

Prediction of selection response for Poisson distributed traits

JL Foulley

*INRA, Station de Génétique Quantitative et Appliquée,
Centre de Recherches de Jouy-en-Josas, 78352 Jouy-en-Josas Cedex, France*

(Received 18 November 1992; accepted 19 February 1993)

Summary – This paper presents a formula to predict expected response to one generation of truncation selection for a Poisson distributed trait under polygenic additive inheritance. The derivation relies on the Poisson-lognormal model and its analysis *via* quasi-likelihood. The formula derived accounts for asymmetry of response. The relationship with the classical formula $R = i\rho\sigma_G$ is explained. Properties are illustrated with an example of sire selection based on progeny test performance.

discrete variable / Poisson distribution / prediction / selection response / asymmetry of response

Résumé – Prédiction de la réponse à la sélection pour des caractères distribués suivant une loi de Poisson. *Cet article présente une formule de prédiction de la réponse à une génération de sélection par troncature pour un caractère distribué suivant une loi de Poisson, sous l'hypothèse d'un déterminisme polygénique additif. Le raisonnement est basé sur le modèle Poisson log normal et son traitement par quasi-vraisemblance. La formule présentée rend compte d'une asymétrie de la réponse. Sa relation avec la formule classique $R = i\rho\sigma_G$ est expliquée. Ses propriétés sont illustrées par un exemple de sélection de mâles en contrôle de descendance.*

variables discrètes / distribution de Poisson / prédiction / réponse à la sélection / asymétrie de la réponse

The theory of response of selection under polygenic inheritance relies basically on the normality assumption and kindred linear model methodology. Discrete traits cannot be properly analyzed using classical procedures (Gianola, 1982) and nonlinear statistical methods have recently been proposed for genetic evaluation of such traits (see *eg* review by Ducrocq, 1990). Moreover, some work has already been done to predict either analytically (Curnow, 1984; Falconer, 1989; Foulley,

1992) or *via* simulation (Danell and Rönningen, 1981), selection response for threshold dichotomous traits. The purpose of this note is to generalize the analytical approach to count traits described by a Poisson distribution which may arise *eg* in reproductive performance (ovulation rate, embryo production, prolificacy): see Perez *et al* (1993) for an application in swine.

The same model as in Foulley *et al* (1987) is postulated. Let Y be the random variable with realized value $y = 0, 1, 2, \dots$. Given λ , the Y 's have independent Poisson distributions with parameter λ , *ie*:

$$\Pr(Y = y) = \lambda^y \exp(-\lambda)/y! \tag{1}$$

Let us assume that the trait is determined by a purely additive genetic model on the transformed scale, *ie* $\ln(\lambda) = \eta + a$ where η is a location parameter, and $a \sim N(0, \sigma_a^2)$ is the genetic value on the transformed scale, assumed to be normally distributed with mean zero and variance σ_a^2 . Following Falconer (1989), the genetic value (g) on the observed scale can be defined as the mean phenotypic value of individuals having the same genotype, *ie*:

$$g = E(Y|a) = \exp(\eta + a). \tag{2}$$

Generally a is not known and one may alternatively condition on \hat{a} , the estimated breeding value (EBV). Then, the expression in [2] has to accommodate the uncertainty in a given \hat{a} ; this is achieved by taking its expectation with respect to the conditional density $p(a|\hat{a})$ of a given \hat{a} , *ie*:

$$E(Y|\hat{a}) = \int_{-\infty}^{+\infty} \exp(\eta + a)p(a|\hat{a})da. \tag{3}$$

If we assume, as usually done in animal breeding, truncation selection on the EBVs of candidates ($\hat{a}_c \geq s$) with proportion $Q = \int_s^{+\infty} p(\hat{a}_c)d\hat{a}_c$ selected ($p(\hat{a}_c)$ being the density of \hat{a} among candidates), one has the following expression for the expected response to 1 generation of upward selection on EBVs:

$$R^{(+)} = \left\{ \int_s^{+\infty} \left[\int_{-\infty}^{+\infty} \exp(\eta + a)p(a|\hat{a}_c)da \right] p(\hat{a}_c)d\hat{a}_c/Q \right\} - \mu \tag{4}$$

where $\mu = \exp[\eta + (\sigma_a^2/2)]$ corresponds to the population mean $E(Y)$ (Foulley and Im, 1993).

