Supplementary Appendices to "Applying Multivariate Discrete Distributions to Genetically Informative Count Data" by R. M. Kirkpatrick & M. C. Neale, submitted to *Behavior Genetics*

Appendix A: The Univariate Poisson and Negative-Binomial Distributions

The Univariate Poisson Distribution

In this section, we briefly review widely known or easily shown properties of the Poisson distribution. Further details may be found in such references as Forbes, Evans, Hastings, & Peacock (2011) and Johnson, Kemp, & Kotz (2005, chapter 4). The Poisson distribution is unimodal, with support on the set of nonnegative integers. It is the limiting distribution of sums of i.i.d. Bernoulli trials, specifically, it is the limiting form of the binomial distribution as index parameter $n \to \infty$, Bernoulli parameter $\wp \to 0$, but the product $n\wp$ is held constant at some value θ. This parameter, θ, is the sole parameter of the univariate Poisson, being both its mean and variance. It is proportional to the rate parameter for a Poisson process, a stochastic process characterized by its "memoryless" property, with interarrival times that follow an Exponential distribution. If $X \sim \text{Pois}(\theta)$, then the PMF of X is

$$p_X(x) = \frac{\theta^x}{x!}\exp(-\theta) \qquad (A1)$$

for x = 0, 1, 2… (zero otherwise) and $\theta > 0$. Though it may be slightly unorthodox, we will define a Poisson RV with $\theta = 0$ as having unit mass on the event X = 0.

The Poisson possesses an "addition rule" (or more formally, a "convolution property"). Suppose $X_1, \dots, X_n$ are independent Poisson RVs, with $X_i \sim \text{Pois}(\theta_i)$, for $i = 1, \dots, n$. Then, $\sum_{i=1}^{n} X_i$ is also a Poisson RV, with parameter equal to $\sum_{i=1}^{n} \theta_i$. This addition rule is critical to the construction of the multivariate Poisson distribution.

Though appealing in its simplicity, the Poisson often poorly approximates the observed distributions of real-data count variables. Perhaps the most common deviation from Poisson distribution is overdispersion, which refers to the distribution's variance exceeding its mean. If the true data-generating distribution is overdispersed, then Poisson-based estimates of variance components cannot even be asymptotically unbiased. We therefore turn our attention to two other univariate distributions, the negative binomial and the Lagrangian Poisson, which allow for overdispersion (relative to the Poisson), and are similar to one another in many respects.

The Univariate Negative Binomial Distribution

In this section, we briefly review widely known or easily shown properties of the negative binomial distribution. Details may be found in Forbes et al. (2011), Johnson et al. (2005, chapter 5), and Cameron & Trivedi (1986).

The negative binomial distribution is unimodal and has support on the set of nonnegative integers. It may be derived from "inverse sampling" of Bernoulli trials. Suppose we are to observe a sequence of i.i.d. Bernoulli trials, each of which has probability of success equal to $\wp$, until we have observed some critical number ν of successes. As is conventional, let $q = 1 - \wp$.

The count of trials ending in failure that precede the νth success is a random variable, *X*, which follows a *Pascal* distribution. In the special case that $\nu = 1$, we are dealing with a *geometric* distribution. The negative binomial distribution is a generalization of the Pascal distribution, in which parameter ν may take non-integer values.

The negative binomial may also be derived as a Gamma mixture of Poisson. Consider a RV *X*, following a Poisson distribution in which the parameter θ itself is a random variable, following a Gamma distribution with shape parameter ν and rate parameter $\mathscr{p} \div \mathscr{q}$. When θ is integrated out of the joint density of *X* and θ, the result is that *X* marginally has a negative binomial distribution with index parameter ν and Bernoulli parameter $\mathscr{p}$. Symbolically, $X \sim \mathrm{NB}(\nu, \mathscr{p})$, with PMF

$$p_X(x) = \binom{\nu + x - 1}{\nu - 1} \mathscr{p}^\nu \mathscr{q}^x = \frac{\Gamma(\nu + x)}{\Gamma(x + 1)\Gamma(\nu)} \mathscr{p}^\nu q^x \qquad (A2)$$

for $x = 0, 1, 2, \ldots$ (zero otherwise), $0 < \mathscr{p} < 1$, and $\nu > 0$. We will define a negative-binomial RV with $\nu = 0$ as one with unit mass on the event $X = 0$. The expectation and variance of the negative binomial are:

$$\mathrm{E}(X) = \frac{\nu q}{\mathscr{p}} \qquad (A3)$$

$$\mathrm{var}(X) = \frac{\nu q}{\mathscr{p}^2} \qquad (A4)$$

Thus, the distribution is obligatorily overdispersed relative to Poisson—its variance always exceeds its mean. Indeed, the Bernoulli parameter $\mathscr{p}$ is the ratio of the mean to the variance.

Like the Poisson, the negative binomial also has an addition rule. Suppose $X_1, \ldots, X_n$ are independent negative-binomial RVs, with $X_i \sim \mathrm{NB}(\nu_i, \mathscr{p})$, for $i = 1, \ldots, n$. Then, $Y = \sum_{i=1}^n X_i$ is also a negative-binomial random variable, with $Y \sim \mathrm{NB}(\sum_{i=1}^n \nu_i\,, \mathscr{p})$.

