

Chapter 2

The quadratic exponential family

Various generalized estimating equations of order 2 (GEE2) to simultaneously estimate the mean and the association structure can be obtained from the pseudo maximum likelihood 2 (PML2) method. PML2 estimation has been introduced by Gourieroux et al. (1984b), and it is based on the quadratic exponential family. In the first section of this chapter, the quadratic exponential family is introduced with a vectorized form of the association parameters. This representation allows the properties of the linear exponential family to transfer to the quadratic exponential family. It also permits a simple derivation and formulation of the GEE2 (Chapt. 7). Selected properties of the quadratic exponential family are derived in Sect. 2.2. Examples illustrate the applicability of the quadratic exponential family (Sect. 2.3). In the final section of this chapter, we discuss the formulation of the joint distribution of dichotomous variables because of its complexity and its importance for applications.

2.1 Definition

Several GEE2 can be derived from PML2 estimation, and the latter is based on the quadratic exponential family. The important difference between the definitions of the linear exponential family (Chapt. 1) and the quadratic exponential family is that some nuisance parameters Ψ are used in the definition of the linear exponential family, while the covariance matrix Σ is used for the quadratic exponential family.

Definition 2.1 (Quadratic exponential family). Let $\mathbf{y} \in \mathbb{R}^T$ be a random vector and $\mathbf{w} = (y_1^2, y_1 y_2, \dots, y_1 y_T, y_2^2, y_2 y_3, \dots, y_T^2)'$, let $\boldsymbol{\mu} \in \Delta \subset \mathbb{R}^T$ be the corresponding mean vector, and let Σ be the respective positive definite $T \times T$ covariance matrix. Furthermore, let $a : \mathbb{R}^T \times \mathbb{R}^{T \times T} \rightarrow \mathbb{R}$, $b : \mathbb{R}^T \rightarrow \mathbb{R}$, $\mathbf{c} : \mathbb{R}^T \times \mathbb{R}^{T \times T} \rightarrow \mathbb{R}^T$, and $\mathbf{j} : \mathbb{R}^T \times \mathbb{R}^{T \times T} \rightarrow \mathbb{R}^{T(T+1)/2}$

be (measurable) functions. The T -dimensional quadratic exponential family with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is given by the set of distributions with density functions

$$f(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \exp\left(\mathbf{c}(\boldsymbol{\mu}, \boldsymbol{\Sigma})'\mathbf{y} + a(\boldsymbol{\mu}, \boldsymbol{\Sigma}) + b(\mathbf{y}) + \mathbf{j}(\boldsymbol{\mu}, \boldsymbol{\Sigma})'\mathbf{w}\right). \quad (2.1)$$

By letting $\boldsymbol{\vartheta} = \mathbf{c}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{\lambda} = \mathbf{j}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ as in Sect. 1.3, this density may be rewritten as

$$f(\mathbf{y}|\boldsymbol{\vartheta}, \boldsymbol{\lambda}) = \exp\left(\boldsymbol{\vartheta}'\mathbf{y} - d(\boldsymbol{\vartheta}, \boldsymbol{\lambda}) + b(\mathbf{y}) + \boldsymbol{\lambda}'\mathbf{w}\right). \quad (2.2)$$

Remark 2.2.

- The representation of the quadratic exponential family in Eq. 2.2 does not immediately open up the term quadratic. However, the function $\mathbf{j}(\boldsymbol{\mu}, \boldsymbol{\Sigma})'\mathbf{w}$ can also be represented by the quadratic form $\mathbf{y}'\mathbf{D}\mathbf{y}$ for a symmetric matrix $\mathbf{D}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ because

$$\mathbf{y}'\mathbf{D}\mathbf{y} = \sum_{t=1}^T \sum_{t'=1}^T y_t y_{t'} [\mathbf{D}]_{tt'} = \sum_{t=1}^T y_t^2 [\mathbf{D}]_{tt} + 2 \sum_{t'>t} y_t y_{t'} [\mathbf{D}]_{tt'} = \mathbf{j}'\mathbf{w},$$

