A Latent Class Model for Bivariate Binary Responses from Twins

Anders Ekholm, Jukka Jokinen, John W. McDonald, Peter W. F. Smith

Abstract

We analyse a 2^4 table reporting the presence or absence of alcohol dependence and depression in both members of 597 pairs of female monozygotic twins. The statistical analysis is based on a latent class model, formulated as a 2x2 table and parametrized by two marginal univariate genetic dispositions and the dependence ratio between these dispositions. The final model, selected after some empirically motivated simplifying assumptions are adopted, has six parameters and fits the data very well. The identifiability and stability of the final model is studied by contour plots of profile log-likelihood functions with two arguments. Numerical results are compared to those obtained when the measure of association between binary responses is the odds ratio rather than the dependence ratio. The conditional probability of alcohol dependence given the disposition for it is approximately 1/3, while the corresponding probability of depression is approximately 2/3. The association between these two latent dispositions is strong, the dependence ratio being approximately 3. An extended model, defining measures of heritability and similarity, for analysing, simultaneously, data on monzygotic and dizygotic twins is proposed.

Southampton Statistical Sciences Research Institute
Methodology Working Paper M06/10

University of Southampton

A Latent Class Model for Bivariate Binary Responses from Twins

Anders Ekholm,[1,*] Jukka Jokinen,[1,] John W. McDonald,[2,***] and Peter W. F. Smith[2,****]**

[1] Department of Mathematics and Statistics,
P.O. Box 68, FI-00014 University of Helsinki, Finland.
[2] Southampton Statistical Sciences Research Institute,
University of Southampton, Southampton, SO17 1BJ, U.K.

[*] *email:* anders.ekholm@helsinki.fi
[**] *email:* jukka.jokinen@helsinki.fi
[***] *email:* j.w.mcdonald@soton.ac.uk
[****] *email:* pws@soton.ac.uk

SUMMARY. We analyse a 2^4 table reporting the presence or absence of alcohol dependence and depression in both members of 597 pairs of female monozygotic twins. The statistical analysis is based on a latent class model, formulated as a 2×2 table and parametrized by two marginal univariate genetic dispositions and the dependence ratio between these dispositions. The final model, selected after some empirically motivated simplifying assumptions are adopted, has six parameters and fits the data very well. The identifiability and stability of the final model is studied by contour plots of profile log-likelihood functions with two arguments. Numerical results are compared to those obtained when the measure of association between binary responses is the odds ratio rather than the dependence ratio. The conditional probability of alcohol dependence given the disposition for it is approximately 1/3, while the corresponding probability of depression is approximately 2/3. The association between these two latent dispositions is strong, the dependence ratio being approximately 3. An extended model, defining measures of heritability and similarity, for analysing, simultaneously, data on monzygotic and dizygotic twins is proposed.

KEY WORDS: Contour plot; Dependence and Odds ratio; Dizygotic and Monozygotic twin; Genetic disposition; Cohen's kappa; Moment parameterization.

1. Presentation and Inspection of the Monozygotic Twin Data

Drton and Richardson (2005) present the data reproduced below as Table 1, referring to Kendler et al. (1992) where, however, these data are not published. Table 1 is a cross-classification of alcohol dependence and depression for 597 female monozygotic twin pairs in $2^4 = 16$ different categories. For twin member i, where $i = 1, 2$, denote by $A_i = 1/0$ the presence (1) or absence

1

(0) of alcohol dependence and denote by $D_i = 1/0$ the presence/absence of depression. We use consistently the following ordering of these four binary responses: (A_1, A_2, D_1, D_2), so that, for example, $\#(0, 0, 1, 1) = 51$, indicates that there are 51 twin pairs in which neither member has alcohol dependence, but both have depression.

<div align="center">

Table 1

Observed counts of 597 female monozygotic twin pairs classified by concordance/discordance with respect to alcohol dependence and depression.

</div>

		$D_1 = 1$		$D_1 = 0$	
		$D_2 = 1$	$D_2 = 0$	$D_2 = 1$	$D_2 = 0$
$A_1 = 1$	$A_2 = 1$	7	4	2	3
$A_1 = 1$	$A_2 = 0$	9	8	4	8
$A_1 = 0$	$A_2 = 1$	10	7	9	15
$A_1 = 0$	$A_2 = 0$	51	92	80	288

The univariate relative frequencies, denoted $\tilde{\mu}$, are from Table 1,

$$\tilde{\mu}_{A_1} = 0.075, \quad \tilde{\mu}_{A_2} = 0.095, \quad \tilde{\mu}_{D_1} = 0.315, \quad \tilde{\mu}_{D_2} = 0.288,$$

indicating that depression is much more common than alcohol dependence. The relative frequencies for twin 1 and twin 2 are closely similar, both for alcohol dependence and for depression. Since there is no indication in Kendler et al. (1992) that the numbering of the twins in a pair carries any information, we will in all modelling assume marginal homogeneity within pairs, i.e., $E(A_1) = E(A_2) = \mu_A$ and $E(D_1) = E(D_2) = \mu_D$.

Regarded as a four-variate binary response, the vector (A_1, A_2, D_1, D_2) has the unusual and challenging feature, that the bivariate marginal distributions (A_1, D_1) and (A_2, D_2) refer to the same individual, but for different variables, i.e., co-morbidity, while (A_1, A_2) and (D_1, D_2) refer to the same variable, but for different individuals, i.e., concordance.

For inspection of the six different marginal bivariate distributions of Table 1, we use the dependence ratio of second order, defined, for A_1 and A_2 as $\tau_{A_1 A_2} = E(A_1 A_2)/\{E(A_1)E(A_2)\}$, that is, the probability that both members suffer from alcohol dependence divided by the same probability assuming that A_1 and A_2 are independent. Ekholm et al. (1995) introduced the dependence ratio; the most important features and comparisons with the more familiar odds ratio are in Ekholm et al. (2003), and Ekholm (2003). The dependence ratios for the six bivariate distributions, computed from the six

marginal 2×2 tables of Table 1, are denoted by $\tilde{\tau}$. Their numerical values are listed below, with a note on what is constant and what is varying for the bivariate distribution in question. For comparison the numerical values of the odds ratio, denoted $\tilde{\chi}$, are listed in parallel,

$$
\begin{aligned}
\tilde{\tau}_{A_1 A_2} &= 3.72, \ \tilde{\chi}_{A_1 A_2} = 6.88, & \text{two twins, one problem: alcohol,} \\
\tilde{\tau}_{D_1 D_2} &= 1.42, \ \tilde{\chi}_{D_1 D_2} = 2.29, & \text{two twins, one problem: depression,} \\
\tilde{\tau}_{A_1 D_2} &= 1.70, \ \tilde{\chi}_{A_1 D_2} = 2.56, & \text{two twins, two problems,} \\
\tilde{\tau}_{A_2 D_1} &= 1.56, \ \tilde{\chi}_{A_2 D_1} = 2.29, & \text{two twins, two problems,} \\
\tilde{\tau}_{A_1 D_1} &= 1.98, \ \tilde{\chi}_{A_1 D_1} = 4.04, & \text{twin 1, two problems,} \\
\tilde{\tau}_{A_2 D_2} &= 1.71, \ \tilde{\chi}_{A_2 D_2} = 2.66, & \text{twin 2, two problems.}
\end{aligned}
$$

The dependence ratio, $\tau_{XY} = 1$ if and only if X and Y are independent. If the association between X and Y is positive, then $\tau_{XY} > 1$, and if the association is negative, then $0 \leq \tau_{XY} < 1$. That $\tilde{\tau}_{A_1 A_2} = 3.72$ signifies that the probability for both members of a monozygotic twin pair being alcohol dependent is 272% higher than under independence. Note that $\tilde{\tau}_{A_1 A_2}$ is by far the highest of the six second order dependence ratios listed above. The maximum possible value for $\tilde{\tau}_{A_1 A_2}$ is $\min\{1/\tilde{\mu}_{A_1}, 1/\tilde{\mu}_{A_2}\}$, which is 10.53. Similarly, all the other five dependence ratios are well below their upper bound. The upper bound of $\tilde{\tau}_{D_1 D_2}$ is 3.17. There is a clear indication that the association between twins is stronger for alcohol dependence than for depression. We investigate this association by constructing a model with separate genetic dispositions for alcohol dependence and depression.

The dependence and the odds ratio rank the six associations in the same order, the odds ratio always being larger, which is a general result to which we return in Section 4 *(iii)*.

Note that both $\tilde{\tau}_{A_1 D_1}$ and $\tilde{\tau}_{A_2 D_2}$ are greater than $\tilde{\tau}_{A_1 D_2}$ and $\tilde{\tau}_{A_2 D_1}$, even if by only a narrow margin. This empirical observation motivates us to allow, in the model, for the possibility that there is an individual effect in developing alcohol dependence and depression, emanating from individual environmental experiences rather than from genetic dispositions.