Appendix B. The Bivariate Poisson Distribution, with Application to Twin Modeling

The bivariate Poisson (Teicher, 1954; Holgate, 1964; Johnson, Kotz, & Balakrishnan, 1997) is constructed as follows. Consider three independent (latent) RVs X_0, X_1, and X_2, where

$$X_0 \sim \text{Pois}(\theta_0)$$
$$X_1 \sim \text{Pois}(\theta_1)$$
$$X_2 \sim \text{Pois}(\theta_2) \qquad (B1)$$

Now, define (observable) RVs Y_1 and Y_2, where

$$Y_1 = X_0 + X_1$$
$$Y_2 = X_0 + X_2 \qquad (B2)$$

Then,

$$Y_1 \sim \text{Pois}(\theta_0 + \theta_1)$$
$$Y_2 \sim \text{Pois}(\theta_0 + \theta_2) \qquad (B3)$$

and Y_1 and Y_2 jointly follow a bivariate Poisson distribution, with $\text{cov}(Y_1, Y_2) = \theta_0$. Since the latent variables X_0, X_1, and X_2 are independent, their joint PMF is

$$\begin{aligned} p_{\mathbf{X}}(x_0, x_1, x_2) &= p_{X0}(x_0) \cdot p_{X1}(x_1) \cdot p_{X2}(x_2) \\ &= p_{X0}(x_0) \cdot p_{X1}(y_1 - x_0) \cdot p_{X2}(y_2 - x_0) \end{aligned} \qquad (B4)$$

Logically, x_0 cannot exceed the smaller of the pair (y_1, y_2). The distribution of Y_1 and Y_2, after marginalizing out X_0, is therefore given by

$$\begin{aligned} p_{\mathbf{Y}}(y_1, y_2) &= \sum_{x_0=0}^{\min(y_1,y_2)} p_{X0}(x_0) \cdot p_{X1}(y_1 - x_0) \cdot p_{X2}(y_2 - x_0) \\ &= \exp(-\theta_0 - \theta_1 - \theta_2) \sum_{x_0=0}^{\min(y_1,y_2)} \frac{\theta_0^{x_0}}{x_0!} \cdot \frac{\theta_1^{y_1 - x_0}}{(y_1 - x_0)!} \cdot \frac{\theta_2^{y_2 - x_0}}{(y_2 - x_0)!} \end{aligned} \qquad (B5)$$

We will here describe our application of the bivariate Poisson distribution to twin modeling in the simplest case, the monophenotype ACE model in a classical twin study. For MZ twins,

$$X_0 \sim \text{Pois}(V_A + V_C)$$
$$X_1, X_2 \sim \text{Pois}(V_E) \qquad (B6)$$

and therefore,

$$Y_1, Y_2 \sim \text{Pois}(V_A + V_C + V_E)$$
$$\text{cov}(Y_1, Y_2) = V_A + V_C \qquad (B7)$$

For DZ twins,

$$X_0 \sim \text{Pois}(V_C + 0.5V_A)$$
$$X_1, X_2 \sim \text{Pois}(V_E + 0.5V_A) \qquad (B8)$$

and therefore,

$$Y_1, Y_2 \sim \text{Pois}(V_A + V_C + V_E)$$
$$\text{cov}(Y_1, Y_2) = 0.5V_A + V_C \qquad (B9)$$

This model is depicted as a path diagram in the figure. The model can be fit to raw data, and maximum-likelihood estimates of variance components V_A, V_C, and V_E can be obtained.

Figure S1. Monophenotype bivariate-Poisson twin model, for MZ (A) and DZ (B) twins.

