Chapter 2
Overview of Regression Models for Cross-Sectional Univariate Categorical Data

2.1 Covariate Free Basic Univariate Multinomial Fixed Effect Models

Let there be K individuals and an individual responds to one of the J categories. For $j = 1,\ldots,J$, let π_j denote the marginal probability that the response of an individual belongs to the jth category so that $\sum_{j=1}^{J} \pi_j = 1$. Suppose that $y_i = [y_{i1},\ldots,y_{ij},\ldots,y_{i,J-1}]'$ denotes the $J-1$ dimensional multinomial response variable of the ith ($i = 1,\ldots,K$) individual such that $y_{ij} = 1$ or 0, with $\sum_{j=1}^{J} y_{ij} = 1$. Further suppose that for a q-dimensional unit vector 1_q, for example,

$$y_i^{(j)} = \delta_{ij} = [01'_{j-1}, 1, 01'_{J-1-j}]'$$

denotes the response of the ith individual that belongs to the jth category for $j = 1,\ldots,J-1$, and

$$y_i^{(J)} = \delta_{iJ} = 01_{J-1}$$

denotes that the response of the ith individual belongs to the Jth category which may be referred to as the reference category. It then follows that

$$P[y_i = y_i^{(j)} = \delta_{ij}] = \pi_j, \text{for all } \ j = 1,\ldots,J. \tag{2.1}$$

For convenience of generalization to the covariate case, we consider

$$\pi_j = \begin{cases} \frac{\exp(\beta_{j0})}{1+\sum_{g=1}^{J-1}\exp(\beta_{g0})} & \text{for } j = 1,\ldots,J-1 \\ \frac{1}{1+\sum_{g=1}^{J-1}\exp(\beta_{g0})} & \text{for } j = J, \end{cases} \tag{2.2}$$

B.C. Sutradhar, *Longitudinal Categorical Data Analysis*, Springer Series in Statistics, DOI 10.1007/978-1-4939-2137-9_2

It then follows that the elements of y_i follow the multinomial distribution given by

$$P[y_{i1},\ldots,y_{ij},\ldots,y_{i,J-1}] = \frac{1!}{y_{i1}!\cdots y_{ij}!\cdots y_{iJ}!}\Pi_{j=1}^{J}\pi_j^{y_{ij}}, \tag{2.3}$$

where $y_{iJ} = 1-\sum_{j=1}^{J-1} y_{ij}$. Now suppose that out of these K independent individuals, $K_j = \sum_{i=1}^{K} y_{ij}$ individuals belong to the jth category for $j = 1,\ldots,J$, so that $\sum_{j=1}^{J} K_j = K$. By an argument similar to that of (2.3), one may write the joint distribution for $\{K_j\}$ with $K_J = K - \sum_{j=1}^{J-1} K_j$, that is, the multinomial distribution for $\{K_j\}$ as

$$\begin{aligned} P[K_1,\, K_2,\ldots,K_j,\, \ldots,K_{J-1}] &= \frac{K!}{K_1!\cdots K_j!\cdots K_J!}\Pi_{j=1}^{J}\pi_j^{\sum_{i=1}^{K} y_{ij}} \\ &= \frac{K!}{K_1!\cdots K_J!}\Pi_{j=1}^{J}\pi_j^{K_j}. \end{aligned} \tag{2.4}$$

In the next section, we provide some basic properties of this multinomial distribution. Inference for the multinomial probabilities through the estimation of the parameters $\beta_{j0}(j = 1,\ldots,J-1)$, along with an example, is discussed in Sect. 2.1.2.

A derivation of the multinomial distribution (2.4):
Suppose that

$$K_j \sim Poi(\mu_j),\ j = 1,\ldots,J,$$

where $Poi(\mu_j)$ denotes the Poisson distribution with mean μ_j, that is,

$$P(K_j|\mu_j) = \frac{\exp(-\mu_j)\mu_j^{K_j}}{K_j!},\ K_j = 0,1,2,\ldots$$

Also suppose that K_j's are independent for all $j = 1,\ldots,J$ It then follows that

$$K = \sum_{j=1}^{J} K_j \sim Poi(\mu = \sum_{j=1}^{J}\mu_j),$$

and conditional on total K, the joint distribution of the counts $K_1,\ldots,K_j,\ldots,K_{J-1}$, has the form

$$P[K_1,\ldots,K_j,\ldots,K_{J-1}|K] = \frac{\Pi_{j=1}^{J}\left[\frac{\exp(-\mu_j)\mu_j^{K_j}}{K_j!}\right]}{\frac{\exp(-\mu)\mu^K}{K!}},$$

where now $K_J = K - \sum_{j=1}^{J-1} K_j$ is known. Now by using $\pi_j = \frac{\mu_j}{\mu}$, one obtains the multinomial distribution (2.4), where $\pi_J = 1 - \sum_{j=1}^{J-1} \pi_j$ is known.

Note that when $K = 1$, one obtains the multinomial distribution (2.3) from (2.4) by using $K_j = y_{ij}$ as a special case.

2.1.1 Basic Properties of the Multinomial Distribution (2.4)

Lemma 2.1.1. *The count variable $K_j (j = 1, \ldots, J-1)$ marginally follows a binomial distribution $B(K_j; K, \pi_j)$, with parameters K and π_j, yielding $E[K_j] = K\pi_j$ and $var[K_j] = K\pi_j(1-\pi_j)$. Furthermore, for $j \neq k, j, k = 1, \ldots, J-1$, $cov[K_j, K_k] = -K\pi_j\pi_k$.*

Proof. Let

$$\xi_1 = \pi_1,\ \xi_2 = [1-\pi_1],\ \xi_3 = [1-\pi_1-\pi_2], \ldots, \xi_{J-1} = [1-\pi_1-\cdots-\pi_{J-2}].$$

By summing over the range of K_{J-1} from 0 to $[K - K_1 - \ldots, K_{J-2}]$, one obtains the marginal multinomial distribution of $K_1, \ldots, K_{J-2}$ from (2.4) as

$$\begin{aligned} P[K_1, \ldots, K_j, \ldots, K_{J-2}] &= \frac{K!}{K_1! \cdots K_j! \cdots \{K - K_1 - \cdots - K_{J-2}\}!} \Pi_{j=1}^{J-2} \pi_j^{K_j} [\xi_{J-1}]^{\{K-K_1-\cdots-K_{J-2}\}} \\ \times \frac{\{K-K_1-\cdots-K_{J-2}\}!}{K_{J-1}!\{K-K_1-\cdots-K_{J-2}-K_{J-1}\}!} &\sum_{K_{J-1}=0}^{K-K_1-\cdots-K_{J-2}} [\frac{\pi_{J-1}}{\xi_{J-1}}]^{K_{J-1}} [1 - \frac{\pi_{J-1}}{\xi_{J-1}}]^{\{(K-K_1-\cdots-K_{J-2})-K_{J-1}\}} \\ = \frac{K!}{K_1! \cdots K_j! \cdots \{K-K_1-\cdots-K_{J-2}\}!} &\Pi_{j=1}^{J-2} \pi_j^{K_j} [\xi_{J-1}]^{\{K-K_1-\cdots-K_{J-2}\}}. \end{aligned} \tag{2.5}$$

By summing, similar to that of (2.5), successively over the range of $K_{J-2}, \ldots, K_2$, one obtains the marginal distribution of K_1 as

$$P[K_1] = \frac{K!}{K_1!\{K-K_1\}!} \pi^{K_1} [1-\pi_1]^{K-K_1}, \tag{2.6}$$

which is a binomial distribution with parameters (K, π_1). Note that this averaging or summing technique to find the marginal distribution is exchangeable. Thus, for any $j = 1, \ldots, J-1$, K_j will have marginally binomial distribution with parameters (K, π_j). This yields the mean and the variance of K_j as in the Lemma.

Next to derive the covariance between K_j and K_k, for convenience we find the covariance between K_1 and K_2. For this computation, following (2.5), we first write the joint distribution of K_1 and K_2 as

$$\begin{aligned} P[K_1, K_2] &= \frac{K!}{K_1!K_2!\{K-K_1-K_2\}!} \Pi_{j=1}^{2} \pi_j^{K_j} [\xi_3]^{\{K-K_1-K_2\}} \\ &= \frac{K!}{K_1!K_2!\{K-K_1-K_2\}!} \Pi_{j=1}^{2} \pi_j^{K_j} [1-\pi_1-\pi_2]^{\{K-K_1-K_2\}}. \end{aligned} \tag{2.7}$$

It then follows that

$$\begin{aligned}
E[K_1K_2] &= \sum_{K_1=0}^{K}\sum_{K_2=0}^{\{K-K_1\}} K_1K_2\frac{K!}{K_1!K_2!\{K-K_1-K_2\}!}\Pi_{j=1}^{2}\pi_j^{K_j}[1-\pi_1-\pi_2]^{\{K-K_1-K_2\}} \\
&= K(K-1)\pi_1\pi_2\sum_{K_1^*=0}^{K-2}\sum_{K_2^*=0}^{\{K-2-K_1^*\}}\frac{\{K-2\}!}{K_1^*!K_2^*!\{K-2-K_1^*-K_2^*\}!} \\
&\quad \times\Pi_{j=1}^{2}\pi_j^{K_j^*}[1-\pi_1-\pi_2]^{\{K-2-K_1^*-K_2^*\}} \\
&= K(K-1)\pi_1\pi_2, \qquad (2.8)
\end{aligned}$$

yielding

$$\text{cov}[K_1,K_2] = E[K_1K_2]-E[K_1]E[K_2] = K(K-1)\pi_1\pi_2-K^2\pi_1\pi_2 = -K\pi_1\pi_2. \qquad (2.9)$$

Now because the multinomial distribution is exchangeable in variables, one obtains $\text{cov}[K_j,K_k] = -K\pi_j\pi_k$, as in the Lemma.

Lemma 2.1.2. *Let*

$$\begin{aligned}
\psi_1 &= \pi_1 \\
\psi_2 &= \frac{\pi_2}{1-\pi_1} \\
\cdots &\quad \cdots\cdots\cdots \\
\psi_{J-1} &= \frac{\pi_{J-1}}{1-\pi_1-\cdots-\pi_{J-2}}. \qquad (2.10)
\end{aligned}$$

Then the multinomial probability function in (2.3) can be factored as

$$B(y_{i1};1,\psi_1)B(y_{i2};1-y_{i1},\psi_2)\cdots B(y_{i,J-1};1-y_{i1}-\cdots-y_{i,J-2},\psi_{J-1}) \qquad (2.11)$$

where $B(x;K^*,\psi)$, for example, represents the binomial probability of x successes in K^* trials when the success probability is ψ in each trial.

Proof. It is convenient to show that (2.11) yields (2.3). Rewrite (2.11) as

$$\begin{aligned}
&[\frac{1!}{y_{i1}!(1-y_{i1})!}\pi_1^{y_{i1}}(1-\pi_1)^{1-y_{i1}}] \\
\times\ &\frac{(1-y_{i1})!}{y_{i2}!(1-y_{i1}-y_{i2})!}[\frac{\pi_2}{1-\pi_1}]^{y_{i2}}[\frac{1-\pi_1-\pi_2}{1-\pi_1}]^{1-y_{i1}-y_{i2}} \\
&\cdots\cdots\cdots \\
\times\ &\frac{(1-y_{i1}-\cdots-y_{i,J-2})!}{y_{i,J-1}!(1-y_{i1}-\cdots-y_{i,J-1})!}[\frac{\pi_{J-1}}{1-\pi_1-\cdots-\pi_{J-2}}]^{y_{i,J-1}}[\frac{1-\pi_1-\cdots-\pi_{J-1}}{1-\pi_1-\cdots-\pi_{J-2}}]^{1-y_{i1}-\cdots-y_{i,J-1}}.
\end{aligned}$$

By some algebras, this reduces to (2.3).

Lemma 2.1.3. *The binomial factorization (2.11) yields the conditional means and variances as follows:*

$$\begin{aligned} E[Y_{i1}] &= \psi_1,\ \ var[Y_{i1}] = \psi_1(1-\psi_1) \\ E[Y_{i2}|y_{i1}] &= (1-y_{i1})\psi_2,\ \ var[Y_{i2}|y_{i1}] = (1-y_{i1})\psi_2(1-\psi_2) \\ \cdots &\quad \cdots\cdots \\ E[Y_{i,J-1}|y_{i1},\cdots,y_{i,J-2}] &= (1-y_{i1}-\cdots-y_{i,J-2})\psi_{J-1} \\ var[Y_{i,J-1}|y_{i1},\cdots,y_{i,J-2}] &= (1-y_{i1}-\cdots-y_{i,J-2})\psi_{J-1}(1-\psi_{J-1}). \end{aligned} \tag{2.12}$$

Example 2.1. Consider the multinomial model (2.4) with $J = 3$ categories. This model is referred to as the trinomial probability model. Suppose that π_1,π_2, and π_3 denote the probabilities that an individual fall into categories 1, 2, and 3, respectively. Also suppose that out of K independent individuals, these three cells were occupied by K_1, K_2, and K_3 individuals so that $K = K_1 + K_2 + K_3$. Let $\psi_1 = \pi_1$ and $\psi_2 = \frac{\pi_2}{1-\pi_1}$. Then, similar to (2.11), it can be shown that the trinomial probability function (2.4) (with $J = 3$) can be factored as the product of two binomial probability functions as given by

$$B(K,K_1;\psi_1)B(K-K_1,K_2;\psi_2).$$

Similar to Lemma 2.1.3, one then obtains the mean and variance of K_2 conditional on K_1 as

$$E[K_2|K_1] = [K-K_1]\psi_2, \text{ and } \mathrm{var}[K_2|K_1] = [K-K_1]\psi_2(1-\psi_2), \tag{2.13}$$

respectively. It then follows that the unconditional mean and variance of K_2 are given by

$$E[K_2] = E_{K_1}E[K_2|K_1] = E_{K_1}[(K-K_1)\psi_2] = [K-K\psi_1]\psi_2 = K(1-\pi_1)\frac{\pi_2}{1-\pi_1} = K\pi_2, \tag{2.14}$$

and

$$\begin{aligned} \mathrm{var}[K_2] &= E_{K_1}[\mathrm{var}\{K_2|K_1\}] + \mathrm{var}_{K_1}[E\{K_2|K_1\}] \\ &= E_{K_1}[\{K-K_1\}\psi_2(1-\psi_2)] + \mathrm{var}_{K_1}[\{K-K_1\}\psi_2] \\ &= K(1-\pi_1)\frac{\pi_2}{1-\pi_1}[\frac{1-\pi_1-\pi_2}{1-\pi_1}] + K\pi_1(1-\pi_1)\frac{\pi_2^2}{(1-\pi_1)^2} \\ &= \frac{K\pi_2}{1-\pi_1}[1-\pi_1-\pi_2+\pi_1\pi_2] \\ &= K\pi_2(1-\pi_2), \end{aligned} \tag{2.15}$$

respectively. Note that these unconditional mean (2.14) and variance (2.15) are the same as in Lemma 2.1.1, but they were derived in a different way than that of Lemma 2.1.1. Furthermore, similar to that of (2.15), the unconditional covariance between K_1 and K_2 may be obtained as

$$\begin{aligned}\mathrm{cov}[K_1,K_2] &= E_{K_1}[\mathrm{cov}\{(K_1,K_2)|K_1\}] + \mathrm{cov}_{K_1}[K_1,E\{K_2|K_1\}] \\ &= \mathrm{cov}_{K_1}[K_1,E\{K_2|K_1\}] \\ &= \mathrm{cov}_{K_1}[K_1,(K-K_1)\psi_2] = -\psi_2\mathrm{var}[K_1] = -K\pi_1\pi_2, \qquad (2.16)\end{aligned}$$

which agrees with the covariance results in Lemma 2.1.

2.1.2 *Inference for Proportion $\pi_j (j = 1,\ldots,J-1)$*

Recall from (2.4) that

$$P[K_1,\ K_2,\ldots,K_j,\ \ldots,K_{J-1}] = \frac{K!}{K_1!\cdots K_J!}\Pi_{j=1}^{J}\pi_j^{K_j}, \qquad (2.17)$$

where π_j by (2.2) has the formula

$$\pi_j = \begin{cases} \frac{\exp(\beta_{j0})}{1+\sum_{g=1}^{J-1}\exp(\beta_{g0})} & \text{for } j = 1,\ldots,J-1 \\ \frac{1}{1+\sum_{g=1}^{J-1}\exp(\beta_{g0})} & \text{for } j = J, \end{cases}$$

(a) Moment estimation for π_j

When K_j for $j = 1,\ldots,J-1$, follow the multinomial distribution (2.17), it follows from Lemma 2.1 that $E[K_j] = K\pi_j$ yielding the moment estimating equation for π_j as

$$K_j - K\pi_j = 0 \text{ subject to the condition } \sum_{j=1}^{J}\pi_j = 1. \qquad (2.18)$$

Because by (2.18), one writes

$$\pi_J = 1 - \sum_{j=1}^{J-1}\pi_j = 1 - \sum_{j=1}^{J-1}\frac{K_j}{K} = \frac{K-\sum_{j=1}^{J-1}K_j}{K} = \frac{K_J}{K},$$

thus, in general, the moment estimator for π_j for all $j = 1,\ldots,J$, has the form

$$\hat{\pi}_{j,MM} = \frac{K_j}{K}. \qquad (2.19)$$

Note however that once the estimation of π_j for $j = 1,\ldots,J-1$ is done, estimation of π_J does not require any new information because $K_J = K - \sum_{j=1}^{J-1} K_j$ becomes known.

(b) Likelihood Estimation of proportion $\pi_j, j = 1,\ldots,J$-1

It follows from (2.17) that the log likelihood function of $\{\pi_j\}$ with $\pi_J = 1 - \sum_{j=1}^{J-1} \pi_j$ is given by

$$logL(\pi_1,\ldots,\pi_{J-1}) = k_0 + \sum_{j=1}^{J} K_j log(\pi_j), \tag{2.20}$$

where k_0 is the normalizing constant free from $\{\pi_j\}$. It then follows that the maximum likelihood (ML) estimator of π_j, for $j = 1,\ldots,J-1$, is the solution of the likelihood equation

$$\frac{\partial logL(\pi_1,\ldots,\pi_{J-1})}{\partial \pi_j} = \frac{K_j}{\pi_j} - \frac{K_J}{1-\sum_{j=1}^{J-1} \pi_j} = 0, \tag{2.21}$$

and is given by

$$\hat{\pi}_{j,ML} = \hat{\pi}_{J,ML} \frac{K_j}{K_J}. \tag{2.22}$$

But, as $\sum_{j=1}^{J} \hat{\pi}_{j,ML} = 1$, it follows from (2.22) that

$$\hat{\pi}_{J,ML} = \frac{K_J}{K},$$

yielding

$$\hat{\pi}_{j,ML} = \frac{K_j}{K} = \frac{K_j}{\sum_{j=1}^{J} K_j} \text{ for } j = 1,\ldots,J-1.$$

Thus, in general, one may write the formula

$$\hat{\pi}_{j,ML} = \frac{K_j}{K} = \frac{K_j}{\sum_{j=1}^{J} K_j}, \tag{2.23}$$

for all $j = 1,\ldots,J$. This ML estimate in (2.23) is the same as the moment estimate in (2.19).

(c) Illustration 2.1

To illustrate the aforementioned ML estimation for the categorical proportion, we, for example, consider a modified version of the health care utilization data, studied by Sutradhar (2011). This data set contains number of physician visits by 180 members of 48 families over a period of 6 years from 1985 to 1990. Various

Table 2.1 Summary statistics of physician visits by four covariates in the health care utilization data for 1985

		Number of Visits					
Covariates	**Level**	**0**	**1**	**2**	**3–5**	**≥6**	**Total**
Gender	Male	28	22	18	16	12	96
	Female	11	5	15	21	32	84
Chronic Condition	No	26	20	15	16	11	88
	Yes	13	7	18	21	33	92
Education Level	< High School	17	5	11	10	15	58
	High School	6	4	4	8	11	33
	> High School	16	18	18	19	18	89
Age	20–30	23	17	14	15	15	84
	31–40	1	1	3	3	3	11
	41–50	4	4	5	12	8	33
	51–65	10	5	8	5	13	41
	66–85	1	0	3	2	5	11

Table 2.2 Categorizing the number of physician visits

Latent number of visits	**Visit category**	**1985 visit**
0	None	$K_1 = 39$
1–2	Few	$K_2 = 60$
3–5	Not so few	$K_3 = 37$
6 or more	High	$K_4 = 44$

covariates such as gender, age, education level, and chronic conditions for each of these 180 members were also collected. The full data set is available in Sutradhar (2011, Appendix 6A). The primary objective of this study was to examine the effects of these covariates on the physician visits by accommodating familial and longitudinal correlations among the responses of the members. To have a feeling about this data set, we reproduce below in Table 2.1, some summary statistics on the physicians visit data for 1985 only.

Suppose that we group the physician visits into $J = 4$ categories as in Table 2.2. In the same table we also give the 1985 health status for 180 individuals.

Note that an individual can belong to one of the four categories with a multinomial probability as in (2.3). Now by ignoring the family grouping, that is, assuming all 180 individuals are independent, and by ignoring the effects of the covariates on the visits, one may use the multinomial probability model (2.17) to fit the data in Table 2.2.

Now by (2.23), one obtains the likelihood estimate for π_j, for $j = 1,\ldots,4$, as

$$\hat{\pi}_{j,ML} = \frac{K_j}{K},$$

where $K = 180$. Thus, for example, for $j = 1$, since, $K_1 = 39$ individuals did not pay any visits to the physician, an estimate (likelihood or moment) for the probability that an individual in St. John's in 1985 belong to category 1 was

$$\hat{\pi}_{1,ML} = \hat{\pi}_{1,MM} = 39/180 = 0.217.$$

That is, approximately 22 out of 100 people did not pay any visits to the physician in St. John's (indicating the size of the group with no health complications) during that year. Note that these naive estimates are bound to change when multinomial probabilities will be modeled involving the covariates. This type of multinomial regression model will be discussed in Sect. 2.2 and in many other places in the book.