and $\mathbf{j} = ([\mathbf{D}]_{11}, 2[\mathbf{D}]_{12}, \dots, 2[\mathbf{D}]_{1T}, [\mathbf{D}]_{22}, 2[\mathbf{D}]_{23}, \dots, [\mathbf{D}]_{TT})'$. Therefore, the exponential family is quadratic because the exponent can be formulated quadratic in $\mathbf{y}$ with coefficient matrix $\mathbf{D}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. This quadratic form of the quadratic exponential family has been used by Gourieroux et al. (1984b). However, the vectorized version is more convenient for deriving some properties of the quadratic exponential family (see next section).

- In contrast to the linear exponential family (Definition 1.1), where second-order moments were treated as nuisance, these are of interest in the quadratic exponential family. Moments of order three and higher are, however, ignored and set to $\mathbf{0}$. This is important for the definition of the joint distribution of a T -variate dichotomous random vector (Sect. 2.4). If readers are interested in higher order exponential families, they may refer, e.g., to Holly et al. (2008).
- We can define the vector $\boldsymbol{\nu} = (\nu_{11}, \nu_{12}, \dots, \nu_{22}, \dots)'$ in analogy to $\boldsymbol{\lambda}$ with $\nu_{tt'} = \sigma_{tt'} + \mu_t \mu_{t'}$, where $\nu_{tt'} = \mathbb{E}(y_t y_{t'})$ denotes the second ordinary moment, and $\sigma_{tt'} = [\boldsymbol{\Sigma}]_{tt'}$ is the tt' th element of the covariance matrix $\boldsymbol{\Sigma}$. $\boldsymbol{\nu}$ will be used below to formulate GEE using the second ordinary moments, which are also termed product moments.

2.2 Selected properties

By condensing $\boldsymbol{\vartheta}$ and $\boldsymbol{\lambda}$ to one column vector and $\mathbf{y}$ and $\mathbf{w}$ to another column vector, the properties of Theorem 1.2 for the linear exponential family can be extended to the quadratic exponential family. For example, mean and variance of the quadratic exponential family with vectorized association parameter $\boldsymbol{\lambda}$ are given by

$$\begin{aligned} \mathbb{E}((\mathbf{y}', \mathbf{w}')') &= \frac{\partial d(\boldsymbol{\vartheta}, \boldsymbol{\lambda})}{\partial (\boldsymbol{\vartheta}', \boldsymbol{\lambda}')'} = \begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\nu} \end{pmatrix}, \quad \text{and} \\ \mathbb{V}\text{ar}((\mathbf{y}', \mathbf{w}')') &= \begin{pmatrix} \mathbb{V}\text{ar}(\mathbf{y}) & \mathbb{C}\text{ov}(\mathbf{y}, \mathbf{w}) \\ \mathbb{C}\text{ov}(\mathbf{w}, \mathbf{y}) & \mathbb{V}\text{ar}(\mathbf{w}) \end{pmatrix} = \frac{\partial (\boldsymbol{\mu}', \boldsymbol{\nu}')}{\partial (\boldsymbol{\vartheta}', \boldsymbol{\lambda}')'} = \begin{pmatrix} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\vartheta}'} & \frac{\partial \boldsymbol{\nu}}{\partial \boldsymbol{\vartheta}'} \\ \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\lambda}'} & \frac{\partial \boldsymbol{\nu}}{\partial \boldsymbol{\lambda}'} \end{pmatrix}. \end{aligned} \quad (2.3)$$

Similarly, Property 1.6, which has been given for the linear exponential family, can be generalized to the quadratic exponential family.

