

Quantifying the genetic contribution to the variability of count traits

Peer-reviewed author version

Oliveira, Izabela R.C.; MOLENBERGHS, Geert; Demétrio, C.G.B.; Dias, Carlos T.S. & Souza, Claudio L. (2015) Quantifying the genetic contribution to the variability of count traits. In: Friedl, Herwig; Wagner, Helga (Ed.). Proceedings of the 30th International Workshop on Statistical Modelling, p. 314-318.

Handle: <http://hdl.handle.net/1942/21022>

Quantifying the genetic contribution to the variability of count traits

Izabela R. C. Oliveira^{1,2}, Geert Molenberghs^{2,3}, Clarice G. B. Demétrio¹,
Carlos T. S. Dias¹, Cláudio L. Souza⁴

Abstract

Heritability is a important concept in animal and plant breeding, as it is in human biological applications. It is quantified based on fitting a model to hierarchical data. For data where linear models can be used, this attribute is conveniently defined as a ratio of variance components. Matters are less simple for non-Gaussian outcomes. The focus here is on count outcomes where extensions of the Poisson model are used to describe the data. Expressions for heritability of count traits are derived using the so-called Poisson combined model, which combines a Poisson outcome distribution with normal as well as gamma random effects, to capture both correlation among repeated observations as well as overdispersion, and admits closed-form expressions for the mean, variances and, hence, ratio of variances. It thus flexibly accommodates overdispersion and within-unit correlation. The proposed methodology is illustrated using maize data from a plant breeding program and compared with the usual, but questionable analysis using linear mixed models.

Keywords: Combined model, Gamma distribution, Generalized linear mixed model, Overdispersion, Poisson distribution, Random effect

1. Introduction

Plant breeding is the purposeful manipulation of plant species to improve certain aspects of plants, so as to perform new roles or enhance existing ones.

¹Department of Exact Sciences, ESALQ, Piracicaba, São Paulo, Brazil

²I-BioStat, Universiteit Hasselt, B-3500 Hasselt, Belgium

³I-BioStat, KU Leuven, B-3000 Leuven, Belgium

⁴Department of Genetics, ESALQ, Piracicaba, São Paulo, Brazil

Heritability, defined as the proportion of the genetic contribution over the total variability in a phenotype, is often of importance in plant and animal breeding. Knowledge of this attribute is useful to quantify the magnitude of improvement in the population and it is used when predicting the outcome of selection practiced among clones, inbred lines, or varieties (?).

The heritability determination is routinely based on hierarchical data of a family-based nature. When the outcomes are normally distributed, linear mixed models (?) are frequently used to estimate the genetic and environmental effects by considering these factors as random terms in the model. For data where these models can be used, the heritability can be quantified as the ratio of the genotypic variance, σ_g^2 say, to the total phenotypic variance, σ_p^2 say.

However, when the trait of interest is not normally distributed and/or it does not follow a linear model, the genetic and environmental random terms are no longer easily separable from the other model terms. This difficulty arises in particular when one deals with count outcomes, which are common in agricultural and livestock studies and are the focus of this paper. One often models such data using generalized linear modeling (?), which covers widely used statistical models, such as Poisson log-linear models for count data.

In empirical research, it has been observed recurrently that the mean-variance relationship for the Poisson model may not be met. As a result, quite a bit of research was devoted to formulate models to deal with this phenomenon, referred to as overdispersion or, sometimes also occurring, underdispersion. We will simply refer to it as overdispersion.

The so-called generalized linear mixed model (GLMM) (??) has gained popularity in discrete hierarchical data modeling. When overdispersion and the need for hierarchical modeling occur simultaneously, the combined models proposed by ? can be used. This model family accommodates overdispersion and clustering through two separate sets of random effects and contains as special cases the GLMM on the one hand, and several overdispersion models, such as the negative-binomial model, on the other. In this paper, we use such models for handling overdispersion and correlated data, while obtaining heritability coefficients based on count traits.

The proposed methodology will be illustrated using data from a study in plant breeding. The data refer to the number of tassel branches in maize progenies. This count trait is of interest in breeding procedures, with smaller numbers considered better. We derive heritability using two approaches: (a)

linear mixed models, a conventional but in principle inadequate analysis, and (b) using Poisson models.

The paper is organized as follows. In Section 2, the motivating case is described with both analyses reported in Section 5. A review of the Poisson combined model for hierarchical and overdispersed count data is the subject of Section 3. We use this combined model to obtain heritability coefficients for count traits, the expressions of which are presented in Section 4.

2. Motivating Case Study

The data considered here are obtained from a square lattice designed experiment, implemented in four different environments, for the selection of maize progenies in a plant breeding program at ESALQ, Piracicaba, Brazil. Forty-nine progenies were replicated twice in each environment, which consists of a combination between a crop season (first *versus* second crop season) and a location (ESALQ *versus* Sertãozinho). The planting during the first crop season usually occurs between September and December, while the planting in the second crop season takes place between January and April. This division of planting periods is common in tropical countries such as Brazil while in temperate regions there is only one planting season.