2. Model for Monozygotic Twins
The model which we use for the statistical analysis of the data in Table 1 is based on three assumptions.
Assumption 1: Each twin pair belongs to exactly one of four latent classes, denoted by $(a, d), (a, \bar{d}), (\bar{a}, d), (\bar{a}, \bar{d})$, indicating presence in the following way:

(a, d) the genetic disposition for alcohol dependence and for depression,

$(a, \bar{d})$ the genetic disposition for alcohol dependence, but not for depression,

$(\bar{a}, d)$ the genetic disposition not for alcohol dependence, but for depression,

$(\bar{a}, \bar{d})$ the genetic disposition not for alcohol dependence, nor for depression.

The probabilities for these four latent classes are expressed by three parameters: the marginal univariate probabilities ν_a and ν_d of the dispositions for, respectively, alcohol dependence and depression, and the dependence ratio between these two dispositions, ξ_{ad}. The ensuing probabilities of the four latent classes and their marginal sums are presented in Table 2.

Assumption 2: We introduce four parameters, denoted α, ϕ, δ and ψ, defined, for $0 < \alpha, \delta < 1$ and $\phi, \psi \geq 0$, to depict the conditional probabilities of the manifest responses given the latent classes:

$$
\begin{aligned}
\mathrm{pr}(A_1 = 1|a) &= \mathrm{pr}(A_2 = 1|a) = \alpha, & (1) \\
\mathrm{pr}(A_1 = 1|\bar{a}) &= \mathrm{pr}(A_2 = 1|\bar{a}) = \phi\alpha, & (2) \\
\mathrm{pr}(D_1 = 1|\bar{d}) &= \mathrm{pr}(D_2 = 1|\bar{d}) = \delta, & (3) \\
\mathrm{pr}(D_1 = 1|d) &= \mathrm{pr}(D_2 = 1|d) = \psi\delta. & (4)
\end{aligned}
$$

Table 2

The bivariate probabilities that a monozygotic twin pair belongs to one of four latent classes and the corresponding marginal probabilities for the genetic dispositions.

Genetic dispositions			
Alcohol	Depression		
dependence	d	$\bar{d}$	Sum
a	$\nu_a\nu_d\xi_{ad}$	$\nu_a(1 - \nu_d\xi_{ad})$	ν_a
$\bar{a}$	$\nu_d(1 - \nu_a\xi_{ad})$	$1 - \nu_a - \nu_d + \nu_a\nu_d\xi_{ad}$	$1 - \nu_a$
Sum	ν_d	$1 - \nu_d$	1

Four features of equations (1) to (4) deserve comments:

(i) They imply marginal homogeneity, for the pairs (A_1, A_2), and (D_1, D_2).

(ii) The probability for alcohol dependence is influenced only by the genetic disposition $a/\bar{a}$, not at all by $d/\bar{d}$, and conversely for depression.

(iii) The model provides for a possible association at the latent level between the dispositions for alcohol dependence and depression; the strength of this latent association is measured by the dependence ratio ξ_{ad}.

4

(iv) The parameter α is the probability of alcohol dependence, given that a twin has the genetic disposition a, while δ is the 'background' probability of depression, given that a twin does not have the genetic disposition d. The reason for this asymmetry is technical convenience: by simply setting $\phi = 0$, the background probability for alcohol dependence, calculated as $\phi\alpha$, is set to zero and the model simplified by removal of one parameter.

Assumption 3: We assume conditional independence between the members of a twin pair, given the pairs genetic structure. Let G be a generic symbol for the genetic structure of a twin pair, so that any realization of G is one of the four latent classes $(a, d), (a, \bar{d}), (\bar{a}, d), (\bar{a}, \bar{d})$. We assume,

$$A_1 \perp\!\!\!\perp A_2 | G, \quad D_1 \perp\!\!\!\perp D_2 | G, \quad A_1 \perp\!\!\!\perp D_2 | G, \quad A_2 \perp\!\!\!\perp D_1 | G. \tag{5}$$

Statements (5) imply conditional independence for four of the six bivariate marginal distributions of (A_1, A_2, D_1, D_2). At the onset we must assume $A_1 \not\perp\!\!\!\perp D_1 | G$ and $A_2 \not\perp\!\!\!\perp D_2 | G$, for these bivariate distributions are based, not only on the genetic disposition of the individual, but also on her unique environmental experiences. To account for individual environmental effects we introduce an eighth parameter, denoted β and defined, for $i, i' = 1, 2$, $i \neq i'$, as

$$\beta = \frac{\mathrm{pr}(A_i = 1, D_i = 1 | G)}{\mathrm{pr}(A_i = 1, D_{i'} = 1 | G)}. \tag{6}$$

Using statements (1) to (6) one can calculate, conditionally on G, all moments of order four or less, for A_1, A_2, D_1, D_2. For example, $E(A_1 A_2 D_1 | G) = \beta \{E(A|G)\}^2 E(D|G)$ and $E(A_1 A_2 D_1 D_2 | G) = \beta^2 \{E(A|G)\}^2 \{E(D|G)\}^2$. The unconditional moments are then obtained as weighted sums of the conditional moments, summing over the four latent classes in Table 2, using the latent class probabilities as weights. From the explicit expressions, in terms of parameters, for all the moments, the probabilities for all cells of Table 1 are easily calculated. For example,

$$\mathrm{pr}(1, 0, 0, 1) = E(A_1 D_2) - E(A_1 A_2 D_2) - E(A_1 D_1 D_2) + E(A_1 A_2 D_1 D_2),$$

and the longest expression is for $(0, 0, 0, 0)$:

$$
\begin{aligned}
\mathrm{pr}(0, 0, 0, 0) \;=\; & 1 - E(A_1) - E(A_2) - E(D_1) - E(D_2) \\
& + E(A_1 A_2) + E(A_1 D_1) + E(A_1 D_2) + E(A_2 D_1) + E(A_2 D_2) + E(D_1 D_2) \\
& - E(A_1 A_2 D_1) - E(A_1 A_2 D_2) - E(A_1 D_1 D_2) - E(A_2 D_1 D_2) \\
& + E(A_1 A_2 D_1 D_2).
\end{aligned}
$$

Thus, we obtain an explicit expression for the log-likelihood function:

$$l(\nu_a, \nu_d, \xi_{ad}, \alpha, \phi, \delta, \psi, \beta | \mathbf{y}) = 7 \log\{\mathrm{pr}(1,1,1,1)\} + \cdots + 288 \log\{\mathrm{pr}(0,0,0,0)\}.$$

The arguments of the log-likelihood function are the eight parameters defined in Table 2 and equations (1) to (4) and it is based on the data $\mathbf{y}$, which are the 16 counts listed in Table 1.

3. Model Fitting and Selection

All programming was performed in R (Ihaka and Gentleman,1996), in particular the maximizing of the log-likelihood function using a quasi-Newton optimiser.

Table 3 reports four model statistics,

(i) the maximum value of the log-likelihood function, denoted $\max\{l(\cdot|\mathbf{y})\}$,

(ii) the number of parameters, denoted $\#(parameters)$,

(iii) the value of Akaike's Information Criterion, denoted AIC,

(iv) the value of the determinant of the correlation matrix for the parameter
  estimates, denoted det(corr),

for four models: Model 1 (all 8), in which all eight parameters, are estimated, Model 2 ($\beta = 1$), in which the value of β is set to 1, Model 3 ($\phi = 0$), in which the value of ϕ is set to 0, and Model 4 ($\beta = 1, \phi = 0$), in which β is set to 1, and ϕ is set to 0.

Table 3

Four model statistics for four models fitted to the data of Table 1.

Model	Model statistics			
	$\max\{l(\cdot|\mathbf{y})\}$	$\#(parameters)$	AIC	det(corr)
Model 1 (all 8)	-1040.8	8	2097.6	0.00035
Model 2 ($\beta = 1$)	-1041.5	7	2097.0	0.00153
Model 3 ($\phi = 0$)	-1041.3	7	2096.5	0.00235
Model 4 ($\beta = 1, \phi = 0$)	-1042.5	6	2096.9	0.00685

Assessed by the likelihood ratio test, none of the models with more parameters fit the data significantly better than Model 4. Model 3 has a smaller AIC value than Model 4, but by so small a margin that it would not be sensible to prefer Model 3. The determinant of the matrix of correlations of the parameter estimates can be used as a global measure of the stability of the estimates. It makes for a fairer comparison of determinants of matrices to use $\det^{1/p}$, where p is the dimension of the matrix. Even after the p-root transformation Model 4 has the highest value, indicating the best determination of the parameters.

