



Received 16 October 1978
Final 10 January 1979

Twin Azygotic Test for the Study of Hereditary Qualitative Traits in Twin Populations

L. Gedda¹, C. Rossi², G. Brenci¹

¹*The Mendel Institute, Rome;* ²*Institute of Probability Calculus, University of Rome*

Following previous formulations of a model of qualitative analysis of twin population data independent of zygosity, a new Bayesian approach has been developed. The present model can be applied to any qualitative genetic trait in twin population data, provided no specific source of variation be introduced by the twin condition, and allows not only estimation of the frequencies of mono- and dizygosity as well as the gene frequencies, but also verification of the trait's mode of inheritance.

Key words: Azygotic test, Bayesian model, Demographic data, Mode of inheritance, Gene frequency, Qualitative analysis, Twins, Zygosity determination

One possible approach in twin research consists in the qualitative analysis of twin data drawn from vital statistics or public health registers, for which zygosity is usually not available. However, zygosity determination is required by most methods of concordance analysis. Following critical revisions of the classic model introduced, among others, by Allen in 1955 [1] and Stern in 1958 [10], a more general model of concordance analysis has just been developed by Allen and Hrubec [2].

The assessment of zygosity is relatively expensive for any twin sample and is practically impossible to perform at the population level. For these reasons, as early as 1962, we started to develop a model of analysis of twin data independent of zygosity — ie, the Twin Azygotic Test.

In the first formulation, 17 years ago [3], the test assumed, for the trait under study, complete penetrance and independent manifestation in dizygotic cotwins. Based on the frequencies of pair concordance or discordance, the frequencies of monozygosity and dizygosity could be derived, as well as the frequency of the trait under study.

In the second formulation, presented at the Third International Congress of Human Genetics in Chicago in 1966 [4], the assumption of complete penetrance was eliminated and the model allowed to calculate, along with the frequencies of mono- and dizygosity, the frequency of the trait and its penetrance. A somewhat similar model was then proposed by Selvin in 1971 [9].

In the third formulation, presented at the Second International Congress on Twin Studies in Washington in 1977 [5], the assumption of independence of manifestation of the trait in dizygotic cotwins was also overcome. The Bayesian approach developed by one of us (C.R.) allowed estimation of the frequencies of mono- and dizygosity as well as the frequencies of the gene(s) responsible for the given trait. The limit of this model consisted in the nonexplicit form of its solutions.

We are now presenting a new development of our model that can be applied to twin population data, provided no specific source of variation be introduced by the twin condition, and to any qualitative genetic trait, allowing 1) estimation of the frequencies of mono- and dizygosity, 2) estimation of the frequencies of the gene(s) responsible for the given trait, and 3) verification of the trait's mode of inheritance. As an example, the model here considers the simple case of an autosomal diallelic trait in the absence of dominance.

The present note is aimed at simply describing the mathematical model. The latter is, however, already being applied to an analysis of congenital diseases based on data drawn from Italian vital statistics. The results of this, and possibly other practical applications of our model, will be the subject of further notes.

MATHEMATICAL MODEL

Consider the Poisson process of twin births. Calling λ the velocity of the process, we obtain the following probability distribution for the number of twin births in a time period t :

$$P_h(t) = e^{-\lambda t} \frac{(\lambda t)^h}{h!}, \quad h = 0, 1, 2, \dots \tag{1}$$

Consider now an autosomal diallelic trait determined, in the absence of dominance, by two alleles, a_1 and a_2 , with frequencies p and q . Calling x , y , and z the phenotypes corresponding to the three genotypes, a_1a_1 , a_1a_2 , and a_2a_2 , respectively, and assuming that the twin condition introduces no further factor of variability, the following probability distribution of phenotypes in a twin population is obtained:

$$\begin{aligned} f_{xx} &= mp^2 + 1/4 dp^2(1 + p)^2 & f_{xy} &= dp^2q(1 + p) \\ f_{yy} &= 2mpq + dpq(1 + pq) & f_{xz} &= 1/2 dp^2q^2 \\ f_{zz} &= mq^2 + 1/4 dq^2(1 + q)^2 & f_{yz} &= dpq^2(1 + q) \end{aligned} \tag{2}$$

where m is the conditional probability that a twin pair is monozygotic and $d = 1 - m$ the probability that it is dizygotic.

We can now consider the following six conditionally independent components of the Poisson process of twin births given by the Poisson processes with velocities:

$$\begin{aligned} \lambda_{xx} &= \lambda f_{xx} & \lambda_{xy} &= \lambda f_{xy} \\ \lambda_{yy} &= \lambda f_{yy} & \lambda_{xz} &= \lambda f_{xz} \\ \lambda_{zz} &= \lambda f_{zz} & \lambda_{yz} &= \lambda f_{yz} \end{aligned} \tag{3}$$

So, we must solve an inferential problem related to the estimation of parameters p , m , λ , given a realization of the six Poisson processes.