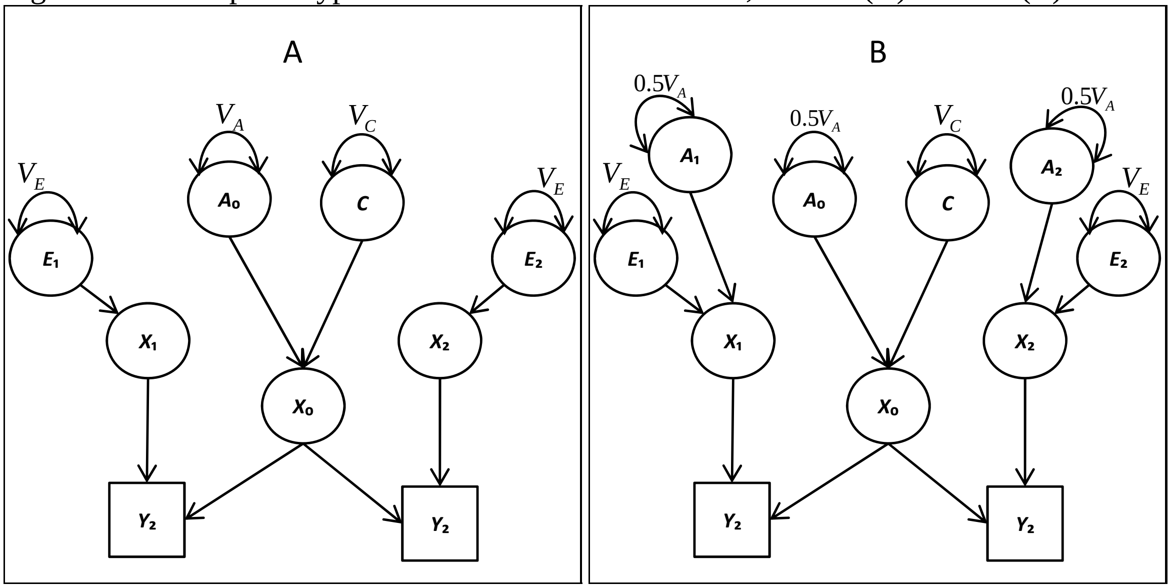

Figure provides visual representation of Eq.s (B6) thru (B9). All unlabeled single-headed arrows have path coefficients of 1.0. The latent variables are Poisson distributed, and therefore their means (not shown) are equal to their variances.

Appendix C. The Trivariate Poisson Distribution

Define three rvs Y_1, Y_2, and Y_3, where

$$\begin{aligned} Y_1 &= T + U_{12} + U_{13} + W_1 \\ Y_2 &= T + U_{12} + U_{23} + W_2 \\ Y_3 &= T + U_{23} + U_{13} + W_3 \end{aligned} \qquad (C1)$$

and T, U_{12}, U_{23}, U_{13}, W_1, W_2, and W_3 are mutually independent rvs such that

$$\begin{aligned} T &\sim \text{Pois}(\theta_0) \\ U_{12} &\sim \text{Pois}(\theta_{12}) \\ U_{23} &\sim \text{Pois}(\theta_{23}) \\ U_{13} &\sim \text{Pois}(\theta_{13}) \\ W_1 &\sim \text{Pois}(\theta_1) \\ W_2 &\sim \text{Pois}(\theta_2) \\ W_3 &\sim \text{Pois}(\theta_3) \end{aligned} \qquad (C2)$$

Then,

$$\begin{aligned} Y_1 &\sim \text{Pois}(\theta_0 + \theta_{12} + \theta_{13} + \theta_1) \\ Y_2 &\sim \text{Pois}(\theta_0 + \theta_{12} + \theta_{23} + \theta_2) \\ Y_3 &\sim \text{Pois}(\theta_0 + \theta_{23} + \theta_{13} + \theta_3) \end{aligned} \qquad (C3)$$

and Y_1, Y_2, and Y_3 jointly follow a trivariate Poisson distribution. The variance matrix of their joint distribution is

$$\mathbf{\Sigma} = \begin{bmatrix} \theta_0 + \theta_{12} + \theta_{13} + \theta_1 & \theta_0 + \theta_{12} & \theta_0 + \theta_{13} \\ \theta_0 + \theta_{12} & \theta_0 + \theta_{12} + \theta_{23} + \theta_2 & \theta_0 + \theta_{23} \\ \theta_0 + \theta_{13} & \theta_0 + \theta_{23} & \theta_0 + \theta_{23} + \theta_{13} + \theta_3 \end{bmatrix} \qquad (C4)$$

The distribution is specified by seven parameters, and $\mathbf{\Sigma}$ has six unique elements. Therefore, it may be parametrized in terms of its variance matrix $\mathbf{\Sigma}$ and the common-to-all component, θ_0. For ease of notation, let $\mathbf{X}$ represent the vector of latent variables, i.e. $\mathbf{X} = [T, U_{12}, U_{23}, U_{13}, W_1, W_2, W_3]^T$. Because the latent variables are independent, their joint PMF is equal to the product of their marginal PMFs:

$$p_{\mathbf{X}}(\mathbf{x}) = p_t(t) \cdot p_{u12}(u_{12}) \cdot p_{u23}(u_{23}) \cdot p_{u13}(u_{13}) \cdot p_{w1}(w_1) \cdot p_{w2}(w_2) \cdot p_{w3}(w_3)$$

Upon rearrangement of and substitution from (C1), we can express w_1, w_2, and w_3 in terms of the observable variables, y_1, y_2, and y_3, and the other latent variables:

$$\begin{aligned} p_{\mathbf{X}}(\mathbf{x}) = p_t(t) \cdot p_{u12}(u_{12}) \cdot p_{u23}(u_{23}) \cdot p_{u13}(u_{13}) \cdot p_{w1}(y_1 - u_{13} - u_{12} - t) \\ \cdot p_{w2}(y_2 - u_{23} - u_{12} - t) \cdot p_{w3}(y_3 - u_{23} - u_{13} - t) \end{aligned}$$