Property 2.3. For any $\boldsymbol{\mu}, \boldsymbol{\mu}_0 \in \Delta$ and for all positive definite $T \times T$ covariance matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_0$,

$$\begin{aligned} \mathbf{c}(\boldsymbol{\mu}, \boldsymbol{\Sigma})' \boldsymbol{\mu}_0 + a(\boldsymbol{\mu}, \boldsymbol{\Sigma}) + \mathbf{j}(\boldsymbol{\mu}, \boldsymbol{\Sigma})' \boldsymbol{\nu}_0 \\ \leq \mathbf{c}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)' \boldsymbol{\mu}_0 + a(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) + \mathbf{j}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)' \boldsymbol{\nu}_0. \end{aligned} \quad (2.4)$$

Equality holds, if and only if $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$.

Proof. Analogously to Property 1.6, Property 2.3 is a direct consequence of Kullback's inequality. $\square$

2.3 Examples for quadratic exponential families

In this section, we give three examples for quadratic exponential families. We start with the univariate normal distribution because the normal distribution is of great importance for PML2 estimation. It is followed by the multivariate normal distribution, and the last example of this section is one given by Gouriéroux et al. (1984b). In the next section, two further examples are given. The reader should note that some standard distributions that belong to two-parameter linear exponential families (see, e.g., Example 1.11) do not belong to the class of quadratic exponential families. A typical example is the Gamma distribution.

Example 2.4 (Univariate normal distribution). The density of the univariate normal distribution has been given in Example 1.10. If Ψ is replaced by $\Sigma = \sigma^2$, one obtains $b(y) = 0$ as well as

$$c(\mu, \Sigma) = \frac{\mu}{\Sigma}, a(\mu, \Sigma) = -\frac{1}{2} \frac{\mu^2}{\Sigma} - \frac{1}{2} \ln(2\pi\Sigma), \quad \text{and} \quad j(\mu, \Sigma) = -\frac{1}{2} \frac{1}{\Sigma}.$$

Example 2.5 (Multivariate normal distribution). The classical example for a distribution belonging to the quadratic exponential family is the T -dimensional normal distribution. Its density has been given in Example 1.12. By replacing Ψ with Σ , we obtain $b(\mathbf{y}) = 0$,

$$j(\mu, \Sigma) = -\frac{1}{2} ([\Sigma^{-1}]_{11}, 2[\Sigma^{-1}]_{12}, \dots, 2[\Sigma^{-1}]_{1T}, [\Sigma^{-1}]_{22}, \dots, [\Sigma^{-1}]_{TT}),$$

$$a(\mu, \Sigma) = -\frac{1}{2} \mu' \Sigma^{-1} \mu - \frac{T}{2} \ln(2\pi) - \frac{1}{2} \ln(\det(\Sigma)), \quad \text{and} \quad c(\mu, \Sigma) = \Sigma^{-1} \mu.$$

Example 2.6 (Discrete distribution on $\{-1, 0, 1\}$). Consider a discrete distribution on $\{-1, 0, 1\}$ with probabilities p_{-1}, p_0 , and p_1 . Its density is given by

$$\tilde{f}(y|p_{-1}, p_0, p_1) = p_{-1}^{y(y-1)/2} p_0^{(1-y^2)} p_1^{y(1+y)/2},$$

subject to $p_{-1} + p_0 + p_1 = 1$, $p_1 - p_{-1} = \mu$, and $p_{-1} + p_1 = \Sigma + \mu^2$. We obtain

$$f(y|\mu, \Sigma) = \exp \left(\frac{y}{2} \ln \left(\frac{\Sigma + \mu^2 + \mu}{\Sigma + \mu^2 - \mu} \right) + \frac{y^2}{2} \ln \left(\frac{(\Sigma + \mu^2 + \mu)(\Sigma + \mu^2 - \mu)}{4 \cdot (1 - \Sigma - \mu^2)^2} \right) \right. \\ \left. + \ln(1 - \Sigma - \mu^2) \right).$$

We complete the example by noting that $a(\mu, \Sigma) = \ln(1 - \Sigma - \mu^2)$, $b(y) = 0$,
 $c(\mu, \Sigma) = \frac{1}{2} \ln \frac{\Sigma + \mu^2 + \mu}{\Sigma + \mu^2 - \mu}$, and $j(\mu, \Sigma) = \frac{1}{2} \ln \frac{(\Sigma + \mu^2 + \mu)(\Sigma + \mu^2 - \mu)}{4 \cdot (1 - \Sigma - \mu^2)^2}$.