2.1.3 Inference for Category Effects β_{j0}, $j = 1, \ldots, J-1$, with $\beta_{J0} = 0$

2.1.3.1 Moment Estimating Equations for $\beta_{j0}(j = 1, \ldots, J-1)$ Using Regression Form

Because

$$E[K_j] = K\pi_j \text{ for } j = 1, \ldots, J-1,$$

with

$$\pi_j = \frac{m_j}{m} = \frac{\exp(\beta_{j0})}{1 + \sum_{j=1}^{J-1} \exp(\beta_{j0})} = \frac{\exp(x_j'\theta)}{\sum_{j=1}^{J} \exp(x_j'\theta)},$$

and π_J has to satisfy the relationship

$$\pi_J = 1 - \sum_{j=1}^{J-1} \pi_j = 1 - \sum_{j=1}^{J-1} \frac{K_j}{K} = \frac{K_J}{K},$$

one needs to solve for $\theta = (\beta_{10}, \ldots, \beta_{j0}, \ldots, \beta_{J-1,0})'$ satisfying

$$K_j - K\pi_j = 0, \text{ for all } j = 1, \ldots, J.$$

For convenience, we express all π_j as functions of θ. We do this by using

$$x_j = (01_{j-1}', 1, 01_{J-1-j}')' \text{ for } j = 1, \ldots, J-1, \text{ and } x_J = 01_{J-1},$$

so that

$$\pi_j = \frac{\exp(x_j'\theta)}{\sum_{j=1}^{J}\exp(x_j'\theta)}, \text{ for all } j = 1,\ldots,J. \tag{2.24}$$

Now solving the moment equations $K_j - K\pi_j = 0$ for $\theta = (\beta_{10},\ldots,\beta_{j0},\ldots,\beta_{J-1,0})'$ is equivalent to solve

$$f(\theta) = X'(y - K\pi) = 0, \tag{2.25}$$

for θ, where $y = (K_1,\ldots,K_j,\ldots,K_J)'$, $\pi = (\pi_1,\ldots,\pi_j,\ldots,\pi_J)'$, and

$$X = \begin{pmatrix} x_1' \\ x_2' \\ \cdot \\ x_{J-1}' \\ x_J' \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdot & 0 & 0 \\ 0 & 1 & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & 0 & 1 \\ 0 & 0 & \cdot & 0 & 0 \end{pmatrix} : J \times J - 1.$$

2.1.3.2 Marginal Likelihood Estimation for β_{j0} ($j = 1,\ldots, J$-1) with $\beta_{J0} = 0$

Note that due to invariance principle of the likelihood estimation method, one would end up with solving the same likelihood estimating equation (2.23) even if one attempts to obtain the likelihood estimating equations for $\beta_{j0},\ j = 1,\ldots,J-1$, directly. We clarify this point through following direct calculations.

Rewrite the multinomial distribution based log likelihood function (2.20) as

$$logL(\pi_1,\ldots,\pi_J) = k_0 + \sum_{j=1}^{J} K_j log(\pi_j),$$

where, by (2.17), π_j has the formulas

$$\pi_j = \begin{cases} \frac{\exp(\beta_{j0})}{1+\sum_{j=1}^{J-1}\exp(\beta_{j0})} & \text{for } j = 1,\ldots,J-1 \\ \frac{1}{1+\sum_{j=1}^{J-1}\exp(\beta_{j0})} & \text{for } j = J. \end{cases}$$

It then follows for $j = 1,\ldots,J-1$, that

$$\frac{\partial logL(\pi_1,\ldots,\pi_J)}{\partial\beta_{j0}} = \sum_{c=1}^{J-1}\left[\frac{K_c}{\pi_c}\right]\frac{\partial\pi_c}{\partial\beta_{j0}} + [\frac{K_J}{\pi_J}]\frac{\partial\pi_J}{\partial\beta_{j0}}, \tag{2.26}$$

where

$$\frac{\partial \pi_c}{\partial \beta_{j0}} = \begin{cases} \frac{\{1+\sum_{c=1}^{J-1}\exp(\beta_{c0})\}\exp(\beta_{j0})-\exp(\beta_{j0})\{\exp(\beta_{j0})\}}{\left[1+\sum_{c=1}^{J-1}\exp(\beta_{c0})\right]^2} = \pi_j(1-\pi_j) & \text{for } c=j \\ \frac{-\exp(\beta_{c0})\{\exp(\beta_{j0})\}}{\left[1+\sum_{c=1}^{J-1}\exp(\beta_{c0})\right]^2} = -\pi_c\pi_j & \text{for } c \neq j,\ c=1,\ldots,J-1. \\ \frac{-\exp(\beta_{j0})}{\left[1+\sum_{c=1}^{J-1}\exp(\beta_{c0})\right]^2} = -\pi_J\pi_j & \text{for } c=J. \end{cases} \tag{2.27}$$

By using (2.27) in (2.26), we then write the likelihood equation for β_{j0} as

$$\frac{\partial logL(\pi_1,\ldots,\pi_J)}{\partial \beta_{j0}} = \sum_{c=1}^{J}[\frac{K_c}{\pi_c}][-\pi_c\pi_j] + \frac{K_j}{\pi_j}[\pi_j] = 0, \tag{2.28}$$

yielding

$$-K\pi_j + K_j = 0, \text{for } j=1,\ldots,J-1, \tag{2.29}$$

which are the same likelihood equations as in (2.23). Thus, in the likelihood approach, similar to the moment approach, one solves the estimating equation (2.25), that is,

$$f(\theta) = X'(y-K\pi) = 0 \tag{2.30}$$

for θ iteratively, so that $f(\hat{\theta}) = 0$.

Further note that because of the definition of π_j given by (2.2) or (2.17), all estimates $\hat{\beta}_{j0}$ for $j=1,\ldots,J-1$ are interpreted comparing their value with $\beta_{J0}=0$.

2.1.3.3 Joint Estimation of $\beta_{10},\ldots,\beta_{j0},\ldots,\beta_{J\text{-}1,0}$ Using Regression Form

The log likelihood function by (2.20) has the form

$$logL(\beta_{10},\ldots,\beta_{(J-1)0}) = k_0 + \sum_{j=1}^{J} K_j log\ \pi_j.$$

We now write $m_j = \exp(\beta_{j0})$ for $j=1,\ldots,J-1$, and $m_J = \exp(\beta_{J0}) = 1$, and $m = \sum_{j=1}^{J} m_j$, and re-express the above log likelihood function as

$$logL(\beta_{10},\ldots,\beta_{(J-1)0}) = k_0 + \sum_{j=1}^{J} K_j[log\ m_j - log\ m]. \tag{2.31}$$

Next for $\theta = (\beta_{10},\ldots,\beta_{j0},\ldots,\beta_{(J-1)0})'$ express $\log m_j$ in linear regression form

$$\log m_j = x_j'\theta \tag{2.32}$$

such that $\log m_j = \beta_{j0}$ for $j = 1,\ldots,J-1$, and $\log m_J = 0$. Note that finding x'_j for all $j = 1,\ldots,J$ is equivalent to write

$$\log \tilde{m} = [\log m_1,\ldots,\log m_j,\ldots,\log m_J]' = X\theta,$$

where the $J \times (J-1)$ covariate matrix X has the same form as in (2.25), i.e.,

$$X = \begin{pmatrix} x'_1 \\ x'_2 \\ \cdot \\ x'_{J-1} \\ x'_J \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdot & 0 & 0 \\ 0 & 1 & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & 0 & 1 \\ 0 & 0 & \cdot & 0 & 0 \end{pmatrix} : J \times J-1. \tag{2.33}$$

It then follows from (2.31) and (2.32) that

$$\begin{aligned} f(\theta) = \frac{\partial logL(\theta)}{\partial\theta} &= \frac{\partial}{\partial\theta}[\sum_{j=1}^{J} K_j x'_j \theta - Klog\, m] \\ &= \sum_{j=1}^{J} K_j x_j - \frac{K}{m}\frac{\partial}{\partial\theta}[\sum_{j=1}^{J} \exp(x'_j\theta)] \\ &= \sum_{j=1}^{J} K_j x_j - \frac{K}{m}\sum_{j=1}^{J} m_j x_j \\ &= \sum_{j=1}^{J} K_j x_j - K\sum_{j=1}^{J} \pi_j x_j, \end{aligned} \tag{2.34}$$

yielding the likelihood estimating equation

$$f(\theta) = X'(y - K\pi) = 0, \tag{2.35}$$

same as (2.30).

2.1.3.3.1 Likelihood Estimates and their Asymptotic Variances

Because the likelihood estimating equations in (2.35) are non-linear, one obtains the estimate of $\theta = (\beta_{10},\ldots,\beta_{j0},\ldots,\beta_{J-1,0})'$ iteratively, so that $f(\hat{\theta}) = 0$. Suppose that $\hat{\theta}_0$ is not a solution for $f(\theta) = 0$, but a trial estimate and hence $f(\hat{\theta}_0) \neq 0$. Next suppose that $\hat{\theta} = \hat{\theta}_0 + h^*$ is the estimate of θ satisfying $f(\hat{\theta}) = f(\hat{\theta}_0 + h^*) = 0$. Now by using the first order Taylor's expansion, one writes

$$f(\hat{\theta}) = f(\hat{\theta}_0 + h^*) = f(\hat{\theta}_0) + h^* f'(\theta)|_{\theta=\hat{\theta}_0} = f(\theta)|_{\theta=\hat{\theta}_0} + (\hat{\theta} - \hat{\theta}_0) f'(\theta)|_{\theta=\hat{\theta}_0} = 0$$

yielding the solution

$$\hat{\theta} = \hat{\theta}_0 - \left[\{f'(\theta)\}^{-1} f(\theta)\right]|_{\theta=\hat{\theta}_0}. \tag{2.36}$$

Further, because

$$\begin{aligned}
\frac{\partial \pi_j}{\partial \theta'} &= \frac{1}{m^2}[m\frac{\partial m_j}{\partial \theta'} - m_j \frac{\partial m}{\partial \theta'}] \\
&= \frac{1}{m^2}[mm_j x'_j - m_j \sum_{j=1}^{J} m_j x'_j] \\
&= \pi_j x'_j - \pi_j \sum_{j=1}^{J} \pi_j x'_j \\
&= \pi_j x'_j - \pi_j \pi' X,
\end{aligned} \tag{2.37}$$

one obtains

$$K\frac{\partial \pi_j}{\partial \theta'} = K[\pi_j x'_j - \pi_j \pi' X]. \tag{2.38}$$

Consequently, it follows from (2.35) that

$$\begin{aligned}
f'(\theta) = -KX'\frac{\partial \pi}{\partial \theta'} &= -KX'\{\text{diag}[\pi_1, \ldots, \pi_J] - \pi\pi'\}X \\
&= -KX'[D_\pi - \pi\pi']X,
\end{aligned} \tag{2.39}$$

and the iterative equation (2.36) takes the form

$$\hat{\theta}(r+1) = \hat{\theta}(r) + \left[\frac{1}{K}[X'\{D_\pi - \pi\pi'\}X]^{-1}X'(y - K\pi)\right]_{\theta=\hat{\theta}(r)}, \tag{2.40}$$

yielding the final estimate $\hat{\theta}$. The covariance matrix of $\hat{\theta}$ has the formula

$$\text{var}(\hat{\theta}) = \frac{1}{K}\left[X'\{D_\pi - \pi\pi'\}X\right]^{-1}. \tag{2.41}$$

2.1.4 Likelihood Inference for Categorical Effects $\beta_{j0}, j = 1, \ldots, J-1$ with $\beta_{J0} = -\sum_{j=1}^{J-1} \beta_{j0}$ Using Regression Form

There exists an alternative modeling for π_j such that $\hat{\beta}_{j0}$ for $j = 1, \ldots, J-1$ are interpreted by using the restriction

$$\sum_{j=1}^{J} \hat{\beta}_{j0} = 0, \text{ that is, } \hat{\beta}_{J0} = -\sum_{j=1}^{J-1} \hat{\beta}_{j0}.$$

As opposed to (2.17), π_j's are then defined as

$$\pi_j = \begin{cases} \frac{\exp(\beta_{j0})}{\sum_{c=1}^{J-1}\exp(\beta_{c0})+\exp(-\sum_{c=1}^{J-1}\beta_{c0})} & \text{for } j = 1,\ldots,J-1 \\ \frac{\exp(-\sum_{c=1}^{J-1}\beta_{c0})}{\sum_{c=1}^{J-1}\exp(\beta_{c0})+\exp(-\sum_{c=1}^{J-1}\beta_{c0})} & \text{for } j = J. \end{cases} \tag{2.42}$$

Now for $m_j = \exp(\beta_{j0})$ for $j = 1,\ldots,J-1$, and $m_J = \exp(-\sum_{c=1}^{J-1}\beta_{c0})$, one may use the linear form $\log m_j = x_j'\theta$, that is,

$$\log \tilde{m} = [\log m_1,\ldots,\log m_j,\ldots,\log m_J]' = X\theta,$$

where, unlike in (2.25) and (2.33), X now is the $J \times (J-1)$ covariate matrix defined as

$$X = \begin{pmatrix} 1 & 0 & \cdot & 0 & 0 \\ 0 & 1 & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & 0 & 1 \\ -1 & -1 & \cdot & -1 & -1 \end{pmatrix}. \tag{2.43}$$

Thus, the likelihood estimating equation has the same form

$$f(\theta) = X'(y - K\pi) = 0 \tag{2.44}$$

as in (2.35), but with covariate matrix X as in (2.43) which is different than that of (2.33).

Note that because $\beta_{J0} = 0$ leads to different covariate matrix X as compared to the covariate matrix under the assumption $\beta_{J0} = -\sum_{j=1}^{J-1}\beta_{j0}$, the likelihood estimates for $\theta = (\beta_{10},\ldots,\beta_{(J-1)0})'$ would be different under these two assumptions.

2.2 Univariate Multinomial Regression Model

2.2.1 Individual History Based Fixed Regression Effects Model

Suppose that a history based survey is done so that in addition to the categorical response status, an individual also provides p covariates information. Let $w_i = [w_{i1},\ldots,w_{is},\ldots,w_{ip}]'$ denote the p-dimensional covariate vector available from the ith $(i = 1,\ldots,K)$ individual. To incorporate this covariate information, the multinomial probability model (2.1)–(2.2) may be generalized as

$$P[y_i = y_i^{(j)} = \delta_{ij}] = \pi_{(i)j} = \begin{cases} \frac{\exp(\beta_{j0}+\beta_j' w_i)}{1+\sum_{g=1}^{J-1}\exp(\beta_{g0}+\beta_g' w_i)} & \text{for } j = 1,\ldots,J-1 \\ \frac{1}{1+\sum_{g=1}^{J-1}\exp(\beta_{g0}+\beta_g' w_i)} & \text{for } j = J, \end{cases} \tag{2.45}$$

(see also Agresti 1990, p. 343, Exercise 9.22) where $\beta_j = [\beta_{j1},\ldots,\beta_{js},\ldots,\beta_{jp}]'$ for $j = 1,\ldots,J-1$. Let

$$\theta^* = [\beta_1^{*\prime},\ldots,\beta_j^{*\prime},\ldots,\beta_{J-1}^{*}{}']', \text{ where } \beta_j^* = [\beta_{j0},\beta_j']'.$$

Then as an extension to (2.4), one may write the likelihood function as

$$L(\theta^*) = L(\beta_1^*,\ldots,\beta_{J-1}^*) = \prod_{i=1}^{K}\prod_{j=1}^{J}\frac{1! \times \left\{\pi_{(i)j}\right\}^{y_{ij}}}{y_{ij}!} \tag{2.46}$$

where $y_{iJ} = \left(1-\sum_{j=1}^{J-1} y_{ij}\right)$ and $\pi_{(i)J} = \left(1-\sum_{j=1}^{J-1}\pi_{(i)j}\right)$. It then follows that the likelihood estimating equation for $\beta_j^* = \left(\beta_{j0},\, \beta_j'\right)'$ for $j = 1,\ldots,J-1$, that is,

$$\begin{aligned} \frac{\partial logL(\theta^*)}{\partial \beta_j^*} &= \frac{\partial}{\partial \beta_j^*}\left[C + \sum_{i=1}^{K}\sum_{g=1}^{J-1} y_{ig}\begin{pmatrix}1\\ w_i\end{pmatrix}'\beta_g^* - \sum_{i=1}^{K} log\left\{1+\sum_{g=1}^{J-1}\begin{pmatrix}1\\ w_i\end{pmatrix}'\beta_g^*\right\}\right] \\ &= \sum_{i=1}^{K}\left[\begin{pmatrix}1\\ w_i\end{pmatrix} y_{ij} - \begin{pmatrix}1\\ w_i\end{pmatrix}\pi_{(i)j}\right] \\ &= \sum_{i=1}^{K}\begin{pmatrix}1\\ w_i\end{pmatrix}\left[y_{ij}-\pi_{(i)j}\right] = 0, \end{aligned} \tag{2.47}$$

leads to the likelihood equation for θ^* as

$$\frac{\partial logL(\theta^*)}{\partial \theta^*} = \sum_{i=1}^{K}\left[I_{J-1}\otimes\begin{pmatrix}1\\ w_i\end{pmatrix}\right]\left[y_i - \pi_{(i)}\right] = 0 \tag{2.48}$$

where $\pi_{(i)} = \left(\pi_{(i)1},\cdots,\pi_{(i)(J-1)}\right)'$ corresponding to $y_i = (y_{i1},\ldots,y_{i(J-1)})'$; w_i is the $p\times 1$ design vector, I_{J-1} is the identity matrix of order $J-1$, and $\otimes$ denotes the Kronecker or direct product. In (2.47), C is a normalizing constant.

This likelihood equation (2.48) may be solved for θ^* by using the iterative equation

Table 2.3 Snoring and heart disease: A frequency table

	Heart disease	
Snoring	**Yes**	**No**
Never	24	1355
Occasionally	35	603
Nearly every night	21	192
Every night	30	224

$$\hat{\theta}^*(r+1) = \hat{\theta}^*(r) + \left[\sum_{i=1}^{K}\left[I_{J-1}\otimes\begin{pmatrix}1\\ w_i\end{pmatrix}\right]\left[\mathrm{diag}[\pi_{i1},\ldots,\pi_{J-1}]-\pi_{(i)}\pi'_{(i)}\right]\left[I_{J-1}\otimes\begin{pmatrix}1\\ w_i\end{pmatrix}\right]'\right]^{-1}_{|\hat{\theta}^*(r)}$$
$$\times\sum_{i=1}^{K}\left[I_{J-1}\otimes\begin{pmatrix}1\\ w_i\end{pmatrix}\right]\left[y_i-\pi_{(i)}\right]_{|\hat{\theta}^*(r)}, \tag{2.49}$$

and the variances of the estimator may be found from the covariance matrix

$$\mathrm{var}[\hat{\theta}^*] = \left[\sum_{i=1}^{K}\left[I_{J-1}\otimes\begin{pmatrix}1\\ w_i\end{pmatrix}\right]\left[\mathrm{diag}[\pi_{i1},\ldots,\pi_{J-1}]-\pi_{(i)}\pi'_{(i)}\right]\left[I_{J-1}\otimes\begin{pmatrix}1\\ w_i\end{pmatrix}\right]'\right]^{-1}. \tag{2.50}$$

Note that in the absence of covariates, one estimates $\theta^* = [\beta_{10},\ldots,\beta_{j0},\ldots,\beta_{J-1.0}]'$. In this case, the estimating equation (2.48) for θ^* reduces to the estimating equation (2.35) for θ, because $\sum_{i=1}^{K} y_{ij} = K_j$ and $\sum_{i=1}^{K}\pi_{(i)j} = \sum_{i=1}^{K}\pi_j = K\pi_j$, for example.

2.2.1.1 Illustration 2.2: Binary Regression Model ($J = 2$) with One Covariate

2.2.1.1 (a) An Existing Analysis (Snoring as a Continuous Covariate with Arbitrary Values)

Consider the heart disease and snoring relationship problem discussed in Agresti (2002, Section 4.2.3, p. 121–123). The data is given in the following Table 2.3.

By treating snoring as an one dimensional $(p = 1)$ fixed covariate $w_i = w_{i1}$ for the ith individual with its values

$$w_i \equiv w_{i1} = 0, 2, 4, 5, \tag{2.51}$$

for snoring never, occasionally, nearly every night, and every night, respectively, and treating the heart disease status as the binary $(J = 2)$ variable and writing

$$y_i = y_{i1} = \begin{cases} 1 \text{ if } i \in \text{yes} \\ 0 \text{ otherwise,}\end{cases}$$

Agresti (2002, Section 4.2.3, p. 121–123), for example, analyzed this 'snoring and heart disease' data by fitting the binary probability model (a special case of the multinomial probability model (2.45))

$$P[y_i = y_i^{(1)}] = P[y_{i1} = 1] = \pi_{(i)1}(w_i) = \frac{\exp(\beta_{10} + \beta_{11} w_i)}{1 + \exp(\beta_{10} + \beta_{11} w_i)}, \tag{2.52}$$

and

$$P[y_i = y_i^{(2)}] = P[y_{i1} = 0] = \pi_{(i)2}(w_i) = \frac{1}{1 + \exp(\beta_{10} + \beta_{11} w_i)}.$$

The binary likelihood is then given by

$$\begin{aligned} L(\theta^*) = L(\beta_{10}, \beta_{11}) &= \Pi_{i=1}^{K} [\pi_{(i)1}(w_i)]^{y_{i1}} [\pi_{(i)2}(w_i)]^{y_{i2}} \\ &= \Pi_{i=1}^{K} [\pi_{(i)1}(w_i)]^{y_{i1}} [1 - \pi_{(i)1}(w_i)]^{1 - y_{i1}}, \end{aligned} \tag{2.53}$$

yielding the log likelihood estimating equations as

$$\frac{\partial log\ L(\theta^*)}{\partial \theta^*} = \frac{\partial log \Pi_{i=1}^{K} \frac{\exp[y_{i1}(w_i^{*\prime}\theta^*)]}{1 + \exp(w_i^{*\prime}\theta^*)}}{\partial \theta^*} = 0, \tag{2.54}$$

where

$$w_i^{*\prime} = (1, w_i), \text{ and } \theta^* = \beta_1^* = (\beta_{10}, \beta_{11})'.$$

This log likelihood equations may be simplified as

$$\begin{aligned} \frac{\partial log\ L(\theta^*)}{\partial \theta^*} &= \sum_{i=1}^{K} y_{i1} w_i^* - \sum_{i=1}^{K} \pi_{(i)1} w_i^* \\ &= \sum_{i=1}^{K} w_i^* [y_{i1} - \pi_{(i)1}] \\ &= \sum_{i=1}^{K} \begin{pmatrix} 1 \\ w_i \end{pmatrix} [y_{i1} - \pi_{(i)1}] = 0. \end{aligned} \tag{2.55}$$

Note that the binary likelihood equation (2.45) is a special case of the multinomial likelihood equation (2.48) with $J = 2$. This equation may be solved for $\hat{\theta}^*$ by using the iterative equation

$$\hat{\theta}^*(r+1) = \hat{\theta}^*(r) + \left[\sum_{i=1}^{K} \begin{pmatrix} 1 \\ w_i \end{pmatrix} \left[\pi_{(i)1}(1 - \pi_{(i)}) \right] \begin{pmatrix} 1 \\ w_i \end{pmatrix}' \right]^{-1}_{|\hat{\theta}^*(r)}$$

$$\times \sum_{i=1}^{K} \begin{pmatrix} 1 \\ w_i \end{pmatrix} \left[y_i - \pi_{(i)1} \right]_{|\hat{\theta}^*(r)}, \tag{2.56}$$

and the variances of the estimator may be found from the covariance matrix

$$\text{var}[\hat{\theta}^*] = \left[\sum_{i=1}^{K} \begin{pmatrix} 1 \\ w_i \end{pmatrix} \left[\pi_{(i)1}(1-\pi_{(i)1}) \right] \begin{pmatrix} 1 \\ w_i \end{pmatrix}' \right]^{-1}. \tag{2.57}$$

For the snoring and heart disease relationship problem, the scalar covariate $(w_i = w_{i1})$ based estimates are given by

$$\hat{\theta}^* = \hat{\beta}_1^* \equiv [\hat{\beta}_{10} = -3.87, \text{ and } \hat{\beta}_{11} = 0.40]'.$$

However using this type of scalar covariate, i.e., $w_i = w_{i1}$ with arbitrary values for snoring levels does not provide actual effects of the snoring on heart disease. Below, we illustrate a categorical covariate based estimation for this problem.

