Acta Genet Med Gemellol 30:217-222 (1981)
The Mendel Institute/Alan R. Liss, Inc.

Short Note

Received 12 January 1981
Final 25 May 1981

Zygosity Test Based on Growth Measures: A Bayesian Approach

C. Rossi,[1] G. Brenci,[2] and R. Ziparo[3]

[1]Department of Mathematics, II University of Rome; [2]The Gregor Mendel Institute of Medical Genetics and Twin Research, Rome; [3]Institute of Physiology, University of Rome

A statistical index is proposed in order to determine twin zygosity based on weight and height measures and a Bayesian test of hypotheses. Some Roman twin data are analyzed on the basis of the Euclidean distance index, which appears to be the best for an easy calculation and for its genetic interpretation based on the formulation of genetic distance.

Key words: Twin zygosity, Twin concordance, Bayesian approach, Genetic distance, Height, Weight

INTRODUCTION

Longitudinal studies of growth measures in twins indicate different patterns in monozygotic (MZ) versus dizygotic (DZ) twins (eg, see [3]). We should therefore like to propose a mathematical model of twin concordance allowing for a probabilistic zygosity determination based on twin longitudinal data. What we shall propose is a suitable distance function between cotwins through which, based on growth measures such as height and weight, MZ and DZ pairs can be discriminated with a high level of probability. We shall then consider growth measures of pairs of unknown zygosity in order to obtain a posterior probability of monozygosity versus dizygosity as a function of the distance index.

MATHEMATICAL MODEL

Many statistical indices of diversity are available as suitable distance functions between individuals of a twin pair based on growth measures, and many papers have been written about the mathematical properties of such functions (eg, [1], and for a review [2]). In the present paper we take into account only the so-called Eu-

0001-5660/81/3003-0217$02.00

clidean normalized distance, which allows good results for our problem in a very easy way. Let us define this function.

Consider two vectors $x <=> (x_1, x_2 \ldots, x_n)$ and $y <=> (y_1, y_2 \ldots, y_n)$ define:

$$d(x,y) = \left\{ \sum_{i=1}^{n} \left(\frac{x_i - y_i}{\frac{(x_i + y_i)}{2}} \right)^2 \right\}^{1/2}$$

The function $d(x,y)$ is the normalized Euclidean distance of the vectors x and y.

Let us now consider the following mathematical model of concordance for biometric traits of MZ and DZ pairs.

Ronald Fisher (1918) proposed that the genotypic expression of "continuous" hereditary traits is a normal random variable X_g with mean m_g and variance σ_g^2, depending on sex and age. A random independent overlapping fluctuation X_e, due to environmental conditions, must be considered and results in a normal random variable with mean m_e and variance σ_e^2. If we define X(n) the random variable that represents height measure at age n, and Y(n) the random variable that represents weight measure at age n, for any individual of a well defined population, we can write:

$$X(n) = X_g(n) + X_e(n)$$
$$Y(n) = Y_g(n) + Y_e(n)$$

where

$$X(n) \sim N\ (E(X_g(n)) + E(X_e(n))),\ \sigma^2(X_g(n)) + \sigma^2(X_e(n)))$$
$$Y(n) \sim N\ (E(Y_g(n)) + E(Y_e(n))),\ \sigma^2(Y_g(n)) + \sigma^2(Y_e(n)))$$

Let us now consider a twin pair and define $X^{(i)}(n)$ and $Y^{(i)}(n)$ $(i = 1,2)$ the height and weight measures of the i-th individual of the pair, at age n. We can write

		for MZ pairs	for DZ pairs
$X^{(1)}(n)$	=	$X_g^{(1)}(n) + X_e^{(1)}(n)$	$X_g^{(1)}(n) + X_e^{(1)}(n)$
$Y^{(1)}(n)$	=	$Y_g^{(1)}(n) + Y_e^{(1)}(n)$	$Y_g^{(1)}(n) + Y_e^{(1)}(n)$
$X^{(2)}(n)$	=	$X_g^{(1)}(n) + X_e^{(2)}(n)$	$X_g^{(2)}(n) + X_e^{(2)}(n)$
$Y^{(2)}(n)$	=	$Y_g^{(1)}(n) + Y_e^{(2)}(n)$	$Y_g^{(2)}(n) + Y_e^{(2)}(n)$

It follows immediately that, for all n:

$$E(d_n(DZ)) > E(d_n(MZ))$$

where

$$d_n(DZ) = \left\{ \left(\frac{X^{(1)}(n) - X^{(2)}(n)}{\frac{X^{(1)}(n) + X^{(2)}(n)}{2}} \right)^2 + \left(\frac{Y^{(1)}(n) - Y^{(2)}(n)}{\frac{Y^{(1)}(n) + Y^{(2)}(n)}{2}} \right)^2 \right\}^{1/2}$$

https://doi.org/10.1017/S0001566000007686 Published online by Cambridge University Press

with the same expression for $d_n(MZ)$.