2.4 The joint distribution of dichotomous random variables

Several representations of the joint distribution of T dichotomous random variables have been given in the literature, and their pros and cons have been extensively discussed (Kauermann, 1997; Liang et al., 1992; Prentice, 1988). In this section, we first consider the joint distribution of two dichotomous random variables and show that this distribution is a member of the quadratic exponential family. Furthermore, we consider the joint distribution of T dichotomous random variables and embed a specific version of it into the quadratic exponential family. In this version, all moments of order three and above are set to 0.

2.4.1 The joint distribution of two dichotomous random variables

The joint distribution of two dichotomous items is often used in applications. The standard parameter to measure the association between the responses is the odds ratio (OR). We therefore define the OR first and next give the definition of the joint distribution. Desirable properties and interpretations of the OR and other measures of association for pairs of binary responses have been described in detail, e.g., by Bishop et al. (1975).

The starting point is the 2×2 Table 2.1, which displays a bivariate random vector $\mathbf{y} = (y_1, y_2)'$, where $\pi_{tt'} = \mathbb{P}(y_1 = t, y_2 = t')$. The means are given by $\mu_1 = \pi_1 = \mathbb{P}(y_1 = 1)$ and $\mu_2 = \pi_2 = \mathbb{P}(y_2 = 1)$. For simplicity, we assume that all probabilities are > 0 and < 1 .

Table 2.1 2×2 table for y_1 and y_2

		y_2		
		0	1	
y_1	0	π_{00}	π_{01}	$1 - \pi_1$
	1	π_{10}	π_{11}	π_1
		$1 - \pi_2$	π_2	1

Definition 2.7 (Odds, odds ratio).

1. The odds of column 1 is $O_1 = \frac{\pi_{00}}{\pi_{10}}$, and the odds of column 2 is $O_2 = \frac{\pi_{01}}{\pi_{11}}$.
2. The odds ratio (OR) τ , also termed cross product ratio, is the ratio of the two odds

$$\tau_{12} = \text{OR}_{12} = \text{OR}(y_1, y_2) = \frac{O_1}{O_2} = \frac{\pi_{00} \pi_{11}}{\pi_{01} \pi_{10}}.$$

Definition 2.8 (Joint distribution of two dichotomous random variables). The joint distribution of two dichotomous items y_1 and y_2 is given by

$$\mathbb{P}(\mathbf{y}) = \mathbb{P}(y_1, y_2) = \exp(y_1 y_2 \ln \pi_{11} + y_1(1 - y_2) \ln \pi_{10} + (1 - y_1)y_2 \ln \pi_{01} + (1 - y_1)(1 - y_2) \ln \pi_{00}). \quad (2.5)$$

Remark 2.9. Equation 2.5 can be rewritten as

$$\mathbb{P}(\mathbf{y}) = \exp(y_1 y_2 [\ln(\pi_{11} \pi_{00}) - \ln(\pi_{10} \pi_{01})] + y_1 [\ln \pi_{10} - \ln \pi_{00}] + y_2 [\ln \pi_{01} - \ln \pi_{00}] + \ln \pi_{00}),$$

and it can therefore be seen that the joint distribution of two dichotomous random variables belongs to the quadratic exponential family. The natural parameters have a loglinear representation and are given by

$$\begin{aligned}\vartheta_1 &= \text{logit}\{\mathbb{P}(y_1 = 1|y_2 = 0)\} = \ln \pi_{10} - \ln \pi_{00}, \\ \vartheta_2 &= \text{logit}\{\mathbb{P}(y_2 = 1|y_1 = 0)\} = \ln \pi_{01} - \ln \pi_{00}, \quad \text{and} \\ \lambda_{12} &= \log \text{OR}(y_1, y_2) = \ln(\pi_{11}\pi_{00}) - \ln(\pi_{10}\pi_{01}).\end{aligned}\tag{2.6}$$

The natural parameters ϑ_t are conditional probabilities, and λ_{12} is an unconditional log OR. The normalization constant is $d(\boldsymbol{\vartheta}, \boldsymbol{\lambda}) = -\ln \pi_{00}$, and $b(\mathbf{y}) = 0$. The parameters λ_{11} and λ_{22} of y_1^2 and y_2^2 are identical to 0 because the variances are completely specified by the means.