2.2.1.1 (b) A Refined Analysis (Snoring as a Fixed Covariate with Four Nominal Levels)

In the aforementioned existing analysis, the snoring status: never, occasionally, nearly every night, every night, has been denoted by a covariate w with values $0, 2, 4,$ and 5, respectively. This is an arbitrary coding and may not correctly reflect the levels. To avoid confusion, in the proposed book, we will represent these $L = 4$ levels of the 'snoring' covariate for the ith individual by three dummy covariates $(p = 3)$ w_{i1}, w_{i2}, w_{i3} with values

$$(w_{i1}, w_{i2}, w_{i3}) = \begin{cases} (1,0,0) \text{ for occasionally snoring, level 1 } (\ell\text{=1}) \\ (0,1,0) \text{ for nearly every night snoring, level 2 } (\ell\text{=2}) \\ (0,0,1) \text{ for every night snoring, level 3 } (\ell\text{=3}) \\ (0,0,0) \text{ for never snoring, level 4 } (\ell\text{=4}) . \end{cases}$$

Now for $j = 1, \ldots, J-1$ with $J = 2$, by using $\beta_{j1}, \beta_{j2}, \beta_{j3}$ as the effects of w_{i1}, w_{i2}, w_{i3}, on an individual's $(i = 1, \ldots, K)$ heart status belonging to jth category, one may fit the probability model (2.45) to this binary data. For convenience, write the model as

$$\pi_{(i)j} = \begin{cases} \frac{\exp(\beta_{j0}+\beta_{j1}w_{i1}+\beta_{j2}w_{i2}+\beta_{j3}w_{i3})}{1+\sum_{g=1}^{J-1}\exp(\beta_{g0}+\beta_{g1}w_{i1}+\beta_{g2}w_{i2}+\beta_{g3}w_{i3})} & \text{for } j = 1 \\ \frac{1}{1+\sum_{g=1}^{J-1}\exp(\beta_{g0}+\beta_{g1}w_{i1}+\beta_{g2}w_{i2}+\beta_{g3}w_{i3})} & \text{for } j = J = 2. \end{cases} \tag{2.58}$$

It is of interest to estimate the parameters $\theta^* = \beta_1^* = (\beta_{10}, \beta_{11}, \beta_{12}, \beta_{13})'$.

After a slight modification, we may use the likelihood equation (2.45) to estimate these parameters. More specifically, by (2.45), the likelihood equation for $\theta^* = \beta_1^*$ now has the form

$$\sum_{i=1}^{K} w_i^* [y_{i1} - \pi_{(i)1}] = \sum_{i=1}^{K} \begin{pmatrix} 1 \\ w_{i1} \\ w_{i2} \\ w_{i3} \end{pmatrix} [y_{i1} - \pi_{(i)1}] = 0. \tag{2.59}$$

This equation may be solved for $\hat{\theta}^*$ iteratively by using

$$\hat{\theta}^*(r+1) = \hat{\theta}^*(r) + \left[\sum_{i=1}^{K} \begin{pmatrix} 1 \\ w_{i1} \\ w_{i2} \\ w_{i3} \end{pmatrix} \left[\pi_{(i)1}(1 - \pi_{(i)}) \right] \begin{pmatrix} 1 \\ w_{i1} \\ w_{i2} \\ w_{i3} \end{pmatrix}' \right]^{-1}_{|\hat{\theta}^*(r)}$$
$$\times \sum_{i=1}^{K} \begin{pmatrix} 1 \\ w_{i1} \\ w_{i2} \\ w_{i3} \end{pmatrix} \left[y_i - \pi_{(i)1} \right]_{|\hat{\theta}^*(r)}, \tag{2.60}$$

and the variances of the estimator may be found from the covariance matrix

$$\text{var}[\hat{\theta}^*] = \left[\sum_{i=1}^{K} \begin{pmatrix} 1 \\ w_{i1} \\ w_{i2} \\ w_{i3} \end{pmatrix} \left[\pi_{(i)1}(1 - \pi_{(i)1}) \right] \begin{pmatrix} 1 \\ w_{i1} \\ w_{i2} \\ w_{i3} \end{pmatrix}' \right]^{-1} \tag{2.61}$$

2.2.2 *Multinomial Likelihood Models Involving One Covariate with $L = p + 1$ Nominal Levels*

Suppose that the $L = p + 1$ levels of a covariate for an individual i may be represented by p dummy covariates as

$$(w_{i1}, \cdots, w_{ip}) \equiv \begin{cases} (\,1,\ 0,\ \cdots,\ 0\,) \longrightarrow \text{Level 1} \\ (\,0,\ 1,\ \cdots,\ 0\,) \longrightarrow \text{Level 2} \\ (\cdots\cdots\cdots\cdots) \\ (\,0,\ 0,\ \cdots,\ 1\,) \longrightarrow \text{Level p} \\ (\,0,\ 0,\ \cdots,\ 0\,) \longrightarrow \text{Level p+1} \end{cases} \tag{2.62}$$

Table 2.4 A notational display for cell counts and probabilities for J categories under each covariate level ℓ

		J **categories of the response variable**					
Covariate level	**Quantity**	1	...	j	...	J	**Total**
1	Cell count	$K_{[1]1}$	...	$K_{[1]j}$	...	$K_{[1]J}$	$K_{[1]}$
	Cell probability	$\pi_{[1]1}$	...	$\pi_{[1]j}$	...	$\pi_{[1]J}$	1
.	.	.	...	.	...	.	.
	.	.	...	.	...	.	.
ℓ	Cell count	$K_{[\ell]1}$	...	$K_{[\ell]j}$	...	$K_{[\ell]J}$	$K_{[\ell]}$
	Cell probability	$\pi_{[\ell]1}$	...	$\pi_{[\ell]j}$	...	$\pi_{[\ell]J}$	1
.	.	.	...	.	...	.	.
	.	.	...	.	...	.	.
$L = p+1$	Cell count	$K_{[p+1]1}$	...	$K_{[p+1]j}$	...	$K_{[p+1]J}$	$K_{[p+1]}$
	Cell probability	$\pi_{[p+1]1}$	...	$\pi_{[p+1]j}$	...	$\pi_{[p+1]J}$	1
	Total count	K_1	...	K_j	...	K_J	K

By (2.45), one may then write the probability for an individual i with covariate at level $\ell(\ell = 1, \ldots, p)$ to be in the jth category as

$$\pi_{[\ell]j} = \pi_{(i\in\ell)j} = \begin{cases} \frac{\exp(\beta_{j0}+\beta_{j\ell})}{1+\sum_{g=1}^{J-1}\exp(\beta_{g0}+\beta_{g\ell})} & \text{for } j = 1, \ldots, J-1 \\ \frac{1}{1+\sum_{g=1}^{J-1}\exp(\beta_{g0}+\beta_{g\ell})} & \text{for } j = J, \end{cases} \tag{2.63}$$

whereas for $\ell = p+1$, these probabilities are written as

$$\pi_{[p+1]j} = \pi_{(i\in(p+1))j} = \begin{cases} \frac{\exp(\beta_{j0})}{1+\sum_{g=1}^{J-1}\exp(\beta_{g0})} & \text{for } j = 1, \ldots, J-1 \\ \frac{1}{1+\sum_{g=1}^{J-1}\exp(\beta_{g0})} & \text{for } j = J. \end{cases} \tag{2.64}$$

Using the level based probability notation from (2.63)–(2.64) into (2.46), one may write the likelihood function as

$$\begin{aligned} L(\theta^*) &= L[(\beta_1^*, \ldots, \beta_j^*, \ldots, \beta_{(J-1)}^*)|y] \\ &= \Pi_{\ell=1}^{p+1} \Pi_{i\in\ell}^{K_{[\ell]}} \frac{1!}{y_{i1}!y_{i2}!\ldots y_{iJ}!} \pi_{[i\in(\ell)]1}^{y_{i1}} \pi_{[i\in(\ell)]2}^{y_{i2}} \cdots, \pi_{[i\in(\ell)]J}^{y_{iJ}}, \end{aligned} \tag{2.65}$$

where $\beta_j^* = (\beta_{j0}, \beta_{j1}, \ldots, \beta_{jp})'$, and $K_{[\ell]}$ denotes the number of individuals with covariate level ℓ so that $\sum_{\ell=1}^{p+1} K_{[\ell]} = K$. Further suppose that $K_{[\ell]j}$ denote the number of individuals those belong to the jth response category with covariate level ℓ so that $\sum_{j=1}^{J} K_{[\ell]j} = K_{[\ell]}$. For convenience of writing the likelihood estimating equations, we have displayed these notations for cell counts and cell probabilities as in the $L \times J$ contingency Table 2.4.

Note that in our notation, the row dimension L refers to the combined L levels for the categorical covariates under consideration, and J refers to the number of categories of a response variable. By this token, in Chap. 4, a contingency table for a bivariate multinomial problem with L level categorical covariates will be referred to as the $L \times R \times J$ contingency table, where J refers to the number of categories of the multinomial response variable Y as in this chapter, and R refers to the number of categories of the other multinomial response variable Z, say. When this notational scheme is used, the contingency Table 2.2 with four categories for Y (physician visit status) but no covariates, has the dimension 1×4. Thus, for a model involving no covariates, $p = 0$, i.e., $L = 1$. Further, when there are, for example, two categorical covariates in the model one with $p_1 + 1$ levels and the other with $p_2 + 1$ levels, one uses $p_1 + p_2$ dummy covariates to represent these $L = (p_1 + 1)(p_2 + 1)$ levels.

Turning back to the likelihood function (2.65), because $y_{ij} = 1$ or 0, with $\sum_{j=1}^{J} y_{ij} = 1$, by using the cell counts from Table 2.4, one may re-express this likelihood function as

$$L(\theta^*) = L[(\beta_1^*, \ldots, \beta_j^*, \ldots, \beta_{(J-1)}^*|y] = \Pi_{\ell=1}^{p+1} (\pi_{[\ell]1})^{K_{[\ell]1}} \ldots (\pi_{[\ell]J})^{K_{[\ell]J}}. \tag{2.66}$$

which one will maximize to estimate the desired parameters in θ^*.

2.2.2.1 Product Multinomial Likelihood Based Estimating Equations with a Global Regression form Using all Parameters

In some situations, it may be appropriate to assume that the cell counts for a given level in Table 2.3 follow a multinomial distribution and the distributions corresponding to any two levels are independent. For example, in a gender related study, male and females may be interviewed separately and hence $K_{[\ell]}$ at ℓth level may be assumed to be known, and they may be distributed in J cells, i.e., J categories, following the multinomial distribution. Note that in this approach K is not needed to be known in advance, rather all values for $K_{[\ell]}$ together yield $\sum_{\ell=1}^{L} K_{[\ell]} = K$. Following Table 2.4, we write this multinomial probability function at level ℓ as

$$f(K_{[\ell]1}, \ldots, K_{[\ell](J-1)}) = \frac{K_{[\ell]}!}{K_{[\ell]1}! \ldots K_{[\ell]J}!} \Pi_{j=1}^{J} [\pi_{[\ell]j}]^{K_{[\ell]j}} = L_\ell, \tag{2.67}$$

yielding the product multinomial function as

$$L(\theta^*) = \Pi_{\ell=1}^{p+1} f(K_{[\ell]1}, \ldots, K_{[\ell](J-1)}) = \Pi_{\ell=1}^{p+1} L_\ell. \tag{2.68}$$

At a given level $\ell(\ell = 1, \ldots, p+1)$, one may then write the probabilities in (2.63)–(2.64) for all $j = 1, \ldots, J$, as

$$\pi_{[\ell]j} = \frac{\exp(x'_{[\ell]j}\theta^*)}{\sum_{g=1}^{J}\exp(x'_{[\ell]g}\theta^*)}, \tag{2.69}$$

where

$$\theta^* = [\beta_1^{*\prime},\ldots,\beta_j^{*\prime},\ldots,\beta_{J-1}^{*}{}']', \text{ with } \beta_j^* = [\beta_{j0},\beta_j']',$$

and $x'_{[\ell]j}$ is the jth $(j=1,\ldots,J)$ row of the $J \times (J-1)(p+1)$ matrix X_ℓ, defined for ℓth level as follows:

$$X_\ell = \begin{pmatrix} x'_{[\ell]1} \\ x'_{[\ell]2} \\ \cdot \\ x'_{[\ell](J-1)} \\ x'_{[\ell]J} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 01'_{\ell-1},1,01'_{p-\ell} & 0 & 01'_p & \cdot & 0 & 01'_p \\ 0 & 01'_p & 1 & 01'_{\ell-1},1,01'_{p-\ell} & \cdot & 0 & 01'_p \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 01'_p & 0 & 01'_p & \cdot & 1 & 01'_{\ell-1},1,01'_{p-\ell} \\ 0 & 01'_p & 0 & 01'_p & \cdot & 0 & 01'_p \end{pmatrix} \text{ for } \ell = 1,\ldots,p$$

$$X_{p+1} = \begin{pmatrix} x'_{[p+1]1} \\ x'_{[p+1]2} \\ \cdot \\ x'_{[p+1](J-1)} \\ x'_{[p+1]J} \end{pmatrix} = \begin{pmatrix} 1 & 01'_p & 0 & 01'_p & \cdot & 0 & 01'_p \\ 0 & 01'_p & 1 & 01'_p & \cdot & 0 & 01'_p \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 01'_p & 0 & 01'_p & \cdot & 1 & 01'_p \\ 0 & 01'_p & 0 & 01'_p & \cdot & 0 & 01'_p \end{pmatrix}. \tag{2.70}$$

Following (2.35), the likelihood function (2.68) yields the likelihood equations

$$\frac{\partial \log L(\theta^*)}{\partial \theta^*} = \sum_{\ell=1}^{p+1} X'_\ell \left[y_{[\ell]} - K_{[\ell]}\pi_{[\ell]}\right] = 0, \tag{2.71}$$

where

$$y_{[\ell]} = [K_{[\ell]1},\ldots,K_{[\ell]j},\ldots,K_{[\ell]J}]' \text{ and } \pi_{[\ell]} = [\pi_{[\ell]1},\ldots,\pi_{[\ell]j},\ldots,\pi_{[\ell]J}]',$$

and X_ℓ matrices for $\ell = 1,\ldots,p+1$ are given as in (2.70).

2.2.2.1.1 Likelihood Estimates and their Asymptotic Variances

Note that the likelihood estimating equation (2.35) was developed for the covariates free cases, that is, for the cases with $p = 0$, whereas the likelihood estimating equation (2.71) is developed for one covariate with $p+1$ levels, represented by p dummy covariates. Thus, the estimating equation (2.71) may be treated as a generalization of the estimating equation (2.35) to the p covariates case. Let $\hat{\theta}^*$ be the solution of $f(\theta^*) = 0$ in (2.71). Assuming that $\hat{\theta}^*_0$ is not a solution for $f(\theta^*) = 0$, but a trial estimate and hence $f(\hat{\theta}^*_0) \neq 0$, by similar calculations as in (2.36), the iterative equation for $\hat{\theta}^*$ is obtained as

$$\hat{\theta}^* = \hat{\theta}^*_0 - \left[\{f'(\theta^*)\}^{-1} f(\theta^*)\right]|_{\theta^*=\hat{\theta}^*_0}. \tag{2.72}$$

Further, by similar calculations as in (2.38), one obtains from (2.71) that

$$K_{[\ell]} \frac{\partial \pi_{[\ell]j}}{\partial \theta^{*\prime}} = K_{[\ell]} \left[\pi_{[\ell]j} x'_{[\ell]j} - \pi_{[\ell]j} \pi'_{[\ell]} X_\ell\right]. \tag{2.73}$$

Consequently, it follows from (2.71) that

$$\begin{aligned} f'(\theta^*) = -\sum_{\ell=1}^{p+1} K_{[\ell]} X'_\ell \frac{\partial \pi_{[\ell]}}{\partial \theta^{*\prime}} &= -\sum_{\ell=1}^{p+1} K_{[\ell]} X'_\ell \left[\mathrm{diag}(\pi_{[\ell]1}, \ldots, \pi_{[\ell]J}) - \pi_{[\ell]} \pi'_{[\ell]}\right] X_\ell \\ &= -\sum_{\ell=1}^{p+1} K_{[\ell]} X'_\ell \left[D_{\pi_{[\ell]}} - \pi_{[\ell]} \pi'_{[\ell]}\right] X_\ell, \end{aligned} \tag{2.74}$$

and the iterative equation (2.72) takes the form

$$\begin{aligned} \widehat{\theta^*}(r+1) = \widehat{\theta^*}(r) &+ \left[\sum_{\ell=1}^{p+1} K_{[\ell]} X'_\ell \left[D_{\pi_{[\ell]}} - \pi_{[\ell]} \pi'_{[\ell]}\right] X_\ell\right]^{-1} \\ &\times \left[\sum_{\ell=1}^{p+1} X'_\ell \left(y_{[\ell]} - K_{[\ell]} \pi_{[\ell]}\right)\right]_{\theta^*=\widehat{\theta^*}(r)}, \end{aligned} \tag{2.75}$$

yielding the final estimate $\widehat{\theta^*}$. The covariance matrix of $\widehat{\theta^*}$ has the formula

$$\mathrm{var}(\widehat{\theta^*}) = \left[\sum_{\ell=1}^{p+1} K_{[\ell]} X'_\ell \{D_{\pi_{[\ell]}} - \pi_{[\ell]} \pi'_{[\ell]}\} X_\ell\right]^{-1}. \tag{2.76}$$

2.2.2.2 Product Multinomial Likelihood Based Estimating Equations with Local (Level Specified) Regression form Using Level Based Parameters

Note that in the last two sections, regression parameters were grouped category wise, that is, $\theta^* = [\beta_1^{*\prime}, \ldots, \beta_j^{*\prime}, \ldots, \beta_{J-1}^{*\prime}]'$, where $\beta_j^* = [\beta_{j0}, \beta_{j1}, \ldots, \beta_{jp}]'$ is formed corresponding to the covariates from all $p+1$ levels under the jth category response. Under product multinomial approach, it however makes more sense to group the parameters for all categories together under a given level $\ell(\ell = 1, \ldots, p+1)$, and write the estimating equations for these parameters of the multinomial distribution corresponding to the level ℓ, and then combine all estimating equations for overall parameters. Thus, we first use

$$\theta_\ell == \begin{cases} (\beta_{10}, \cdots, \beta_{J-1,0}, \beta_{1\ell}, \cdots, \beta_{J-1,\ell})' = (\beta_0', \beta_\ell')' : 2(J-1) \times 1, & \text{for } \ell = 1, \ldots, p \\ (\beta_{10}, \cdots, \beta_{J-1,0})' = \beta_0 : (J-1) \times 1 & \text{for } \ell = p+1, \end{cases}$$

and define

$$log\, m_{\ell j} = \tilde{x}'_{[\ell]j} \theta_\ell$$

satisfying the probability formulas

$$\pi_{[\ell]j} = \frac{m_{\ell j}}{\sum_{j=1}^J m_{\ell j}} = \frac{\exp(\tilde{x}'_{[\ell]j} \theta_\ell)}{\sum_{j=1}^J \exp(\tilde{x}'_{[\ell]j} \theta_\ell)}$$

in (2.63)–(2.64) for all $j = 1, \ldots, J$ at a given level ℓ. In regression form, it is equivalent to construct the $J \times 2(J-1)$ dummy covariate matrix $\tilde{X}_\ell$ for $\ell = 1, p$, and $J \times (J-1)$ dummy covariate matrix $\tilde{X}_{p+1}$, so that

$$\log \tilde{m}_\ell = [\log m_{\ell 1}, \cdots, \log m_{\ell j}, \cdots, \log m_{\ell J}]' = \tilde{X}_\ell \theta_\ell.$$

It follows that $\tilde{X}_\ell$ must have the form

$$\tilde{X}_\ell = \begin{pmatrix} 1\ 0\ 0 \cdot 0\ 1\ 0 \cdot 0 \\ 0\ 1\ 0 \cdot 0\ 0\ 1 \cdot 0 \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ 0\ 0\ 0 \cdot 1\ 0\ 0 \cdot 1 \\ 0\ 0\ 0 \cdot 0\ 0\ 0 \cdot 0 \end{pmatrix} : J \times 2(J-1), \text{ for } \ell = 1, \ldots, p \tag{2.77}$$

$$\tilde{X}_\ell = \begin{pmatrix} 1\ 0\ 0 \cdot 0 \\ 0\ 1\ 0 \cdot 0 \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ 0\ 0\ 0 \cdot 1 \\ 0\ 0\ 0 \cdot 0 \end{pmatrix} : J \times (J-1), \text{ for } \ell = p+1. \tag{2.78}$$

By similar calculations as in (2.35) (no covariate case), it follows from (2.67) (for covariate level ℓ) that the likelihood equation for θ_ℓ has the form

$$f(\theta_\ell) = \tilde{X}'_\ell(y_{[\ell]} - K_{[\ell]}\pi_{[\ell]}) = 0 \tag{2.79}$$

where $\tilde{X}_\ell$ have the forms as in (2.77) and (2.78), and

$$y_{[\ell]} = [K_{[\ell]1}, \ldots, K_{[\ell]j}, \ldots, K_{[\ell]J}]' \text{ and } \pi_{[\ell]} = [\pi_{[\ell]1}, \ldots, \pi_{[\ell]j}, \ldots, \pi_{[\ell]J}]'.$$

We then write a vector of distinct parameters, say θ, collecting them from all levels and append the estimating equation (2.79) for θ_ℓ to the final estimating equation for θ simply by using the chain rule of derivatives. In the present case for a single categorical covariate with $p+1$ levels, the θ vector can be written as

$$\theta = [\beta'_0, \beta'_1, \ldots, \beta'_\ell, \ldots, \beta'_p]' : (J-1)(p+1) \times 1,$$

with

$$\beta_0 = (\beta_{10}, \ldots, \beta_{j0}, \ldots, \beta_{(J-1)0})' \text{ and } \beta_\ell = (\beta_{1\ell}, \ldots, \beta_{j\ell}, \ldots, \beta_{(J-1)\ell})' \text{ for } \ell = 1, \ldots, p,$$

and by appending (2.79), the likelihood estimating equation for θ has the form

$$\begin{aligned} f(\theta) &= \sum_{\ell=1}^{p+1} [\frac{\partial \theta'_\ell}{\partial \theta} \tilde{X}'_\ell(y_{[\ell]} - K_{[\ell]}\pi_{[\ell]}) \\ &= \sum_{\ell=1}^{p+1} Q_\ell \tilde{X}'_\ell(y_{[\ell]} - K_{[\ell]}\pi_{[\ell]}) = 0, \end{aligned} \tag{2.80}$$

where Q_ℓ, for $\ell = 1, \ldots, p$, is the $(p+1)(J-1) \times 2(J-1)$ matrix and of dimension $(p+1)(J-1) \times (J-1)$ for $\ell = p+1$. These coefficient matrices are given by

$$Q_\ell = \begin{pmatrix} I_{J-1} & 0_{(J-1)\times(J-1)} \\ 0_{(\ell-1)(J-1)\times(J-1)} & 0_{(\ell-1)(J-1)\times(J-1)} \\ 0_{(J-1)\times(J-1)} & I_{J-1} \\ 0_{(p-\ell)(J-1)\times(J-1)} & 0_{(p-\ell)(J-1)\times(J-1)} \end{pmatrix} \text{ for } \ell = 1, \ldots, p$$

$$Q_{p+1} = \begin{pmatrix} I_{J-1} \\ 0_{(p)(J-1)\times(J-1)} \end{pmatrix}.$$

2.2.2.3 Illustration 2.3 (Continuation of Illustration 2.2): Partitioning the Product Binary ($J = 2$) Likelihood into Four Groups Corresponding to Four Nominal Levels of the Snoring Covariate

Note that Table 2.3 shows that the data were collected from 2,484 independent individuals. Because the individual status was recorded with regard to both snoring and heart disease problems, it is reasonable to consider the snoring status and heart disease status as two response variables. One would then analyze this data set by using a bivariate multinomial model to be constructed by accommodating the correlation between two multinomial response variables. This will be discussed in Chap. 5.