We can denote by \mathbf{a} the random vector $\mathbf{a} \rightleftharpoons (a_{xx}, a_{yy}, a_{zz}, a_{xy}, a_{xz}, a_{yz})$, the components of which are the observed absolute frequencies of twin pairs xx , yy , zz , xy ,

xz, and yz, respectively. If we introduce the a priori probability density of parameters p, m, λ, which can be considered independent random variables:

$$\phi_{p,m,\lambda}(p, m, \lambda) = \phi_1(p)\phi_2(m)\phi_3(\lambda) \tag{4}$$

we can calculate the a posteriori probability density of the parameters, given a realization of vector **a**, during a given time unit:

$$\begin{aligned} \phi_{p,m,\lambda}(p, m, \lambda | \mathbf{a} = \mathbf{h}) &= \phi_{p,m,\lambda}(p, m, \lambda | a_{xx} = h_1, a_{yy} = h_2, a_{zz} = h_3, \\ &\quad a_{xy} = h_4, a_{xz} = h_5, a_{yz} = h_6) \\ &= k_1 P(a_{xx} = h_1, a_{yy} = h_2, a_{zz} = h_3, a_{xy} = h_4, \\ &\quad a_{xz} = h_5, a_{yz} = h_6 | p, m, \lambda) \phi_1(p)\phi_2(m)\phi_3(\lambda) \end{aligned} \tag{5}$$

where

$$k_1 = \left[\int_0^{+\infty} \int_0^1 \int_0^1 P(a_{xx} = h_1, a_{yy} = h_2, a_{zz} = h_3, a_{xy} = h_4, a_{xz} = h_5, a_{yz} = h_6 | p, m, \lambda) \phi_1(p)\phi_2(m)\phi_3(\lambda) dp dm d\lambda \right]^{-1}$$

Let us now explicate Eq. 5 in the following particular case:

$$\begin{aligned} \phi_1(p) &= 1 \quad (0 \leq p \leq 1) \\ \phi_2(m) &= 1 \quad (0 \leq m \leq 1) \\ \phi_3(\lambda) &= k_2 \lambda^\alpha e^{-\beta\lambda} \quad (0 \leq \lambda < +\infty) \end{aligned} \tag{6}$$

where $k_2 = \left(\int_0^{+\infty} \lambda^\alpha e^{-\beta\lambda} d\lambda \right)^{-1}$

Replacing Eq. 6 in Eq. 5, for the conditional independence of the components of vector **a**, we obtain:

$$\phi_{p,m,\lambda}(p, m, \lambda | \mathbf{a} = \mathbf{h}) = k_1 k_2 \frac{e^{-(1+\beta)\lambda} \lambda^{\alpha + \sum_{i=1}^6 h_i}}{h_1! h_2! h_3! h_4! h_5! h_6!} (f_{xx}^{h_1} f_{yy}^{h_2} f_{zz}^{h_3} f_{xy}^{h_4} f_{xz}^{h_5} f_{yz}^{h_6}) \tag{7}$$

which can be written as follows:

$$\phi_{p,m,\lambda}(p, m, \lambda | \mathbf{a} = \mathbf{h}) = k_3 g_\lambda(\lambda | \mathbf{a} = \mathbf{h}) f_{p,m}(p, m | \mathbf{a} = \mathbf{h}) \tag{7'}$$

It follows that the random vector (p, m) is also a posteriori independent of λ, while parameters p and m are no longer independent conditionally to **a**.

Generally, we are interested in estimating p and m, while λ can be considered a disturbance parameter. So we can calculate the marginal joint density of vector (p, m):

$$\phi_{p,m}(p, m | \mathbf{a} = \mathbf{h}) = k_3 f_{p,m}(p, m | \mathbf{a} = \mathbf{h}) \int_0^{+\infty} g_\lambda(\lambda | \mathbf{a} = \mathbf{h}) d\lambda = k_4 f_{p,m}(p, m | \mathbf{a} = \mathbf{h}) \tag{8}$$

where $k_4 = \left[\int_0^1 \int_0^1 f_{p,m}(p, m | \mathbf{a} = \mathbf{h}) dp dm \right]^{-1}$ and

$$\begin{aligned} f_{p,m}(p, m | \mathbf{a} = \mathbf{h}) &= [mp^2 + \frac{1}{4} dp^2 (1 + p^2)]^{h_1} \cdot [2mpq + dpq(1 + pq)]^{h_2} \\ &\quad \cdot [mq^2 + \frac{1}{4} dq^2 (1 + q^2)]^{h_3} \cdot [dp^2 q(1 + p)]^{h_4} \cdot [\frac{1}{2} dp^2 q^2]^{h_5} \cdot [dpq^2(1 + q)]^{h_6} \end{aligned} \tag{9}$$

which is a polynomial $P(p, m)$ of degree $d = 5 \sum_{i=1}^6 h_i$. The calculation of the integral, which gives k_4 , is now trivial.