Finally, in Eq. 2.6 the joint distribution is formulated using conditional logits. In the next section, a representation in the marginal moments $(\pi_1, \pi_2, \lambda_{12})'$ will be considered.

To derive GEE for dichotomous items using the quadratic exponential family, we require the following functional relationship between the second ordinary moments $\pi_{11} = \mathbb{E}(y_1 y_2)$ and the OR τ_{12} in Chapt. 7. This functional relationship has been given, e.g., by Bishop et al. (1975):

$$\mathbb{E}(y_1 y_2) = \pi_{11} = \begin{cases} \frac{f_{12} - \sqrt{f_{12}^2 - 4\tau_{12}(\tau_{12} - 1)\pi_1\pi_2}}{2(\tau_{12} - 1)} & \text{if } \tau_{12} \neq 1, \\ \tau_{12}\pi_1\pi_2 & \text{if } \tau_{12} = 1, \end{cases}\tag{2.7}$$

where $f_{12} = (1 - (1 - \tau_{12})(\pi_1 + \pi_2))$. By using

$$\tau_{12} = \frac{\pi_{11}(1 - \pi_1 - \pi_2 + \pi_{11})}{(\pi_1 - \pi_{11})(\pi_2 - \pi_{11})} \quad \text{and} \quad \varrho = \frac{\pi_{11} - \pi_1\pi_2}{\sqrt{\pi_1(1 - \pi_1)\pi_2(1 - \pi_2)}},$$

a one-to-one functional relation between the OR τ and the correlation coefficient $\varrho = \text{Corr}(y_1, y_2)$ can also be established. As a result, the OR can therefore also be written as a function of the means and the correlation coefficient, and the correlation coefficient can be written as a function of the means and the OR.

2.4.2 The joint distribution of T dichotomous random variables

The joint distribution of two binary random variables can be extended easily to T dichotomous random variables. Under the restrictive assumptions that all third and higher order moments equal 0, this distribution also belongs to

the quadratic exponential family. Three different parameterizations are often used in practice for this model, the loglinear parameterization, the marginal parameterization using contrasts of odds ratios, and the marginal parameterization using the correlation coefficient as the measure of association.

Definition 2.10. Consider T dichotomous random variables $y_1, \dots, y_T$, which are summarized to a vector $\mathbf{y}$. The joint distribution in the loglinear parameterization is given by

$$\begin{aligned} \mathbb{P}(\mathbf{y}) = \exp \bigg(& \sum_{t=1}^T y_t \vartheta_t + \sum_{t < t'} y_t y_{t'} \lambda_{tt'} \\ & + \sum_{t < t' < t''} y_t y_{t'} y_{t''} \zeta_{tt't''} + \dots + y_1 y_2 \dots y_T \zeta_{1\dots T} - d(\boldsymbol{\vartheta}, \boldsymbol{\lambda}, \boldsymbol{\zeta}) \bigg), \end{aligned} \quad (2.8)$$

for $\boldsymbol{\zeta}' = (\zeta_{tt't''}, \dots, \zeta_{1\dots T})$ with ϑ_t being logits of conditional probabilities

$$\vartheta_t = \text{logit}\{\mathbb{P}(y_t = 1 | y_{t'} = 0, t \neq t')\} = \ln \frac{\mathbb{P}(y_t = 1 | y_{t'} = 0, t \neq t')}{\mathbb{P}(y_t = 0 | y_{t'} = 0, t \neq t')}.$$