Statistical Distribution of $d_n(DZ)$ and $d_n(MZ)$

Based on growth measures of 28 MZ and 28 DZ twin pairs obtained at n = 0, 3, 6, 9, and 12 months of age, $d_n(MZ)$ and $d_n(DZ)$ values have been derived (Table 1). Statistical mean and variance have then been calculated so that:

$$D_n(DZ) = \frac{d_n(DZ) - E(d_n(DZ))}{\sigma(d_n(DZ))}$$

and

$$D_n(MZ) = \frac{d_n(MZ) - E(d_n(MZ))}{\sigma(d_n(MZ))}$$

have a probability density function t_{27} and then, for instance (using the table of Student's t),

$$P(d_n(DZ) \in (E(d_n(DZ)) \pm 2.0526(d_n(DZ))) = 0.99 =$$

$$= P(d_n(MZ) \in (E(d_n(MZ)) \pm 2.0526(d_n(MZ))).$$

We could now apply standard hypothesis testing theory to any twin pair and get a probabilistic determination of zygosity based on the index d_n and an approximation of the likelihood functions given by the normal densities with means given by $\bar{d}_n(DZ)$ and $\bar{d}_n(MZ)$.

A Test From a Bayesian Point of View

Let us now consider a Bayesian statistical test for zygosity determination.

If we call P(MZ) the prior probability that any twin pair be MZ and P(DZ) = 1 − P(MZ) the prior probability that any twin pair be DZ[1], we can write:

$$P(MZ/d_n) = \frac{f(d_n/MZ)P(MZ)}{f(d_n/MZ)P(MZ) + g(d_n/DZ)P(DZ)}$$

and

$$P(DZ/d_n) = \frac{g(d_n/DZ)P(DZ)}{f(d_n/MZ)P(MZ) + g(d_n/DZ)P(DZ)},$$

where $f(d_n/MZ)$ is the probability density function of the random variable $d_n(MZ)$ and $g(d_n/DZ)$ the probability density function of $d_n(DZ)$, which have been approximated by normal densities with parameters estimated on the basis of our statistical data.

[1]That can be statistically estimated. Based on the twin data available at the Mendel Institute, we obtained the following estimations: P(MZ) = 0.29, P(DZ) = 0.71.

TABLE 1. Statistical Distribution of $d_n(DZ)$ and $d_n(MZ)$ at $n = 0, 3, 6, 9$, and 12 Months of Age