2.2.2.3.1 Product Binomial Approach

If one is, however, interested to examine the effect of snoring levels on the heart disease status, then the same data set may be analyzed by conditioning on the snoring levels and fitting a binary distribution at a given snoring level. This leads to a product binomial model that we use in this section to fit this snoring and heart disease data. To be specific, following the notations from Sect. 2.2.1.1(b), let $K_{[\ell]}$ be the number of individuals at ℓth snoring level. The responses of these individuals are distributed into two categories with regard to the heart disease problem. Thus the two cell counts at level ℓ will follow a binomial distribution. More specifically, because the $L = p+1 = 4$ levels are non-overlapping, one may first rewrite the product binary likelihood function (2.65) as

$$L^*[(\beta_{10},\beta_{11},\beta_{12},\beta_{13})|y] = \Pi_{\ell=1}^{p+1}\Pi_{i=1}^{K_{[\ell]}}\frac{1!}{y_{i1}!y_{i2}!}\pi_{[i\in(\ell)]1}^{y_{i1}}\pi_{[i\in(\ell)]2}^{y_{i2}}, \tag{2.81}$$

where

$$y_{i2} = 1-y_{i1} \text{ and } \pi_{[i\in(\ell)]2} = 1-\pi_{[i\in(\ell)]1}.$$

Further note that because $\pi_{[i\in(\ell)]j}$ is not a function of i any more, without any loss of generality we denote this by $\pi_{[\ell]j}$, for $j = 1,2$. Also suppose that $\pi_{[\ell]2} = 1-\pi_{[\ell]1}$, and $K_{[\ell]1}+K_{[\ell]2} = K_{[\ell]}$, where $\sum_{i=1}^{K_\ell} y_{i1} = K_{[\ell]1}$. When these notations are used, the binary likelihood function from (2.81) at a given level ℓ reduces to the binomial distribution

$$f(K_{[\ell]1}) = \frac{K_{[\ell]}!}{K_{[\ell]1}!K_{[\ell]2}!}(\pi_{[\ell]1})^{K_{[\ell]1}}(\pi_{[\ell]2})^{K_{[\ell]2}}, \tag{2.82}$$

for all $\ell = 1,\ldots,p+1$, yielding the product binomial likelihood as

$$L[(\beta_{10}, \beta_{11}, \beta_{12}, \beta_{13})|K] = \Pi_{\ell=1}^{p+1} \frac{K_{[\ell]}!}{K_{[\ell]1}!K_{[\ell]2}!} (\pi_{[\ell]1})^{K_{[\ell]1}} (\pi_{[\ell]2})^{K_{[\ell]2}}. \tag{2.83}$$

This product binomial (2.83) likelihood function may be maximized to obtain the likelihood estimates for the parameters involved, i.e., for $\beta_{10}, \beta_{11}, \beta_{12}$, and β_{13}.

2.2.2.3.1 (a) Estimating Equations: Global Regression Approach

Because in this example $J = 2$ and $p+1 = 4$, it follows from (2.69) that

$$\theta^* = (\beta_{10}, \beta_{11}, \beta_{12}, \beta_{13})',$$

and by following (2.70), one writes

$$X_1 = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$X_2 = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$X_3 = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$X_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Thus, by (2.71), the estimating equation for θ^* has the form

$$\sum_{\ell=1}^{p+1} X'_\ell \left[y_{[\ell]} - K_{[\ell]} \pi_{[\ell]} \right] = 0, \tag{2.84}$$

where

$$y_{[\ell]} = \begin{pmatrix} K_{[\ell]1} \\ K_{[\ell]2} \end{pmatrix} \text{ and } \pi_{[\ell]} = \begin{pmatrix} \pi_{[\ell]1} \\ \pi_{[\ell]2} \end{pmatrix}.$$

Note that for this heart disease and snoring relationship problem, the data and probabilities in terms of global parameters $\theta^* = (\beta_{10}, \beta_{11}, \beta_{12}, \beta_{13})'$ are given by
Level 1 (Occasional snoring):
Response count: $K_{[1]1} = 35,\ K_{[1]2} = 603,\ K_{[1]} = 638.$
Probabilities: $\pi_{[1]1} = \frac{\exp(\beta_{10}+\beta_{11})}{1+\exp(\beta_{10}+\beta_{11})}$, and $\pi_{[1]2} = \frac{1}{1+\exp(\beta_{10}+\beta_{11})}$.

Global regression form: $\pi_{[1]1} = \frac{\exp(x'_{[1]1}\theta^*)}{\sum_{g=1}^{2}\exp(x'_{[1]g}\theta^*)}$, and $\pi_{[1]2} = \frac{\exp(x'_{[1]2}\theta^*)}{\sum_{g=1}^{2}\exp(x'_{[1]g}\theta^*)}$, where $x'_{[1]1}$, for example, is the first row vector of the above written $X_1 : 2 \times 4$ matrix.
Level 2 (Nearly every night snoring):
Response count: $K_{[2]1} = 21,\ K_{[2]2} = 192,\ K_{[2]} = 213.$
Probabilities: $\pi_{[2]1} = \frac{\exp(\beta_{10}+\beta_{12})}{1+\exp(\beta_{10}+\beta_{12})}$, and $\pi_{[2]2} = \frac{1}{1+\exp(\beta_{10}+\beta_{12})}$.
Global regression form: $\pi_{[2]1} = \frac{\exp(x'_{[2]1}\theta^*)}{\sum_{g=1}^{2}\exp(x'_{[2]g}\theta^*)}$, and $\pi_{[2]2} = \frac{\exp(x'_{[2]2}\theta^*)}{\sum_{g=1}^{2}\exp(x'_{[2]g}\theta^*)}$.
Level 3 (Every night snoring):
Response count: $K_{[3]1} = 30,\ K_{[3]2} = 224,\ K_{[3]} = 254.$
Probabilities: $\pi_{[3]1} = \frac{\exp(\beta_{10}+\beta_{13})}{1+\exp(\beta_{10}+\beta_{13})}$, and $\pi_{[3]2} = \frac{1}{1+\exp(\beta_{10}+\beta_{13})}$.
Global regression form: $\pi_{[3]1} = \frac{\exp(x'_{[3]1}\theta^*)}{\sum_{g=1}^{2}\exp(x'_{[3]g}\theta^*)}$, and $\pi_{[3]2} = \frac{\exp(x'_{[3]2}\theta^*)}{\sum_{g=1}^{2}\exp(x'_{[3]g}\theta^*)}$.
Level 4 (Never snoring):
Response count: $K_{[4]1} = 24,\ K_{[4]2} = 1355,\ K_{[4]} = 1379.$
Probabilities: $\pi_{[4]1} = \frac{\exp(\beta_{10})}{1+\exp(\beta_{10})}$, and $\pi_{[4]2} = \frac{1}{1+\exp(\beta_{10})}$.
Global regression form: $\pi_{[4]1} = \frac{\exp(x'_{[4]1}\theta^*)}{\sum_{g=1}^{2}\exp(x'_{[4]g}\theta^*)}$, and $\pi_{[4]2} = \frac{\exp(x'_{[4]2}\theta^*)}{\sum_{g=1}^{2}\exp(x'_{[4]g}\theta^*)}$.

2.2.2.3.1 (b) Estimating Equations: Local Regression Approach

For convenience we rewrite the binary probabilities under all four levels as
Level 1 (Occasional snoring):
Probabilities: $\pi_{[1]1} = \frac{\exp(\beta_{10}+\beta_{11})}{1+\exp(\beta_{10}+\beta_{11})}$, and $\pi_{[1]2} = \frac{1}{1+\exp(\beta_{10}+\beta_{11})}$.
Local regression parameters: $\theta_1 = (\beta_{10}, \beta_{11})'$.
Local regression form: $\pi_{[1]1} = \frac{\exp(\tilde{x}'_{[1]1}\theta_1)}{\sum_{g=1}^{2}\exp(\tilde{x}'_{[1]g}\theta_1)}$, and $\pi_{[1]2} = \frac{\exp(\tilde{x}'_{[1]2}\theta_1)}{\sum_{g=1}^{2}\exp(\tilde{x}'_{[1]g}\theta_1)}$, yielding the $\tilde{X}_1 : J \times 2(J-1)$ matrix (see (2.78)) as

$$\tilde{X}_1 = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}.$$

Level 2 (Nearly every night snoring):
Probabilities: $\pi_{[2]1} = \frac{\exp(\beta_{10}+\beta_{12})}{1+\exp(\beta_{10}+\beta_{12})}$, and $\pi_{[2]2} = \frac{1}{1+\exp(\beta_{10}+\beta_{12})}$.
Local regression parameters: $\theta_2 = (\beta_{10}, \beta_{12})'$.
Local regression form: $\pi_{[2]1} = \frac{\exp(\tilde{x}'_{[2]1}\theta_2)}{\sum_{g=1}^{2}\exp(\tilde{x}'_{[2]g}\theta_2)}$, and $\pi_{[2]2} = \frac{\exp(\tilde{x}'_{[2]2}\theta_2)}{\sum_{g=1}^{2}\exp(\tilde{x}'_{[2]g}\theta_2)}$, yielding the $\tilde{X}_2 : J \times 2(J-1)$ matrix (see (2.78)) as

$$\tilde{X}_2 = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}.$$

Level 3 (Every night snoring):
Probabilities: $\pi_{[3]1} = \frac{\exp(\beta_{10}+\beta_{13})}{1+\exp(\beta_{10}+\beta_{13})}$, and $\pi_{[3]2} = \frac{1}{1+\exp(\beta_{10}+\beta_{13})}$.
Local regression parameters: $\theta_3 = (\beta_{10}, \beta_{13})'$.
Local regression form: $\pi_{[3]1} = \frac{\exp(\tilde{x}'_{[3]1}\theta_3)}{\sum_{g=1}^{2}\exp(\tilde{x}'_{[3]g}\theta_3)}$, and $\pi_{[3]2} = \frac{\exp(\tilde{x}'_{[3]2}\theta_3)}{\sum_{g=1}^{2}\exp(\tilde{x}'_{[3]g}\theta_3)}$, yielding the $\tilde{X}_3 : J \times 2(J-1)$ matrix (see (2.78)) as

$$\tilde{X}_3 = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}.$$

Level 4 (Never snoring):
Probabilities: $\pi_{[4]1} = \frac{\exp(\beta_{10})}{1+\exp(\beta_{10})}$, and $\pi_{[4]2} = \frac{1}{1+\exp(\beta_{10})}$.
Local regression parameters: $\theta_4 = (\beta_{10})$.
Local regression form: $\pi_{[4]1} = \frac{\exp(\tilde{x}'_{[4]1}\theta_4)}{\sum_{g=1}^{2}\exp(\tilde{x}'_{[4]g}\theta_4)}$, and $\pi_{[4]2} = \frac{\exp(\tilde{x}'_{[4]2}\theta_4)}{\sum_{g=1}^{2}\exp(\tilde{x}'_{[4]g}\theta_4)}$, yielding the $\tilde{X}_4 : J \times (J-1)$ matrix (see (2.79)) as

$$\tilde{X}_4 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

The likelihood estimating equation for

$$\theta_\ell = \begin{cases} (\beta_0', \beta_\ell')' = (\beta_{10}, \beta_{1\ell})' : 2(J-1)\times 1; & \text{for } \ell = 1, \ldots, p, \\ \beta_{10} & \text{for } \ell = p+1 = 4, \end{cases}$$

by (2.79), has the form

$$\tilde{X}'_\ell (y_{[\ell]} - K_{[\ell]}\pi_{[\ell]}) = 0,$$

for $\ell = 1, \ldots, 4$. Next in this special case with

$$\theta = (\beta_0', \beta_1', \ldots, \beta_p')' = (\beta_{10}, \beta_{11}, \beta_{12}, \beta_{13})'$$

the estimating equation for this parameter θ, by (2.80), has the form

$$\sum_{\ell=1}^{4} Q_\ell \tilde{X}'_\ell (y_{[\ell]} - K_{[\ell]}\pi_{[\ell]}) = 0, \tag{2.85}$$

where

$$Q_\ell = \frac{\partial \theta'_\ell}{\partial \theta},$$

for all $\ell = 1,\ldots,4$, have the forms

$$Q_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, Q_2 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}, Q_3 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}, Q_4 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

2.2.2.3.1 (c) Equivalence of the Likelihood Equations (2.59)(i), (2.84)(ii), and (2.85)(iii)

(i) The estimating equation (2.59) has the form

$$\sum_{i=1}^{K} \begin{pmatrix} 1 \\ w_{i1} \\ w_{i2} \\ w_{i3} \end{pmatrix} [y_{i1} - \pi_i] = 0,$$

which, by using

$$\begin{aligned}
&\sum_{i=1}^{K} y_{i1} = K_{[\cdot]1}, \ \sum_{i=1}^{K} \pi_{(i)1} = \sum_{\ell=1}^{4} K_{[\ell]}\pi_{[\ell]1} \\
&\sum_{i=1}^{K} y_{i1} z_{i1} = K_{[1]1}, \ \sum_{i=1}^{K} w_{i1}\pi_{(i)1} = K_{[1]}\pi_{[1]1} \\
&\sum_{i=1}^{K} y_{i1} w_{i2} = K_{[2]1}, \ \sum_{i=1}^{K} w_{i2}\pi_{(i)1} = K_{[2]}\pi_{[2]1} \\
&\sum_{i=1}^{K} y_{i1} w_{i3} = K_{[3]1}, \ \sum_{i=1}^{K} w_{i3}\pi_{(i)1} = K_{[3]}\pi_{[3]1},
\end{aligned} \tag{2.86}$$

reduces to

$$\begin{pmatrix} K_{[\cdot]1} - \sum_{\ell=1}^{4} K_{[\ell]}\pi_{[\ell]1} \\ K_{[1]1} - K_{[1]}\pi_{[1]1} \\ K_{[2]1} - K_{[2]}\pi_{[2]1} \\ K_{[3]1} - K_{[3]}\pi_{[3]1} \end{pmatrix} = 0. \tag{2.87}$$

(ii) Next, the estimating equation in (2.84) has the form

$$\sum_{\ell=1}^{p+1} X'_\ell \left[y_{[\ell]} - K_{[\ell]}\pi_{[\ell]} \right] = 0,$$

which, for convenience, we re-express as

$$\begin{pmatrix}1&0\\1&0\\0&0\\0&0\end{pmatrix}\begin{pmatrix}K_{[1]1}-K_{[1]}\pi_{[1]1}\\K_{[1]2}-K_{[1]}\pi_{[1]2}\end{pmatrix}+\begin{pmatrix}1&0\\0&0\\1&0\\0&0\end{pmatrix}\begin{pmatrix}K_{[2]1}-K_{[2]}\pi_{[2]1}\\K_{[2]2}-K_{[2]}\pi_{[2]2}\end{pmatrix}$$
$$+\begin{pmatrix}1&0\\0&0\\0&0\\1&0\end{pmatrix}\begin{pmatrix}K_{[3]1}-K_{[3]}\pi_{[3]1}\\K_{[3]2}-K_{[3]}\pi_{[3]2}\end{pmatrix}+\begin{pmatrix}1&0\\0&0\\0&0\\0&0\end{pmatrix}\begin{pmatrix}K_{[4]1}-K_{[4]}\pi_{[4]1}\\K_{[4]2}-K_{[4]}\pi_{[4]2}\end{pmatrix}=0. \quad (2.88)$$

After a simple algebra, (2.88) reduces to (2.87).

(iii) Further, the estimating equation (2.85) has the form

$$\sum_{\ell=1}^{p+1}Q_\ell\tilde{X}'_\ell(y_{[\ell]}-K_{[\ell]}\pi_{[\ell]})=0.$$

Now to see that this estimating equation is the same as (2.88), one has to simply verify that $Q_\ell\tilde{X}'_\ell=X'_\ell$ for $\ell=1,\ldots,p+1$. As the algebra below shows, this equality holds. Here

$$Q_1\tilde{X}'_1=\begin{pmatrix}1&0\\0&1\\0&0\\0&0\end{pmatrix}\begin{pmatrix}1&0\\1&0\end{pmatrix}=\begin{pmatrix}1&0\\1&0\\0&0\\0&0\end{pmatrix}=X'_1,$$

$$Q_2\tilde{X}'_2=\begin{pmatrix}1&0\\0&0\\0&1\\0&0\end{pmatrix}\begin{pmatrix}1&0\\1&0\end{pmatrix}=\begin{pmatrix}1&0\\0&0\\1&0\\0&0\end{pmatrix}=X'_2,$$

$$Q_3\tilde{X}'_3=\begin{pmatrix}1&0\\0&0\\0&0\\0&1\end{pmatrix}\begin{pmatrix}1&0\\1&0\end{pmatrix}=\begin{pmatrix}1&0\\0&0\\0&0\\1&0\end{pmatrix}=X'_3,$$

$$Q_4\tilde{X}'_4=\begin{pmatrix}1\\0\\0\\0\end{pmatrix}\begin{pmatrix}1&0\end{pmatrix}=\begin{pmatrix}1&0\\0&0\\0&0\\0&0\end{pmatrix}=X'_4.$$

Hence, as expected, all three estimating equations are same. Note that the estimating equation (2.59) requires individual level information, whereas the estimating equations (2.84) and (2.85) are based on grouped or contingency type data. Between

(2.84) and (2.85), it is easier to construct the $\tilde{X}_\ell$ matrices in (2.85) as coefficients of the local level parameters than constructing similar matrices X_ℓ for (2.84) corresponding to global parameters. However, unlike in (2.85), there is no need of constructing the chain derivative matrices Q_ℓ, in (2.84). Thus, it is up to the users to choose between (2.84) and (2.85). In this book we will mostly follow the global parameters based estimating equation (2.84).

2.2.2.3.1 (d) Illustration 2.3 Continued: Application of the Product Binomial Model to the Snoring and Heart Disease Problem

Forming the $X_\ell(\ell = 1,\ldots,4)$ matrices and writing the probabilities in global regression form as in Sect. 2.2.2.3.1(a), and using the 4×2 cross-table data from Table 2.3, we now solve the likelihood estimating equation (2.85) (see also (2.71)) using the iterative equation (2.75). To be specific, because the observed probabilities under category one (having heart disease) are relatively much smaller as compared to those under category two, starting with an initial value of $\beta_{10,0} = -3.0$ and small positive initial values for other parameters ($\beta_{11,0} = \beta_{12,0} = \beta_{13,0} = 0.10$), the iterative equation (2.75) yielded converged estimates for these four parameters in five iterations. These estimates were then used in (2.76) to compute the estimated variances and pair-wise covariances of the estimators. The estimates and their corresponding estimated standard errors are given in Table 2.5 below.

Note that as the snoring status is considered to be a fixed covariate (as opposed to a response variable) with four levels, the heart disease status of an individual follow a binary distribution at a given level. For example, it is clear from Sect. 2.2.2.3.1(a) that an individual belonging to level 4, i.e., who snores every night (see Table 2.3), has the probability $\pi_{[4]1} = \frac{\exp(\beta_{10})}{1+\exp(\beta_{10})}$ for having a heart disease.

Table 2.5 Parameter estimates for the snoring and heart disease data of Table 2.3

	Regression parameters			
Quantity	β_{10}	β_{20}	β_{30}	β_{40}
Estimate	−4.034	1.187	1.821	2.023
Standard error	0.206	0.269	0.309	0.283

Table 2.6 Observed and estimated probabilities for the snoring and heart disease data

	Heart disease			
	Yes		No	
Snoring level	**Observed**	**Estimated**	**Observed**	**Estimated**
Occasionally	0.055	0.055	0.945	0.945
Nearly every night	0.099	0.099	0.901	0.901
Every night	0.118	0.118	0.882	0.882
Never	0.017	0.017	0.983	0.983

These probabilities at all four levels may now be estimated by using the parameter estimates from Table 2.5. These estimated probabilities along with their respective observed probabilities are shown in Table 2.6.

Notice from the results in Table 2.6 that there is no difference between the observed and estimated probabilities at any snoring level. This result is expected because of the fact that the product binomial model is constructed with four independent regression parameters to fit data in four independent cells. This type of models are known as saturated models. In summary, the product binomial model and the estimation of its parameters by using the likelihood approach appear to be perfect for both fitting and interpretation of the data. Note that the observed and estimated probabilities appear to support that as the snoring level increases the probability for an individual to have a heart disease gets larger.