Now $\hat{a} = \theta\hat{a}_c$ where θ is twice Malecot's coefficient of kinship between the offspring (in which response is measured) and the candidate for selection (here $\theta = 1$ in monoecious populations and $\theta = 1/2$ in dioecious populations with selection in 1 sex); then $p(a|\hat{a}_c) = p(a|\hat{a})$ and $p(\hat{a}_c)d\hat{a}_c = p(\hat{a})d\hat{a}$ and an alternative expression for $R^{(+)}$ is:

$$R^{(+)} = \left\{ \int_{\theta s}^{+\infty} \left[\int_{-\infty}^{+\infty} \exp(\eta + a)p(a|\hat{a})da \right] p(\hat{a})d\hat{a}/Q \right\} - \mu \tag{5}$$

A further simplification can be achieved in cases where evaluations are based on sufficient data for the joint distribution of (a, \hat{a}) is assumed to be normal, ie:

$$a|\hat{a} \sim N[\hat{a}, (1 - \rho^2)\sigma_a^2], \tag{6a}$$

$$\hat{a} \sim N[0, \rho^2\sigma_a^2] \tag{6b}$$

where ρ^2 is the accuracy of genetic evaluation of a .

Then, expression [3] arising in [4] can be written explicitly from the moment generating function of the normal distribution defined in [6a], ie:

$$E(Y|\hat{a}) = \exp(\eta + \hat{a})\exp[(1 - \rho^2)\sigma_a^2/2]. \tag{7}$$

Now,

$$\exp(\hat{a})p(\hat{a})d\hat{a} = \exp(\sigma_a^2/2)\phi(u)du$$

with $u = (\hat{a} - \sigma_a^2)/\sigma_a$, ($\phi(\cdot)$ being the normal density function).

Using this expression and [7] in formula [4] makes the integration analytically feasible, which gives:

$$R^{(+)} = \mu\{\Phi(\omega + \sigma_a)/Q\} - 1\}, \tag{8}$$

where:

$$\omega = -\theta s/\sigma_{\hat{a}} = -s/\sigma_{\hat{a}_c} = \Phi^{-1}(Q),$$

$\sigma_{\hat{a}_c} = \rho_c\sigma_a$ being the standard deviation of EBVs in candidates, and $y = \Phi(x)$ and $x = \Phi^{-1}(y)$ the cumulative density function for the standardized normal distribution and its inverse function respectively.

If use is made of the more commonly tabulated quantity $L(x) = \Pr[N(0, 1) > x]$ (Johnson and Kotz 1972), formula [8] is then expressed as:

$$R^{(+)} = \mu\{[L(\omega^* - \sigma_a)/Q] - 1\}, \tag{9}$$

where $\omega^* = -\omega$.

Response to downward selection $R^{(-)}$ can easily be obtained by taking in [4] the sum from $-\infty$ to $-s$ which results in:

$$R^{(-)} = \mu\{[\Phi(\omega - \sigma_a)/Q] - 1\} \tag{10}$$

or,

$$R^{(-)} = \mu\{[L(\omega^* + \sigma_a)/Q] - 1\}. \tag{11}$$

$\Psi(\sigma_a) = \Phi(\omega + \sigma_a) - \Phi(\omega)$ is not an odd function, ie $\Psi(-\sigma_a) \neq -\Psi(\sigma_a)$, since:

$$\Psi[(-1)^p\sigma_a] = (-1)^p\sigma_a\phi(\omega) - (\sigma_a^2/2)\omega\phi(\omega) + o(\sigma_a^3). \tag{12}$$