The second-order moments are conditional log ORs

$$\begin{aligned} \lambda_{tt'} &= \log \text{OR}(y_t, y_{t'} | y_{t''} = 0) \\ &= \ln \frac{\mathbb{P}(y_t = 1, y_{t'} = 1 | y_{t''} = 0)}{\mathbb{P}(y_t = 0, y_{t'} = 1 | y_{t''} = 0)} - \ln \frac{\mathbb{P}(y_t = 1, y_{t'} = 0 | y_{t''} = 0)}{\mathbb{P}(y_t = 0, y_{t'} = 0 | y_{t''} = 0)} \end{aligned}$$

with $t'' \neq t, t'$.

Remark 2.11.

- In Definition 2.10, second-order moments are conditional log ORs, while they were unconditional in Eq. 2.6.
- If all parameters of order three and above are set to 0, thus $\boldsymbol{\zeta} = \mathbf{0}$, the joint distribution of T binary random variables belongs to the quadratic exponential family.

An alternative parameterization of the joint distribution of T dichotomous items is in terms of marginal rather than fully conditional distributions. In applications, the first two moments are of primary interest, and these are initially specified by

$$\mathbb{P}(y_t = 1) = \mu_t, \quad \text{and} \quad \text{OR}(y_t, y_{t'}) = \tau_{tt'}. \quad (2.9)$$

The parameterization of the full joint distribution can be completed in several ways, e.g., in terms of contrasts of conditional ORs (Liang et al., 1992):

$$\begin{aligned}
\zeta_{tt't''} &= \ln \text{OR}(y_t, y_{t'} | y_{t''} = 1) - \ln \text{OR}(y_t, y_{t'} | y_{t''} = 0), \\
\zeta_{rr'tt'} &= \ln \text{OR}(y_t, y_{t'} | y_r = 1, y_{r'} = 1) - \ln \text{OR}(y_t, y_{t'} | y_r = 1, y_{r'} = 0) \\
&\quad - \ln \text{OR}(y_t, y_{t'} | y_r = 0, y_{r'} = 1) + \ln \text{OR}(y_t, y_{t'} | y_r = 0, y_{r'} = 0), \\
\zeta_{t_1 \dots t_S} &= \sum (-1)^{\sum_{s=3}^S y_{t_s} + S-2} \ln \text{OR}(y_{t_1}, y_{t_2} | y_{t_3}, \dots, y_{t_S}).
\end{aligned} \tag{2.10}$$

Here, the sum is taken over the 2^{S-2} possible combinations of $(y_{t_3}, \dots, y_{t_S})$. Another completion of the joint distribution uses the following higher order moments:

$$\zeta_{tt't''} = \ln \frac{\pi_{111} \pi_{000} \pi_{100} \pi_{010}}{\pi_{101} \pi_{011} \pi_{110} \pi_{000}}, \tag{2.11}$$

$$\zeta_{rr'tt'} = \ln \frac{\pi_{1111} \pi_{0011} \pi_{1100} \pi_{0000} \pi_{1010} \pi_{0110} \pi_{1001} \pi_{0101}}{\pi_{1110} \pi_{1101} \pi_{1011} \pi_{0111} \pi_{1000} \pi_{0100} \pi_{0010} \pi_{0001}}, \tag{2.12}$$

$$\begin{aligned}
\varsigma_{t_1 \dots t_S} &= (-1)^S \ln \prod_{\{(t_1, \dots, t_S) : \sum_{s=1}^S t_s = 2n, n \in \mathbb{N}\}} \pi_{t_1 \dots t_S} + \\
&\quad (-1)^{S+1} \ln \prod_{\{(t_1, \dots, t_S) : \sum_{s=1}^S t_s = 2n+1, n \in \mathbb{N}\}} \pi_{t_1 \dots t_S}.
\end{aligned} \tag{2.13}$$

Kauermann (1997) has shown that the loglinear and the marginal parameterization have a 1:1 correspondence, although higher order marginal and log-linear parameters differ. The most important difference between the loglinear and marginal parameters is in the interpretation of the first two moments. In the loglinear parameterization, they can be interpreted as conditional probabilities of y_t given $y_{t'}$, or as conditional log ORs of $y_t, y_{t'}$ given $y_{t''}$.