We have thus obtained every possible information on vector (p, m) conditionally to \mathbf{a} and, if we want, we can obtain point estimations for parameters p and m . Let us consider the marginal densities of p and m :

$$\phi_1(p | \mathbf{a} = \mathbf{h}) = k_4 \int_0^1 f_{p,m}(p, m | \mathbf{a} = \mathbf{h}) dm$$

$$\phi_2(m | \mathbf{a} = \mathbf{h}) = k_4 \int_0^1 f_{p,m}(p, m | \mathbf{a} = \mathbf{h}) dp$$

Following Lindley [6, 7], we obtain the following point estimations:

$$\hat{p} = \int_0^1 p \phi_1(p | \mathbf{a} = \mathbf{h}) dp$$

$$\hat{m} = \int_0^1 m \phi_2(m | \mathbf{a} = \mathbf{h}) dm$$

Alternatively, following O’Hagan [8], we can consider the point estimations given by the joint or marginal modes of the posterior joint or marginal densities. This way of calculating point estimations for p and m is equivalent to the classic maximum likelihood estimation, since we have uniform a priori densities for p and m , and λ is independent of p and m , conditionally to $\mathbf{a} = \mathbf{h}$. On the contrary, Bayesian point estimations are quite different from the classic ones. Consider, for instance, the following a priori distributions:

$$\phi'_1(p) = A_1 p^A (1 - p)^B \quad (0 \leq p \leq 1)$$

$$\phi'_2(m) = A_2 m^C (1 - m)^D \quad (0 \leq m \leq 1) \tag{10}$$

$$\phi'_3(\lambda) = \phi_3(\lambda) = k_3 \lambda^{\alpha} e^{-\beta \lambda} \quad (0 \leq \lambda < +\infty)$$

which is the case of beta a priori densities.

In such a case we obtain the following a posteriori marginal joint density for vector (p, m) :

$$\phi'_{p,m}(p, m | \mathbf{a} = \mathbf{h}) = k_4 A_1 A_2 f_{p,m}(p, m | \mathbf{a} = \mathbf{h}) p^A (1 - p)^B m^C (1 - m)^D \tag{11}$$

and, if A, B, C, D are integers, then $\phi'_{p,m}(p, m | \mathbf{a} = \mathbf{h})$ is a polynomial $P(p, m)$ of degree $g = 5 \sum_{i=1}^6 h_i + A + B + C + D$, every integral can be easily calculated, and we can use Lindley’s or O’Hagan’s point estimations. If A, B, C, D are not all integers, we can use the mode of the joint density of vector (p, m) as point estimation, which can be calculated by differentiating $\phi'_{p,m}(p, m | \mathbf{a} = \mathbf{h})$ with respect to p and m , and then use numerical standard methods to obtain the roots, or we can use numerical standard methods to approximate the integrals which give Lindley’s estimations.

REFERENCES

1. Allen G (1955): Comments on the analysis of twin samples. *Acta Genet Med Gemellol* 4:143–160.
2. Allen G, Hrubec Z (1979): Twin concordance: A more general model. *Acta Genet Med Gemellol* 28:3–13.

3. Gedda L, Brenci G (1962): Proposta del test gemellare azigotico. *Acta Genet Med Gemellol* 11:1–8.
4. Gedda L, Brenci G (1966): Theoretical models in twin research. *Acta Genet Med Gemellol* 15: 219–223.
5. Gedda L, Brenci G, Rossi C (1978): Twin models in population genetics. In Nance WE, Allen G, Parisi P (eds): “Twin Research, Part B: Biology and Epidemiology.” Proceedings of the Second International Congress on Twin Studies, Washington 1977. New York: Alan R Liss, pp 149–152.
6. Lindley DV (1971): The estimation of many parameters. In Godambe VP, Sprott DA (eds): “Foundations of Statistical Inference.” New York: Holt, Rinehart & Winston.
7. Lindley DV (1971). Bayesian statistics: A review. SIAM, Regional Conference Series in Applied Mathematics.
8. O’Hagan A (1976). On posterior joint and marginal modes. *Biometrika* 63:329–333.
9. Selvin S (1970). Concordance in a twin population model. *Acta Genet Med Gemellol* 19:584–590.
10. Stern K (1958). The ratio of monozygotic to dizygotic affected twins and the frequencies of affected twins in unselected data. *Acta Genet Med Gemellol* 7:313–320.

Correspondence: Professor Luigi Gedda, The Mendel Institute, Piazza Galeno 5, 00161 Rome, Italy.