Based on these considerations we selected Model 4 as our final model. From a substantive point of view, it is notable that the estimates for β, by Models 1 and 3 are, respectively 1.12 and 1.16, in both cases with standard errors 0.11. That the individual environmental effect is so small, in this data set, could be a result of the majority of these twins having lived in the same or closely similar circumstances, rather than being a confirmation of the dominant effect of the genetic disposition. The estimates of $\mathrm{pr}(A = 1|\bar{a})$ from Models 1 and 2 are, respectively, 0.024 and 0.028. Setting, in the final model, the background probability of alcohol dependence to 0, is not because we believe that having the genetic disposition a is a necessary prerequisite for alcohol dependence. But the values 0.024 and 0.028 are so small, compared to the estimate of the background probability for depression, $\mathrm{pr}(D = 1|\bar{d}) = 0.20$, that it is sensible to increase the stability of the model by eliminating the parameter $\mathrm{pr}(A = 1|\bar{a})$.

Table 4 reports the parameter estimates with standard errors, Table 5 the correlations of the estimates, Table 6 the fitted counts alongside the observed ones and Table 7 the estimates of the probabilities of the latent classes corresponding to the algebraic forms of Table 2.

Table 4

The parameter estimates with standard errors of Model 4, fitted to the observed counts reported in Table 1.

Parameter	Estimate	Standard error
ν_a	0.272	0.050
ν_d	0.215	0.087
ξ_{ad}	2.943	0.676
α	0.314	0.060
δ	0.199	0.031
ψ	3.381	0.428

The background probability for alcohol dependence is set to zero in Model 4 and $\hat{\mathrm{pr}}(A = 1|a) = \hat{\alpha} = 0.314$, while the estimate of the background probability for depression is $\hat{\mathrm{pr}}(D = 1|\bar{d}) = \hat{\delta} = 0.199$ and the probability for depression given the genetic disposition is $\hat{\mathrm{pr}}(D = 1|d) = \hat{\psi}\hat{\delta} = 0.673$.

Table 5

The correlations of the parameter estimates of Model 4.

	$\hat{\nu}_a$	$\hat{\nu}_d$	$\hat{\xi}_{ad}$	$\hat{\alpha}$	$\hat{\delta}$	$\hat{\psi}$
$\hat{\nu}_a$	1.00	0.04	-0.08	-0.84	0.00	0.00
$\hat{\nu}_d$	·	1.00	-0.82	0.00	-0.84	0.01
$\hat{\xi}_{ad}$	·	·	1.00	0.00	0.76	-0.22
$\hat{\alpha}$	·	·	·	1.00	0.00	0.00
$\hat{\delta}$	·	·	·	·	1.00	-0.47
$\hat{\psi}$	·	·	·	·	·	1.00

Table 6

Observed and fitted counts of 597 female monozygotic twin pairs classified by concordance/discordance with respect to alcohol dependence and depression. The italicized fitted counts are for Model 4.

		$D_1 = 1$		$D_1 = 0$	
		$D_2 = 1$	$D_2 = 0$	$D_2 = 1$	$D_2 = 0$
$A_1 = 1$	$A_2 = 1$	7 *4.84*	4 *3.16*	2 *3.16*	3 *4.84*
$A_1 = 1$	$A_2 = 0$	9 *10.58*	8 *6.92*	4 *6.92*	8 *10.58*
$A_1 = 0$	$A_2 = 1$	10 *10.58*	7 *6.92*	9 *6.92*	15 *10.58*
$A_1 = 0$	$A_2 = 0$	51 *50.99*	92 *86.01*	80 *86.01*	288 *287.99*

The likelihood ratio test statistic for Model 4 against the saturated model has the value 7.98 on 10 degrees of freedom corresponding to $p = 0.63$, indicating that Model 4 fits the data very well.

Table 7

The fitted latent class probabilities for Model 4 corresponding to the algebraic expressions given in Table 2.

Genetic dispositions			
Alcohol	Depression		
dependence	d	$\bar{d}$	Sum
a	0.172	0.100	0.272
$\bar{a}$	0.043	0.685	0.728
Sum	0.215	0.785	1

Two out of three (0.685) female monozygotic twin pairs have neither disposition and more than half (0.546) of those that have at least one, have both. This latter fact is a reflection of $\hat{\xi}_{ad} = 2.94$, indicating strong association between the genetic dispositions for alcohol dependence and depression.

4. Model validation
In the framework of likelihood-based inference, a complete understanding of the features of the selected model can only be reached by studying the surface $l(\nu_a, \nu_d, \xi_{ad}, \alpha, \delta, \psi|\mathbf{y})$ as an object in an Euclidian space of seven dimensions. We have to be content with subspaces of, at most, three dimensions and concentrate on features of most interest from the subject-matter point of view. We group these issues under three headings.

(i) Label switching. It turns out, that the log-likelihood function has two maxima with values -1042.5. Table 8 lists the two sets of arguments for the two maxima, with arg(max 1) being the maximum likelihood estimates, reported in Table 4, and arg(max 2) the coordinates of the second maximum.

Table 8

Two different sets of arguments both giving the log-likelihood function of Model 4 the maximum value -1042.5, see also Table 4.

Parameter	ν_a	ν_d	ξ_{ad}	α	δ	ψ
arg(max 1)	0.272	0.215	2.943	0.314	0.199	3.381
arg(max 2)	0.272	0.785	0.468	0.314	0.674	0.296

We denote the coordinates of the second maximum by $v_1, \ldots, v_6$ and drop, for simplicity, the hats on the maximum likelihood estimates. Note first, that $v_1 = 0.272 = \nu_a = \text{pr}(a)$ and $v_4 = 0.314 = \alpha = \text{pr}(A = 1|a)$, so that ν_a and α appear in both solutions. A clue to understanding the relations between the differing coordinates is that $v_2 = 0.785 = 1 - 0.215 = 1 - \nu_d$, which implies that the labels for d and $\bar{d}$ have been switched, so that $v_2 = \text{pr}(\bar{d})$. Accordingly, $v_3 = 0.468 = \text{pr}(a, \bar{d})/\{\text{pr}(a)\text{pr}(\bar{d})\} = 0.100/(0.272 \cdot 0.785)$, see Table 7. Moreover, $v_5 = 0.674 = \psi\delta = \text{pr}(D = 1|d)$ and, finally, $v_6 = 0.296 = 1/3.381 = 1/\psi$. The iterative algorithm that we used for finding the maximum likelihood estimates converges to arg(max 2) when the initial values for the arguments is set, for example, to $(0.3, 0.7, 0.5, 0.3, 0.7, 0.3)$. There would not be any scientific sense in exchanging ξ_{ad} for $\xi_{a\bar{d}}$, therefore, choosing arg(max 1) as the maximum likelihood estimates is undisputable. Some of the contour plots discussed next, are designed so that the location of the second maximum is shown.

(ii) Identifiability and stability. There are in Table 5 four pairs of parameter estimates with disturbingly high correlations. Reading across rows, they are for the pairs (1), $(\hat{\nu}_a, \hat{\alpha})$, (2), $(\hat{\nu}_d, \hat{\xi}_{ad})$, (3), $(\hat{\nu}_d, \hat{\delta})$, (4), $(\hat{\xi}_{ad}, \hat{\delta})$. To understand what these high correlations tell about the shape of the log-likelihood

surface in the vincinity of the maximum, we use profile log-likelihood functions with two arguments. The profile log-likelihood function for pair (1), denoted $l_P(\nu_a, \alpha)$, is defined as,

$$l_P(\nu_a, \alpha) = l\{\nu_a, \hat{\nu}_d(\nu_a, \alpha), \hat{\xi}_{ad}(\nu_a, \alpha), \alpha, \hat{\delta}(\nu_a, \alpha), \hat{\psi}(\nu_a, \alpha)\}, \tag{7}$$

where $\hat{\nu}_d(\nu_a, \alpha)$, $\hat{\xi}_{ad}(\nu_a, \alpha)$, $\hat{\delta}(\nu_a, \alpha)$, $\hat{\psi}(\nu_a, \alpha)$ indicates that we have, for any pair of fixed values (ν_a, α), maximized the log-likelihood function with respect to the remaining four arguments. The profile log-likelihood function of two arguments is, thus, a synthetic function, composed of maxima of the log-likelihood with respect to the other four arguments. For studying the shape of $l(\cdot)$, as a function of two arguments, the other four arguments are eliminated by maximisation, rather than by integration as in Bayesian inference. Generally, the maximum of the profile log-likelihood coincides with the maximum of the log-likelihood, in particular, $\mathrm{argmax}\{l_P(\nu_a, \alpha)\} = \mathrm{argmax}\{l(\nu_a, \nu_d, \xi_{ad}, \alpha, \delta, \psi)\}$.