Five plants were randomly selected from each plot and several phenotypic characteristics were measured, including the kernel-row number per ear and the number of tassel branches. The latter is a characteristic related to drought tolerance; the smaller the number of branches, the better. This is because the plant does not need to move so many photo-assimilates there rather than move to the ear. On the other hand, the kernel-row number is a component of production, with higher values preferred.

The key research questions are what proportion of the variability of each of these traits is due to the genetic effect, i.e., the heritability for this population, and what progenies present the best predicted values with respect to these characteristics. As a summary of the data with respect to the number of tassel branches, Figure 1 shows the frequency plots, over both replications, over all progenies, and in each environment. In general, the number of branches varies from 7 to 29 and the distribution is approximately symmetric.

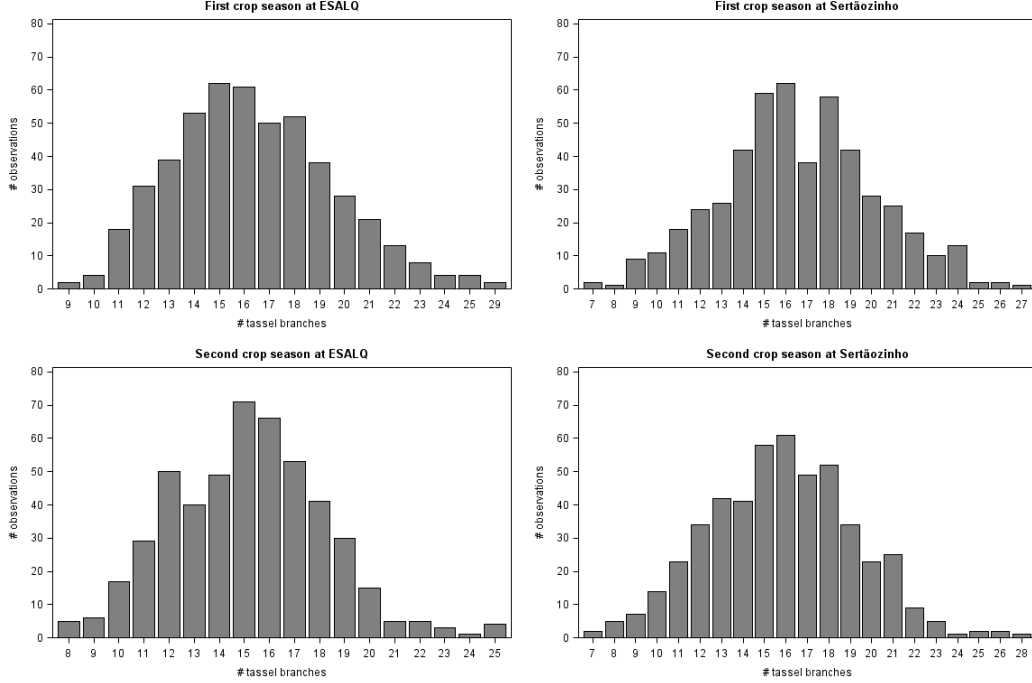


Figure 1: *Maize Data. Frequency plots for number of tassel branches, over both replications, over all progenies, and in each environment.*

3. An Extended Poisson Model to Handle Hierarchical and Overdispersed Data

In certain applications of standard generalized linear models, it is found that the data exhibit overdispersion, i.e., the variability is greater than predicted by the mean-variance relationship inherent in the model formulation. A number of models have been proposed for handling this phenomenon, especially in the Poisson case (??). Some also handle the rarer case of underdispersion.

An elegant formulation is through a two-stage approach. In the univariate Poisson case, we assume that $Y_i|\lambda_i \sim \text{Pois}(\lambda_i)$ and then that λ_i is a random variable with $E(\lambda_i) = \mu_i$ and $\text{Var}(\lambda_i) = \sigma_i^2$. Then it follows that

$$\begin{aligned} E(Y_i) &= E[E(Y_i|\lambda_i)] = E(\lambda_i) = \mu_i, \\ \text{Var}(Y_i) &= E[\text{Var}(Y_i|\lambda_i)] + \text{Var}[E(Y_i|\lambda_i)] = E(\lambda_i) + \text{Var}(\lambda_i) = \mu_i + \sigma_i^2. \end{aligned}$$

It is common to assume a gamma distribution for the random effects λ_i ,

leading to the so-called negative-binomial model.

This model can be extended to the case of repeated measurements. We then assume a hierarchical data structure where Y_{ij} denotes the j th outcome measured for cluster i , ($i = 1, \dots, N; j = 1, \dots, n_i$) and \mathbf{Y}_i is the n_i -dimensional vector of all measurements available for cluster i . The scalar λ_i becomes a vector $\boldsymbol{\lambda}_i = (\lambda_{i1}, \dots, \lambda_{in_i})'$, with $E(\boldsymbol{\lambda}_i) = \boldsymbol{\mu}_i$ and $\text{Var}(\boldsymbol{\lambda}_i) = \Sigma_i$. Then, $E(\mathbf{Y}_i) = \boldsymbol{\mu}_i$ and $\text{Var}(\mathbf{Y}_i) = M_i + \Sigma_i$, where M_i is a diagonal matrix with the vector $\boldsymbol{\mu}_i$ along the main diagonal. The diagonal structure of M_i reflects the conditional independence assumption, that is, all dependence between measurements on the same unit stems from the random effects.

