{aligned} h_0^2 &= \frac{1}{(h\beta_2\alpha^2)^2} [a_1 + \sum_{i=1}^{m-1} (a_{i+1} - a_i) m_i]^2 / \\ &\quad [\sum_{i=1}^m \alpha^2 \Pi_i (1 - \Pi_i) - 2 \sum_{i=1}^m \sum_{j=1}^m a_i a_j \Pi_i \Pi_j] \end{aligned} \quad (16)$$

and when there are two response categories, equation (16) becomes

$$h_0^2 = \frac{1}{(h\beta_2\alpha^2)^2} [a_1 + (a_2 - a_1)m_1]^2 / \Pi(1 - \Pi)(a_2 - a_1)^2, \quad (17)$$

which is not invariant to the scoring procedure, unless $a_1 = 0$, and reduces to

$$h_0^2 = \frac{1}{(h\beta_2\alpha^2)^2} m_1^2 / \Pi(1 - \Pi)$$

when $a_1 = 0$ and $a_2 = 1$. Clearly, equation (16) depends on the threshold values, which in turn depend on β_1 and β_2 since t_{j-1}^* and t_j^* must be obtained from the density function of γ . Since in the exponential distribution $\text{Var}(e^*) = \beta_1^2$ and $\text{Var}(g^*) = \beta_2^2$, the choice of $\beta_1 = (1 - h^2)^{-.5}$ and $\beta_2 = h$ is natural. Even in this case, it seems impossible to calculate the threshold values without knowledge of heritability in the underlying scale.

LITERATURE CITED

- DEMPSTER, E. R. and I. M. LERNER, 1950 Heritability of threshold characters. *Genetics* **35**: 212-236.
- WRIGHT, S., 1920 The relative importance of heredity and environment in determining the piebald pattern of guinea pigs. *Proc. Natl. Acad. Sci. U.S.* **6**: 320-332. —, 1926 A frequency curve adapted to variation in percentage occurrence. *J. Amer. Statist. Assoc.* **21**: 162-178. —, 1934a An analysis of variability in number of digits in an inbred strain of guinea pigs. *Genetics* **19**: 506-536. —, 1934b The results of crosses between inbred strains of guinea pigs, differing in number of digits. *Genetics* **19**: 537-551.

Corresponding editor: J. F. KIDWELL