Fig. 1 depicts the contours of constant value of $l_P(\nu_a, \alpha)$. The point, plotted as a tiny circle inside the closed contour marked '-1043', is the $\mathrm{argmax}\{l_P(\nu_a, \alpha)\} = (\hat{\nu}_a, \hat{\alpha}) = (0.272, 0.314)$, at which $l_P(0.272, 0.314) = -1042.5$. It is clear from Fig. 1 that the profile log-likelihood has a ridge in the direction of the principal axis of the '-1043'-contour. The heuristic explanation for this ridge is that $\mathrm{pr}(A = 1) = \mathrm{pr}(a)\mathrm{pr}(A = 1|a) = \nu_a\alpha$. From inspection of $l_P(\nu_a, \alpha)$, the bad news is that a small perturbation of the data could lead to considerable changes, in opposite directions, of the maximum likelihood estimates for ν_a and α. The good news is, that there is a unique highest point on the ridge and ν_a and α are properly identified, though not estimated in a very stable way.

Fig. 2 depicts the contours of constant height of $l_P(\xi_{ad}, \nu_d)$. The striking difference to Fig. 1 is, that both maxima show up, since the range for the ξ_{ad}-axis is $(0, 5)$. The part of Fig. 2 of subject-matter interest is, of course, $1 < \xi_{ad} < 5$. The '$y = c/x$'-shape of the ridge, inside the '-1043'-contour emanates from the fact that $\xi_{ad} = \mathrm{pr}(a, d)/(\nu_a \nu_d) \propto c/\nu_d$. Fig. 3, being for $l_P(\nu_d, \delta)$, resembles closely Fig. 1 and Fig. 4, being for $l_P(\xi_{ad}, \delta)$, resembles Fig. 2. Finally, as a contrast, Fig. 5 presents the contour plot for $l_P(\psi, \alpha)$, in which case $\mathrm{corr}(\hat{\psi}, \hat{\alpha}) = 0.00$. Not only is the ridge less pronounced, but since it is parallel to the abscissa the correlation is zero.

In conclusion, we note that for all four high correlations at least one of the parameters involved, belongs to the group (ν_a, ν_d, ξ_{ad}), which regulate
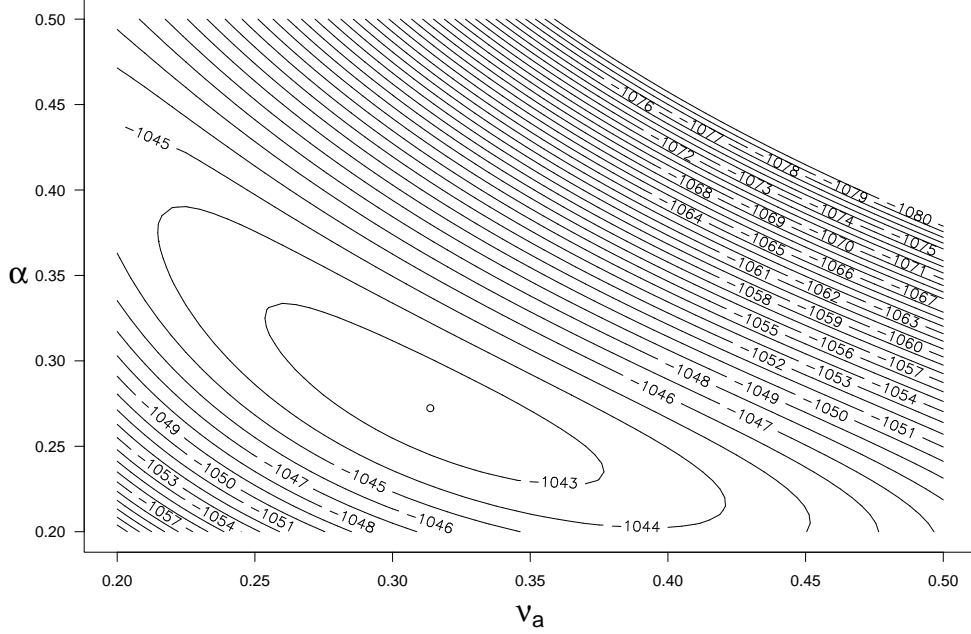
Figure 1: Contour plot of profile log-likelihood of ν_a and α, $l_P(\nu_a, \alpha)$

the latent structure, while all correlations are small between the parameters in the group $(\alpha,\ \delta,\ \psi)$, which regulate the conditional probabilities of the observable events. The problem of instability of estimates is, therefore, rather an intrinsic feature of the model we propose, than of bad luck with the particular data set we analyse.

(iii) Odds ratio or dependence ratio? Commentators have voiced preference for the well established odds ratio, over the more recently introduced dependence ratio (Ekholm et al., 1995), on the grounds that the odds ratio is invariant to the scoring and orthogonal to the marginal probabilities. To be specific we denote the odds ratio for Table 2 by χ_{ad} and the cell probabilities, reading across rows, by $\pi_1, \pi_2, \pi_3, \pi_4$. Using the expressions from Table 2 we find,

$$\chi_{ad} \equiv \frac{\pi_1 \pi_4}{\pi_2 \pi_3} = \frac{\xi_{ad}(1 - \nu_a - \nu_d + \xi_{ad}\nu_a\nu_d)}{(1 - \nu_a\xi_{ad})(1 - \nu_d\xi_{ad})}. \tag{8}$$

It follows from (8) that

$$\chi_{ad} - 1 = \frac{\xi_{ad} - 1}{(1 - \nu_a\xi_{ad})(1 - \nu_d\xi_{ad})}. \tag{9}$$
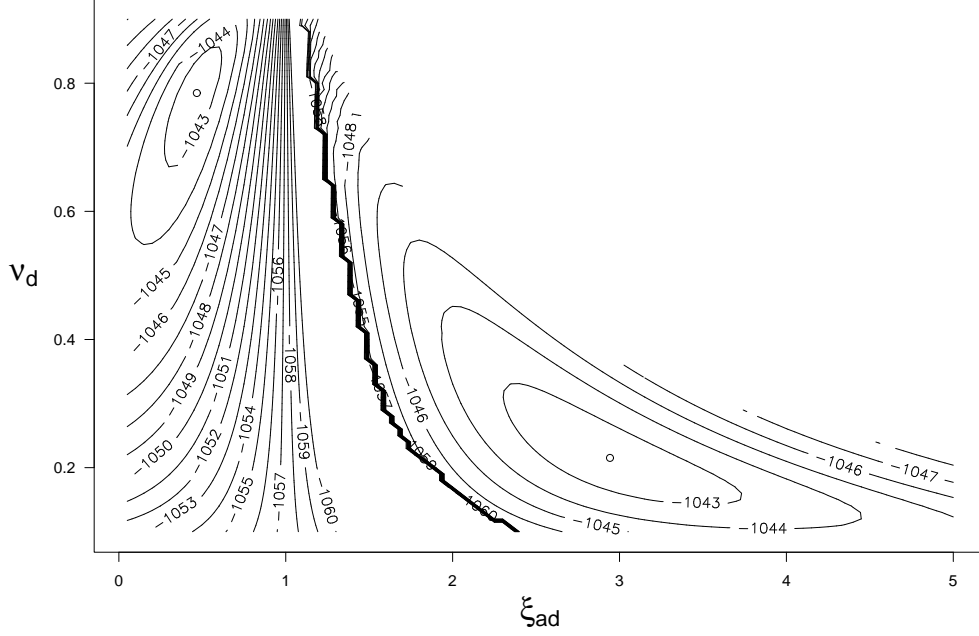
11

Figure 2: Contour plot of profile log-likelihood of ξ_{ad} and ν_d, $l_P(\xi_{ad}, \nu_d)$

From (9) one readily deduces that

$$|\xi_{ad} - 1| \leq |\chi_{ad} - 1|, \tag{10}$$

with equality only under independence, when $\xi_{ad} = \chi_{ad} = 1$. The dependence ratio is bounded towards zero association by the odds ratio.

The odds ratio is, indeed, invariant to the scoring if one reverses the scores for both dispositions, that is, applying both switches $a \to \bar{a}$ and $d \to \bar{d}$, then $\chi_{\bar{a}\bar{d}} = \chi_{ad}$, while $\xi_{\bar{a}\bar{d}} \neq \xi_{ad}$. But if only the switch $d \to \bar{d}$ is applied, while a is untouched as in subsection *(i)*, then the odds ratio is not invariant; instead $\chi_{a\bar{d}} = 1/\chi_{ad}$. Therefore, parametrising Table 2 by $(\nu_a, \nu_d, \chi_{ad})$ does not prevent the appearance of two maxima for the log-likelihood.