To obtain the joint PMF of $\mathbf{Y} = [Y_1, Y_2, Y_3]^T$, all that remains is to marginalize T, U_{12}, U_{23}, and U_{13} out of the expression:

$$
\begin{aligned}
p_{\mathbf{Y}}(\mathbf{y}) &= \sum_{t=0}^{\min(\mathbf{y})} \sum_{u_{12}=0}^{\min(y_1,y_2)-t} \sum_{u_{13}=0}^{\min(y_1-u_{12},y_3)-t} \sum_{u_{23}=0}^{\min(y_2-u_{12},y_3-u_{13})-t} p_{\mathbf{X}}(\mathbf{x}) \\
&= \sum_{t=0}^{\min(\mathbf{y})} p_t(t) \sum_{u_{12}=0}^{\min(y_1,y_2)-t} p_{u12}(u_{12}) \sum_{u_{13}=0}^{\min(y_1-u_{12},y_3)-t} p_{u13}(u_{13}) \sum_{u_{23}=0}^{\min(y_2-u_{12},y_3-u_{13})-t} p_{u23}(u_{23}) \cdot p_{w1}(y_1-u_{13}-u_{12}-t) \\
&\qquad \cdot p_{w2}(y_2-u_{23}-u_{12}-t) \cdot p_{w3}(y_3-u_{23}-u_{13}-t) \qquad (C5) \\
&= \exp(-\theta_0-\theta_{12}-\theta_{23}-\theta_{13}-\theta_1-\theta_2-\theta_3) \cdot \theta_1^{y_1}\theta_2^{y_2}\theta_3^{y_3} \\
&\qquad \cdot \sum_{t=0}^{\min(\mathbf{y})} \frac{\theta_0^t}{t!} \sum_{u_{12}=0}^{\min(y_1,y_2)-t} \frac{\theta_{12}^{u_{12}}}{u_{12}!} \sum_{u_{13}=0}^{\min(y_1-u_{12},y_3)-t} \frac{\theta_{13}^{u_{13}}}{u_{13}!} \sum_{u_{23}=0}^{\min(y_2-u_{12},y_3-u_{13})-t} Q
\end{aligned}
$$

where

$$
Q = \frac{\theta_{23}^{u_{23}}\theta_1^{-u_{13}-u_{12}-t}\theta_2^{-u_{23}-u_{12}-t}\theta_3^{-u_{13}-u_{23}-t}}{u_{23}!\,(y-u_{13}-u_{12}-t)!\,(y-u_{23}-u_{12}-t)!\,(y-u_{13}-u_{23}-t)!}
$$

We note that the form in ($C5$) holds generally for trivariate discrete distributions constructed via latent-variate reduction.

Appendix D. Diphenotype and Triphenotype Twin Analysis with Multivariate Poisson

Let $\mathbf{Y} = [Y_{11}, Y_{12}, Y_{21}, Y_{22}]^T$ denote a multivariate-Poisson random vector. The first of the two subscripts distinguishes twin #1 and twin #2 in a pair from one another, whereas the second subscript distinguishes one phenotype from the other. For example, Y_{21} would represent twin #2's scores on phenotype #1. Define the elements of **Y** as follows:

$$\begin{aligned} Y_{11} &= T + U_1 + V_1 + W_1 \\ Y_{12} &= T + U_1 + V_2 + W_2 \\ Y_{21} &= T + U_2 + V_1 + W_3 \\ Y_{22} &= T + U_2 + V_2 + W_4 \end{aligned} \qquad (D1)$$

where

$$\begin{aligned} T &\sim \text{Pois}(\theta_t) \\ U_1 &\sim \text{Pois}(\theta_{u1}) \\ U_2 &\sim \text{Pois}(\theta_{u2}) \\ V_1 &\sim \text{Pois}(\theta_{v1}) \\ V_2 &\sim \text{Pois}(\theta_{v2}) \\ W_1 &\sim \text{Pois}(\theta_{w1}) \\ W_2 &\sim \text{Pois}(\theta_{w2}) \\ W_3 &\sim \text{Pois}(\theta_{w3}) \\ W_4 &\sim \text{Pois}(\theta_{w4}) \end{aligned} \qquad (D2)$$

Intuitively, T represents what is common to both traits and both twins, and accounts for the cross-trait cross-twin covariance; U_1 and U_2, what is common to both traits within a given twin (#1 or #2); V_1 and V_2, what contributes to the cross-twin covariance within a given trait (#1 or #2); and the W's, what is unique to a particular twin on a particular trait. Most of the time, twin data are of an intraclass nature and it is arbitrary which twin is #1 or #2. Then, it can be assumed that $\theta_{u1} = \theta_{u2} \equiv \theta_u$, that $\theta_{w1} = \theta_{w3} \equiv \phi_1$, and that $\theta_{w2} = \theta_{w4} \equiv \phi_2$. With this assumption, the covariance matrix of **Y** is

$$var(\mathbf{Y}) = \mathbf{\Sigma}$$
$$= \begin{bmatrix} \theta_t + \theta_u + \theta_{v1} + \phi_1 & \theta_t + \theta_u & \theta_t + \theta_{v1} & \theta_t \\ \theta_t + \theta_u & \theta_t + \theta_u + \theta_{v2} + \phi_2 & \theta_t & \theta_t + \theta_{v2} \\ \theta_t + \theta_{v1} & \theta_t & \theta_t + \theta_u + \theta_{v1} + \phi_1 & \theta_t + \theta_u \\ \theta_t & \theta_t + \theta_{v2} & \theta_t + \theta_u & \theta_t + \theta_u + \theta_{v2} + \phi_2 \end{bmatrix} \qquad (D3)$$