Alternatively, this repeated version of the overdispersion model can be combined with normal random effects in the linear predictor. Such models, proposed also by ? and ?, for the count case, will be discussed next.

? specified a model for count data combining ideas from the overdispersion models and models with normal random effects. Later, ? proposed a broad class of generalized linear models where the binary, count, and time-to-event cases were given particular emphasis. These models, named combined models, accommodate overdispersion and clustering through two separate sets of random effects and produce models with only random effects and models with only overdispersion as special cases.

? showed that the count models allow for closed-form expressions for the mean vector and variance-covariance matrix. As highlighted by the authors, the derivation of such closed forms has important implications because they admit, for example, explicit correlation expressions. This aspect was examined by ? for Poisson-type models for count data.

In line with ?, ? specified a model for repeated Poisson data with overdispersion:

$$Y_{ij} \sim \text{Poi}(\theta_{ij}\lambda_{ij}), \quad (1)$$

$$\lambda_{ij} = \exp(x'_{ij}\boldsymbol{\beta} + z'_{ij}\mathbf{b}_i), \quad (2)$$

$$\mathbf{b}_i \sim N(\mathbf{0}, D), \quad (3)$$

$$E(\boldsymbol{\theta}_i) = E[(\theta_{i1}, \dots, \theta_{in_i})'] = \boldsymbol{\Phi}_i, \quad (4)$$

$$\text{Var}(\boldsymbol{\theta}_i) = \Sigma_i. \quad (5)$$

Notice that the normal random effect to capture correlation among repeated observations is placed in the linear predictor while the random effect to accommodate overdispersion acts multiplicatively in the mean of the variable.

The θ_{ij} can be assumed to follow a gamma model producing a Poisson-gamma-normal model or, equivalently, a negative-binomial-normal model. When the gamma distribution is chosen, it is assumed that the components θ_{ij} of $\boldsymbol{\theta}_i$ are independent. In this case, Σ_i reduces to a diagonal matrix. It should be noted that it is possible to allow for general covariance structures; this is not considered further here.

Then, regarding the overdispersion random effects, three instances could be of interest: (1) the random-effects θ_{ij} are independent; (2) the θ_{ij} are allowed to be dependent; (3) they are equal to each other and hence reduce to $\theta_{ij} = \theta_i$. Independent θ_{ij} imply that the use is strictly confined to capture additional overdispersion, i.e., not captured by the normal random effects. In contrast, when they are allowed to be correlated, they offer a way to model, for example, serial correlation.

The marginal mean vector and variance-covariance matrix were derived by ? and are reproduced in Appendix A. We considered this model to calculate the genetic contribution to the total variability of the traits of interest, that is, the heritability. The derivation of such measure is presented in the following section.

4. Derivation of Heritability for Count Data

Consider the Poisson-Gamma-Normal model and its variance presented in (A.1). Also, without loss of generality, set $E(\boldsymbol{\theta}_i) = 1$. The variance is

$$\text{Var}(\mathbf{Y}_i) = \mu_{ij} + \mu_{ij}(P_{i,jj} - 1)\mu_{ij}, \quad (6)$$

where

$$\mu_{ij} = \exp\left(x'_{ij}\boldsymbol{\beta} + \frac{1}{2}z'_{ij}Dz_{ij}\right) = \mu_{0ij}\mu_{1ij},$$

with notation as in (1)–(5), and

$$P_{i,jj} = \mu_{1ij}(\sigma_{i,jj} + 1)\mu_{1ij} = \mu_{1ij}^2(\sigma_{i,jj} + 1).$$

The non-genetic contribution over the total variability is:

$$\xi_{ij} = \frac{1 + \mu_{0ij}[(\sigma_{i,jj} + 1) - 1]}{\mu_{1ij}\{1 + \mu_{0ij}\mu_{1ij}[\mu_{1ij}^2(\sigma_{i,jj} + 1) - 1]\}}. \quad (7)$$

The ratio in (7) places the variance presented in (6), for $D = 0$, in the numerator and the full variance of the combined model in the denominator. The heritability, that is, the proportion of the total variability related to the genetic effect is:

$$H_{ij}^2 = 1 - \xi_{ij}. \quad (8)$$

The full variance is the phenotypic variance, σ_p^2 , and the obtained heritability is at individual level. Because of the mean-variance relationship, ξ_{ij} and hence H_{ij}^2 depends on the mean, thus also on the covariates.