Expressing the cell probabilities of Table 2 by $(\nu_a, \nu_d, \chi_{ad})$ is a more cumbersome task, than expressing them by (ν_a, ν_d, ξ_{ad}). To go from given values of ν_a, ν_d, χ_{ad} to the corresponding $\pi_1, \pi_2, \pi_3, \pi_4$, one needs first to find the root, satisfying $0 \leq \pi_1 \leq 1$, of the quadratic equation,

$$(\chi_{ad} - 1)\pi_1^2 - \{(\chi_{ad} - 1)(\nu_a + \nu_d) + 1\}\pi_1 + \chi_{ad}\nu_a\nu_d = 0,$$
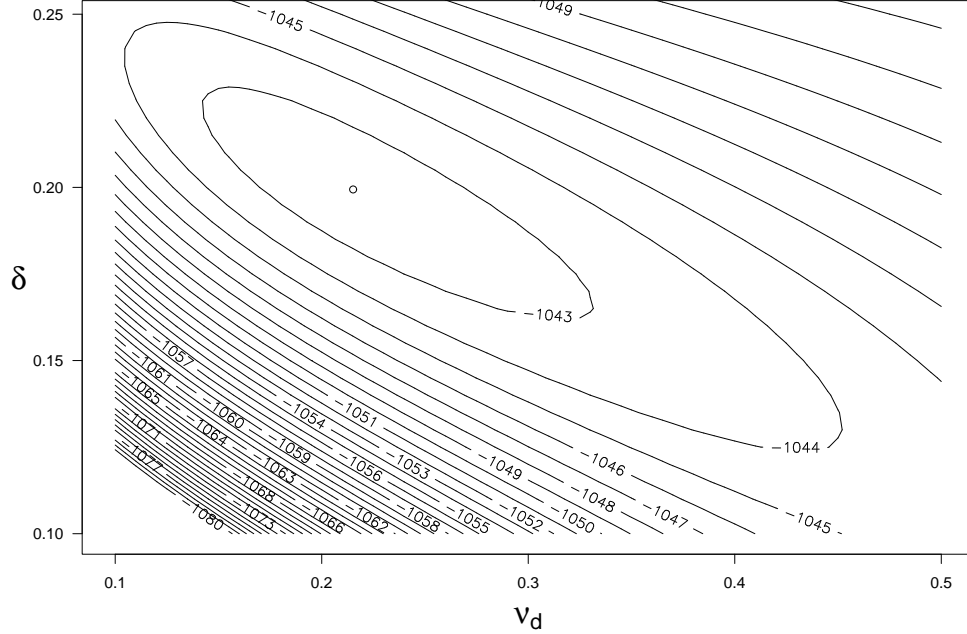
Figure 3: Contour plot of profile log-likelihood of ν_d and δ, $l_P(\nu_d, \delta)$

and then use the linear relations $\pi_1 + \pi_2 = \nu_a$, $\pi_1 + \pi_3 = \nu_d$, and $\pi_1 + \pi_2 + \pi_3 + \pi_4 = 1$. Parametrising a $k \times k$ table, where $k > 2$, by marginal probabilities and odds ratios, the required root of the corresponding high degree polynomial can only be found iteratively. We will encounter such a 4×4 table when extending our model to the case of dizygotic twins in Section 5. For a general $k \times k$ table the expressions for the cell probabilities, in terms of marginal probabilities and dependence ratios, are simple and explicit.

The odds ratio, χ_{ad}, is orthogonal to the marginal probabilities, ν_a and ν_d, in the following sense: denoting by $\tilde{\nu}_a$, $\tilde{\nu}_d$ and $\tilde{\chi}_{ad}$ the maximum likelihood estimates of these three parameters from any 2×2 table of observed frequencies, one finds that, generally, $\mathrm{corr}(\tilde{\nu}_a, \tilde{\chi}_{ad}) = \mathrm{corr}(\tilde{\nu}_d, \tilde{\chi}_{ad}) = 0$, while $\mathrm{corr}(\tilde{\nu}_a, \tilde{\xi}_{ad})$ and $\mathrm{corr}(\tilde{\nu}_d, \tilde{\xi}_{ad})$ are not generally zero. The formulas for $\mathrm{corr}(\tilde{\nu}_a, \tilde{\xi}_{ad})$ and $\mathrm{corr}(\tilde{\nu}_d, \tilde{\xi}_{ad})$ are derived in Ekholm (2003, Append.). Since in Table 5, $\mathrm{corr}(\hat{\nu}_d, \hat{\xi}_{ad}) = -0.82$ and $\mathrm{corr}(\hat{\nu}_a, \hat{\xi}_{ad}) = -0.08$, it is of interest to check if the orthogonality of the odds ratio carries over to the situation where the 2×2 table is latent, instead of directly observed. As another piece of model validation we, therefore, fitted Model 4 replacing ξ_{ad} with three alternatives, in turn, $\log \xi_{ad}$, χ_{ad} and $\log \chi_{ad}$. The results from comparing Model 4, which is the ξ_{ad}-version to the $\log \xi_{ad}$-version, the χ_{ad}-version and
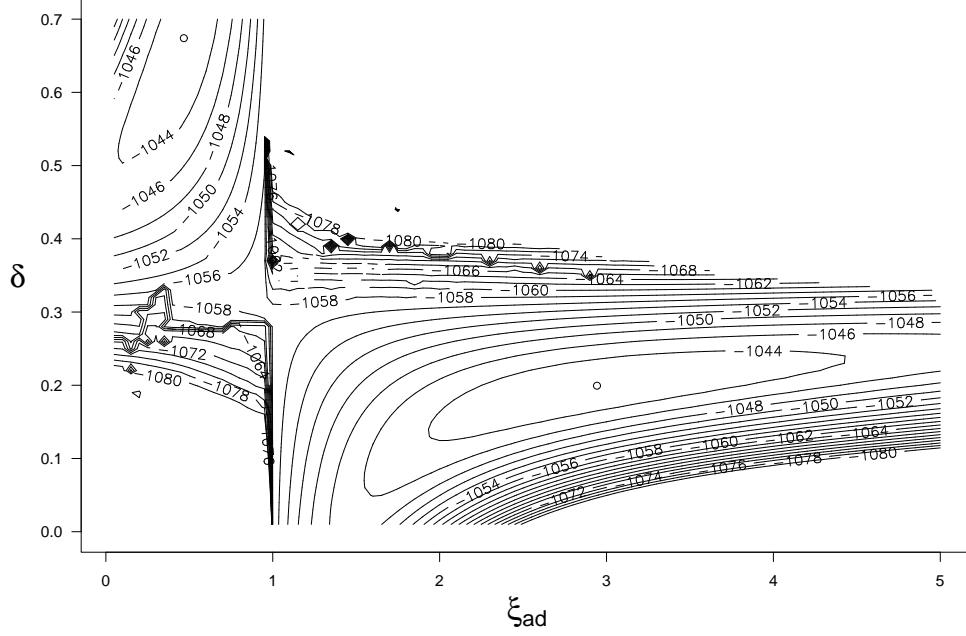
13

Figure 4: Contour plot of profile log-likelihood of ξ_{ad} and δ, $l_P(\xi_{ad}, \delta)$

the $\log \chi_{ad}$-version are:

(1) The alterations to the estimates and standard errors for $\nu_a, \nu_d, \alpha, \delta, \psi$ are negligible.

(2) The alterations to AIC and to the fitted frequencies are negligible.

(3) The alterations to all correlations of estimates, not involving the measure of latent association, are negligible.

(4) The estimates, with standard errors in parentheses, of ξ_{ad}, $\log \xi_{ad}$, χ_{ad} and $\log \chi_{ad}$ are, respectively, 2.94 (0.68), 1.08 (0.23), 27.59 (35.40) and 3.32 (1.29). Note that $\log 2.94 = 1.08$ and $\log 27.59 = 3.32$, satisfying the invariance of maximum likelihood estimates to parameter transformations: $\widehat{\log \xi_{ad}} = \log \hat{\xi}_{ad}$ and $\widehat{\log \chi_{ad}} = \log \hat{\chi}_{ad}$. We use, for esthetic reasons, the notations $\log \hat{\xi}_{ad}$ and $\log \hat{\chi}_{ad}$.