This 4×4 matrix $\mathbf{\Sigma}$ can be written in terms of the 2×2 matrices **A**, **C**, and **E**. Matrix **A** can be defined in terms of a singular matrix $\mathbf{A}_s = \begin{bmatrix} a_s & a_s \\ a_s & a_s \end{bmatrix}$ and a diagonal matrix $\mathbf{A}_d = \begin{bmatrix} a_{d1} & 0 \\ 0 & a_{d2} \end{bmatrix}$, as $\mathbf{A} = \mathbf{A}_s + \mathbf{A}_d$; likewise for **C** and **E**. The construction in terms of **A**, **C**, and **E** will illustrate the differences between the covariance matrix for MZ twins, $\mathbf{\Sigma}_{MZ}$, and for DZ twins, $\mathbf{\Sigma}_{DZ}$:

$$\boldsymbol{\Sigma}_{MZ} = \begin{bmatrix} \mathbf{A}+\mathbf{C}+\mathbf{E} & \mathbf{A}+\mathbf{C} \\ \mathbf{A}+\mathbf{C} & \mathbf{A}+\mathbf{C}+\mathbf{E} \end{bmatrix}$$
$$\boldsymbol{\Sigma}_{DZ} = \begin{bmatrix} \mathbf{A}+\mathbf{C}+\mathbf{E} & 0.5\mathbf{A}+\mathbf{C} \\ 0.5\mathbf{A}+\mathbf{C} & \mathbf{A}+\mathbf{C}+\mathbf{E} \end{bmatrix} \quad (D4)$$

With all this in mind, for MZ twins:

$$\theta_t = a_s + c_s$$
$$\theta_u = e_s$$
$$\theta_{v1} = a_{d1} + c_{d1}$$
$$\theta_{v2} = a_{d2} + c_{d2}$$
$$\phi_1 = e_{d1}$$
$$\phi_2 = e_{d2} \quad (D5)$$

And for DZ twins

$$\theta_t = 0.5a_s + c_s$$
$$\theta_u = 0.5a_s + e_s$$
$$\theta_{v1} = 0.5a_{d1} + c_{d1}$$
$$\theta_{v2} = 0.5a_{d2} + c_{d2}$$
$$\phi_1 = 0.5a_{d1} + e_{d1}$$
$$\phi_2 = 0.5a_{d2} + e_{d2} \quad (D6)$$

The joint pmf of $\mathbf{X} = [T, U_1, U_2, V_1, V_2, W_1, W_2, W_3, W_4]^T$ is

$$p_{\mathbf{x}}(\mathbf{x}) = p_t(t) \cdot p_{u1}(u_1) \cdot p_{u2}(u_2) \cdot p_{v1}(v_1) \cdot p_{v2}(v_2) \cdot p_{w1}(w_1) \cdot p_{w2}(w_2) \cdot p_{w3}(w_3) \cdot p_{w4}(w_4) \quad (D7)$$

Recall that:

$$w_1 = y_{11} - t - u_1 - v_1$$
$$w_2 = y_{12} - t - u_1 - v_2$$
$$w_3 = y_{21} - t - u_2 - v_1$$
$$w_4 = y_{22} - t - u_2 - v_2 \quad (D8)$$

Upon substitution,

$$p_{\mathbf{x}}(\mathbf{x}) = p_t(t) \cdot p_{u1}(u_1) \cdot p_{u2}(u_2) \cdot p_{v1}(v_1) \cdot p_{v2}(v_2) \cdot$$
$$p_{w1}(y_{11} - t - u_1 - v_1) \cdot p_{w2}(y_{12} - t - u_1 - v_2) \cdot p_{w3}(y_{21} - t - u_2 - v_1) \cdot p_{w4}(y_{22} - t - u_2 - v_2) \quad (D9)$$

To simplify notation, let $\mathbf{y} = [y_{11}, y_{12}, y_{21}, y_{22}]^T$. Then,

$$p_{\mathbf{y}}(\mathbf{y}) = \sum_{t=0}^{\min(\mathbf{y})} \sum_{v_2=0}^{\min(y_{12},y_{22})-t} \sum_{v_1=0}^{\min(y_{11},y_{21})-t} \sum_{u_2=0}^{\min(y_{21}-v_1,y_{22}-v_2)-t} \sum_{u_1=0}^{\min(y_{11}-v_1,y_{12}-v_2)-t} p_{\mathbf{x}}(\mathbf{x})$$

$$= \sum_{t=0}^{\min(\mathbf{y})} p_t(t) \sum_{v_2=0}^{\min(y_{12},y_{22})-t} p_{v2}(v_2) \sum_{v_1=0}^{\min(y_{11},y_{21})-t} p_{v1}(v_1) \sum_{u_2=0}^{\min(y_{21}-v_1,y_{22}-v_2)-t} p_{u2}(u_2)\cdot p_{w3}(y_{21}-t-u_2-v_1)$$
$$\cdot p_{w4}(y_{22}-t-u_2-v_2) \sum_{u_1=0}^{\min(y_{11}-v_1,y_{12}-v_2)-t} p_{u1}(u_1)\cdot p_{w1}(y_{11}-t-u_1-v_1)\cdot p_{w2}(y_{12}-t-u_1-v_2)$$