A specific case arises when there is no overdispersion. In such case, the non-genetic contribution can be derived from the Poisson-Normal model or simply by setting $\sigma_{i,jj} = 0$ in (7). The contribution then is

$$\xi_{PN,ij} = \frac{1}{\mu_{1ij}[1 + \mu_{0ij}\mu_{1ij}(\mu_{1ij}^2 - 1)]}, \quad (9)$$

and the heritability is calculated from replacing ξ_{ij} by $\xi_{PN,ij}$ in (8).

It is not uncommon to model other sources of variability, such as the environmental effect. This effect can be included as a random term in model (1)–(5), which slightly changes the terms:

$$\begin{aligned} \lambda_{ij} &= \exp(x'_{ij}\boldsymbol{\beta} + z'_{1ij}\mathbf{v}_i + z'_{2ij}\mathbf{w}_i), \\ \mathbf{v}_i &\sim N(\mathbf{0}, D_1), \\ \mathbf{w}_i &\sim N(\mathbf{0}, D_2), \end{aligned}$$

where \mathbf{v}_i and \mathbf{w}_i are the genetic and environmental effects, respectively. Then, the variance is:

$$\text{Var}(Y_{ij}) = \mu_{ij} + \mu_{ij}(P_{i,jj} - 1)\mu_{ij},$$

where

$$\mu_{ij} = \exp\left(x'_{ij}\beta + \frac{1}{2}z'_{1ij}D_1z_{1ij} + \frac{1}{2}z'_{2ij}D_2z_{2ij}\right) = \mu_{0ij}\mu_{1ij}\mu_{2ij}$$

and

$$P_{i,jj} = \mu_{1ij}\mu_{2ij}(\sigma_{i,jj} + 1)\mu_{1ij}\mu_{2ij} = \mu_{1ij}^2\mu_{2ij}^2(\sigma_{i,jj} + 1).$$

In this case, the contribution from overdispersion and from the random effect \mathbf{w}_i over the total variability is:

$$\xi_{ij} = \frac{1 + \mu_{0ij}\mu_{2ij}[\mu_{2ij}(\sigma_{i,jj} + 1) - 1]}{\mu_{1ij}\{1 + \mu_{0ij}\mu_{1ij}\mu_{2ij}[\mu_{1ij}^2\mu_{2ij}^2(\sigma_{i,jj} + 1) - 1]\}}. \quad (10)$$

When there is no overdispersion, (10) reduces to:

$$\xi_{PN,ij} = \frac{1 + \mu_{0ij}\mu_{2ij}(\mu_{2ij} - 1)}{\mu_{1ij}[1 + \mu_{0ij}\mu_{1ij}\mu_{2ij}(\mu_{1ij}^2\mu_{2ij}^2 - 1)]}. \quad (11)$$

The heritability values are obtained from applying (10) or (11) to (8).

Notice that the ratios ξ_{ij} and $\xi_{PN,ij}$ are not free of the marginal mean function. In practice, therefore, one should compute some useful summaries of the values H_{ij} , given that they depend on the means at measurement j for unit i . Of course, when covariates are limited to a few factors with a limited number of levels, like is the case in our applications, the heritability coefficients will only be dependent on these.

5. Data Analysis

The data introduced in Section 2 were analyzed using two approaches. First, we used linear models, which are frequently used for genetic evaluation, referred to as ‘conventional analysis.’ In spite of its convenience, we consider it less adequate because it does not do full justice to the type of data collected. Second, we used the appropriate methodology from the previous section. From the estimates obtained under both approaches, we calculated the heritability for the number of tassel branches in this population.

It is noteworthy that, in this work, we do not model the progeny-environment interaction that can exist in this kind of study. So, all analyses were performed separately for each environment. Furthermore, we opted to work with the observations only at the plot level, which is the lowest and most informative hierarchical level, and did not estimate the block and replication nested effects of the lattice design.

The following linear predictor was considered in both approaches:

$$\eta_{ij} = \beta_0 + b_i,$$

where β_0 is an effect common to all observations, b_i is the genetic random effect of the i th progeny (49 progenies), and $b_i \sim N(0, \sigma_g^2)$. In the linear models, a random residual effect, ε_{ij} say, completes the model specification, where $\varepsilon_{ij} \sim N(0, \sigma^2)$.

5.1. The Conventional Linear Analysis

In this section, we will analyze the number of tassel branches considering the linear fixed-effects and linear mixed-effects models. Results from fitting both models are displayed in Table 1. Unsurprisingly, the goodness-of-fit of the mixed models is higher than that of univariate linear regression. From the estimated variance components and using the classical definition of heritability at the individual level, $H^2 = \sigma_g^2 / (\sigma_r^2 + \sigma_g^2)$, we calculated the genetic contribution in each one of the environments.

The heritability values are $H^2 = 0.1931$ for the first crop season at ESALQ, $H^2 = 0.3864$ for the first crop season at Sertãozinho, $H^2 = 0.3508$ for the second crop season at ESALQ, and $H^2 = 0.266$ for the second crop season at Sertãozinho.