Remark that when the same snoring data is analyzed by using the snoring as a covariate with arbitrary codes, as it was done in Sect. 2.2.1.1(a) following Agresti (2002, Section 4.2.3, p. 121–123), one obtains the estimated probabilities for an individual to have a heart disease as

$$0.049,\ 0.094,\ 0.134,\ 0.020$$

based on individual's corresponding snoring level: occasional; nearly every night; every night; or never. Agresti (2002, Table 4.2) reported these probabilities as

$$0.044,\ 0.093,\ 0.132,\ 0.021,$$

which are slightly different. In any case, these estimated probabilities, as opposed to the estimated probabilities shown in Table 2.6, appear to be far apart from the corresponding observed probabilities under the 'yes' heart disease category. Thus, it is recommended not to use any modeling approach based on arbitrary coding for the fixed categorical covariates.

2.2.2.4 Illustrations Using Multinomial Regression Models Involving Responses with $J > 2$ Categories Along with One Two Levels Categorical Covariate

2.2.2.4.1 Illustration 2.4: Analysis of $2 \times J(=3)$ Aspirin and Heart Attacks Data Using Product Multinomial Approach

To illustrate the application of product multinomial model (2.68)–(2.69) we revisit here the aspirin use and heart attack data set earlier described by Agresti (2002, Section 2.1.1), for example, using a full multinomial (or log linear model) approach. We reproduce the data set below in Table 2.7. We describe and analyze this data using product multinomial approach.

This data set was originally recorded from a report on the relationship between aspirin use and heart attacks by the Physicians Health Study Research Group at

Table 2.7 Cross-classification of aspirin use and myocardial infarction

	Myocardial infarction			
	Fatal Attack	Non-Fatal Attack	No Attack	Total
Aspirin	5	99	10,933	11,037
Placebo	18	171	10,845	11,034
Total	23	270	21,778	22,071

Harvard Medical School. The Physicians Health Study was a 5 year randomized study of whether regular aspirin intake reduces mortality from cardiovascular disease. A physician participating in the study took either one aspirin tablet or a placebo, every other day, over the 5 year study period. This was a blind study for the participants as they did not know whether they were taking aspirin or a placebo for all these 5 years. By considering the heart attack status as one multinomial response variable with three categories (fatal, non-fatal, and no attacks) and the treatment as another multinomial response variable with two categories (placebo and aspirin use), Agresti (2002) used a full multinomial approach and described the association (correlation equivalent) between the two variables through computing certain odds ratios. In notation, let $z_i = (z_{i1}, \ldots, z_{ir}, \ldots, z_{i,R-1})'$ be the second multinomial response, but, with R categories, so that when z_i is realized at the rth category, one writes

$$z_i^{(r)} = (01'_{r-1}, 1, 01'_{R-1-r})', \text{ for } r = 1, \ldots, R-1; \text{ and } z_i^{(R)} = 01'_{R-1}.$$

Then many existing approaches write the joint probabilities, for example, for the aspirin use and heart attack data, as

$$\begin{aligned}\pi_{rj} &= P[z_i = z_i^{(r)}, y_i = y_i^{(j)}], \text{for all } i = 1, \ldots, 22071,\ r = 1, 2,\ j = 1, \ldots, 3 \\ &= \frac{\exp(\alpha_r + \beta_j + \phi_{rj})}{\sum_{r=1}^{2} \sum_{j=1}^{3} \exp(\alpha_r + \beta_j + \phi_{rj})} \\ &= \frac{m_{rj}}{\sum_{r=1}^{2} \sum_{j=1}^{3} m_{rj}} = \frac{m_{rj}}{m},\end{aligned} \tag{2.89}$$

where α_r is the rth category effect of the z variable, β_j is the jth category effect of the y variable, and ϕ_{rj} is the corresponding interaction effect of y and z variables, on any individual. These parameters are restricted by the dependence of the last category of each variable on their remaining independent categories. Thus, in this example, one may use

$$\alpha_2 = \beta_3 = \phi_{13} = \phi_{21} = \phi_{22} = \phi_{23} = 0,$$

and fit the full multinomial model to the data in Table 2.7 by estimating the parameters

$$\alpha_1, \beta_1, \beta_2, \phi_{11}, \text{ and } \phi_{12}.$$

The estimation is achieved by maximizing the full multinomial likelihood

$$L(\theta^*) = \Pi_{r=1}^{2} \Pi_{j=1}^{3} \pi_{rj}^{K_{rj}}, \tag{2.90}$$

with respect to $\theta^* = (\alpha_1, \beta_1, \beta_2, \phi_{11}, \phi_{12})'$, where K_{rj} is the number of individuals in the (r, j)th cell in Table 2.7, for example, $K_{12} = 99$.

This full multinomial approach, that is, considering the treatment as a response variable, lacks justification. This can be understood simply by considering a question that, under the study condition, can the response of one randomly chosen individual out of 22,071 participants belong to one of the six cells in the Table 2.7. This is not possible, because, even though, the placebo pill or aspirin was chosen by some one for a participant with a prior probability, the treatment was made fixed for an individual participant for the whole study period. Thus, treatment variable here must be considered as a fixed regression covariate with two levels. This prompted one to reanalyze this data set by using the product multinomial model (2.68)–(2.69) by treating heart attack status as the multinomial response variable only and the treatment as a categorical covariate with two levels. By this token, for both cross-sectional and longitudinal analysis, this book emphasizes on appropriate modeling for the categorical data by distinguishing categorical covariates from categorical responses.

Product multinomial global regression approach:

Turning back to the analysis of the categorical data in Table 2.7, following (2.69) we first write the multinomial probabilities at two levels of the treatment covariate as follows. Note that in notation of the model (2.69), for this heart attack and aspirin use data, we write $J = 3$ and $p + 1 = 2$, and

$$\theta^* = (\beta_{10}, \beta_{11}, \beta_{20}, \beta_{21})'.$$

When the model (2.68)–(2.69) is compared with (2.90)–(2.91), α_1 from the latter model is not needed. Also, even though

$$\beta_{10}, \beta_{20}, \beta_{11}, \beta_{21}$$

in the model (2.69) are, respectively, equivalent to the notations

$$\beta_1, \beta_2, \phi_{11}, \phi_{12}$$

of the model (2.90), they do not have, however, the same interpretation. This is because, β_{11} and β_{21} in (2.69) are simply regression effects of the covariate level 1 on first two categories, whereas ϕ_{11} and ϕ_{12} in (2.90) are treated to be association or odds ratio parameters. But, there is a definition problem with these odds ratio parameters in this situation, because treatment here cannot represent a response variable.

Now for the product multinomial model (2.68)–(2.69), one writes the level based $\{J \times (J-1)(p+1)\} \equiv \{3 \times 4\}$ covariate matrices as

$$X_1 = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$X_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Then the cell probabilities and their forms in terms of the global parameters $\theta^* = (\beta_{10}, \beta_{11}, \beta_{20}, \beta_{21})'$ are given by
Level 1 (Aspirin user):
Response count: $K_{[1]1} = 5,\ K_{[1]2} = 99,\ K_{[1]3} = 10,933,\ K_{[1]} = 11,037.$
Probabilities:

$$\pi_{[1]1} = \frac{\exp(\beta_{10} + \beta_{11})}{1 + \sum_{g=1}^{2} \exp(\beta_{g0} + \beta_{g1})},\ \pi_{[1]2} = \frac{\exp(\beta_{20} + \beta_{21})}{1 + \sum_{g=1}^{2} \exp(\beta_{g0} + \beta_{g1})},$$
$$\pi_{[1]3} = \frac{1}{1 + \sum_{g=1}^{2} \exp(\beta_{g0} + \beta_{g1})}. \tag{2.91}$$

Global regression form:

$$\pi_{[1]1} = \frac{\exp(x'_{[1]1}\theta^*)}{\sum_{j=1}^{3} \exp(x'_{[1]j}\theta^*)},\ \pi_{[1]2} = \frac{\exp(x'_{[1]2}\theta^*)}{\sum_{j=1}^{3} \exp(x'_{[1]j}\theta^*)},$$
$$\pi_{[1]3} = \frac{\exp(x'_{[1]3}\theta^*)}{\sum_{j=1}^{3} \exp(x'_{[1]j}\theta^*)},$$

where $x'_{[1]2}$, for example, is the second row vector of the above written $X_1 : 3 \times 4$ matrix.
Level 2 (Placebo user):
Response count: $K_{[2]1} = 18,\ K_{[2]2} = 171,\ K_{[2]3} = 10,845,\ K_{[2]} = 11,034.$
Probabilities:

$$\pi_{[2]1} = \frac{\exp(\beta_{10})}{1 + \sum_{g=1}^{2} \exp(\beta_{g0})},\ \pi_{[2]2} = \frac{\exp(\beta_{20})}{1 + \sum_{g=1}^{2} \exp(\beta_{g0})},$$
$$\pi_{[2]3} = \frac{1}{1 + \sum_{g=1}^{2} \exp(\beta_{g0})}. \tag{2.92}$$

Table 2.8 Parameter estimates for the treatment and heart attack status data of Table 2.7

Quantity	Regression parameters			
	β_{10}	β_{11}	β_{20}	β_{21}
Estimate	−6.401	−1.289	−4.150	−0.555
Standard error	0.2360	0.5057	0.0771	0.1270

Global regression form:

$$\pi_{[2]1} = \frac{\exp(x'_{[2]1}\theta^*)}{\sum_{j=1}^{3}\exp(x'_{[2]j}\theta^*)},\ \pi_{[2]2} = \frac{\exp(x'_{[2]2}\theta^*)}{\sum_{j=1}^{3}\exp(x'_{[2]j}\theta^*)},$$

$$\pi_{[2]3} = \frac{\exp(x'_{[2]3}\theta^*)}{\sum_{j=1}^{3}\exp(x'_{[2]j}\theta^*)}.$$

Now following (2.71) and using the iterative equation (2.75), we solve the product multinomial based likelihood estimating equation

$$\sum_{\ell=1}^{2} X'_{\ell}\left[y_{[\ell]} - K_{[\ell]}\pi_{[\ell]}\right] = 0, \tag{2.93}$$

where

$$y_{[\ell]} = \begin{pmatrix} K_{[\ell]1} \\ K_{[\ell]2} \\ K_{[\ell]3} \end{pmatrix} \text{ and } \pi_{[\ell]} = \begin{pmatrix} \pi_{[\ell]1} \\ \pi_{[\ell]2} \\ \pi_{[\ell]3} \end{pmatrix}.$$

These estimates and their corresponding standard errors computed by using (2.76) are reported in Table 2.8.

In order to interpret these parameter estimates, notice from the formulas from the probabilities under level 2 (placebo group) that the values of β_{10} and β_{20} would determine the probabilities of a placebo user individual to be in the 'fatal attack' or 'non-fatal attack' group, as compared to $\beta_{30} = 0$ used for probability for the same individual to be in the reference group, that is, in the 'no attack' group. To be specific, when the large negative values of $\beta_{10}(= -6.401)$ and $\beta_{20}(= -4.150)$ are compared to $\beta_{30} = 0$, it becomes clear by (2.92) that the probability of a placebo user to be in the 'no attack' group is very large, as expected, followed by the probabilities for the individual to be in the 'non-fatal' and fatal groups, respectively. Further because the value of $\beta_{10} + \beta_{11}$ would determine the probability of an aspirin user in the 'fatal attack' group, the negative value of $\beta_{11}(= -1.289)$ shows that an aspirin user has smaller probability than a placebo user to be in the 'fatal attack' group. Other estimates can be interpreted similarly. Now by using these estimates from Table 2.8, the estimates for all three categorized multinomial probabilities in (2.91) under aspirin user treatment level, and in (2.92) under placebo user treatment level, may be computed. These estimated probabilities along

Table 2.9 Observed and estimated multinomial probabilities for the treatment versus heart attack status data of Table 2.7

		Myocardial Infarction			
	Proportion/	Fatal	Non-Fatal	No	
	Probability	Attack	Attack	Attack	Total
Aspirin	Observed	0.00045	0.00897	0.99058	1.00
	Estimated	0.00045	0.00897	0.99058	1.00
Placebo	Observed	0.00163	0.01550	0.98287	1.00
	Estimated	0.00163	0.01550	0.98287	1.00

with their counterpart (observed proportions) are displayed in Table 2.9. Similar to Table 2.6 for the snoring and heart disease problem data, it is clear from Table 2.9 that the observed and estimated probabilities are same. This happens because the four independent parameters, namely β_{10}, β_{11}, β_{20}, and β_{21} are used to define the probabilities in (2.91)–(2.92) (see also (2.68)–(2.69)) to fit four independent observations, two under aspirin user treatment level and another two under the placebo user treatment level. Thus a saturated model is fitted through solving the corresponding optimal likelihood estimating equations (2.93), and the parameter estimates shown in Table 2.8 are consistent and highly efficient (details are not discussed here).

Remark that the estimates of the regression parameters under two (independent) categories shown in Table 2.8 were obtained by applying the product multinomial estimating equation (2.93) to the observed data given in the contingency Table 2.7. However, because the data in this table are clearly laid out under each of the two treatment levels, one may easily reconstruct the individual level response and covariate information without any identity of the individual. Suppose that the treatment covariate is defined as

$$w_i = \begin{cases} 1 \text{ for aspirin taken by the } i\text{th individual} \\ 0 \text{ otherwise,} \end{cases}$$

and the multinomial response of this individual is given by

$$y_i = \begin{cases} (1,0)' \text{ if this ith individual had fatal attack} \\ (0,1)' \text{ if this ith individual had non fatal attack} \\ (0,0)' \text{ otherwise, i.e., if this ith individual had no attack.} \end{cases}$$

Consequently, one may directly solve the individual history based multinomial likelihood estimating equation (2.48) to obtain the same estimates (as in Table 2.8) of the regression parameters involved in the probability model (2.45).

Turning back to the results shown in Table 2.9, it is clear that the estimated proportion of individuals whose heart attack was either fatal or non-fatal is shown to

Table 2.10 Cross-classification of gender and physician visit

	Physician visit status				
Gender	None	Few	Not so few	High	Total
Male	28	40	16	12	96
Female	11	20	21	32	84
Total	38	62	36	44	180

be $(0.00045 + 0.00897) = 0.00942$ for the aspirin group, and $(0.00164 + 0.01562) = 0.01726$ for the placebo group, indicating the advantage of using aspirin as opposed to using none. This type of comparison is also available in Agresti (1990, Section 2.2.4, page 17), but by using only observed data. Later on it was emphasized in Agresti (2002, Section 2.1.1) for the comparison of the distribution of responses under each treatment level, but unlike in this section, no model was fitted.

2.2.2.4.2 Analysis of Physician Visits Data with $J = 4$ Categories

We continue to illustrate the application of the product multinomial likelihood models now by considering the physician visits data described in Table 2.1. We consider the physician visits in four categories: none, few, not so few, and high visits, as indicated in Table 2.2, whereas there were three categories for heart attack status in the treatment versus heart attack data considered in the last section. To be specific, for the physician visits data, we will fit the product multinomial models to examine the marginal effects of (1) gender; (2) chronic disease; and (3) education levels, on the physician visits, in the following Sects. 2.2.2.4.2(a), (b), and (c), respectively.

2.2.2.4.2 (a) Illustration 2.5: Analysis for Gender Effects on Physician Visits

To examine the gender effects on the physician visits we use the data from Table 2.1 and display them in the 2×4 contingency Table 2.10. For convenience of fitting the product multinomial model (2.67)–(2.69) to this data, for each gender level we write the multinomial observation, probabilities under each of four categories, and the global regression form along with corresponding covariate matrix, as follows:
Level 1 (Male):
Response count: $K_{[1]1} = 27,\ K_{[1]2} = 42,\ K_{[1]3} = 15,\ K_{[1]4} = 12,\ K_{[1]} = 96.$
Probabilities:

$$\pi_{[1]1} = \frac{\exp(\beta_{10} + \beta_{11})}{1 + \sum_{g=1}^{3} \exp(\beta_{g0} + \beta_{g1})},\ \pi_{[1]2} = \frac{\exp(\beta_{20} + \beta_{21})}{1 + \sum_{g=1}^{3} \exp(\beta_{g0} + \beta_{g1})},$$
$$\pi_{[1]3} = \frac{\exp(\beta_{30} + \beta_{31})}{1 + \sum_{g=1}^{3} \exp(\beta_{g0} + \beta_{g1})},\ \pi_{[1]4} = \frac{1}{1 + \sum_{g=1}^{3} \exp(\beta_{g0} + \beta_{g1})}. \quad (2.94)$$

Global regression form:

$$\pi_{[1]1} = \frac{\exp(x'_{[1]1}\theta^*)}{\sum_{j=1}^{4}\exp(x'_{[1]j}\theta^*)},\ \pi_{[1]2} = \frac{\exp(x'_{[1]2}\theta^*)}{\sum_{j=1}^{4}\exp(x'_{[1]j}\theta^*)},$$
$$\pi_{[1]3} = \frac{\exp(x'_{[1]3}\theta^*)}{\sum_{j=1}^{4}\exp(x'_{[1]j}\theta^*)},\ \pi_{[1]4} = \frac{\exp(x'_{[1]4}\theta^*)}{\sum_{j=1}^{4}\exp(x'_{[1]j}\theta^*)},$$

where

$$\theta^* = (\beta_{10}, \beta_{11}, \beta_{20}, \beta_{21}, \beta_{30}, \beta_{31})',$$

and $x'_{[1]3}$, for example, is the third row vector of the $X_1 : 4 \times 6$ matrix given by

$$X_1 = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{2.95}$$

Level 2 (Female):
Response count: $K_{[2]1} = 11,\ K_{[2]2} = 20,\ K_{[2]3} = 21,\ K_{[2]4} = 32,\ K_{[2]} = 84.$
Probabilities:

$$\pi_{[2]1} = \frac{\exp(\beta_{10})}{1+\sum_{g=1}^{3}\exp(\beta_{g0})},\ \pi_{[2]2} = \frac{\exp(\beta_{20})}{1+\sum_{g=1}^{3}\exp(\beta_{g0})},$$
$$\pi_{[2]3} = \frac{\exp(\beta_{30})}{1+\sum_{g=1}^{3}\exp(\beta_{g0})},\ \pi_{[2]4} = \frac{1}{1+\sum_{g=1}^{3}\exp(\beta_{g0})}. \tag{2.96}$$

Global regression form:

$$\pi_{[2]1} = \frac{\exp(x'_{[2]1}\theta^*)}{\sum_{j=1}^{4}\exp(x'_{[2]j}\theta^*)},\ \pi_{[2]2} = \frac{\exp(x'_{[2]2}\theta^*)}{\sum_{j=1}^{4}\exp(x'_{[2]j}\theta^*)},$$
$$\pi_{[2]3} = \frac{\exp(x'_{[2]3}\theta^*)}{\sum_{j=1}^{4}\exp(x'_{[2]j}\theta^*)},\ \pi_{[2]4} = \frac{\exp(x'_{[2]4}\theta^*)}{\sum_{j=1}^{4}\exp(x'_{[2]j}\theta^*)},$$

where θ^* remains the same as

$$\theta^* = (\beta_{10}, \beta_{11}, \beta_{20}, \beta_{21}, \beta_{30}, \beta_{31})',$$

and $x'_{[2]3}$, for example, is the third row vector of the $X_2 : 4 \times 6$ matrix given by

Table 2.11 Parameter estimates for the gender and physician visit status data of Table 2.10

	Regression parameters					
Quantity	β_{10}	β_{11}	β_{20}	β_{21}	β_{30}	β_{31}
Estimate	−1.068	1.915	−0.470	1.674	−0.421	0.709
Standard error	0.3495	0.4911	0.2850	0.4354	0.2808	0.4740

$$X_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{2.97}$$

Now following (2.71) and using the iterative equation (2.75), we solve the product multinomial based likelihood estimating equation

$$\sum_{\ell=1}^{2} X'_\ell \left[y_{[\ell]} - K_{[\ell]} \pi_{[\ell]} \right] = 0, \tag{2.98}$$

where

$$y_{[\ell]} = \begin{pmatrix} K_{[\ell]1} \\ K_{[\ell]2} \\ K_{[\ell]3} \\ K_{[\ell]4} \end{pmatrix} \text{ and } \pi_{[\ell]} = \begin{pmatrix} \pi_{[\ell]1} \\ \pi_{[\ell]2} \\ \pi_{[\ell]3} \\ \pi_{[\ell]3} \end{pmatrix}.$$

These estimates and their corresponding standard errors computed by using (2.76) are reported in Table 2.11.

Notice from (2.96) that the estimates of β_{10}, β_{20}, and β_{30} indicate the relative probability for a female to be in none, few, and not so few categories, respectively, as compared to the probability for high category determined by $\beta_{40} = 0$ (by assumption).

Because all three estimates are negative, the estimate for β_{10} being large negative, it follows that a female has the highest probability to be in 'high visit' group and smallest probability to be in the 'none' (never visited) group. By the same token, it follows from (2.94) that the largest value for $\beta_{20} + \beta_{21} = 1.204$ estimate as compared to its reference value 0.0 indicates that a male has the highest probability to be in the 'few visits' group. These probabilities can be verified from Table 2.12 where we have displayed the estimated as well as observed probabilities. In summary, the estimated probabilities in Table 2.12 show that a female visits the physician for more number of times as compared to a male. These results are in agreement with those of health care utilization study reported in Sutradhar (2011, Section 4.2.8) where

Table 2.12 Observed and estimated multinomial probabilities for the gender versus physician visits data of Table 2.10

		Physician visit status				
Gender	Probability	None	Few	Not so few	High	Total
Male	Observed	0.2917	0.4166	0.1667	0.1250	1.0
	Estimated	0.2917	0.4166	0.1667	0.1250	1.0
Female	Observed	0.1309	0.2381	0.2500	0.3810	1.0
	Estimated	0.1309	0.2381	0.2500	0.3810	1.0

Table 2.13 Cross-classification of chronic condition and physician visit

	Physician visit status				
Chronic condition	None	Few	Not so few	High	Total
Yes	13	25	21	33	92
No	26	35	16	11	88
Total	39	60	37	44	180

the actual number of visits (as opposed to visit category) were analyzed by fitting a familial/clustered model using the so-called generalized quasi-likelihood (GQL) estimation approach.

2.2.2.4.2 (b) Illustration 2.6: Analysis for Chronic Condition Effects on Physician Visits

To examine the chronic condition effects on the number of visits, we first display the physician visit data in the form a contingency (cross-classified) table. More specifically, the 2×4 cross-classified Table 2.13 shows the distribution of the number of the respondents under four visit categories at a given chronic condition level. The chronic condition covariate has two levels. One of the levels represents the individuals with no chronic disease, and the individuals with one or more chronic disease have been assigned to the other group (level). Note that because both Tables 2.10 and 2.13 contain one categorical covariate with two levels, the probability models for the data in Table 2.13 would be the same as that of Table 2.10. The only difference is in the names of the levels. For this reason we do not reproduce the probability formulas and the form of X_ℓ matrices. However because the data are naturally different we write them in notation as follows:
Chronic condition level 1 (Yes):
Response count: $K_{[1]1} = 13,\ K_{[1]2} = 25,\ K_{[1]3} = 21,\ K_{[1]4} = 33,\ K_{[1]} = 92.$
Chronic condition level 2 (No):
Response count: $K_{[2]1} = 25,\ K_{[2]2} = 37,\ K_{[2]3} = 15,\ K_{[2]4} = 11,\ K_{[2]} = 88.$
We then solve the product multinomial likelihood estimating equation (2.98). The estimates of the regression parameters involved in the probability formulas (2.94) and (2.96) for the cross-classified data in Table 2.13 are given in Table 2.14. The estimates probabilities are shown in Table 2.15.