or alternately, and if we do not make the ranges of summation explicit in the notation,

$$p_{\mathbf{y}}(\mathbf{y}) = \exp(-\theta_t-\theta_{v2}-\theta_{v1}-2\theta_u-2\phi_1-2\phi_2)\phi_1^{y_{11}+y_{21}}\phi_2^{y_{12}+y_{22}}$$
$$\cdot \sum_t \frac{\theta_t^t(\phi_1\phi_2)^{-2t}}{t!}\sum_{v_2}\frac{(\theta_{v2}\phi_2^{-2})^{v_2}}{v_2!}\sum_{v_1}\frac{(\theta_{v1}\phi_1^{-2})^{v_1}}{v_1!}\sum_{u_2} Q \sum_{u_1} R$$

where

$$Q = \frac{(\theta_u\phi_1\phi_2)^{u_2}}{u_2!\,(y_{21}-t-u_2-v_1)!\,(y_{22}-t-u_2-v_2)!}$$

and

$$R = \frac{(\theta_u\phi_1\phi_2)^{u_1}}{u_1!\,(y_{11}-t-u_1-v_1)!\,(y_{12}-t-u_1-v_2)!}$$

Now, in the triphenotype case, let $\mathbf{Y} = [Y_{11}, \mathrm{Y}_{12}, \mathrm{Y}_{13}, \mathrm{Y}_{21}, \mathrm{Y}_{22}, \mathrm{Y}_{23}]^T$ denote a multivariate-Poisson random vector. As before, the first of the two subscripts distinguishes twin #1 and twin #2 in a pair from one another, whereas the second subscript distinguishes one phenotype from the other. Define the elements of **Y** as follows:

$$\begin{aligned}
Y_{11} &= S + T_1 + U_{1,12} + 0 + U_{1,13} + V_{11} + W_{1,12} + 0 + W_{1,13}\\
Y_{12} &= S + T_1 + U_{1,12} + U_{1,23} + 0 + V_{12} + W_{2,12} + W_{2,23} + 0\\
Y_{13} &= S + T_1 + 0 + U_{1,23} + U_{1,13} + V_{13} + 0 + W_{3,23} + W_{3,13}\\
Y_{21} &= S + T_2 + U_{2,12} + 0 + U_{2,13} + V_{21} + W_{1,12} + 0 + W_{1,13}\\
Y_{22} &= S + T_2 + U_{2,12} + U_{2,23} + 0 + V_{22} + W_{2,12} + W_{2,23} + 0\\
Y_{23} &= S + T_2 + 0 + U_{2,23} + U_{2,13} + V_{23} + 0 + W_{3,23} + W_{3,13} \qquad (D10)
\end{aligned}$$

The latent variables are distributed as follows:

$S \sim \mathrm{Pois}(\theta_s)$

$T_1, T_2 \sim \mathrm{Pois}(\theta_t)$

$$
\begin{aligned}
&U_{1,12}, U_{2,12} \sim \text{Pois}(\theta_{u12}) \\
&U_{1,23}, U_{2,23} \sim \text{Pois}(\theta_{u23}) \\
&U_{1,13}, U_{2,13} \sim \text{Pois}(\theta_{u13}) \\
&V_{11}, V_{21} \sim \text{Pois}(\theta_{v1}) \\
&V_{12}, V_{22} \sim \text{Pois}(\theta_{v2}) \\
&V_{13}, V_{23} \sim \text{Pois}(\theta_{v2}) \\
&W_{1,12}, W_{2,12} \sim \text{Pois}(\theta_{w12}) \\
&W_{2,23}, W_{3,23} \sim \text{Pois}(\theta_{w23}) \\
&W_{1,13}, W_{3,13} \sim \text{Pois}(\theta_{w13})
\end{aligned}
\tag{D11}
$$

So, there are 21 latent variables in all, 6 of which (the *V*'s) would get redefined in terms of the observable variables, with the remaining 15 marginalized out of the expression to get the PMF of **Y** (which we do not write out here).

The covariance matrix of **Y** is 6 × 6, but it can be constructed from 3 × 3 matrices **A**, **C**, and **E**, which are similar to one another in structure. For instance,

$$
\mathbf{A} = \begin{bmatrix} a_s + a_{12} + a_{13} & a_s + a_{12} & a_s + a_{13} \\ a_s + a_{12} & a_s + a_{12} + a_{13} & a_s + a_{23} \\ a_s + a_{13} & a_s + a_{23} & a_s + a_{13} + a_{23} \end{bmatrix} \tag{D12}
$$

and likewise for **C**, though **E** requires the addition of a diagonal matrix $E_d = \begin{bmatrix} e_{d1} & 0 & 0 \\ 0 & e_{d2} & 0 \\ 0 & 0 & e_{d3} \end{bmatrix}$. For MZ twins, the parameter substitutions are:

$$
\begin{aligned}
&\theta_s = a_s + c_s \\
&\theta_t = e_s \\
&\theta_{u12} = e_{12} \\
&\theta_{u23} = e_{23} \\
&\theta_{u13} = e_{13} \\
&\theta_{v1} = e_{d1} \\
&\theta_{v2} = e_{2d} \\
&\theta_{v3} = e_{3d} \\
&\theta_{w12} = a_{12} + c_{12} \\
&\theta_{w23} = a_{23} + c_{23} \\
&\theta_{w13} = a_{13} + c_{13}
\end{aligned}
\tag{D13}
$$