5.2. The Proposed Poisson-type Analysis

In this section, we will analyze the number of tassel branches considering (1) the Poisson model (P--), (2) the Poisson Normal (P-N), (3) the Poisson-Gamma model (PG-), (4) the Poisson-Gamma-Normal model (PGN). Results from fitting these models are displayed in Table 2.

The estimates of the Gamma parameter were very high in all environments, demonstrating the absence of overdispersion in this application. Again, as expected, in all cases, the goodness-of-fit of the (P-N) models is higher than that of the (P--) models. So, the (P-N) estimates were used in (9) and (8) to obtain heritability.

The estimated non-genetic contributions and heritability values are $\xi_{PN} = 0.9219$ and $H^2 = 0.0781$ for the first crop season at ESALQ, $\xi_{PN} = 0.7828$ and $H^2 = 0.2172$ for the first crop season at Sertãozinho, $\xi_{PN} = 0.8635$ and $H^2 = 0.1365$ for the second crop season at ESALQ and $\xi_{PN} = 0.8586$, and $H^2 = 0.1414$ for the second crop season at Sertãozinho.

Another question of interest in plant breeding programs is to identify the best progeny with respect to the characteristic under investigation, and, for this, we plotted the empirical Bayes estimates for the random effects (Figure 2). In all environments, progeny 45 has the lowest predicted number of tassel branches, except for the second crop season at ESALQ, where progeny 49 showed superior performance.

Note that the heritability values calculated from the (P-N) models are smaller than those obtained from the linear-mixed model analysis. Close

Table 1: *Maize study. Parameter estimates and standard errors for the regression coefficients in (1) the linear model (LM) and (2) linear mixed model (LMM), considering the trait number of tassel branches in the four environments.*

Effect	Par.	(LM)	(LMM)
1. First crop season at ESALQ			
Intercept	β_0	16.359 (0.146)	16.359 (0.241)
Var. of progenies	σ_g^2		2.013 (0.579)
Var. residual	σ_r^2	10.410 (0.665)	8.397 (0.566)
-2log-likelihood		2538.5	2493.2
AIC		2542.5	2499.2
2. First crop season at Sertãozinho			
Intercept	β_0	16.571 (0.162)	16.571 (0.344)
Var. of progenies	σ_g^2		4.995 (1.171)
Var. residual	σ_r^2	12.931 (0.826)	7.935 (0.534)
-2log-likelihood		2644.8	2502.9
AIC		2648.8	2508.9
3. Second crop season at ESALQ			
Intercept	β_0	15.157 (0.138)	15.157 (0.282)
Var. of progenies	σ_g^2		3.292 (0.789)
Var. residual	σ_r^2	9.381 (0.599)	6.090 (0.410)
-2log-likelihood		2487.5	2366.8
AIC		2491.5	2372.8
4. Second crop season at Sertãozinho			
Intercept	β_0	15.818 (0.154)	15.818 (0.284)
Var. of progenies	σ_g^2		3.105 (0.803)
Var. residual	σ_r^2	11.687 (0.747)	8.582 (0.578)
-2log-likelihood		2595.2	2518.9
AIC		2599.2	2524.9

attention should be given to this point: a misspecification of the data distribution can lead to overestimated random effects and, hence, result in erroneously high heritability values.

It is clear that the original structure of the lattice design was not considered and a completely randomized design model was used. For purposes of illustration, it was an acceptable choice. By doing so, the block and replication effects were not modeled and their variation was added to the variance component σ_r^2 , which is overestimated. As a result, the estimated heritability is underestimated. Even the heritabilities for the linear mixed model are

underestimated; they are higher than those considering the Poisson distribution.

To illustrate the effect of model misspecification in the estimation of genetic quantities, we also analyzed the kernel-row number per ear. The frequency plots and the estimates of the models considered are presented in Appendix C. For this trait, there is a significant effect of the ear diameter covariate, considering the linear mixed models in all environments and the Poisson model in the first crop season at Sertãozinho. Also for this environment, the (PG-) and (PGN) models failed to converge, but this is not a surprise due to the high estimates of the Gamma parameter and null estimates of random effects.

If one assumes the normal distribution for this count trait, the heritability values are $H^2 = 0.0806$ for the first crop season at ESALQ, $H^2 = 0.1640$ for the first crop season at Sertãozinho, $H^2 = 0.2936$ for the second crop season at ESALQ and $H^2 = 0.1590$ for the second crop season at Sertãozinho. On the other hand, considering the Poisson models, there are neither significant overdispersion nor significant normal random effects and, hence, the heritability is not significantly different from 0.

The amount of genetic variation determines the rate of change of a trait under selection and if there is no genetic variation, there is no response to selection (?).

6. Concluding Remarks

In this paper, we have derived a principled expression for heritability, based on hierarchical count data. To these data, Poisson-based mixed models as well as and linear mixed models have been fitted and compared. Although conventional, we consider inadequate the use of linear (mixed) models for count data and showed that, using this approach, the genetic random effects can be overestimated, leading to incorrect heritability values. Of course, when counts would be very large and at least approximately normally distributed, the above judgment can be relaxed.