(5) The estimates $\hat{\xi}_{ad}$ and $\log \hat{\xi}_{ad}$ have identical correlations with $\hat{\nu}_a, \hat{\nu}_d, \hat{\alpha}, \hat{\delta}, \hat{\psi}$ and similarly $\hat{\chi}_{ad}$ and $\log \hat{\chi}_{ad}$ have identical correlations, but the correlations of $\hat{\xi}_{ad}$ and of $\hat{\chi}_{ad}$ with $\hat{\nu}_a$, $\hat{\nu}_d$, $\hat{\alpha}$, $\hat{\delta}$, $\hat{\psi}$ are different, being respectively, -0.08, -0.82, 0.00, 0.76, -0.22 and 0.48, -0.48, -0.47, 0.52, -0.26.

(6) When exchanging ξ_{ad} for χ_{ad} the summary measure of stability, det(corr), increases by a factor of 1.57 from 0.00685 to 0.01074.
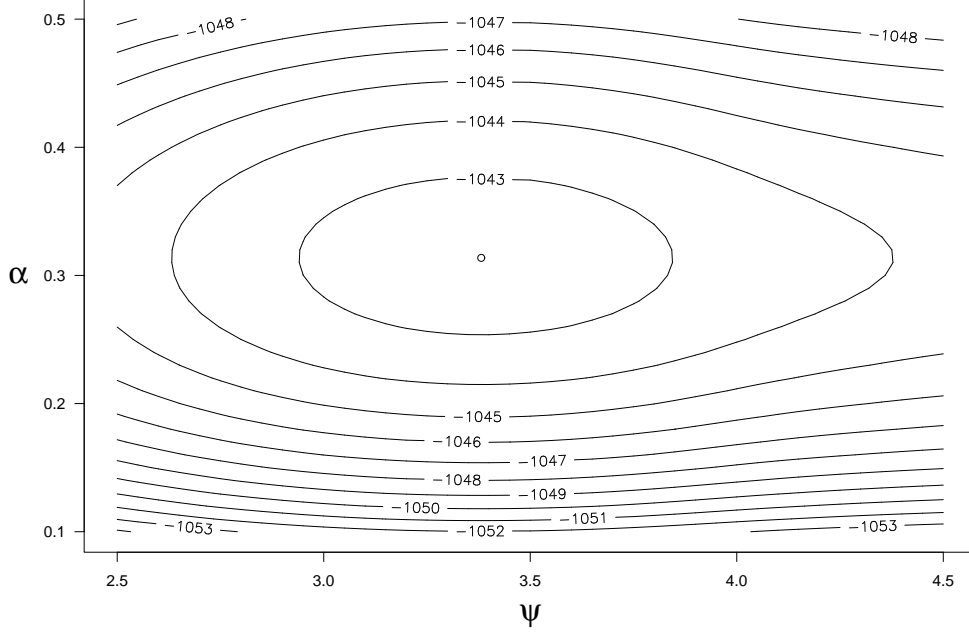
Figure 5: Contour plot of profile log-likelihood of ψ and α, $l_P(\psi, \alpha)$

The results (1) to (3) show that the subject matter conclusions from the odds ratio version of Model (4) are the same as from the dependence ratio version, for all other parameters except for the association between the latent dispositions. Results (5) and (6) demonstrate that χ_{ad} is, in the present setting, not orthogonal to the marginal probabilities, but $\hat{\chi}_{ad}$ is considerably less entangled than $\hat{\xi}_{ad}$ with the estimates of the other five parameters.

Result (4), however, implies that the smaller entanglement comes with $\hat{\chi}_{ad}$, itself, taking a very large value and having a dramatically large standard error. For further confirmation we plot in Fig. 6 the profile log-likelihood functions for the four alternative measures of the latent association, namely $l_P(\xi_{ad})$, $l_P(\log \xi_{ad})$, $l_P(\chi_{ad})$ and $l_P(\log \chi_{ad})$. The horizontal lines in Fig. 6 are at level, $\max\{l_P(\cdot)\} - 1.92$, which, in case of a quadratic log-likelihood function, delimits a 95% confidence interval, see Barndorff-Nielsen and Cox (1994, p. 90). Note that the scale of the bottom two panels is ten-fold that of the top two panels. Fig. 6 shows that no likelihood-based intervals can be constructed for χ_{ad} or $\log \chi_{ad}$. In contrast, $l_P(\xi_{ad})$ and, in particular, $l_P(\log \xi_{ad})$, are well behaved and the latter symmetric enough that a likelihood interval can be given a confidence interpretation. The 95% confidence interval, based on $l_P(\log \xi_{ad})$, for the log-dependence ratio is, $0.607 \leq \log \xi_{ad} \leq 1.548$, and from there, by exponentiation, for the dependence ratio, $1.83 \leq \xi_{ad} \leq 4.70$.
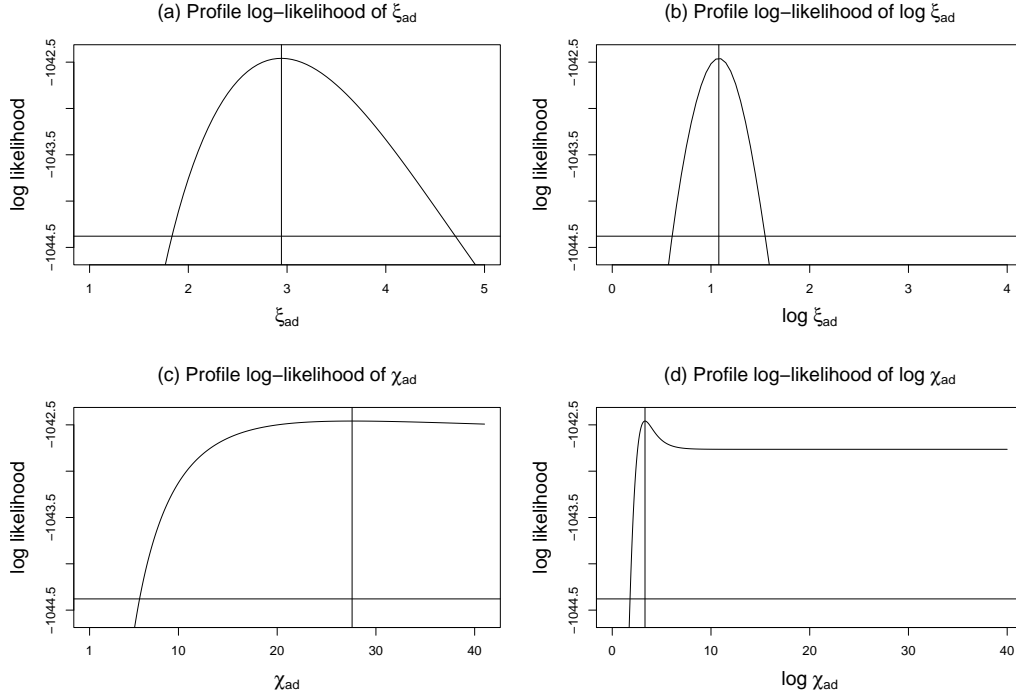
15

Figure 6: Profile log-likelihood for four measures of association between the latent variables. The horizontal line corresponds to $\max\{l_P(\cdot)\} - 1.92$.

From the inequality (10) it was expected that $\xi_{ad} < \chi_{ad}$, but it is surprising that the probabilities of $(a, \bar{d})$ and $(\bar{a}, d)$ are small enough to make χ_{ad} explode beyond usefulness. The intuitive background for the inequality (10) is that ξ_{ad} measures the strength of association between the events a and d, while χ_{ad} measures the strength of association between two binary random variates $Y_1 = a/\bar{a}$ and $Y_2 = d/\bar{d}$. The event specificity of ξ_{ad} makes it asymmetric and bounds it by finite limits, see equation (19). The dependence ratio has been criticized for having finite bounds, but here they are very helpful. Moreover, we show in Section 5, that boundedness of the dependence ratio enables formulating models for monozygotic and dizygotic twins in a common framework. To sum up the issue of odds ratio *versus* dependence ratio, the latter has in the present framework four advantages: (i), the interpretation is simpler, (ii), the generalization to the 4×4 table for dizygotic twins is neater, (iii), the boundedness by finite limits is useful, (iv), the standard error of the estimate is smaller. At least in the present example, the dependence ratio works better than the odds ratio.

16

5. Model for Dizygotic Twins

Monozygotic twins are identical genetically, so that if we tabulate two monozygotic twins, denoted, respectively, MZ1 and MZ2, against each other for the four latent classes of genetic dispositions, $(a,d), (a,\bar{d}), (\bar{a},d), (\bar{a},\bar{d})$, we get Table 9. For brevity, we use in Table 9 and henceforth, in parallel, the following two notations for the probabilities of the latent classes, $\pi_1 = \nu_a \nu_d \xi_{ad}$, $\pi_2 = \nu_a(1 - \nu_d \xi_{ad})$, $\pi_3 = \nu_d(1 - \nu_a \xi_{ad})$ and $\pi_4 = 1 - \nu_a - \nu_d + \nu_a \nu_d \xi_{ad}$.