It follows therefrom that formulae [8] and [10] (or [9] and [11]) account for an asymmetrical pattern of response. However, using the Taylor expansion in [12] at the first order leads to:

$$R^{(+)} \cong \mu\sigma_a\phi(\omega)/\Phi(\omega). \tag{13}$$

Now, the ratio $\phi(\omega)/\Phi(\omega)$ is the selection intensity (i), and $\mu\sigma_{\hat{a}} = \rho\mu\sigma_a$ where $\mu\sigma_a$ is the additive genetic standard deviation (σ_G) on the observed scale (see Foulley and Im, 1993; formula 21), so that formula [13] reduces to:

$$R^{(+)} \cong i\rho\sigma_G, \tag{14}$$

which is the very well known formula to predict expected genetic change after one generation of selection for a normal trait. The same reasoning applies as well to $R^{(-)}$ giving the negative of [14].

If selection is applied in males (m) and females (f), formula [8] becomes:

$$R^{(+)} = \mu\{[\Phi(\omega_m + 1/2\sigma_{\hat{a}_{c,m}})\Phi(\omega_f + 1/2\sigma_{\hat{a}_{c,f}})/Q_m Q_f] - 1\}, \tag{15}$$

where Q_m and Q_f are the proportions selected in males and females respectively with $\omega_m = \Phi^{-1}(Q_m)$ and $\omega_f = \Phi^{-1}(Q_f)$ and, $\sigma_{\hat{a}_{c,m}}$ and $\sigma_{\hat{a}_{c,f}}$ the standard deviation of EBV's in male and female candidates respectively.

A numerical illustration was carried out (table I) dealing with an example of selection of progeny-tested sires ($n = 125$ offspring per male) on their EBVs (transformed scale) for a Poisson distributed trait with mean $\mu = 1.5, 2.0, 3.0$ and 8.0 and selection rate $Q = 0.05, 0.50$ and 0.95 .

Table I. Expected response to one generation of selection of progeny-tested sires on the estimated breeding value (transformed scale) for a Poisson distributed trait^a.

Method	Selection rate	Mean of the trait (μ)			
		1.5	2.0	3.0	8.0
Upward selection	0.05	0.686	0.777	0.928	1.454
	0.50	0.226	0.262	0.321	0.525
	0.95	0.026	0.031	0.039	0.067
Downward selection ^b	0.05	0.502 (30.9)	0.592 (26.9)	0.744 (22.0)	1.270 (13.5)
	0.50	0.226	0.262	0.321	0.525
	0.95	0.036	0.041	0.049	0.077
$i\rho\sigma_G^c$	0.05	0.588 [-14.3]	0.679 [-12.5]	0.832 [-10.4]	1.359 [-6.5]
	0.50	0.227 [0.6]	0.263 [0.4]	0.322 [0.3]	0.526 [0.1]
	0.95	0.031 [17.1]	0.036 [14.6]	0.044 [11.8]	0.072 [7.0]

^a h^2 assumed to be equal to 0.20 and progeny group size equal to 125; ^b figures in parentheses (.) give percentage of asymmetry calculated as $100 \times (\text{upward-downward absolute responses})/\text{mean of upward and downward responses}$ for a fraction selected $Q = 0.05$, results being the same for $1 - Q = 0.95$ and nul for $Q = 1/2$; ^c figures within brackets [.] give percent relative errors, *ie* $100 \times (i\rho\sigma_G\text{-predicted value})/\text{predicted value}$ for upward selection of a fraction Q , results being exactly the same for downward selection of a fraction $(1 - Q)$.

EBVs are assumed to be computed *via* an approximate quasi-likelihood approach as described in Foulley and Im (1993). In that case, $\rho^2 = n/[4(n+k)]$ where $k = (4/h^2) - 1$, and $\theta^2 = 1/4$ represents the squared genetic relationship between sires among which selection is performed and offspring in which response is measured. The numerical application was made with a heritability coefficient h^2 assumed constant (here $h^2 = 0.20$) and σ_a^2 was calculated as $\sigma_a^2 = h^2/[\mu(1-h^2)]$ according to Foulley *et al* (1987) and Foulley and Im (1993). Other assumptions might have been made such as *eg* a constant coefficient of genetic variation (*ie* $\sigma_g/\mu = \sigma_a$ constant). Values of the normal CDF values were obtained using formulae of Ducrocq and Colleau (1986).