The so-called combined model was used in the Poisson approach. It brings together a generalized linear model for count data with both normal and gamma random effects, thus accommodating correlation and overdispersion.. Importantly, the standard generalized linear mixed model is a special case of the combined model, implying that the derivations reported here also apply to this commonly encountered GLMM case.

Table 2: *Maize study. Parameter estimates and standard errors for the regression coefficients in (1) the purely Poisson model (P-), (2) the Poisson-Normal model (P-N), (3) the Poisson-Gamma model (PG-), and (4) the Poisson-Gamma-Normal model (PGN), considering the trait number of tassel branches in the four environments.*

Effect	Par.	(P-)	(P-N)	(PG-)	(PGN)
1. First crop season at ESALQ					
Intercept	β_0	2.795 (0.011)	2.793 (0.015)	2.795 (0.011)	2.793 (0.015)
Gamma param.	α			1000.020 (580.310)	7638.150 (11268)
Var. of progenies	σ_g^2		0.005(0.002)		0.005 (0.002)
-2log-likelihood		2573.6	2564.4	-28771	-28783
AIC		2575.6	2568.4	-28767	-28777
2. First crop season at Sertãozinho					
Intercept	β_0	2.808 (0.011)	2.800 (0.021)	2.808 (0.011)	2.800 (0.021)
Gamma param.	α			5982.350 (10830)	7735.860 (11025)
Var. of progenies	σ_g^2		0.016 (0.005)		0.016 (0.005)
-2log-likelihood		2658.3	2595.4	-29.356	-29419
AIC		2660.3	2599.4	-29352	-29413
3. Second crop season at ESALQ					
Intercept	β_0	2.719 (0.012)	2.713 (0.019)	2.719 (0.012)	2.713 (0.019)
Gamma param.	α			6786.550 (10475)	7791.040 (10969)
Var. of progenies	σ_g^2		0.010 (0.004)		0.010 (0.004)
-2log-likelihood		2532.4	2501.6	-25526	-25556
AIC		2534.4	2505.6	-25522	-25550
4. Second crop season at Sertãozinho					
Intercept	β_0	2.761 (0.011)	2.756 (0.018)	2.761 (0.011)	2.756 (0.018)
Gamma param.	α			6008.200 (10200)	7660.940 (12063)
Var. of progenies	σ_g^2		0.010 (0.003)		0.010 (0.003)
-2log-likelihood		2614.3	2584.9	-27301	-27331
AIC		2616.3	2588.9	-27297	-27325

The combined model and its GLMM sub-model admit closed-form expressions for means, variances, and higher-order moments. As a result, variance ratios have explicit expressions too. The heritability coefficient is sufficiently simple and appealing, in particular in special cases, to be of practical value.

We want to reiterate that, in these models, heritability is a function rather than a constant. At first sight, this is a drawback. However, it is a consequence from the mean-variance relationship in the models considered. If the model fits the data well, it can also be claimed to be a feature of the data. Practically, heritability changes with the effects present in the predictor functions. Evidently, one can summarize the functions in a variety of ways, using averages, medians, quartiles, ranges, etc.

Acknowledgments

Geert Molenberghs gratefully acknowledges support from IAP research Network P7/06 of the Belgian Government (Belgian Science Policy). This

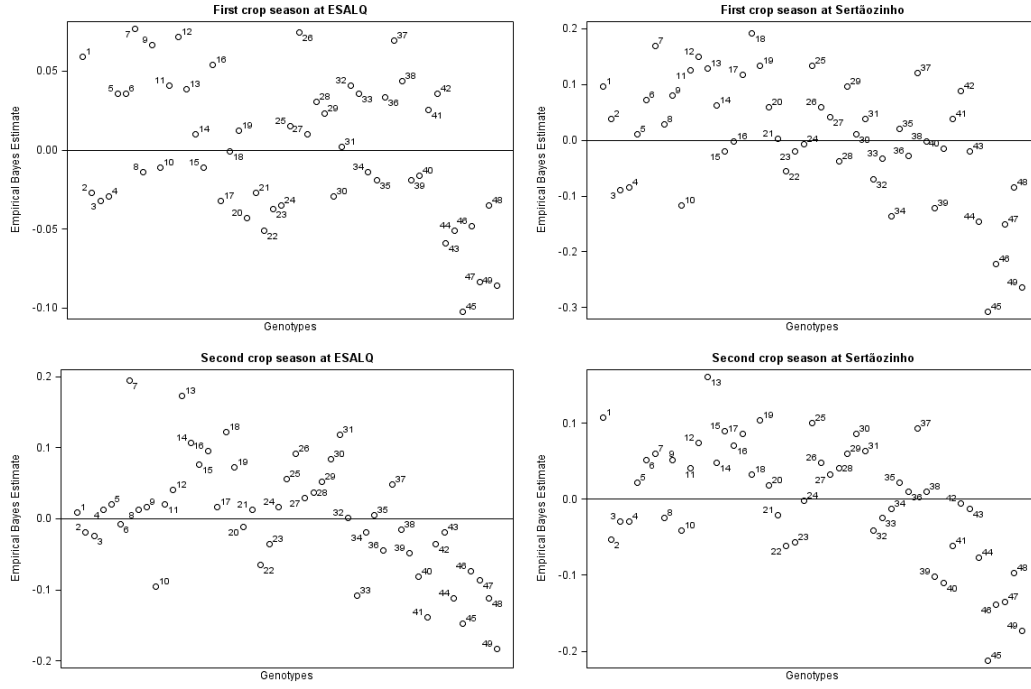


Figure 2: *Maize Data. Empirical Bayes Estimates for number of tassel branches in each environment.*

work was partially supported by grants from CNPq, National Council for Scientific and Technological Development, Brazilian science funding agency.