For a pair of dizygotic twins the table corresponding to Table 9 no longer has zeros in the off-diagonal positions. To derive the joint probabilities that dizygotic twin 1 belongs to one of the four latent classes and dizygotic twin 2 to the same or another, we define two four-valued nominal responses Z_i, for $i = 1, 2$, where Z_i indicates to which latent class twin i belongs. Also, we define a third four-valued nominal response P with the same four outcomes, where P denotes the pool of genetic dispositions that the dizygotic twins can inherit from their parents. We need not speculate whether the mechanisms of inheritance of these four dispositions is additive or dominant, geared to a single or a multitude of loci, gender linked or not, etc. The one assumption we find necessary is that the marginal distribution of the pair of parents over the four latent classes, $(a,d), (a,\bar{d}), (\bar{a},d), (\bar{a},\bar{d})$, is the same as that of their children.

Table 9

The probabilities that monozygotic twins belong to one of the 16 latent classes formed by cross-classifying the classes introduced in Table 2.

MZ1	MZ2 (a,d)	$(a,\bar{d})$	$(\bar{a},d)$	$(\bar{a},\bar{d})$	Sum
(a,d)	$\nu_a \nu_d \xi_{ad}$	0	0	0	π_1
$(a,\bar{d})$	0	$\nu_a(1 - \nu_d \xi_{ad})$	0	0	π_2
$(\bar{a},d)$	0	0	$\nu_d(1 - \nu_a \xi_{ad})$	0	π_3
$(\bar{a},\bar{d})$	0	0	0	$1 - \nu_a - \nu_d + \nu_a \nu_d \xi_{ad}$	π_4
Sum	π_1	π_2	π_3	π_4	1

Assumption 4: The marginal distribution of the latent variable P is the same as that of the latent variables Z_i, for $i = 1, 2$,

$$
\begin{aligned}
\pi_1 &= \text{pr}\{P = (a,d)\} = \text{pr}\{Z_i = (a,d)\} = \nu_a \nu_d \xi_{ad}, & (11)\\
\pi_2 &= \text{pr}\{P = (a,\bar{d})\} = \text{pr}\{Z_i = (a,\bar{d})\} = \nu_a(1 - \nu_d \xi_{ad}),\\
\pi_3 &= \text{pr}\{P = (\bar{a},d)\} = \text{pr}\{Z_i = (\bar{a},d)\} = \nu_d(1 - \nu_a \xi_{ad}),\\
\pi_4 &= \text{pr}\{P = (\bar{a},\bar{d})\} = \text{pr}\{Z_i = (\bar{a},\bar{d})\} = 1 - \nu_a - \nu_d + \nu_a \nu_d \xi_{ad}.
\end{aligned}
$$

The three parameters ν_a, ν_d and ξ_{ad} are, of course, the same as for monozygotic twins from the same population. To complement these three parameters to a saturated parametrisation for the bivariate joint distribution of (P, Z_i), $i = 1, 2$, we define six dependence ratios, denoted by θ, and require that the distribution of (P, Z_i), in addition to being marginally homogenous, is symmetric in the sense that, for $h, k = 1, \ldots, 4, h \neq k$,

$$\mathrm{pr}(P = h, Z_i = k) = \mathrm{pr}(P = k, Z_i = h). \tag{12}$$

The main interest focus on three intra-class dependence ratios depicting the association between parents' and their childrens' latent class, for the three classes indicating presence of at least one of the genetic dispositions.

Assumption 5: We define for $k = 1, 2, 3$ the intra-class dependence ratios as,

$$\theta_{kk} = \frac{\mathrm{pr}(P = k, Z_i = k)}{\mathrm{pr}(P = k) \cdot \mathrm{pr}(Z_i = k)}, \tag{13}$$

and for $h, k = 1, 2, 3, h \neq k$ the cross-class dependence ratios,

$$\theta_{hk} = \frac{\mathrm{pr}(P = h, Z_i = k)}{\mathrm{pr}(P = h) \cdot \mathrm{pr}(Z_i = k)}. \tag{14}$$

It follows from restriction (12) that $\theta_{hk} = \theta_{kh}$ and from equations (11) to (14) that the bivariate probabilities of (P, Z_i), for $i = 1, 2$, are as listed in Table 10, where, for brevity the following notations are used:

$$\begin{aligned}
\sigma_1 &= \pi_1\theta_{11} + \pi_2\theta_{12} + \pi_3\theta_{13}, \quad \sigma_2 = \pi_1\theta_{21} + \pi_2\theta_{22} + \pi_3\theta_{23}, \tag{15}\\
\sigma_3 &= \pi_1\theta_{31} + \pi_2\theta_{32} + \pi_3\theta_{33}, \quad \sigma = 1 - \pi_1(2 - \sigma_1) - \pi_2(2 - \sigma_2) - \pi_3(2 - \sigma_3).
\end{aligned}$$

Table 10

The probabilities that a dizygotic twin belongs to the same or a different latent class as her parents.

Latent class of parents	Latent class of twin $i = 1, 2$				
	(a, d)	$(a, \bar{d})$	$(\bar{a}, d)$	$(\bar{a}, \bar{d})$	Sum
(a, d)	$\pi_1^2\theta_{11}$	$\pi_1\pi_2\theta_{12}$	$\pi_1\pi_3\theta_{13}$	$\pi_1(1 - \sigma_1)$	π_1
$(a, \bar{d})$	as for (1,2)	$\pi_2^2\theta_{22}$	$\pi_2\pi_3\theta_{23}$	$\pi_2(1 - \sigma_2)$	π_2
$(\bar{a}, d)$	as for (1,3)	as for (2,3)	$\pi_3^2\theta_{33}$	$\pi_3(1 - \sigma_3)$	π_3
$(\bar{a}, \bar{d})$	as for (1,4)	as for (2,4)	as for (3,4)	σ	π_4
Sum	π_1	π_2	π_3	π_4	1

The entries in the fourth row and fourth column of Table 10 are calculated by subtraction to make the sums fit the marginal probabilities (11). Under the restrictions (11) and (12), Table 10 is saturated by these nine parameters; three parameters, ν_a, ν_d, ξ_{ad} regulate the marginal probabilities of the latent classes and six parameters, $\theta_{hk}, h, k = 1, 2, 3$, regulate the association between the latent classes of parents and their offspring. The former three parameters are common for monozygotic and dizygotic twins, but the later six parameters are identifiable only for dizygotic twins.

Monozygotic twins are conceived by one sperm fertilizing one egg, which splits in two identical cells, from which two genetically identical individuals develop, while dizygotic twins are conceived by two different sperms fertilizing two different eggs, from which two genetically nonidentical individuals develop. The translation to the language of probability calculus is that two dizygotic twins represent two independent draws from the pool of genetic dispositions that their parents provide. We formulate the genetic origin of dizygotic twins as follows.

Assumption 6: The indicators of the genetic make up of two dizygotic twins are conditionally independent given the parents genetic make up,

$$Z_1 \perp\!\!\!\perp Z_2 | P. \tag{16}$$

It follows from Assumption 6 that, for $h, k = 1, \ldots, 4$,

$$\text{pr}(Z_1 = h, Z_2 = k) = \sum_{j=1}^{4} \pi_j \text{pr}(Z_1 = h | P = j) \text{pr}(Z_2 = k | P = j). \tag{17}$$

Exploiting (17) the joint class probabilities for dizygotic twins are readily calculated from Table 10. These probabilities are listed in Table 11, which is to be compared with Table 9. For brevity the following notation is used, for $i = 1, 2$,

$$
\begin{aligned}
\lambda_{k|j} &= \text{pr}(Z_i = k | P = j) = \pi_k \theta_{jk}, \quad j, k = 1, 2, 3, \tag{18} \\
\lambda_{4|j} &= \text{pr}(Z_i = 4 | P = j) = 1 - \sigma_j, \quad j = 1, 2, 3, \\
\lambda_{j|4} &= \text{pr}(Z_i = j | P = 4) = \frac{\pi_j (1 - \sigma_j)}{\pi_4}, \quad j = 1, 2, 3, \quad \text{and} \quad \lambda_{4|4} = \frac{\sigma}{\pi_4}.
\end{aligned}
$$

Table 11

The probabilities that a pair of dizygotic twins belongs to one of the 16
different latent classes introduced in Table 8.