Results in table I show an asymmetric pattern of response such that, as expected, $|R^{(+) }|$ is larger than $|R^{(-) }|$ for a fraction selected $Q < 1/2$, and the opposite for $Q > 1/2$. Quantitatively, asymmetry turns out to be far from negligible. In the case considered here, asymmetry measured as the relative difference between upward and downward responses varies between 31 and 14% for μ ranging between 1.5 and 8.0 respectively, at $Q = 0.05$ or 0.95. As expected, the degree of asymmetry decreases here with the mean level. More generally, calculations show that asymmetry increases with progeny group size and heritability. This phenomenon makes prediction based on the usual formula $i\rho\sigma_G$ questionable at low μ values and for high and low values of the fraction selected Q : relative errors amount to 10 to 17% for μ between 1.5 and 3 and $Q = 0.05$ or 0.95.

The range in μ covers practical situations encountered in animal breeding for ovulation rate and prolificacy as observed *eg* in sheep ($\mu = 1.5$ to 3.0), rabbit and pig ($\mu = 8.0$), as well as for superovulation rate ($\mu = 8.0$), embryo production ($\mu = 3.0$) and development after transplantation ($\mu = 3.0$) in cattle. However, for some traits such as ovulation rate or litter size with low mean values, it would be better to use a truncated Poisson distribution with 0 excluded (Foulley and Im, 1987).

One main criticism against this approach lies in the assumption of normality for EBVs calculated with a log link function. As already discussed by Foulley (1992) and Gilmour *et al* (1985) for genetic evaluation of all-or-none traits with a probit transformation, this assumption is likely to be realistic if information per candidate is large enough (as in family selection); otherwise (*eg*, individual selection with few records per individual), one would have to make some adjustment in computing the expectation of $E(Y|\hat{a})$ over the real distribution of \hat{a} . Simulation studies would be required to clarify that point.

The main advantage of this approach consists of predicting an asymmetric pattern of response for Poisson distributed traits at high (or low) selection rates which was not the case with the usual formula $i\rho\sigma_G$. However, the adequacy of the Poisson-lognormal model as an appropriate tool to describe a multifactorial determinism for such traits can also be questioned. Alternative representations might be envisaged involving for instance models of multiplicative gene action (Cockerham, 1959; Dillmann, 1992). Predictions derived here might be useful to test the relevance of such models experimentally.

ACKNOWLEDGMENTS

The authors are grateful to V Ducrocq, M Perez Enciso and one anonymous reviewer for their helpful comments and criticisms.

REFERENCES

- Cockerham CC (1959) Partition of hereditary variance for various genetic models. *Genetics* 44, 1141-1148
- Curnow R (1984) Progeny testing for all-or-none traits when a multifactorial model applies. *Biometrics* 40, 375-382
- Dannel O, Rönningen K (1981) All-or-none traits in index selection. *J Anim Breed Genet* 98, 235-284
- Dempster ER, Lerner IM (1950) Heritability of threshold characters. *Genetics* 35, 212-236
- Dillmann C (1992) Organisation de la variabilité génétique chez les plantes: modélisation des effets d'interactions. PhD thesis, INA-PG, Paris
- Ducrocq V (1990) Estimation of genetic parameters arising in nonlinear models. In: *4th World Congr Genetics Applied to Livestock Production: Edinburgh, 23-27 July 1990, vol 13* (Hill WG, Thompson R, Wooliams JA, eds) p 419-428
- Ducrocq V, Colleau JJ (1986) Interests in quantitative genetics of Dutt's and Deak's methods for numerical computation of multivariate normal probability integrals. *Genet Sel Evol* 18, 447-474
- Falconer DS (1989) *Introduction to Quantitative Genetics*. Longman, London, 3rd edn
- Foulley JL (1992) Prediction of selection response for threshold dichotomous traits. *Genetics* 132, 1187-1194
- Foulley JL, Im S (1993) A marginal quasi likelihood approach to the analysis of Poisson variables with generalized linear mixed models. *Genet Sel Evol* 25, 101-107
- Foulley JL, Gianola D, Im S (1987) Genetic evaluation of traits distributed as Poisson-binomial with reference to reproductive characters. *Theor Appl Genet* 73, 870-877
- Gianola D (1982) Theory and analysis of threshold characters. *J Anim Sci* 54, 1079-1096
- Gilmour A, Anderson RD, Rae A (1985) The analysis of binomial data by a generalized linear mixed model. *Biometrika* 72, 593-599
- Johnson NL, Kotz S (1972) *Distributions in Statistics: Continuous Multivariate Distributions*. John Wiley and Sons, NY
- Perez Enciso M, Tempelman RJ, Gianola D (1993) A comparison between linear and Poisson models for litter size in Iberian pigs. *Livest Prod Sci* 35, 303-316