The pros and cons of both parameterizations have been discussed by Liang et al. (1992). The major advantage of the marginal model over the loglinear model is its reproducibility. Thus, if $\mathbf{y}$ satisfies a marginal model, then any subset of $\mathbf{y}$ also does. Hence, the interpretation of the marginal parameters is independent of the length T of $\mathbf{y}$. One drawback of the marginal parameterization is that extensive computations are required for parameter estimation, see, e.g., Fitzmaurice and Laird (1993). Even more important, the parameter space is restricted (see next section), and this restriction has immediate consequences for choosing the weight matrices for GEEs with dichotomous dependent variables (see Sect. 7.3.4).

Another formulation of the joint distribution is based on correlations. Specifically, Bahadur (1961) has shown that the joint distribution of T dichotomous random variables can be written as

$$\begin{aligned}
\mathbb{P}(y_1, \dots, y_T) &= \prod_{t=1}^T \left(\pi_t^{y_t} (1 - \pi_t)^{1-y_t} \right) \left(1 + \sum_{t < t'} \varrho_{tt'} z_t z_{t'} \right. \\
&\quad \left. + \sum_{t < t' < t''} \varrho_{tt't''} z_t z_{t'} z_{t''} + \dots + \varrho_{1\dots T} z_1 z_2 \dots z_T \right),
\end{aligned} \tag{2.14}$$

where $z_i = (y_i - \mu_i)/\sigma_i$ is the standardized variable with $\sigma_i = \sqrt{\mu_i(1 - \mu_i)}$, $\varrho_{tt'} = \text{Corr}(y_t, y_{t'}) = \mathbb{E}(z_t z_{t'})$, $\varrho_{tt't''} = \mathbb{E}(z_t z_{t'} z_{t''})$, etc. The Bahadur representation is similar to the representation in the marginal OR, and the corresponding parameter space is restricted, too.

2.4.3 Restriction of the parameter space in marginal models

When using the parameterization in the correlation coefficient, the parameter space is restricted for $T \geq 2$. Consider the 2×2 Table 2.1 with means $\pi_1 = \pi_{10} + \pi_{11}$ and $\pi_2 = \pi_{01} + \pi_{11}$. Then, the probability π_{11} is restricted to

$$\max(0, \pi_1 + \pi_2 - 1) \leq \pi_{11} \leq \min(\pi_1, \pi_2).$$

Because the correlation coefficient

$$\varrho = \frac{\pi_{11} - \pi_1 \pi_2}{\sqrt{\pi_1(1 - \pi_1) \cdot \pi_2(1 - \pi_2)}}$$

directly depends on the restricted π_{11} , ϱ is also restricted. Specifically, ϱ is constrained by (Prentice, 1988)

$$\max \left\{ -\sqrt{\frac{\pi_1 \pi_2}{\varpi_1 \varpi_2}}, -\sqrt{\frac{\varpi_1 \varpi_2}{\pi_1 \pi_2}} \right\} \leq \varrho \leq \min \left\{ \sqrt{\frac{\pi_1 \varpi_2}{\varpi_1 \pi_2}}, \sqrt{\frac{\varpi_1 \pi_2}{\pi_1 \varpi_2}} \right\}, \quad (2.15)$$

where $\varpi_t = 1 - \pi_t$. Similar restrictions hold for moments $\varrho_{t_1 \dots t_s}$ of order three and above.

Examples for the restrictions are given in Table 2.2. Extreme restrictions occur when one probability is approaching the boundary of the parameter space, while the other is approximately 0.5. For example, if $\pi_1 = 0.01$ and $\pi_2 = 0.5$, ϱ is bounded by ± 0.1 . It is bounded by ± 0.03 if π_1 is 0.001.