DZ1 is in class	DZ2 is in class				
	(a,d) $Z_2 = 1$	$(a,\bar{d})$ $Z_2 = 2$	$(\bar{a},d)$ $Z_2 = 3$	$(\bar{a},\bar{d})$ $Z_2 = 4$	Sum
(a,d) $Z_1 = 1$	$\sum_{j=1}^4 \pi_j \lambda_{1|j}^2$	$\sum_{j=1}^4 \pi_j \lambda_{1|j}\lambda_{2|j}$	$\sum_{j=1}^4 \pi_j \lambda_{1|j}\lambda_{3|j}$	$\sum_{j=1}^4 \pi_j \lambda_{1|j}\lambda_{4|j}$	π_1
$(a,\bar{d})$ $Z_1 = 2$	as for $(1,2)$	$\sum_{j=1}^4 \pi_j \lambda_{2|j}^2$	$\sum_{j=1}^4 \pi_j \lambda_{2|j}\lambda_{3|j}$	$\sum_{j=1}^4 \pi_j \lambda_{2|j}\lambda_{4|j}$	π_2
$(\bar{a},d)$ $Z_1 = 3$	as for $(1,3)$	as for $(2,3)$	$\sum_{j=1}^4 \pi_j \lambda_{3|j}^2$	$\sum_{j=1}^4 \pi_j \lambda_{3|j}\lambda_{4|j}$	π_3
$(\bar{a},\bar{d})$ $Z_1 = 4$	as for $(1,4)$	as for $(2,4)$	as for $(3,4)$	$\sum_{j=1}^4 \pi_j \lambda_{4|j}^2$	π_4
Sum	π_1	π_2	π_3	π_4	1

Three important special cases of Table 11 are:
(1) Collapse to the monozygotic case. Consider two binary variables, $Y_i = 1/0$, for $i = 1, 2$, with $\mathrm{E}(Y_i) = \mu_i$, where $0 \leq \mu_i \leq 1$ and dependence ratio θ defined by $\mathrm{pr}(Y_1 = Y_2 = 1) = \theta \mu_1 \mu_2$. The lower and upper bounds for the dependence ratio are,

$$\max(0, \mu_1^{-1} + \mu_2^{-1} - \mu_1^{-1}\mu_2^{-1}) \leq \theta \leq \min(\mu_1^{-1}, \mu_2^{-1}). \qquad (19)$$

For a plot, see Ekholm (2003, p. 17). Suppose that for the dizygotic twins in Table 11, $\pi_k < 1/2$, for $k = 1, 2, 3$. Recall from Table 7 that this condition is satisfied for the monozygotic twins studied in Section 3. It follows from the bounds (19) that the maxima of the intra-class and the minima of the cross-class dependence ratios are, for $h, k = 1, 2, 3, h \neq k$,

$$\max(\theta_{kk}) = \pi_k^{-1}, \quad \min(\theta_{hk}) = 0. \qquad (20)$$

When, simultaneously, the intra-class dependence ratios reach their upper bound and the cross-class dependence ratios their lower bound, then Table 11 collapses to Table 9. Accordingly, the parametrization, introduced here for the dizygotic case, has as its extreme the monozygotic case.

For comparing the heritability of the three dispositions, (a, d), $(a, \bar{d})$ and $(\bar{a}, d)$ with each other, we propose the three normalized intra-class dependence ratios λ_{kk}, for $k = 1, 2, 3$ and $i = 1, 2$,

$$\lambda_{kk} = \theta_{kk}/\max(\theta_{kk}) = \pi_k \theta_{kk} = \mathrm{pr}(Z_i = k | P = k). \qquad (21)$$

Note that the normalized dependence ratio, λ_{kk}, is also the conditional probability that a child belongs to class k, given that his/her parents belong to

k, for $k = 1, 2, 3$. As a summary measure of the similarity of twins for the genetic complex, composed of the four latent classes (a, d), $(a, \bar{d})$, $(\bar{a}, d)$ and $(\bar{a}, \bar{d})$, we propose to use Cohen's kappa, κ, (Agresti, 2002, p. 434), which measures the excess of similarity over that expected by chance. The definition, in terms of the indicators Z_i, for $i = 1, 2$, is

$$\kappa = \frac{\text{pr}(Z_1 = Z_2) - \text{pr}(Z_1 = Z_2 | Z_1 \perp\!\!\!\perp Z_2)}{1 - \text{pr}(Z_1 = Z_2 | Z_1 \perp\!\!\!\perp Z_2)}, \tag{22}$$

where $\text{pr}(Z_1 = Z_2) = \sum_{k=1}^{4} \sum_{j=1}^{4} \pi_j \lambda_{k|j}^2$ and $\text{pr}(Z_1 = Z_2 | Z_1 \perp\!\!\!\perp Z_2) = \sum_{j=1}^{4} \pi_j^2$. For twins, or more generally siblings, one would expect $\theta_{kk} \geq 1$, for $k = 1, 2, 3$, whence the range of κ is $0 \leq \kappa \leq 1$. If and only if $\theta_{kk} = 1/\pi_k$ and $\theta_{hk} = 0$, for $h, k = 1, 2, 3, h \neq k$, then $\kappa = 1$ and the twins are identical with respect to the four latent classes.

(2) *Complete independence.* If all six dependence ratios of Table 10 are equal to one, then Table 11 collapses to the independence case, $\kappa = 0$ and no heritability of the latent dispositions is detectable. One would expect this model to be adequate only for the case of children born to different pairs of parents, but of equal age and reared together.

(3) *Positive association intra-class and independence cross-class.* The saturated model presented in Table 10 adds six parameters to the model, used for empirical analysis of monozygotic twins, which has six parameters. The total number of parameters of the model, we propose for analysis of data on bivariate binary responses from both monozygotic and dizygotic twins, is thereby 12. On the other hand the combined data would have 32 observed frequencies, so that 12 parameters is not too close to saturation. A way of imposing restrictions on the θ-parameters, which has a touch of genetic motivation, is to put the cross-class dependence ratios to one, but retaining the intra-class dependence ratios as parameters to be estimated. Moreover, if the three intra-class dependence ratios can, without loss of fit, be restricted to be equal, then only one new parameter is added to the six parameters introduced for monozygotic twins.

As a numerical example of the relation between θ and κ, suppose that the marginal probabilities are as in Table 7, $\pi_1 = 0.172$, $\pi_2 = 0.1$, $\pi_3 = 0.043$ and $\pi_4 = 0.685$, and $\theta_{hk} = 1$, for $h, k = 1, 2, 3, h \neq k$. Suppose futher that, for $k = 1, 2, 3$, $\theta_{kk} = \theta$, then κ is of the form $c(\pi_1, \pi_2, \pi_3)(\theta - 1)^2$, where the expression for $c(\cdot)$ is complicated and of little interest. For the values $\theta = 1, 2, 3, 4, 5$, respectively, $\kappa = 0$, 0.033, 0.133, 0.300, 0.533.

21

6. Conclusion

To conclude and tie up with the empirical analysis in Section 3, we propose that a statistical analysis of a data set on alcohol dependence and depression, comprising both monozygotic and dizygotic twins, proceeds by fitting a model with, at most, the following 12 parameters:

(i) $\nu_a, \nu_d, \xi_{ad}, \alpha, \delta, \psi$ having the same definition for both sets of twin data,

(ii) $\theta_{hk}, h, k = 1, 2, 3$, being defined only for the data set of dizygotic twins.

Assumption 3 in Section 2 of conditional independence between twins, given the pairs genetic structure G, is upheld for dizygotic twins too, but with the crucial modification that for dizygotic twins G consists of 16 classes, with, at most ten different probabilities, as listed in Table 11.

REFERENCES

Agresti A (2002) Categorical Data Analysis, 2nd ed. Chichester: Wiley.

Barndorff-Nielsen OE and Cox DR (1994) Inference and Asymptotics. London: Chapman & Hall.

Drton M and Richardson TS (2005) Binary models for marginal independence. *Technical Report No. 474, Dept. of Statistics, Univ. of Washington.*

Ekholm A (2003) Comparing the odds and the dependence ratios. In *Statistics, Econometrics and Society. Essays in Honour of Leif Nordberg.* R. Höglund, M. Jäntti, and G. Rosenqvist (eds), 13-25. Helsinki: Statistics Finland.

Ekholm A, Smith PWF and McDonald JW (1995) Marginal regression analysis of a multivariate binary response. *Biometrika* **82,** 847-854.

Ekholm A, Jokinen J, McDonald JW and Smith PWF (2003) Joint regression and association modeling of longitudinal ordinal data. *Biometrics* **59,** 795-803.

Ihaka R and Gentleman R (1996) R: a language for data analysis and graphics. *Journal of Computational Graphics and Statistics* **5,** 299-314.

Kendler KS, Neale MC, Kessler RC, Heath AC and Eaves LJ (1992) A population-based twin study of major depression in women. *Archives of Genetic Psychiatry* **49,** 257-266.