For DZ twins:

$$
\begin{aligned}
&\theta_s = c_s + 0.5a_s \\
&\theta_t = 0.5a_s + e_s \\
&\theta_{u12} = e_{12} + 0.5a_{12} \\
&\theta_{u23} = e_{23} + 0.5a_{23} \\
&\theta_{u13} = e_{13} + 0.5a_{13} \\
&\theta_{v1} = e_{d1} \\
&\theta_{v2} = e_{d2} \\
&\theta_{v3} = e_{d3} \\
&\theta_{w12} = c_{12} + 0.5a_{12} \\
&\theta_{w23} = c_{23} + 0.5a_{23} \\
&\theta_{w13} = c_{13} + 0.5a_{13}
\end{aligned} \tag{D14}
$$

Merely evaluating the PMF (which we do not present explicitly) for a single observation can be computationally demanding; the limiting factor is the minimum of $\mathbf{y}$. For example, suppose all six elements were equal to 8. Then, marginalizing out the latent variables would entail computing a sum over 9^{15}—in excess of 100 trillion—terms.

Appendix E. On Dispersion Parameters and Twin Modeling

This Appendix uses the same notation as in the "Monophenotype Twin Modeling" section of the main text. Let us assume that latent variables X_0, X_1, and X_2 are Lagrangian Poisson. Suppose that for MZ twins:

$E(X_0) = 2.5$
$\text{var}(X_0) = 5$
$E(X_1) = E(X_2) = 2.5$
$\text{var}(X_1) = \text{var}(X_2) = 5$

Then, X_0, X_1, and X_2 are $\text{LGP}(1.77,\ 0.293)$. Therefore, Y_1 and Y_2 are marginally $\text{LGP}(3.53, 0.293)$, with expectation 5 and variance 10.

Now, suppose that for DZ twins,

$E(X_0) = 2.5$
$\text{var}(X_0) = 2.5$
$E(X_1) = E(X_2) = 2.5$
$\text{var}(X_1) = \text{var}(X_2) = 7.5$

Then, Y_1 and Y_2 have expectation 5 and variance 10 for DZ twins, just as they do for MZ twins. But,

$X_0 \sim \text{LGP}(2.5, 0)$
$X_1, X_2 \sim \text{LGP}(1.44, 0.423)$

and since X_1 and X_2 do not have the same value for parameter λ as X_0, Y_1 and Y_2 are not LGP. Even though the phenotypic distribution has the same mean and variance for both MZ and DZ twins, it is not the same for both zygosity groups. This can be clearly seen by computing, say, the third central moment for the two zygosity groups. Per Consul & Famoye (2006), the third central moment of a LGP distribution is

$$\mu_3 = \theta(1 + 2\lambda)(1 - \lambda)^{-5}$$

For MZ twins, this would be

$$\mu_3 = 3.53(1 + 2 \cdot 0.293)(1 - 0.293)^{-5} = 31.7$$

For DZ twins, because X_0, X_1, and X_2 are independent, the third central moment of (say) Y_1 would be the sum of the third central moments of X_0 and X_1:

$$\mu_3 = 2.5(1 + 2 \cdot 0)(1 - 0)^{-5} + 1.44(1 + 2 \cdot 0.423)(1 - 0.423)^{-5} = 44.1$$

To help ensure that the phenotypic distribution be the same for both zygosity groups, we require that parameter λ have the same value for latent variables X_0, X_1, and X_2, for both MZ and DZ twins. This has the additional advantage that the resulting marginals of Y_1 and Y_2 are themselves LGP. When (Y_1, Y_2) is instead modeled as bivariate negative binomial, we impose the analogous restriction on the latent variables' dispersion parameters $\wp$, and the marginals of Y_1 and Y_2 are then negative binomial.

At the urging of an anonymous referee, we have considered a possible case where the strict equality of dispersion parameters might be relaxed. This case is easier to present using PGFs instead of PMFs, so we will use the bivariate negative binomial distribution here, since its PGF is simpler than that of the LGP. Specifically, the common term X_0 would have dispersion parameter $\wp_0$, for both MZ and DZ twins, and the unique terms, X_1 and X_2, would have dispersion parameter $\wp_1$, again for both MZ and DZ twins.