$d_0(MZ)$	$d_0(DZ)$	$d_3(MZ)$	$d_3(DZ)$	$d_6(MZ)$	$d_6(DZ)$	$d_9(MZ)$	$d_9(DZ)$	$d_{12}(MZ)$	$d_{12}(DZ)$
.280	.223	.186	.216	.167	.211	.194	.233	.120	.166
.077	.083	.129	.112	.054	.093	.106	.293	.050	.059
.044	.074	.035	.033	.046	.051	.013	.067	.041	.041
.049	.223	.014	.215	.112	.245	.046	.077	.013	.267
.255	.068	.146	.090	.050	.019	.104	.071	.017	.048
.250	.044	.103	.075	.024	.042	.034	.109	.023	.190
.043	.148	.034	.092	.017	.120	.032	.164	.022	.083
.074	.199	.000	.130	.030	.111	.274	.102	.013	.209
.087	.139	.018	.175	.030	.123	.083	.061	.024	.134
.148	.224	.049	.172	.091	.084	.074	.143	.083	.031
.174	.314	.139	.116	.396	.203	.021	.117	.087	.159
.174	.120	.420	.079	.078	.052	.012	.018	.044	.027
.109	.220	.075	.119	.008	.072	.062	.055	.102	.136
.068	.513	.041	.053	.142	.045	.022	.089	.024	.112
.024	.239	.018	.107	.015	.094	.013	.108	.034	.138
.043	.183	.091	.041	.046	.048	.105	.124	.146	.075
.122	.076	.084	.085	.109	.045	.179	.045	.171	.000
.188	.304	.077	.251	.029	.108	.042	.113	.041	.108
.056	.037	.181	.063	.051	.239	.054	.179	.040	.161
.326	.350	.054	.102	.158	.025	.021	.120	.022	.044
.164	.115	.138	.087	.033	.120	.012	.116	.072	.140
.293	.176	.044	.055	.038	.119	.044	.082	.019	.217
.051	.143	.038	.000	.050	.019	.035	.031	.052	.101
.067	.039	.039	.116	.016	.071	.014	.087	.025	.075
.066	.000	.010	.136	.017	.076	.038	.046	.029	.065
.137	.128	.062	.000	.029	.063	.036	.100	.052	.132
.073	.196	.078	.155	.084	.193	.027	.041	.048	.037
.313	.000	.020	.070	.922	.068	.019	.024	.030	.056
$E(d_0(MZ)) = 0.134$		$E(d_3(MZ)) = 0.083$		$E(d_6(MZ)) = 0.069$		$E(d_9(MZ)) = 0.061$		$E(d_{12}(MZ)) = 0.052$	
$\sigma^2(d_0(MZ)) = 0.008$		$\sigma^2(d_3(MZ)) = 0.007$		$\sigma^2(d_6(MZ)) = 0.006$		$\sigma^2(d_9(MZ)) = 0.004$		$\sigma^2(d_{12}(MZ)) = 0.002$	
$E(d_0(DZ)) = 0.164$		$E(d_3(DZ)) = 0.105$		$E(d_6(DZ)) = 0.099$		$E(d_9(DZ)) = 0.101$		$E(d_{12}(DZ)) = 0.108$	
$\sigma^2(d_0(DZ)) = 0.013$		$\sigma^2(d_3(DZ)) = 0.004$		$\sigma^2(d_6(DZ)) = 0.004$		$\sigma^2(d_9(DZ)) = 0.004$		$\sigma^2(d_{12}(DZ)) = 0.004$	

TABLE 2. Posterior Probabilities of Zygosity Based on Growth Measures at 0, 3, 6, 9, and 12 Months of Age

Function	X	P(MZ/X)	P(DZ/X)
d_0	.05880	.36062	.63938
	.07060	.36341	.63659
	.09710	.36405	.63595
	.10210	.36329	.63671
	.10550	.36262	.63738
	.13330	.35237	.64763
	.14770	.34380	.65620
	.17060	.32587	.67413
	.22590	.26399	.73601
	.28830	.17673	.82327
d_3	.05880	.27884	.72116
	.07060	.26149	.73851
	.10210	.23144	.76856
	.10550	.22947	.77053
	.13330	.22170	.77830
	.14770	.22334	.77666
	.17060	.23408	.76592
	.22590	.30860	.69140
	.28830	.50352	.49648
d_6	.05880	.28807	.71193
	.07060	.26943	.73057
	.10210	.23358	.76642
	.10550	.23079	.76921
	.13330	.21491	.78509
	.14770	.21119	.78881
	.17060	.21123	.78877
	.22590	.24294	.75706
	.28830	.34830	.65170
d_9	.05880	.33774	.66226
	.07060	.31187	.68813
	.10210	.24855	.75145
	.10550	.24225	.75775
	.13330	.19492	.80508
	.14770	.17330	.82670
	.17060	.14290	.85710
	.22590	.08751	.91249
	.28830	.04887	.95113
d_{12}	.05880	.43591	.56409
	.07060	.38688	.61312
	.10210	.23650	.76350
	.10550	.22036	.77964
	.13330	.10705	.89295
	.14770	.06652	.93348
	.17060	.02724	.97276
	.22590	.00171	.99829
	.28830	.00003	.99997

The statistical test is based on such posterior distributions rather than only on likelihood functions. Posterior probabilities for twin pairs of unknown zygosity, based on functions d_0, d_3, d_6, d_9, d_{12}, are given in Table 2.

Much more could be said on the basis of the following comprehensive index:

$$d_{0N} = \left\{ \sum_{i=0}^{N} d_n^2 \right\}^{1/2}$$

which takes into account more information from growth measures.

REFERENCES

1. De Finetti B (1974): "Theory of Probability." New York: John Wiley.
2. Leti G (1979): "Distanze e Indici Statistici." Rome: La Goliardica.
3. Wilson RS (1979): Analysis of longitudinal twin data. Basic model and applications to physical growth measures. Acta Genet Med Gemellol 28:93–105.

Correspondence: Professa Carla Rossi, The Mendel Institute, Piazza Galeno 5, 00161 Rome, Italy.