Table 2.14 Parameter estimates for the chronic condition and physician visit status data of Table 2.13

Quantity	Regression parameters					
	β_{10}	β_{11}	β_{20}	β_{21}	β_{30}	β_{31}
Estimate	0.860	−1.792	1.157	−1.435	0.375	−0.827
Standard error	0.3597	0.4864	0.3457	0.4356	0.3917	0.4810

Table 2.15 Observed and estimated multinomial probabilities for the chronic condition versus physician visits data of Table 2.13

		Physician visit status				
Chronic condition	Probability	None	Few	Not so few	High	Total
Yes	Observed	0.1413	0.2717	0.2283	0.3587	1.0
	Estimated	0.1413	0.2717	0.2283	0.3587	1.0
No	Observed	0.2955	0.3977	0.1818	0.1250	1.0
	Estimated	0.2955	0.3977	0.1818	0.1250	1.0

Notice from (2.96) that the estimates of β_{10}, β_{20}, and β_{30}, would indicate the relative probability for an individual with no chronic disease to be in none, few, and not so few categories, respectively, as compared to the probability for being in high category determined by $\beta_{40} = 0$ (by assumption). Consequently,

$$\hat{\beta}_{20} = 1.157 > \hat{\beta}_{10} > \hat{\beta}_{30} > \beta_{40} = 0$$

indicates that an individual with no chronic disease has higher probability of paying no visits or a few visits, as compared to paying higher number of visits, which is expected. By the same token,

$$[\hat{\beta}_{10} + \hat{\beta}_{11} = 0.860 - 1.792 = -0.932] < [\hat{\beta}_{30} + \hat{\beta}_{31}] < [\hat{\beta}_{20} + \hat{\beta}_{21}] < 0,$$

indicates that an individual with chronic disease has higher probability of paying larger number of visits. Note however that these estimates also indicate that irrespective of chronic condition a considerably large number of individuals pay at least a few visits, which may be natural or due to other covariate conditions.

2.2.2.4.2 (c) Illustration 2.7: Analysis for Education Level Effects on Physician Visits

We continue to illustrate the application of product multinomial approach now by examining the marginal effects of education status of an individual on the physician visit. Three levels of education are considered, namely low (less than high school education), medium (high school education), and high (more than high school education). The cross-classified data for education level versus physician visits, obtained from Table 2.1, are shown in Table 2.16.

Table 2.16 Cross-classification of education level and physician visit

	Physician visit status				
Education level	None	Few	Not so few	High	Total
Low (< High School)	17	16	10	15	58
Medium (High School)	6	8	8	11	33
High (> High School)	16	36	19	18	89
Total	39	60	37	44	180

Notice that it is appropriate to consider the education level at a given year (1985) as a fixed covariate. Here this covariate has three levels and two dummy covariates can be used to represent these three levels of education. Now to construct the likelihood estimating equation (2.71), following (2.69), we, for convenience, present the probabilities and their regression form as follows.

Level 1 (Low education):

Response count: $K_{[1]1} = 17,\ K_{[1]2} = 16,\ K_{[1]3} = 10,\ K_{[1]4} = 15,\ K_{[1]} = 58.$

Probabilities:

$$\pi_{[1]1} = \frac{\exp(\beta_{10}+\beta_{11})}{1+\sum_{g=1}^{3}\exp(\beta_{g0}+\beta_{g1})},\ \pi_{[1]2} = \frac{\exp(\beta_{20}+\beta_{21})}{1+\sum_{g=1}^{3}\exp(\beta_{g0}+\beta_{g1})},$$
$$\pi_{[1]3} = \frac{\exp(\beta_{30}+\beta_{31})}{1+\sum_{g=1}^{3}\exp(\beta_{g0}+\beta_{g1})},\ \pi_{[1]4} = \frac{1}{1+\sum_{g=1}^{3}\exp(\beta_{g0}+\beta_{g1})}. \quad (2.99)$$

Global regression form: For

$$\theta^* = (\beta_{10},\beta_{11},\beta_{12},\beta_{20},\beta_{21},\beta_{22},\beta_{30},\beta_{31},\beta_{32})',$$

the above probabilities may be re-expressed as

$$\pi_{[1]1} = \frac{\exp(x'_{[1]1}\theta^*)}{\sum_{j=1}^{4}\exp(x'_{[1]j}\theta^*)},\ \pi_{[1]2} = \frac{\exp(x'_{[1]2}\theta^*)}{\sum_{j=1}^{4}\exp(x'_{[1]j}\theta^*)},$$
$$\pi_{[1]3} = \frac{\exp(x'_{[1]3}\theta^*)}{\sum_{j=1}^{4}\exp(x'_{[1]j}\theta^*)},\pi_{[1]4} = \frac{\exp(x'_{[1]4}\theta^*)}{\sum_{j=1}^{4}\exp(x'_{[1]j}\theta^*)},$$

where $x'_{[1]3}$, for example, is the third row vector of the $X_1 : 4\times 9$ matrix given by

$$X_1 = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \quad (2.100)$$

Level 2 (Medium education):
Response count: $K_{[2]1} = 6,\ K_{[2]2} = 8,\ K_{[2]3} = 8,\ K_{[2]4} = 11,\ K_{[2]} = 33.$
Probabilities:

$$\pi_{[2]1} = \frac{\exp(\beta_{10}+\beta_{12})}{1+\sum_{g=1}^{3}\exp(\beta_{g0}+\beta_{g2})},\ \pi_{[2]2} = \frac{\exp(\beta_{20}+\beta_{22})}{1+\sum_{g=1}^{3}\exp(\beta_{g0}+\beta_{g2})},$$
$$\pi_{[2]3} = \frac{\exp(\beta_{30}+\beta_{32})}{1+\sum_{g=1}^{3}\exp(\beta_{g0}+\beta_{g2})},\ \pi_{[2]4} = \frac{1}{1+\sum_{g=1}^{3}\exp(\beta_{g0}+\beta_{g2})}. \quad (2.101)$$

Global regression form:

$$\pi_{[2]1} = \frac{\exp(x'_{[2]1}\theta^*)}{\sum_{j=1}^{4}\exp(x'_{[2]j}\theta^*)},\ \pi_{[2]2} = \frac{\exp(x'_{[2]2}\theta^*)}{\sum_{j=1}^{4}\exp(x'_{[2]j}\theta^*)},$$
$$\pi_{[2]3} = \frac{\exp(x'_{[2]3}\theta^*)}{\sum_{j=1}^{4}\exp(x'_{[2]j}\theta^*)},\pi_{[2]4} = \frac{\exp(x'_{[2]4}\theta^*)}{\sum_{j=1}^{4}\exp(x'_{[2]j}\theta^*)},$$

where θ^* remains the same, but the covariate matrix X_2 is given by

$$X_2 = \begin{pmatrix} 1&0&1&0&0&0&0&0&0 \\ 0&0&0&1&0&1&0&0&0 \\ 0&0&0&0&0&0&1&0&1 \\ 0&0&0&0&0&0&0&0&0 \end{pmatrix}. \quad (2.102)$$

Level 3 (High education):
Response count: $K_{[3]1} = 16,\ K_{[2]2} = 36,\ K_{[3]3} = 19,\ K_{[3]4} = 18,\ K_{[3]} = 89.$
Probabilities:

$$\pi_{[3]1} = \frac{\exp(\beta_{10})}{1+\sum_{g=1}^{3}\exp(\beta_{g0})},\ \pi_{[3]2} = \frac{\exp(\beta_{20})}{1+\sum_{g=1}^{3}\exp(\beta_{g0})},$$
$$\pi_{[3]3} = \frac{\exp(\beta_{30})}{1+\sum_{g=1}^{3}\exp(\beta_{g0})},\ \pi_{[3]4} = \frac{1}{1+\sum_{g=1}^{3}\exp(\beta_{g0})}. \quad (2.103)$$

Global regression form:

$$\pi_{[3]1} = \frac{\exp(x'_{[3]1}\theta^*)}{\sum_{j=1}^{4}\exp(x'_{[3]j}\theta^*)},\ \pi_{[3]2} = \frac{\exp(x'_{[3]2}\theta^*)}{\sum_{j=1}^{4}\exp(x'_{[3]j}\theta^*)},$$
$$\pi_{[3]3} = \frac{\exp(x'_{[3]3}\theta^*)}{\sum_{j=1}^{4}\exp(x'_{[3]j}\theta^*)},\pi_{[3]4} = \frac{\exp(x'_{[3]4}\theta^*)}{\sum_{j=1}^{4}\exp(x'_{[3]j}\theta^*)},$$

Table 2.17 Parameter estimates for the education level and physician visit status data of Table 2.16

	Regression parameters								
Quantity	β_{10}	β_{11}	β_{12}	β_{20}	β_{21}	β_{22}	β_{30}	β_{31}	β_{32}
Estimate	−0.118	0.243	−0.488	0.693	−0.629	−1.012	0.054	−0.460	−0.373
Standard error	0.3436	0.4935	0.6129	0.2887	0.4610	0.5470	0.3289	0.5243	0.5693

where θ^* remains the same, but the covariate matrix X_3 is given by

$$X_3 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{2.104}$$

Now following (2.71) and using the iterative equation (2.75), we solve the product multinomial based likelihood estimating equation

$$\sum_{\ell=1}^{2} X'_\ell \left[y_{[\ell]} - K_{[\ell]} \pi_{[\ell]} \right] = 0, \tag{2.105}$$

where

$$y_{[\ell]} = \begin{pmatrix} K_{[\ell]1} \\ K_{[\ell]2} \\ K_{[\ell]3} \\ K_{[\ell]4} \end{pmatrix} \text{ and } \pi_{[\ell]} = \begin{pmatrix} \pi_{[\ell]1} \\ \pi_{[\ell]2} \\ \pi_{[\ell]3} \\ \pi_{[\ell]4} \end{pmatrix}.$$

These estimates and their corresponding standard errors computed by using (2.76) are reported in Table 2.17. Further by using these estimates in the probability formulas in (2.99), (2.101), and (2.103), we compute the estimated probabilities, which are same as the corresponding observed probabilities. For the sake of completeness, these probabilities are displayed in Table 2.18.

Notice from (2.103) that the estimates of β_{10}, β_{20}, and β_{30} indicate the relative probability for an individual with high education to be in none, few, and not so few categories, respectively, as compared to the probability for high category determined by $\beta_{40} = 0$ (by assumption). A large positive value of $\hat{\beta}_{20} = 0.693$ as compared to $\beta_{40} = 0$ shows that a large proportion of individuals belonging to the high education group paid a few visits to the physician. Similarly, for the individuals with medium level education, the negative values of $\hat{\beta}_{j0} + \hat{\beta}_{j2}$ for $j = 1,2,3$, such as $\hat{\beta}_{20} + \hat{\beta}_{22} = (0.693 - 1.012) = -0.319$, as compared to 0 show that a large proportion of individuals in this group paid high visits to the physician. On the contrary, by using (2.99), the largest positive value of $\hat{\beta}_{10} + \hat{\beta}_{11} = (-0.118 + 0.243) = 0.125$ indicates that a large proportion of individuals in the low education group did not pay any

Table 2.18 Observed and estimated multinomial probabilities for the education level versus physician visits data of Table 2.16

		Physician visit status				
Education level	Probability	None	Few	Not so few	High	Total
Low	Observed	0.2931	0.2759	0.1724	0.2586	1.0
	Estimated	0.2931	0.2759	0.1724	0.2586	1.0
Medium	Observed	0.1819	0.2424	0.2424	0.3333	1.0
	Estimated	0.1819	0.2424	0.2424	0.3333	1.0
High	Observed	0.1798	0.4045	0.2135	0.2022	1.0
	Estimated	0.1798	0.4045	0.2135	0.2022	1.0

visits. Thus, in general, most of the individuals in the low education group paid no visits to the physician, whereas most of the individuals with higher education paid a moderate number of visits (few visits). These categorical data based results agree, in general, with those reported in Sutradhar (2011, Section 4.2.8) based on original counts. However, the present categorical data based analysis naturally reveals more detailed pattern of visits.

2.2.3 Multinomial Likelihood Models with $L = (p+1)(q+1)$ Nominal Levels for Two Covariates with Interactions

Let there be two categorical covariates, one with $p+1$ levels and the other with $q+1$ levels. Following (2.45), for an individual i, we use the p-dimensional vector $w_i = [w_{i1},\ldots,w_{is},\ldots,w_{ip}]'$ containing p dummy covariates to represent $p+1$ levels of the first categorical covariate, and the q-dimensional vector $v_i = [v_{i1},\ldots,v_{im},\ldots,v_{iq}]'$ containing q dummy covariates to represent the $q+1$ levels of the second categorical covariate. Further, let $w_i(v_i)$ be a pq-dimensional nested covariate vector with v_i nested within w_i. That is,

$$w_i(v_i) = [w_{i1}v_{i1},\ldots,w_{i1}v_{iq},w_{i2}v_{11},\ldots,w_{is}v_{im},\ldots,w_{ip}v_{iq}]'.$$

Similar to (2.45), one may then write the probability for the response of the ith individual to be in the jth $(j = 1,\ldots,J)$ category as

$$P[y_i = y_i^{(j)} = \delta_{ij}] = \pi_{(i)j} = \begin{cases} \frac{\exp(\beta_{j0}+\beta_j' w_i+\xi_j' v_i+\phi_j^{*\prime} w_i(v_i))}{1+\sum_{g=1}^{J-1}\exp(\beta_{g0}+\beta_g' w_i+\xi_g' v_i+\phi_g^{*\prime} w_i(v_i))} & \text{for } j = 1,\ldots,J-1 \\ \frac{1}{1+\sum_{g=1}^{J-1}\exp(\beta_{g0}+\beta_g' w_i+\xi_g' v_i+\phi_g^{*\prime} w_i(v_i))} & \text{for } j = J, \end{cases} \tag{2.106}$$

where $\beta_j = [\beta_{j1},\ldots,\beta_{js},\ldots,\beta_{jp}]'$, $\xi_j = [\xi_{j1},\ldots,\xi_{jm},\ldots,\xi_{jq}]'$, and ϕ_j^* be the pq-dimensional vector of interaction effects of the covariates defined as

$$\phi_j^* = [\phi_{j11}^*,\ldots,\phi_{j1q}^*,\phi_{j21}^*,\ldots,\phi_{jsm}^*,\ldots,\phi_{jpq}^*]'.$$

Note that in (2.89), the interaction effects of two response variables are denoted by $\{\phi_{rj}\}$ whereas in (2.106), $\{\phi^*_{jum}\}$ represent the interaction effects of two covariates on the response $y_i = y_i^{(j)}$. Thus, a clear difference is laid out so that one does not use the same model to deal with contingency tables between two responses, and the contingency tables between one response and one or two or more categorical covariates. To be more explicit, one must use the probabilities in (2.89) to construct a full multinomial model, whereas the probabilities in (2.106) should be used to construct the product multinomial model.

Note that the $p+1$ levels corresponding to the covariate vector w_i may be formed as

$$(w_{i1},\cdots,w_{ip}) \equiv \begin{cases} (\,1,\, 01'_{p-1}) & \longrightarrow \text{Level 1} \\ (\,01'_{1},\, 1,\, 01'_{p-2}) & \longrightarrow \text{Level 2} \\ (\cdots\cdots\cdots\cdots) & \\ (\,01'_{\ell_1-1},\, 1,\, 01'_{p-\ell_1}) & \longrightarrow \text{Level } \ell_1 \\ (\cdots\cdots\cdots\cdots) & \\ (\,01'_{p-1},\, 1\,) & \longrightarrow \text{Level p} \\ (\,01'_{p}) & \longrightarrow \text{Level p+1} \end{cases} \tag{2.107}$$

Similarly, the $q+1$ levels corresponding to the covariate vector v_i may be formed as

$$(v_{i1},\cdots,v_{iq}) \equiv \begin{cases} (\,1,\, 01'_{q-1}) & \longrightarrow \text{Level 1} \\ (\,01'_{1},\, 1,\, 01'_{q-2}) & \longrightarrow \text{Level 2} \\ (\cdots\cdots\cdots\cdots) & \\ (\,01'_{\ell_2-1},\, 1,\, 01'_{q-\ell_2}) & \longrightarrow \text{Level } \ell_2 \\ (\cdots\cdots\cdots\cdots) & \\ (\,01'_{q-1},\, 1\,) & \longrightarrow \text{Level q} \\ (\,01'_{q}) & \longrightarrow \text{Level q+1} \end{cases} \tag{2.108}$$

Consequently, by (2.106), we may write the level based probabilities for an individual i, with covariates at level (ℓ_1,ℓ_2), to be in the jth category as

$$\begin{aligned}\pi_{[\ell_1,\ell_2]j} &= \pi_{(i\in\{\ell_1,\ell_2\})j} \\ &= \begin{cases} \frac{\exp(\beta_{j0}+\beta_{j\ell_1}+\xi_{j\ell_2}+\phi^*_{j,\ell_1\ell_2})}{1+\sum_{g=1}^{J-1}\exp(\beta_{g0}+\beta_{g\ell_1}+\xi_{g\ell_2}+\phi^*_{g,\ell_1\ell_2})} & \text{for } j=1,\ldots,J-1; \ell_1=1,\ldots,p; \ell_2=1,\ldots,q \\ \frac{1}{1+\sum_{g=1}^{J-1}\exp(\beta_{g0}+\beta_{g\ell_1}+\xi_{g\ell_2}+\phi^*_{g,\ell_1\ell_2})} & \text{for } j=J;\ \ell_1=1,\ldots,p; \ell_2=1,\ldots,q, \end{cases}\end{aligned} \tag{2.109}$$

$$\pi_{[\ell_1,q+1]j} = \pi_{(i\in\{\ell_1,q+1\})j} = \begin{cases} \frac{\exp(\beta_{j0}+\beta_{j\ell_1})}{1+\sum_{g=1}^{J-1}\exp(\beta_{g0}+\beta_{g\ell_1})} & \text{for } j=1,\ldots,J-1;\ell_1=1,\ldots,p \\ \frac{1}{1+\sum_{g=1}^{J-1}\exp(\beta_{g0}+\beta_{g\ell_1})} & \text{for } j=J;\ \ell_1=1,\ldots,p, \end{cases} \tag{2.110}$$

$$\pi_{[p+1,\ell_2]j} = \pi_{(i\in\{p+1,\ell_2\})j} = \begin{cases} \frac{\exp(\beta_{j0}+\xi_{j\ell_2})}{1+\sum_{g=1}^{J-1}\exp(\beta_{g0}+\xi_{g\ell_2})} & \text{for } j=1,\ldots,J-1;\ell_2=1,\ldots,q \\ \frac{1}{1+\sum_{g=1}^{J-1}\exp(\beta_{g0}+\xi_{g\ell_2})} & \text{for } j=J;\ \ell_2=1,\ldots,q, \end{cases} \tag{2.111}$$

and

$$\pi_{[p+1,q+1]j} = \pi_{(i\in\{p+1,q+1\})j} = \begin{cases} \frac{\exp(\beta_{j0})}{1+\sum_{g=1}^{J-1}\exp(\beta_{g0})} & \text{for } j=1,\ldots,J-1 \\ \frac{1}{1+\sum_{g=1}^{J-1}\exp(\beta_{g0})} & \text{for } j=J. \end{cases} \tag{2.112}$$

Next we display the observed cell counts in notation under all J categories and covariates level (ℓ_1,ℓ_2). Note that by transforming the rectangular levels to a real valued level, that is, using the relabeling formula $\{\ell \equiv [(\ell_1-1)(q+1)+\ell_2], \ell_1=1,\ldots,p+1; \ell_2=1,\ldots,q+1\}$, one may still use the Table 2.4 after a slight adjustment to display the cell counts in the present setup with two covariates. The cell counts with level adjustment are shown in Table 2.19.

Note that even though the cell probabilities in Tables 2.4 and 2.19 are denoted by the same notation $\pi_{[\ell]j}$, they are however different. The difference lies in the form of global regression parameter θ^*. To be more specific, the probabilities in Table 2.4 follow the formulas in (2.63)–(2.64), which were further re-expressed by (2.69) as

$$\pi_{[\ell]j} = \frac{\exp(x'_{[\ell]j}\theta^*)}{\sum_{g=1}^{J}\exp(x'_{[\ell]g}\theta^*)},$$

with

$$\theta^* = [\beta_1^{*\prime},\ldots,\beta_j^{*\prime},\ldots,\beta_{J-1}^{*}{}']', \text{ where } \beta_j^* = [\beta_{j0},\beta_j']'.$$

Note that once θ^* is written, the row vector $x'_{[\ell]j}$ becomes specified from the probability formulas. Now because $\pi_{[\ell]j}$ in Table 2.19 represent the two covariates level based probabilities defined in (2.109)–(2.112), the global regression parameters are different than θ^* in (2.69).

Table 2.19 A notational display for cell counts and probabilities for J categories under covariates level $(\ell_1, \ell_2) \longrightarrow$ Newlevel $\ell = (\ell_1 - 1)(\mathrm{q}+1) + \ell_2$

			***J* response categories**			
Covariates level	**New level**	**Quantity**	**1**	...	***J***	**Total**
$(1,1)$	1	Count	$K_{[1]1}$	...	$K_{[1]J}$	$K_{[1]}$
		Probability	$\pi_{[1]1}$	...	$\pi_{[1]J}$	1
.	.	.	.	...	.	.
		.	.	...	.	.
$(1,q+1)$	$q+1$	Count	$K_{[q+1]1}$	...	$K_{[q+1]J}$	$K_{[q+1]}$
		Probability	$\pi_{[q+1]1}$	...	$\pi_{[q+1]J}$	1
$(2,1)$	$q+2$	Count	$K_{[q+2]1}$	...	$K_{[q+2]J}$	$K_{[q+2]}$
		Probability	$\pi_{[q+2]1}$	...	$\pi_{[q+2]J}$	1
.	.	.	.	...	.	.
		.	.	...	.	.
(ℓ_1,ℓ_2)	ℓ	Count	$K_{[\ell]1}$	...	$K_{[\ell]J}$	$K_{[\ell]}$
		Probability	$\pi_{[\ell]1}$	...	$\pi_{[\ell]J}$	1
.	.	.	.	...	.	.
		.	.	...	.	.
$(p+1,q+1)$	$(p+1)(q+1)$	Count	$K_{[(p+1)(q+1)]1}$	...	$K_{[(p+1)(q+1)]J}$	$K_{[(p+1)(q+1)]}$
		Probability	$\pi_{[(p+1)(q+1)]1}$	...	$\pi_{[(p+1)(q+1)]J}$	1

Let θ^{**} denote the regression parameters used in two covariates level based probabilities in (2.109)–(2.112). To be specific, we write

$$\theta^{**} = [\beta_1^{**\prime}, \ldots, \beta_j^{**\prime}, \ldots, \beta_{J-1}^{**}{}^{\prime}]' : \{(J-1)(p+1)(q+1)\} \times 1, \tag{2.113}$$

where

$$\beta_j^{**} = [\beta_{j0},\, \beta_j',\, \xi_j',\, \phi^{*\prime}{}_j]' : \{(p+1)(q+1)\} \times 1,$$

with

$$\begin{aligned} \beta_j' &= [\beta_{j1}, \ldots, \beta_{js}, \ldots, \beta_{jp}] \\ \xi_j' &= [\xi_{j1}, \ldots, \xi_{jm}, \ldots, \xi_{jq}] \\ \phi^{*\prime}{}_j &= [\phi^*_{j,11}, \ldots, \phi^*_{j,1q}, \phi^*_{j,21}, \ldots, \phi^*_{j,sm}, \ldots, \phi^*_{j,pq}], \end{aligned}$$

by (2.106).