Appendix A. Model Elements for the Poisson Combined Model

The mean and variance expressions for the Poisson combined model (1)–(5) were presented by ?. The mean vector $\boldsymbol{\mu}_i = E(\mathbf{Y}_i)$ has components

$$\mu_{ij} = \phi \exp \left(x'_{ij} \boldsymbol{\beta} + \frac{1}{2} z'_{ij} D z_{ij} \right)$$

and the variance-covariance matrix is given by

$$\text{Var}(\mathbf{Y}_i) = M_i + M_i(P_i - J_{n_i})M_i, \quad (\text{A.1})$$

where M_i is a diagonal matrix with the vector $\boldsymbol{\mu}_i$ along the diagonal and the $(j, k)^{th}$ element of P_i equals

$$p_{i,jk} = \exp \left(\frac{1}{2} z'_{ij} D z_{ik} \right) \frac{\sigma_{i,jk} + \phi_{ij} \phi_{ik}}{\phi_{ij} \phi_{ik}} \exp \left(\frac{1}{2} z'_{ik} D z_{ij} \right).$$

Note that $\sigma_{i,jk}$ is the (j, k) th element of Σ_i .

These expressions also produce their simplified counterparts for the special cases, including the Poisson-normal model and the Poisson model. For instance, when only normal random effects are present, the mean vector components slightly simplify:

$$\mu_{ij} = \exp \left(x'_{ij} \boldsymbol{\beta} + \frac{1}{2} z'_{ij} D z_{ij} \right),$$

and the variance-covariance matrix is

$$\text{Var}(\mathbf{Y}_i) = M_i + M_i(e^{Z_i D Z'_i} - J_{n_i})M_i.$$

Appendix B. SAS Implementation

The SAS programs, using the procedure NLMIXED, for the linear mixed model and Poisson models are as follows.

```
proc nlmixed data=env1 corr;
title 'NTB: Linear mixed model';
parms beta0=16.36 sigma=3.23;
mean = beta0 + b;
model ntb ~ normal(mean,sigma*sigma);
```

```

random b ~ normal(0,d*d) subject=gen;
estimate 'Variance Prog.' d*d;
estimate 'Variance Residual' sigma*sigma;
run;

```

The special case of the Poisson model simply follows from removing the RANDOM statement and the b effect in the linear predictor.

```

proc nlmixed data=env1 tech=NRRIDG qpoints=50 corr;
title 'NTB: Poisson-Normal model';
parms beta0=2.7948;
eta = beta0 + b;
lambda = exp(eta);
model ntb ~ poisson(lambda);
random b ~ normal(0,d*d) subject=gen;
estimate 'Variance Prog.' d*d;
run;

```

The above program makes use of the built-in Poisson likelihood. The Poisson-Gamma-Normal model, or the combined model, need to be implemented using the so-called *general likelihood*, i.e., the user defined likelihood feature. The resulting program is as follows:

```

proc nlmixed data=env1 tech=NRRIDG qpoints=50;
title 'NTB: Poisson-Gamma-Normal model';
parms beta0=2.79 alpha=200 d=0.06775;
eta = beta0 + b;
lambda = exp(eta);
ll = lgamma(alpha + ntb) - lgamma(alpha) + ntb*log(1/alpha) -
    (ntb + alpha)*log(1 + (1/alpha)*lambda) + ntb*eta;
model ntb ~ general(ll);
random b ~ normal(0,d*d) subject=gen;
estimate 'Variance Prog.' d*d;
run;

```

The special case of the Poisson-Gamma model or, equivalently, the negative binomial model, follows from removing the RANDOM statement in the previous program.