Table 2.2 Restrictions of the correlation coefficient in a 2×2 setting given the marginal means π_1 and π_2

		π_2		
		0.1	0.3	0.5
π_1	0.1	-0.11;1.00	-0.22;0.51	-0.33;0.33
	0.3		-0.43;1.00	-0.65;0.65
	0.5			-1.00;1.00

Marginal moments other than the correlation coefficient are also restricted. Specifically, the OR is restricted for binary dependent variables if $T \geq 3$. For example, let the parameters $(\pi_1, \pi_2, \pi_3, \tau_{12}, \tau_{13})$ be fixed, and $T = 3$. $\mathbb{E}(y_1 y_2)$

and $\mathbb{E}(y_1 y_3)$ can be determined through these parameters. For example, the domain of $\mathbb{E}(y_2 y_3)$ is restricted to

$$\max \{0, \mathbb{E}(y_1 y_2) + \mathbb{E}(y_1 y_3) - \pi_1\} \leq \mathbb{E}(y_2 y_3) \leq \min \{\pi_2, \pi_3\}. \quad (2.16)$$

As a result, the domain of τ_{23} is also restricted because

$$\tau_{23} = \frac{\mathbb{E}(y_2 y_3)(1 - \pi_2 - \pi_3 + \mathbb{E}(y_2 y_3))}{(\pi_2 - \mathbb{E}(y_2 y_3))(\pi_3 - \mathbb{E}(y_2 y_3))}.$$

Analogous restrictions exist for moments of order greater than three.

In summary, the OR parameterization can be used without any constraints for two dichotomous random variables, while the parameter space of the correlation coefficient is already restricted in this case. If more than two dichotomous random variables are considered, the parameter space is restricted in all marginal parameterizations. For greater flexibility in the parameter space, higher order moments should be added to analyze correlated dichotomous random variables.

As a last example for the restriction of the parameter space, consider the joint distribution of T binary responses $\mathbf{y} = (y_1, \dots, y_T)'$ in the Bahadur representation (Eq. 2.14) with all moments $\varrho_{t_1 \dots t_s}$ of three and above being set to 0:

$$\mathbb{P}(\mathbf{y}) = \prod_{t=1}^T \left(\pi_t^{y_t} (1 - \pi_t)^{1-y_t} \right) \left(1 + \sum_{t < t'} \varrho_{tt'} \frac{(y_t - \pi_t)}{\sqrt{\pi_t(1 - \pi_t)}} \frac{(y_{t'} - \pi_{t'})}{\sqrt{\pi_{t'}(1 - \pi_{t'})}} \right).$$

For simplicity, let $\pi_t = \pi > 0$, $\varrho_{tt'} = \varrho$, $T = 2k$, $k \in \mathbb{N}$. Then, a sequence of m zeros and m ones has probability (Prentice, 1988)

$$\pi^m (1 - \pi)^m \left(1 + \varrho \left(\frac{1}{2} m(m-1) \frac{1-\pi}{\pi} + \frac{1}{2} m(m-1) \frac{\pi}{1-\pi} - m^2 \right) \right),$$

yielding $\varrho \leq |m^2 - \frac{1}{2} m(m-1) \frac{1-\pi}{\pi} - \frac{1}{2} m(m-1) \frac{\pi}{1-\pi}|^{-1}$. For example, $\varrho \leq |\frac{1}{m}|$ if $\pi = \frac{1}{2}$. Thus, the impact of the restrictions increases with the number of observations T per cluster.

A summary of the permissible ranges of dependent dichotomous variables has been given by Chaganty and Deng (2007). The effect of ignoring the bounds has been nicely illustrated by Sabo and Chaganty (2010). A detailed discussion how these restrictions may be overcome can be found in Ziegler and Vens (2010); also see Shults (2011).



<http://www.springer.com/978-1-4614-0498-9>

Generalized Estimating Equations

Ziegler, A.

2011, XV, 144 p., Softcover

ISBN: 978-1-4614-0498-9