Chronique des livres

Dictionnaire de génétique (avec index anglais - français), sous la direction de JC Sournia, édité en 1991 par le Conseil international de la langue française, 352 pages, ISBN: 2-85319-231-8, 240 FF

Ce *Dictionnaire de génétique* est le fruit du travail d'une équipe de scientifiques français dont la diversité des horizons, allant de l'université à l'enseignement supérieur agronomique, garantit une couverture aussi large que possible de cette discipline, aujourd'hui devenue très vaste et très polymorphe. La liste des domaines couverts comporte en effet 15 secteurs qui vont de la cytogénétique aux ressources génétiques, en passant par la génétique des animaux domestiques et des plantes cultivées, la génétique évolutive, la génétique humaine, la génétique microbienne, la génétique moléculaire, la génétique quantitative, l'immuno-génétique, etc.

Au total, environ 2 700 termes et syntagmes sont recensés. Chacun d'eux est accompagné des informations suivantes :

- la traduction anglaise (un index placé à la fin de l'ouvrage permet de faire le passage inverse, de l'anglais au français) ;
- le ou les domaines concernés ;
- un symbole éventuel, par exemple cM pour centimorgan ;
- une ou plusieurs définitions succinctes ;
- des remarques explicatives complémentaires ;
- des renvois à d'autres termes du dictionnaire utiles à consulter sur le sujet.

A ces informations s'ajoutent quelques schémas explicatifs (par exemple à anticorps, *cis-trans*, coefficient de consanguinité, croisement, dendrogramme, etc.) et 6 annexes relatives au code génétique (annexe I) et à des aspects de reproduction végétale (annexes II et III) et de génétique humaine (annexes IV à VI).

La couverture d'ensemble est remarquablement à jour, sachant qu'elle correspond à la génétique de 1990. La comparaison avec le *Dictionnaire de génétique* de P L'Héritier - publié en 1978 chez Masson et dont, à lire la préface, les auteurs ne semblent pas avoir eu connaissance - est révélatrice de l'évolution de cette science, dans la mesure où la cote des mots peut en être un reflet. Si le gène reste toujours dominant dans le vocabulaire des généticiens (avec plus de 40 citations, contre une cinquantaine chez L'Héritier, où cependant le nombre total de termes est moindre), on note que la séquence lui dispute aujourd'hui la suprématie, puisqu'elle intervient dans 40 syntagmes, alors qu'elle n'avait pas obtenu son droit d'entrée pour le dictionnaire de 1978. Le généticien quantitatif, de son côté, trouvera dans ce dictionnaire d'abondantes références à la sélection (citée plus de 40 fois). Il y trouvera aussi son BLUP (suivi d'une définition qui ne fera sans doute pas l'unanimité) et son modèle animal, ce dernier distingué fort justement du modèle animal des généticiens humains.

La clarté et la précision des définitions témoignent du soin qui a présidé à leur formulation. De plus, les auteurs courent au-devant des contestations éventuelles en soulignant, à juste titre, que la génétique est une science neuve, dont la terminologie n'est pas encore fixée ni universellement admise. Cela est surtout vrai pour les secteurs les plus récemment développés, comme la génétique moléculaire.