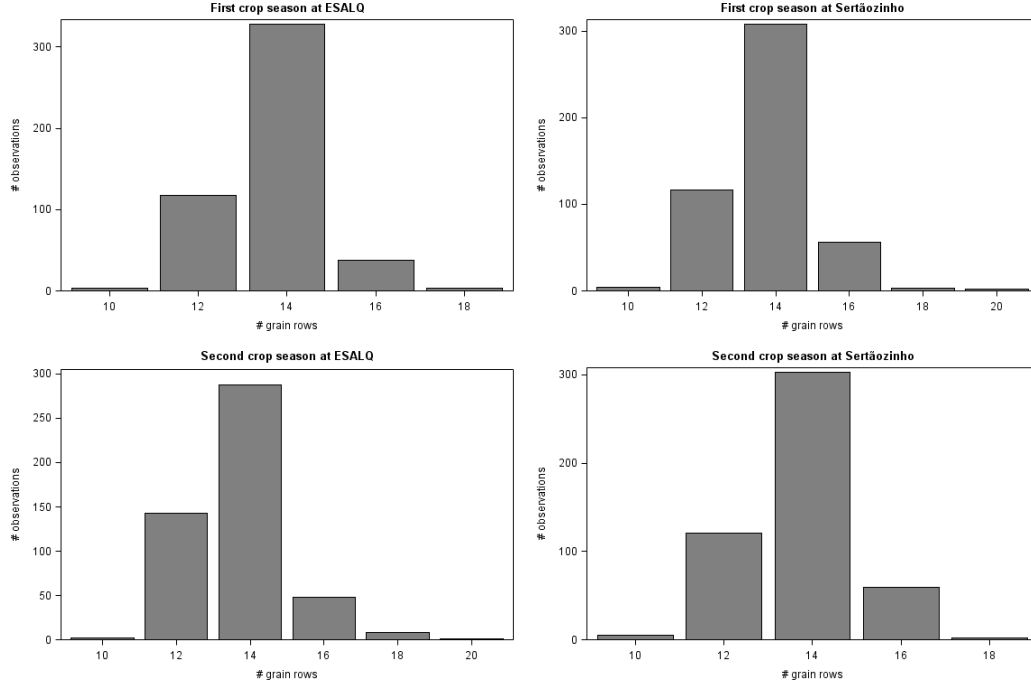


Figure C.3: *Maize Data*. Frequency plots for kernel-row number per ear, over both replications, over all progenies and in each environment.

Appendix C. Results for Kernel-row Number

Figure C.3 shows the frequency plots for the characteristic kernel-row number per ear, over both replications, over all progenies and in each environment. This count is always an even number that varies from 10 to 20 in this population. The estimates of the linear mixed model and Poisson models are displayed in Table C.3.

Table C.3: *Maize study. Parameter estimates and standard errors for the regression coefficients in (1) the linear mixed model (LMM), (2) the purely Poisson model (P-), (3) the Poisson-Normal model (P-N), (4) the Poisson-Gamma model (PG-), and (5) the Poisson-Gamma-Normal model (PGN), considering the trait kernel-row number per ear in the four environments.*

Effect	Par.	(LMM)	(P-)	(P-N)	(PG-)	(PGN)
1. First crop season at ESALQ						
Intercept	β_0	10.059 (0.939)	2.615 (0.012)	2.615 (0.012)	2.615 (0.012)	2.615 (0.012)
Ear diameter	β_1	0.739 (0.192)				
Gamma param.	α				9116.39 (11350)	1200.00 (539.77)
Var. of progenies	σ_g^2	0.107 (0.048)		1.15×10^{-14}		1.18×10^{-14}
Var. residual	σ_r^2	1.219 (0.082)				
-2log-likelihood		1518.6	2236.1	2236.1	-21636	-21632
AIC		1526.6	2238.1	2240.1	-21632	-21626
2. First crop season at Sertãozinho						
Intercept	β_0	8.505 (0.997)	2.156 (0.200)	2.156 (0.200)		
Ear diameter	β_1	1.096 (0.207)	0.097 (0.042)	0.097 (0.042)		
Gamma param.	α					
Var. of progenies	σ_g^2	0.260 (0.081)		6.51×10^{-27}		
Var. residual	σ_r^2	1.325 (0.089)				
-2log-likelihood		1581.8	2245.1	2245.1		
AIC		1589.8	2249.1	2251.1		
3. Second crop season at ESALQ						
Intercept	β_0	12.236 (0.983)	2.615 (0.012)	2.615 (0.012)	2.615 (0.012)	2.615 (0.012)
Ear diameter	β_1	0.307 (0.209)				
Gamma param.	α				9131.74 (11593)	800.00 (300.44)
Var. of progenies	σ_g^2	0.534 (0.136)		3.42×10^{-17}		2.51×10^{-12}
Var. residual	σ_r^2	1.285 (0.087)				
-2log-likelihood		1593.9	2251.3	2251.3	-21646	-21640
AIC		1601.9	2253.3	2255.3	-21642	-21634
4. Second crop season at Sertãozinho						
Intercept	β_0	9.967 (1.064)	2.619 (0.012)	2.619 (0.012)	2.619 (0.012)	2.619 (0.012)
Ear diameter	β_1	0.779 (0.220)				
Gamma param.	α				8410.39 (10023)	1000.00 (414.27)
Var. of progenies	σ_g^2	0.248 (0.077)		1.39×10^{-16}		8.75×10^{-15}
Var. residual	σ_r^2	1.312 (0.088)				
-2log-likelihood		1575.6	2245.9	2245.9	-21772	-21767
AIC		1583.6	2247.9	2249.9	-21768	-21761