Suppose that for MZ twins,

$$X_0 \sim \mathrm{NB}(\nu_{M0}, \mathcal{p}_0)$$
$$X_1, X_2 \sim \mathrm{NB}(\nu_{M1}, \mathcal{p}_1)$$

and for DZ twins,

$$X_0 \sim \mathrm{NB}(\nu_{D0}, \mathcal{p}_0)$$
$$X_1, X_2 \sim \mathrm{NB}(\nu_{D1}, \mathcal{p}_1)$$

Then, the marginal phenotypic distribution would not be negative binomial, but that of a convolution of independent negative-binomial RVs having different dispersion parameters. For MZ twins, the marginal PGF (i.e., for Y_1 and Y_2) would be

$$G_{Y_M}(z) = \left(\frac{\mathcal{p}_0}{1 - \mathcal{q}_0 z}\right)^{\nu_{M0}} \cdot \left(\frac{\mathcal{p}_1}{1 - \mathcal{q}_1 z}\right)^{\nu_{M1}}$$

and for DZ twins,

$$G_{Y_D}(z) = \left(\frac{\mathcal{p}_0}{1 - \mathcal{q}_0 z}\right)^{\nu_{D0}} \cdot \left(\frac{\mathcal{p}_1}{1 - \mathcal{q}_1 z}\right)^{\nu_{D1}}$$

If we set the two PGFs equal to one another, then with a bit of algebraic manipulation it can be shown that, under the given conditions, the phenotypic distribution will be equal for MZ and DZ twins, provided that the constraint

$$\left(\frac{\mathcal{p}_0}{1 - \mathcal{q}_0 z}\right)^{\nu_{M0} - \nu_{D0}} \cdot \left(\frac{\mathcal{p}_1}{1 - \mathcal{q}_1 z}\right)^{\nu_{M1} - \nu_{D1}} = 1$$

is satisfied.

Appendix F: The Bivariate log(*y*+1)-Normal Distribution

Suppose $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$, and that X_1 and X_2 are jointly bivariate normal, with correlation ρ. Define

$Y_1 = \exp(X_1) - 1$, and therefore $X_1 = \log(Y_1 + 1)$
$Y_2 = \exp(X_2) - 1$, and therefore, $X_2 = \log(Y_2 + 1)$

We will call the joint distribution of Y_1 and Y_2 bivariate log(*y*+1)-normal. The Jacobian of the substitution of Y_1 and Y_2 for X_1 and X_2 is

$$J(Y_1, Y_2) = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} = \begin{vmatrix} \frac{1}{y_1 + 1} & 0 \\ 0 & \frac{1}{y_2 + 1} \end{vmatrix}$$

$$= \frac{1}{(y_1 + 1)(y_2 + 1)} \qquad (F1)$$

The PDF of the bivariate log(*y*+1)-normal is then

$$f_{\mathbf{Y}}(y_1, y_2) = \frac{1}{2\pi(y_1 + 1)(y_2 + 1)\sigma_1\sigma_2\sqrt{1 - \rho^2}} \cdot \exp\left(\frac{-D}{2[1 - \rho^2]}\right) \qquad (F2)$$

for $(y_1, y_2) \in (-1, \infty)^2$, where

$$D = \left(\frac{\log(y_1 + 1) - \mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{\log(y_1 + 1) - \mu_1}{\sigma_1}\right)\left(\frac{\log(y_2 + 1) - \mu_2}{\sigma_2}\right) + \left(\frac{\log(y_2 + 1) - \mu_2}{\sigma_2}\right)^2$$

The marginal distribution of, say, Y_1 has p.d.f

$$f_{Y1}(y_1) = \frac{1}{(y_1 + 1)\sigma_1\sqrt{2\pi}} \cdot \exp\left(\frac{-[\log(y_1 + 1) - \mu_1]^2}{2\sigma_1^2}\right) \qquad (F3)$$

for $y_1 \in (-1, \infty)$.

The parameters μ_1 and μ_2 are the log-scale means, σ_1^2 and σ_2^2 are the log-scale variances, and the log-scale covariance equals $\rho\sigma_1\sigma_2$.

The joint distribution of $\exp(X_1)$ and $\exp(X_2)$ is an ordinary bivariate lognormal distribution. Therefore, the joint distribution of Y_1 and Y_2 is an ordinary bivariate lognormal distribution shifted downwardly by 1. As the reader is likely aware, shifting a distribution will correspondingly shift its means, but will not change its variances and covariances. From the moments of the ordinary bivariate lognormal distribution (Forbes et al., 2011; Balakrishnan & Lai, 2009), we have the following:

$$E(Y_1) = \exp(\mu_1 + 0.5\sigma_1^2) - 1$$
$$E(Y_2) = \exp(\mu_2 + 0.5\sigma_2^2) - 1$$
$$\mathrm{var}(Y_1) = \exp(2\mu_1 + \sigma_1^2)\,[\exp(\sigma_1^2) - 1]$$
$$\mathrm{var}(Y_2) = \exp(2\mu_2 + \sigma_2^2)\,[\exp(\sigma_2^2) - 1]$$
$$\mathrm{cov}(Y_1, Y_2) = \exp(0.5\sigma_1^2 + 0.5\sigma_2^2 + \mu_1 + \mu_2)\,[\exp(\rho\sigma_1\sigma_2) - 1] \qquad (F4)$$

We have confirmed the identities in $(F4)$ by (rather laboriously) deriving $E(Y_1^2)$, $E(Y_2^2)$, and $E(Y_1 Y_2)$ from the univariate and bivariate PDFs, though in the interest of brevity we do not reproduce the derivation here.