Consequently, at level

$$\ell = (\ell_1 - 1)(q+1) + \ell_2,\ \ell_1 = 1, \ldots, p+1;\ \ell_2 = 1, \ldots, q+1,$$

all probabilities defined in (2.109)–(2.112) may be expressed as

$$\pi_{[\ell]j} = \frac{\exp(x'_{[\ell]j}\theta^{**})}{\sum_{g=1}^{J}\exp(x'_{[\ell]g}\theta^{**})}, \tag{2.114}$$

where $x'_{[\ell]j} :\ 1 \times (J-1)(p+1)(q+1)$ is the jth $(j = 1, \ldots, J)$ row vector of the $X_\ell :\ J \times (J-1)(p+1)(q+1)$ matrix at the ℓth level $(\ell = 1, \ldots, (p+1)(q+1))$. We construct this jth row vector of the X_ℓ matrix in four groups as follows.

Group 1: $\ell = \{(\ell_1 - 1)q + \ell_2\}$ for $\ell_1 = 1, \ldots, p$; $\ell_2 = 1, \ldots, q$

$$\begin{aligned} x'_{[\ell]j} &= x'_{[(\ell_1-1)q+\ell_2]j} \\ &= \begin{cases} \Big[01'_{(j-1)(p+1)(q+1)}, \\ \{1, 01'_{\ell_1-1}, 1, 01'_{p-\ell_1}, 01'_{\ell_2-1}, 1, 01'_{q-\ell_2}, 01'_{(\ell_1-1)q+\ell_2-1}, 1, 01'_{pq-[(\ell_1-1)q+\ell_2]}\}, \\ 01'_{(J-1-j)(p+1)(q+1)}\Big], \ \text{for } j = 1, \ldots, J-1 \\ 01'_{(J-1)(p+1)(q+1)}, \ \text{for } j = J. \end{cases} \end{aligned} \tag{2.115}$$

Group 2: $\ell = \{(\ell_1 - 1)(q+1) + (q+1)\}$ for $\ell_1 = 1, \ldots, p$

$$\begin{aligned} x'_{[\ell]j} &= x'_{[(\ell_1-1)(q+1)+(q+1)]j} \\ &= \begin{cases} \Big[01'_{(j-1)(p+1)(q+1)}, \\ \{1, 01'_{\ell_1-1}, 1, 01'_{p-\ell_1}, 01'_{q}, 01'_{pq}, \}, \\ 01'_{(J-1-j)(p+1)(q+1)}\Big], \ \text{for } j = 1, \ldots, J-1 \\ 01'_{(J-1)(p+1)(q+1)}, \ \text{for } j = J. \end{cases} \end{aligned} \tag{2.116}$$

Group 3: $\ell = \{p(q+1) + \ell_2\}$ for $\ell_2 = 1, \ldots, q$

$$\begin{aligned} x'_{[\ell]j} &= x'_{[p(q+1)+\ell_2]j} \\ &= \begin{cases} \Big[01'_{(j-1)(p+1)(q+1)}, \\ \{1, 01'_{p}, 01'_{\ell_2-1}, 1, 01'_{q-\ell_2}, 01'_{pq}, \}, \\ 01'_{(J-1-j)(p+1)(q+1)}\Big], \ \text{for } j = 1, \ldots, J-1 \\ 01'_{(J-1)(p+1)(q+1)}, \ \text{for } j = J. \end{cases} \end{aligned} \tag{2.117}$$

Group 4: $\ell = \{(p+1)(q+1)\}$

$$x'_{[\ell]j} = x'_{[(p+1)(q+1)]j}$$

$$= \begin{cases} \left[01'_{(j-1)(p+1)(q+1)},\right. \\ \{1, 01'_p, 01'_q, 01'_{pq},\}, \\ \left. 01'_{(J-1-j)(p+1)(q+1)}\right], & \text{for } j = 1, \ldots, J-1 \\ 01'_{(J-1)(p+1)(q+1)}, & \text{for } j = J. \end{cases} \tag{2.118}$$

Now by replacing θ^* with θ^{**} in (2.67)–(2.68), by similar calculations as in (2.71), one obtains the likelihood equations for θ^{**} as

$$\frac{\partial \log L(\theta^{**})}{\partial \theta^{**}} = \sum_{\ell=1}^{(p+1)(q+1)} X'_\ell \left[y_{[\ell]} - K_{[\ell]}\pi_{[\ell]} \right] = 0, \tag{2.119}$$

where

$$y_{[\ell]} = [K_{[\ell]1}, \ldots, K_{[\ell]j}, \ldots, K_{[\ell]J}]' \text{ and } \pi_{[\ell]} = [\pi_{[\ell]1}, \ldots, \pi_{[\ell]j}, \ldots, \pi_{[\ell]J}]',$$

with probabilities are being given by (2.114) or equivalently by (2.109)–(2.112), and furthermore X_ℓ matrices for $\ell = 1, \ldots, (p+1)(q+1)$ are given as in (2.115)–(2.118).

Note that after slight adjustment in notation, one may use the iterative equation (2.75) to solve this likelihood equation in (2.119). To be specific, the iterative equation to solve (2.119) for the final estimate for θ^{**} is given by

$$\begin{aligned} \widehat{\theta^{**}}(r+1) = \widehat{\theta^{**}}(r) + \left[\sum_{\ell=1}^{(p+1)(q+1)} K_{[\ell]} X'_\ell \left[D_{\pi_{[\ell]}} - \pi_{[\ell]}\pi'_{[\ell]} \right] X_\ell \right]^{-1} \\ \times \left[\sum_{\ell=1}^{(p+1)(q+1)} X'_\ell \left(y_{[\ell]} - K_{[\ell]}\pi_{[\ell]} \right) \right]_{\theta^{**} = \widehat{\theta^{**}}(r)}, \end{aligned} \tag{2.120}$$

where $D_{\pi_{[\ell]}} = \text{diag}[\pi_{[\ell]1}, \ldots, \pi_{[\ell]j}, \ldots, \pi_{[\ell]J}]$. Furthermore, the covariance matrix of $\widehat{\theta^{**}}$ has the formula

$$\text{var}(\widehat{\theta^{**}}) = \left[\sum_{\ell=1}^{(p+1)(q+1)} K_{[\ell]} X'_\ell \{ D_{\pi_{[\ell]}} - \pi_{[\ell]}\pi'_{[\ell]} \} X_\ell \right]^{-1}. \tag{2.121}$$

2.2.3.1 Illustration 2.8: Analysis for the Effects of Both Gender and Chronic Condition on the Physician Visits

The marginal effects of gender and chronic condition on the physician visits were discussed in Sects. 2.2.2.4.2(a) and (b), respectively. To illustrate the product multinomial model for a response variable (physician visit) based on two categorical covariates, discussed in the last section, we now consider gender and chronic

Table 2.20 $2 \times 2 \times 4$ contingency table for the physician visit data corresponding to gender and chronic condition of the individuals

		Physician visit status				
Gender	Chronic condition	None	Few	Not so few	High	Level total (K_ℓ)
Male	One or more	8	13	9	10	40
	None	20	27	7	2	56
Female	One or more	5	12	12	23	52
	None	6	8	9	9	32

Table 2.21 Cell counts and probabilities for $J = 4$ physician visit categories under covariates level (ℓ_1, ℓ_2) for $\ell_1 = 1, 2$; and $\ell_2 = 1, 2$

			Physician visit status				
Covariates level	New level (ℓ)	Quantity	None	Few	Not so few	High	Level total $(K_{[\ell]})$
(1,1)	1	Count	8	13	9	10	40
		Probability	$\pi_{[1]1}$	$\pi_{[1]2}$	$\pi_{[1]3}$	$\pi_{[1]4}$	1.0
(1,2)	2	Count	20	27	7	2	56
		Probability	$\pi_{[2]1}$	$\pi_{[2]2}$	$\pi_{[2]3}$	$\pi_{[2]4}$	1.0
(2,1)	3	Count	5	12	12	23	52
		Probability	$\pi_{[3]1}$	$\pi_{[3]2}$	$\pi_{[3]3}$	$\pi_{[3]4}$	1.0
(2,2)	4	Count	6	8	9	9	32
		Probability	$\pi_{[4]1}$	$\pi_{[4]2}$	$\pi_{[4]3}$	$\pi_{[4]4}$	1.0

condition as two covariates and examine their marginal as well as joint (interaction between the two covariates) effects on the physician visit. For the purpose, following the Table 2.19, we first present the observed counts as in the $2 \times 2 \times 4$ contingency Table 2.20. Note that this contingency table is not showing the distribution for three response variables, rather, it shows the distribution of one response variable at different marginal and joint levels for the two covariates. Consequently, it is appropriate to use the product multinomial approach to analyze the data of this Table 2.20.

Further to make the cell probability formulas clear and precise, we use the data from Table 2.20 and put them in Table 2.21 along with probabilities following the format of Table 2.19.

Next, we write the formulas for the probabilities in Table 2.21 in the form of (2.109)–(2.112), and also in global regression form as follows:

Level 1 (Group 1) (based on $\ell_1 = 1, \ell_2 = 1$):

Probabilities:

$$\pi_{[1]1} = \frac{\exp(\beta_{10} + \beta_{11} + \xi_{11} + \phi^*_{1,11})}{1 + \sum_{g=1}^{3} \exp(\beta_{g0} + \beta_{g1} + \xi_{g1} + \phi^*_{g,11})}, \quad \pi_{[1]2} = \frac{\exp(\beta_{20} + \beta_{21} + \xi_{21} + \phi^*_{2,11})}{1 + \sum_{g=1}^{3} \exp(\beta_{g0} + \beta_{g1} + \xi_{g1} + \phi^*_{g,11})},$$

$$\pi_{[1]3} = \frac{\exp(\beta_{30}+\beta_{31}+\xi_{31}+\phi^*_{3,11})}{1+\sum_{g=1}^{3}\exp(\beta_{g0}+\beta_{g1}+\xi_{g1}+\phi^*_{g,11})},\ \pi_{[1]4} = \frac{1}{1+\sum_{g=1}^{3}\exp(\beta_{g0}+\beta_{g1}+\xi_{g1}+\phi^*_{g,11})}. \tag{2.122}$$

Global regression form:

$$\pi_{[1]1} = \frac{\exp(x'_{[1]1}\theta^{**})}{\sum_{j=1}^{4}\exp(x'_{[1]j}\theta^{**})},\ \pi_{[1]2} = \frac{\exp(x'_{[1]2}\theta^{**})}{\sum_{j=1}^{4}\exp(x'_{[1]j}\theta^{**})},$$
$$\pi_{[1]3} = \frac{\exp(x'_{[1]3}\theta^{**})}{\sum_{j=1}^{4}\exp(x'_{[1]j}\theta^{**})},\pi_{[1]4} = \frac{\exp(x'_{[1]4}\theta^{**})}{\sum_{j=1}^{4}\exp(x'_{[1]j}\theta^{**})},$$

where

$$\theta^{**} = (\beta_{10},\beta_{11},\xi_{11},\phi^*_{1,11},\beta_{20},\beta_{21},\xi_{21},\phi^*_{2,11},\beta_{30},\beta_{31},\xi_{31},\phi^*_{3,11},)',$$

and $x'_{[1]3}$, for example, is the third row vector of the $X_1 : 4 \times 12$ matrix given by

$$X_1 = \begin{pmatrix} 1'_4 & 01'_4 & 01'_4 \\ 01'_4 & 1'_4 & 01'_4 \\ 01'_4 & 01'_4 & 1'_4 \\ 01'_4 & 01'_4 & 01'_4 \end{pmatrix}. \tag{2.123}$$

Level 2 (Group 2) (based on $\ell_1 = 1$):
Probabilities:

$$\pi_{[2]1} = \frac{\exp(\beta_{10}+\beta_{11})}{1+\sum_{g=1}^{3}\exp(\beta_{g0}+\beta_{g1})},\ \pi_{[2]2} = \frac{\exp(\beta_{20}+\beta_{21})}{1+\sum_{g=1}^{3}\exp(\beta_{g0}+\beta_{g1})},$$
$$\pi_{[2]3} = \frac{\exp(\beta_{30}+\beta_{31})}{1+\sum_{g=1}^{3}\exp(\beta_{g0}+\beta_{g1})},\ \pi_{[2]4} = \frac{1}{1+\sum_{g=1}^{3}\exp(\beta_{g0}+\beta_{g1})}. \tag{2.124}$$

Global regression form:

$$\pi_{[2]1} = \frac{\exp(x'_{[2]1}\theta^{**})}{\sum_{j=1}^{4}\exp(x'_{[2]j}\theta^{**})},\ \pi_{[2]2} = \frac{\exp(x'_{[2]2}\theta^{**})}{\sum_{j=1}^{4}\exp(x'_{[2]j}\theta^{**})},$$
$$\pi_{[2]3} = \frac{\exp(x'_{[2]3}\theta^{**})}{\sum_{j=1}^{4}\exp(x'_{[2]j}\theta^{**})},\pi_{[2]4} = \frac{\exp(x'_{[2]4}\theta^{**})}{\sum_{j=1}^{4}\exp(x'_{[2]j}\theta^{**})},$$

where θ^{**} is the same as above, that is,

$$\theta^{**} = (\beta_{10},\beta_{11},\xi_{11},\phi^*_{1,11},\beta_{20},\beta_{21},\xi_{21},\phi^*_{2,11},\beta_{30},\beta_{31},\xi_{31},\phi^*_{3,11},)',$$

and $x'_{[2]3}$, for example, is the third row vector of the $X_2 : 4 \times 12$ matrix given by

$$X_2 = \begin{pmatrix} 1'_2 & 01'_2 & 01'_4 & 01'_4 \\ 01'_4 & 1'_2 & 01'_2 & 01'_4 \\ 01'_4 & 01'_4 & 1'_2 & 01'_2 \\ 01'_2 & 01'_2 & 01'_4 & 01'_4 \end{pmatrix}. \tag{2.125}$$

Level 3 (Group 3) (based on $\ell_2 = 1$):
Probabilities:

$$\pi_{[3]1} = \frac{\exp(\beta_{10} + \xi_{11})}{1 + \sum_{g=1}^{3} \exp(\beta_{g0} + \xi_{g1})},\ \pi_{[3]2} = \frac{\exp(\beta_{20} + \xi_{21})}{1 + \sum_{g=1}^{3} \exp(\beta_{g0} + \xi_{g1})},$$
$$\pi_{[3]3} = \frac{\exp(\beta_{30} + \xi_{31})}{1 + \sum_{g=1}^{3} \exp(\beta_{g0} + \xi_{g1})},\ \pi_{[3]4} = \frac{1}{1 + \sum_{g=1}^{3} \exp(\beta_{g0} + \xi_{g1})}. \tag{2.126}$$

Global regression form:

$$\pi_{[3]1} = \frac{\exp(x'_{[3]1}\theta^{**})}{\sum_{j=1}^{4} \exp(x'_{[3]j}\theta^{**})},\ \pi_{[3]2} = \frac{\exp(x'_{[3]2}\theta^{**})}{\sum_{j=1}^{4} \exp(x'_{[3]j}\theta^{**})},$$
$$\pi_{[3]3} = \frac{\exp(x'_{[3]3}\theta^{**})}{\sum_{j=1}^{4} \exp(x'_{[3]j}\theta^{**})},\ \pi_{[3]4} = \frac{\exp(x'_{[3]4}\theta^{**})}{\sum_{j=1}^{4} \exp(x'_{[3]j}\theta^{**})},$$

where $x'_{[3]3}$, for example, is the third row vector of the $X_3 : 4 \times 12$ matrix given by

$$X_3 = \begin{pmatrix} 1 & 0 & 1 & 0 & 01'_4 & 01'_4 \\ 01'_4 & 1 & 0 & 1 & 0 & 01'_4 \\ 01'_4 & 01'_4 & 1 & 0 & 1 & 0 \\ 01'_4 & 01'_4 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{2.127}$$

Level 4 (Group 4)
Probabilities:

$$\pi_{[4]1} = \frac{\exp(\beta_{10})}{1 + \sum_{g=1}^{3} \exp(\beta_{g0})},\ \pi_{[4]2} = \frac{\exp(\beta_{20})}{1 + \sum_{g=1}^{3} \exp(\beta_{g0})},$$
$$\pi_{[4]3} = \frac{\exp(\beta_{30})}{1 + \sum_{g=1}^{3} \exp(\beta_{g0})},\ \pi_{[4]4} = \frac{1}{1 + \sum_{g=1}^{3} \exp(\beta_{g0})}. \tag{2.128}$$

Global regression form:

$$\pi_{[4]1} = \frac{\exp(x'_{[4]1}\theta^{**})}{\sum_{j=1}^{4} \exp(x'_{[4]j}\theta^{**})},\ \pi_{[4]2} = \frac{\exp(x'_{[4]2}\theta^{**})}{\sum_{j=1}^{4} \exp(x'_{[4]j}\theta^{**})},$$

Table 2.22 Parameter estimates for the gender and chronic condition versus physician visit status data of Table 2.21

	Regression parameters							
Quantity	β_{10}	β_{11}	ξ_{11}	$\phi^*_{1,11}$	β_{20}	β_{21}	ξ_{21}	$\phi^*_{2,11}$
Estimate	−0.405	2.708	−1.121	−1.405	−0.118	2.720	−.533	−1.808
Standard error	0.5270	0.9100	0.7220	1.1385	0.4859	0.8793	0.6024	1.0377
Quantity	β_{30}	β_{31}	ξ_{31}	$\phi^*_{3,11}$				
Estimate	0.000	1.253	−0.651	−0.708				
Standard error	0.4714	0.9301	0.5908	1.0968				

Table 2.23 Estimated/observed probabilities corresponding to the data given in $2 \times 2 \times 4$ contingency Table 2.20

		Physician visit status			
Gender	Chronic condition	None	Few	Not so few	High
Male	One or more	0.2000	0.3250	0.2250	0.2500
	None	0.3572	0.4821	0.1250	0.0357
Female	One or more	0.0962	0.2308	0.2308	0.4422
	None	0.1874	0.2500	0.2813	0.2813

$$\pi_{[4]3} = \frac{\exp(x'_{[4]3}\theta^{**})}{\sum_{j=1}^{4}\exp(x'_{[4]j}\theta^{**})}, \pi_{[4]4} = \frac{\exp(x'_{[4]4}\theta^{**})}{\sum_{j=1}^{4}\exp(x'_{[4]j}\theta^{**})},$$

where $x'_{[4]3}$, for example, is the third row vector of the $X_4 : 4 \times 12$ matrix given by

$$X_4 = \begin{pmatrix} 1 & 01'_3 & 01'_4 & 01'_4 \\ 01'_4 & 1 & 01'_3 & 01'_4 \\ 01'_4 & 01'_4 & 1 & 01'_3 \\ 01'_4 & 01'_4 & 01'_3 & 0 \end{pmatrix}. \tag{2.129}$$

Using the gender and chronic condition versus physician visits data from Table 2.21, we now solve the likelihood estimating equation (2.119), i.e.,

$$\frac{\partial \log L(\theta^{**})}{\partial \theta^{**}} = \sum_{\ell=1}^{(p+1)(q+1)} X'_\ell \left[y_{[\ell]} - K_{[\ell]}\pi_{[\ell]}\right] = 0,$$

for θ^{**}. The estimates for all components in θ^{**} along with their standard errors are given in Table 2.22.

Now by using the regression estimates from Table 2.22 into the probability formulas (2.112), (2.124), (2.126), and (2.128), one obtains the estimated probabilities as in Table 2.23. The estimated and observed probabilities are the same.

We now interpret the estimates of the parameters from Table 2.22. Because at level 4, i.e., for a female with no chronic disease, the category probabilities for the first three categories are determined by the respective estimates of β_{10}, β_{20}, and β_{30}, as compared to the conventional value $\beta_{40} = 0$, it is easier to interpret their role first. For example, $\hat{\beta}_{30} = 0.0$ shows that an individual in this group has the same probability to be in the third (not so few) or fourth (high) physician visits category. Further, smaller negative value for $\hat{\beta}_{20} = -0.118$ as compared to $\hat{\beta}_{10} = -0.405$ shows that the individual in this group has a much higher probability to pay a few visits to the physician as opposed to paying no visits at all.

Next the values of $(\hat{\beta}_{j0} + \hat{\xi}_{j1})$ for $j = 1,2,3$, as compared to $\beta_{40} + \xi_{41} = 0.0$ would determine relative probability of an individual at level 3 (group 3) to be in the jth category. Note that group 3 individuals are female with one or more chronic disease. For example, the small negative and equal values of $\hat{\beta}_{20} + \hat{\xi}_{21} = -0.651 = \hat{\beta}_{30} + \hat{\xi}_{31}$ as compared to large negative value of $\hat{\beta}_{10} + \hat{\xi}_{11} = -1.526$ indicate that a female with chronic disease has increasing probabilities to pay more visit to the physicians. But a male with chronic disease, i.e., an individual belonging to group 1 (level 1), has smaller probability to pay a high physician visit. This follows from relatively large positive value of $\hat{\beta}_{20} + \hat{\beta}_{21} + \hat{\xi}_{21} + \hat{\phi}^*{}_{2,11} = 0.261$ as compared to small negative value of $\hat{\beta}_{30} + \hat{\beta}_{31} + \hat{\xi}_{31} + \hat{\phi}^*{}_{3,11} = -0.106$, and $\beta_{40} + \beta_{41} + \xi_{41} + \phi^*{}_{4,11} = 0.0$.

2.3 Cumulative Logits Model for Univariate Ordinal Categorical Data

There are situations where the categories for a response may also be ordinal by nature. For example, when the individuals in a study are categorized to examine their agreement or disagreement on a policy issue with, say, three political groups A, B, and C, these three categories are clearly nominal. However, in the treatment versus heart attack status data analyzed in Sect. 2.2.2.4.1, the three categories accommodating the heart attack status, namely no attack, non-fatal, and fatal attacks, can also be treated as ordinal categories. Similarly, in the physician visit study in Sect. 2.2.2.4.2, four physician visit status, namely none, few, not so few, and high visits, can be treated as ordinal categories. Now because of this additional ordinal property of the categorical responses, one may collapse the $J > 2$ categories in a cumulative fashion into two $(J' = 2)$ categories and use simpler binary model to fit such collapsed data. Note however that there will be various binary groups depending on which category in the middle is used as a cut point. This approach is referred to as the cumulative logits model approach and we discuss this alternative modeling of the categorical data in this section provided the categories also exhibit order in them.

2.3.1 Cumulative Logits Model Involving One Covariate with $L = p + 1$ Levels

Suppose that similar to Sect. 2.2.2, $\pi_{[\ell]j}$ denotes the probability for an individual i with covariate at level $\ell(\ell = 1, \ldots, p+1)$ to be in the jth category, but because categories are ordered, one may collapse the J categories based multinomial model to a binary model with

$$F_{[\ell]j} = \sum_{c=1}^{j} \pi_{[\ell]c}$$

representing the probability for the binary response to be in any categories between 1 and j, and

$$1 - F_{[\ell]j} = \sum_{c=j+1}^{J} \pi_{[\ell]c}$$

representing the probability for the binary response to be in any categories beyond j. Consequently, unlike in Sect. 2.2.2, instead of modeling individual category based probabilities $\pi_{[\ell]j}$, one may model the binary probability $F_{[\ell]j}$ by using the linear logits relationship

$$L_{[\ell]j} = log\left[\frac{1 - F_{[\ell]j}}{F_{[\ell]j}}\right] = \begin{cases} \alpha_{j0} + \alpha_{j\ell} & \text{for } j = 1, \ldots, J-1;\ \ell = 1, \ldots, p \\ \alpha_{j0} & \text{for } j = 1, \ldots, J-1;\ \ell = p+1. \end{cases} \tag{2.130}$$

We refer to this model (2.130) as the logit model 1 (LM1). Also, three other logit models are considered in the next section with relation to a real life data example.

Note that for $j = 1, \ldots, J-1$, the logit relationship in (2.130) is equivalent to write

$$1 - F_{[\ell]j} = \begin{cases} \frac{\exp(\alpha_{j0}+\alpha_{j\ell})}{1+\exp(\alpha_{j0}+\alpha_{j\ell})} & \text{for } \ell = 1, \ldots, p \\ \frac{\exp(\alpha_{j0})}{1+\exp(\alpha_{j0})} & \text{for } \ell = p+1. \end{cases} \tag{2.131}$$

Remark that the logits in (2.130) satisfy the monotonic constraint given in the following lemma.

Lemma 2.3.1. *The logits in (2.130) satisfy the monotonic property*

$$L_{[\ell]1} \geq L_{[\ell]2} \geq \ldots \geq L_{[\ell](J-1)}. \tag{2.132}$$

Proof. Since

$$F_{[\ell]1} \leq F_{[\ell]2} \leq \ldots \leq F_{[\ell](J-1)},$$

and

$$(1-F_{[\ell]1}) \geq (1-F_{[\ell]2}) \geq \ldots \geq (1-F_{[\ell](J-1)}),$$

one obtains

$$\frac{F_{[\ell]1}}{1-F_{[\ell]1}} \leq \frac{F_{[\ell]2}}{1-F_{[\ell]2}} \leq \ldots \leq \frac{F_{[\ell](J-1)}}{1-F_{[\ell](J-1)}}.$$

Hence the lemma follows because $L_{[\ell]j} = log\left[\frac{1-F_{[\ell]j}}{F_{[\ell]j}}\right]$ for all $j = 1,\ldots,J-1$.

2.3.1.1 Weighted Least Square Estimation for the Parameters of the Cumulative Logits Model (2.130)

We describe this estimation technique in the following steps.

Step 1. Writing the logits in linear regression form

Let $F(\pi)$ be a vector consisting of all possible logits, where π represents all $J(p+1)$ individual cell probabilities. That is,

$$F = F(\pi) = \left[L_1', \ldots, L_\ell', \ldots, L_{p+1}'\right]' : (J-1)(p+1) \times 1, \tag{2.133}$$

where L_ℓ is the vector of $J-1$ logits given by

$$L_\ell = \left[L_{[\ell]1}, \ldots, L_{[\ell]j}, \ldots, L_{[\ell](J-1)}\right]', \tag{2.134}$$

with $L_{[\ell]j}$ defined as in (2.130). Note that these logits for $j = 1,\ldots,J-1$ are functions of all J individual probabilities $\pi_{[\ell]1}, \ldots, \pi_{[\ell]J}$ at the covariate level ℓ.

Now define the regression parameters vector α as

$$\alpha = \left[\alpha_0', \alpha_1', \ldots, \alpha_\ell', \ldots, \alpha_p'\right]', \tag{2.135}$$

where

$$\alpha_0 = [\alpha_{10}, \ldots, \alpha_{(J-1)0}]' \text{ and } \alpha_\ell = [\alpha_{1\ell}, \ldots, \alpha_{(J-1)\ell}]',$$

for $\ell = 1,\ldots,p$. Next by using (2.135) and (2.130), one may then express the logits vector (2.133) in the linear regression form as

$$F = X\alpha, \tag{2.136}$$

where

$$X = \begin{bmatrix} I_{J-1} \; X_1 \\ I_{J-1} \; X_2 \\ \cdot \quad \cdot \\ I_{J-1} \; X_p \\ I_{J-1} \; X_{p+1} \end{bmatrix} : (J-1)(p+1) \times (J-1)(p+1), \tag{2.137}$$

with

$$X_\ell = \begin{pmatrix} x'_{[\ell]1} \\ x'_{[\ell]2} \\ \cdot \\ x'_{[\ell](J-1)} \end{pmatrix} : (J-1) \times (J-1)p, \text{ for } \ell = 1, \ldots, p+1 \tag{2.138}$$

where, for $j = 1, \ldots, J-1$,

$$\begin{aligned} x'_{[\ell]j} &= \left(01'_{(\ell-1)(J-1)} \; 01'_{j-1} \; 1 \; 01'_{J-1-j} \; 01'_{(p-\ell)(J-1)} \right) \text{for } \ell = 1, \ldots, p \\ x'_{[p+1]j} &= 01'_{p(J-1)}. \end{aligned} \tag{2.139}$$

Step 2. Formulation of $F(\pi)$ in terms of π

Write

$$\pi = [\pi'_{[1]}, \ldots, \pi'_{[\ell]}, \ldots, \pi'_{[p+1]}]' : J(p+1) \times 1, \tag{2.140}$$

where at covariate level ℓ, as in Sect. 2.2.2, all J cell probabilities are denoted by $\pi_{[\ell]}$, that is,

$$\pi_{[\ell]} = [\pi_{[\ell]1}, \ldots, \pi_{[\ell]j}, \ldots, \pi_{[\ell]J}]',$$

$\pi_{[\ell]j}$ being the probability for the response of an individual with ℓth level covariate information to be in the jth category.

Notice from (2.130) that $L_{[\ell]j}$ has the form

$$\begin{aligned} L_{[\ell]j} &= log \left[\frac{1 - F_{[\ell]j}}{F_{[\ell]j}} \right] \\ &= log \left[\frac{\sum_{c=j+1}^{J} \pi_{[\ell]c}}{\sum_{c=1}^{j} \pi_{[\ell]c}} \right] \end{aligned}$$

$$= \left[log\{\sum_{c=j+1}^{J} \pi_{[\ell]c}\} - log\{\sum_{c=1}^{j} \pi_{[\ell]c}\} \right]. \tag{2.141}$$

Consequently, L_ℓ defined in (2.134) can be expressed as

$$\begin{aligned} L_\ell &= [L_{[\ell]1}, \ldots, L_{[\ell](J-1)}]' \\ &= [\left(log\{\sum_{c=2}^{J} \pi_{[\ell]c}\} - log\{\sum_{c=1}^{1} \pi_{[\ell]c}\} \right), \ldots, \left(log\{\sum_{c=J}^{J} \pi_{[\ell]c}\} - log\{\sum_{c=1}^{J-1} \pi_{[\ell]c}\} \right)]' \\ &= M^* log\,(A^* \pi_{[\ell]}), \end{aligned} \tag{2.142}$$

where $\pi_{[\ell]}$ is defined by (2.140), and A^* and K^* have the forms:

$$A^* = \begin{bmatrix} 1'_1 & 01'_{J-1} \\ 01'_1 & 1'_{J-1} \\ 1'_2 & 01'_{J-2} \\ 01'_2 & 1'_{J-2} \\ \cdot & \cdot \\ \cdot & \cdot \\ 1'_{J-1} & 01'_1 \\ 01'_{J-1} & 1'_1 \end{bmatrix} : 2(J-1) \times J, \tag{2.143}$$

and

$$M^* = \begin{bmatrix} -1 & 1 & 01'_2 & 01'_{2(J-3)} \\ 01'_2 & -1 & 1 & 01'_{2(J-3)} \\ \cdot & \cdot & \cdot & \cdot \\ 01'_{2(J-3)} & 01'_2 & -1 & 1 \end{bmatrix} : (J-1) \times 2(J-1), \tag{2.144}$$

respectively. Now by using (2.142), it follows from (2.133) that

$$
\begin{aligned}
F = F(\pi) = \begin{pmatrix} L_1 \\ L_2 \\ \cdot \\ \cdot \\ L_\ell \\ \cdot \\ \cdot \\ L_{p+1} \end{pmatrix} &= \begin{pmatrix} M^* log\,(A^*\pi_{[1]}) \\ M^* log\,(A^*\pi_{[2]}) \\ \cdot \\ \cdot \\ M^* log\,(A^*\pi_{[\ell]}) \\ \cdot \\ \cdot \\ M^* log\,(A^*\pi_{[p+1]}) \end{pmatrix} \\
&= [I_{p+1} \otimes M^*]\, log\, [(I_{p+1} \otimes A^*)\,\pi] \\
&= M log\,(A\pi),
\end{aligned}
\tag{2.145}
$$

where $\pi = [\pi'_{[1]}, \ldots, \pi'_{[\ell]}, \ldots, \pi'_{[p+1]}]'$, and '$\otimes$' denotes the direct or Kronecker product.

Step 3. Forming a 'working' linear model

Using notations from Step 2 (2.145) in (2.136) under Step 1, one consequently solves α satisfying

$$F = F(\pi) = M log\,(A\pi) = X\alpha. \tag{2.146}$$

Note that this (2.146) is not an estimating equation yet as π is unknown in practice. This means the model (population average) Eq. (2.146) does not involve any data. However, by using the observed proportion p for π, one may write an approximate (working) linear regression model with correlated errors as follows:

$$F(p) \approx F(\pi) + \frac{\partial F(\pi)}{\partial \pi'}[p - \pi] = F(\pi) + \varepsilon,\ (J-1)(p+1) \times 1 \tag{2.147}$$

where ε may be treated as an error vector. Next, because for a given ℓ, the cell counts $\{K_{[\ell]j},\ j = 1, \ldots, J\}$ follow the multinomial probability distribution (2.67) [see also Table 2.4 for data display], it follows that

$$E[p_{[\ell]j}] = E[\frac{K_{[\ell]j}}{K_{[\ell]}}] = \pi_{[\ell]j}, \text{ for all j and } \ell,$$

that is $E[p] = \pi$, where π is defined by (2.140), and p is the corresponding observed proportion vector, with $p_{[\ell]} = [p_{[\ell]j}, \ldots, p_{[\ell]j}, \ldots, p_{[\ell]J}]'$. It then follows that

$$
\begin{aligned}
E[\varepsilon] &= 0, \\
\mathrm{cov}[\varepsilon] &= \left[\frac{\partial F(\pi)}{\partial \pi'}\right] \mathrm{cov}(p) \left[\frac{\partial F'(\pi)}{\partial \pi}\right]
\end{aligned}
$$

$$= \left[\frac{\partial F(\pi)}{\partial \pi'}\right] V \left[\frac{\partial F'(\pi)}{\partial \pi}\right] = \Sigma_\varepsilon \text{ (say)}, \; (J-1)(p+1) \times (J-1)(p+1). \quad (2.148)$$

Note that the approximation in (2.147) follows from the so-called Taylor's series expansion for $F(p)$. To be specific, for $u = 1, \ldots, (J-1)(p+1)$, the uth component of $F(p)$ may be expanded in Taylor's series form as

$$\begin{aligned} F_u(p) &= F_u(\pi) + (p-\pi)' \frac{\partial F_u(\pi)}{\partial \pi} \\ &\quad + \varepsilon^*_{u,K}(||p_{[1]} - \pi_{[1]}||, \ldots, ||p_{[\ell]} - \pi_{[\ell]}||, \ldots, ||p_{[p+1]} - \pi_{[p+1]}||), \end{aligned} \quad (2.149)$$

where for $K = \sum_{\ell=1}^{p+1} K_{[\ell]}$, $\varepsilon^*_{u,K}(\cdot)$ is a higher order remainder term in the Taylor's expansion, and it is a function of Euclidian distances

$$||p_{[\ell]} - \pi_{[\ell]}|| = \sqrt{\sum_{j=1}^{J} [p_{[\ell]j} - \pi_{[\ell]j}]^2}, \text{ for all } \ell = 1, \ldots, p+1.$$

Further note that when $\min_\ell\{K_{[\ell]}\} \to \infty$ it can be shown that

$$\varepsilon^*_{u,K}(\cdot) \to 0 \text{ in probability} \quad (2.150)$$

(see, for example, Rao (1973, p. 387); Bishop et al. (1975, Sec. 14.6) for details on this convergence property). Thus, for all $u = 1, \ldots, (J-1)(p+1)$, and using (2.150), one obtains the approximate linear relationship (2.147) from (2.149). Finally by using (2.146), one may fit the linear model

$$\begin{aligned} F(p) &= F(\pi) + \varepsilon \\ &= X\alpha + \varepsilon, \end{aligned} \quad (2.151)$$

(see also Grizzle et al. (1969), Haberman (1978, pp. 64–77)) where $F(p) = M log\,(Ap)$ with M and A as given by (2.145), and the error vector ε has the zero mean vector and covariance matrix Σ_ε as given by (2.148).

Step 4. WLS (weighted least square) estimating equation

Consequently, one may write the WLS estimating equation for α as

$$X'\Sigma_\varepsilon^{-1}[F(p) - X\alpha] = 0, \quad (2.152)$$

and obtain the WLS estimator of α as

$$\hat{\alpha}_{WLS} = [X'\Sigma_\varepsilon^{-1}X]^{-1}X'\Sigma_\varepsilon^{-1}F(p). \quad (2.153)$$

For similar use of the WLS approach in fitting models to ordinal data, one may be referred to Semenya and Koch (1980) and Semenya et al. (1983) (see also Agresti (1984, Section 7.2, Appendix A.2); Koch et al. (1992)). For computation convenience, one may further simplify Σ_ε from (2.148) as

$$\begin{aligned}
\Sigma_\varepsilon = \mathrm{cov}[F(p)] &= \left[\frac{\partial F(\pi)}{\partial \pi'}\right] V \left[\frac{\partial F'(\pi)}{\partial \pi}\right] \\
&= \left[\frac{\partial M log\, A\pi}{\partial \pi'}\right] V \left[\frac{\partial M log\, A\pi}{\partial \pi'}\right]' \\
&= MD^{-1}AVA'D^{-1}M' = QVQ',\ \text{(say)},
\end{aligned} \tag{2.154}$$

where $D = \mathrm{diag}[A\pi] : 2(J-1)(p+1) \times 2(J-1)(p+1)$, $A\pi : 2(J-1)(p+1) \times 1$ being given by (2.145). Hence, using Σ_ε from (2.154) into (2.153), one may re-express $\hat{\alpha}_{WLS}$ as

$$\hat{\alpha}_{WLS} = \left[X'(QVQ')^{-1}X\right]^{-1} X'(QVQ')^{-1}F(p), \tag{2.155}$$

with $F(p) = M log\ Ap$. Note that to compute $\hat{\alpha}_{WLS}$ by (2.155), one requires to replace the D matrix by its unbiased estimate $\hat{D} = \mathrm{diag}[Ap]$. Next, because, $\mathrm{cov}[F(p)] = QVQ'$ by (2.154), by treating D as a known matrix, one may compute the covariance of the WLS estimator of α as

$$\mathrm{cov}[\hat{\alpha}_{WLS}] = \left[X'(QVQ')^{-1}X\right]^{-1}, \tag{2.156}$$

which can be estimated by replacing π with p, that is,

$$\hat{\mathrm{cov}}[\hat{\alpha}_{WLS}] = \left[X'(QVQ')^{-1}X\right]^{-1}_{|\pi=p}. \tag{2.157}$$

Further note that the V matrix in (2.154)–(2.157) has the block diagonal form given by

$$V = \bigoplus_{\ell=1}^{p+1} [\mathrm{cov}(p_{[\ell]})] : (p+1)J \times (p+1)J, \tag{2.158}$$

where

$$\mathrm{cov}(p_{[\ell]}) = \frac{1}{K_{[\ell]}} \left[\mathrm{diag}[\pi_{[\ell]1}, \ldots, \pi_{[\ell]j}, \ldots, \pi_{[\ell]J}] - \pi_{[\ell]}\pi'_{[\ell]}\right]. \tag{2.159}$$

Table 2.24 Cross-classification of gender and physician visit along with observed proportions

	Physician visit status				
Gender	None	Few	Not so few	High	Total
Male	28	40	16	12	96
Cell proportion	0.2917	0.4166	0.1667	0.1250	1.0
Female	11	20	21	32	84
Cell proportion	0.1309	0.2381	0.2500	0.3810	1.0

2.3.1.1.1 Illustration 2.9: Weighted Least Square Fitting of the Cumulative Logits Model to the Gender Versus Physician Visit Data

Recall the physician visit status data for male and female from Table 2.10. For convenience, we redisplay these data along with observed proportions as in the following Table 2.24. Note that the physician visit status can be treated as ordinal categorical. However, among others, this data set was analyzed in Sect. 2.2.2.4 by applying the product multinomial likelihood approach discussed in Sects. 2.2.2.1 and 2.2.2.2, where categories were treated to be nominal. As discussed in last section, when categories are treated to be ordinal, one may fit the cumulative probability ratios based logits model to analyze such data. The logit models are different than standard multinomial models used for the analysis of nominal categorical data. We now follow the logit model and inferences discussed in the last section to reanalyze the gender versus physician visit status data shown in Table 2.24.

We first write the observed proportion vector p as

$$p = [p'_{[1]}, p'_{[2]}]', \tag{2.160}$$

with

$$\begin{aligned} p_{[1]} &= [p_{[1]1}, p_{[1]2}, p_{[1]3}, p_{[1]4}]' = [0.2917,\ 0.4166,\ 0.1667,\ 0.1250]' \\ p_{[2]} &= [p_{[2]1}, p_{[2]2}, p_{[2]3}, p_{[2]4}]' = [0.1309,\ 0.2381,\ 0.2500,\ 0.3810]'. \end{aligned}$$

Next we follow the steps from the previous section and formulate the matrices and vectors to compute $\hat{\alpha}$ by (2.155).

Step 1. Constructing $F(\pi) = X\alpha$ under LM 1

To define X and α, we write the vector of logits by (2.133) as

$$F(\pi) = [L'_1, L'_2]' : 6 \times 1, \tag{2.161}$$

with

$$\begin{aligned} L_1 &= [L_{[1]1}, L_{[1]2}, L_{[1]3}]' \\ L_2 &= [L_{[2]1}, L_{[2]2}, L_{[2]3}]', \end{aligned}$$

where by (2.130)

$$\begin{aligned} L_{[1]1} &= \alpha_{10} + \alpha_{11} \\ L_{[1]2} &= \alpha_{20} + \alpha_{21} \\ L_{[1]3} &= \alpha_{30} + \alpha_{31}, \end{aligned}$$

and

$$\begin{aligned} L_{[2]1} &= \alpha_{10} \\ L_{[2]2} &= \alpha_{20} \\ L_{[2]3} &= \alpha_{30}, \end{aligned}$$

producing α by (2.135) as

$$\begin{aligned} \alpha &= [\alpha_0', \alpha_1']' \\ &= [\alpha_{10}, \alpha_{20}, \alpha_{30}, \alpha_{11}, \alpha_{21}, \alpha_{31}]'. \end{aligned} \tag{2.162}$$

Now to express $F(\pi)$ in (2.161) as $F(\pi) = X\alpha$ with α as in (2.162), one must write the 6×6 matrix X as

$$X = \begin{bmatrix} 1\,0\,0\,1\,0\,0 \\ 0\,1\,0\,0\,1\,0 \\ 0\,0\,1\,0\,0\,1 \\ 1\,0\,0\,0\,0\,0 \\ 0\,1\,0\,0\,0\,0 \\ 0\,0\,1\,0\,0\,0 \end{bmatrix} \tag{2.163}$$

This matrix satisfies the notations from (2.137) to (2.139).

Note that as indicated in the last section, we also consider three other logit models as follows:

LM2. Instead of using the model (2.130), one may use different restriction on the level effect parameters and write the logit model as

$$L_{[\ell]j} = log\left[\frac{1-F_{[\ell]j}}{F_{[\ell]j}}\right] = \begin{cases} \alpha_{j0}+\alpha_{j\ell} & \text{for } j=1,\ldots,J-1;\ \ell=1,\ldots,p \\ \alpha_{j0}-\sum_{\ell=1}^{p}\alpha_{j\ell} & \text{for } j=1,\ldots,J-1;\ \ell=p+1. \end{cases} \tag{2.164}$$

yielding six logits for the gender versus physician visit data as

$$\begin{aligned} L_{[1]1} &= \alpha_{10}+\alpha_{11} \\ L_{[1]2} &= \alpha_{20}+\alpha_{21} \\ L_{[1]3} &= \alpha_{30}+\alpha_{31}, \end{aligned}$$

and

$$\begin{aligned} L_{[2]1} &= \alpha_{10}-\alpha_{11} \\ L_{[2]2} &= \alpha_{20}-\alpha_{21} \\ L_{[2]3} &= \alpha_{30}-\alpha_{31}. \end{aligned}$$

For

$$\begin{aligned} \alpha &= [\alpha_0', \alpha_1']' \\ &= [\alpha_{10}, \alpha_{20}, \alpha_{30}, \alpha_{11}, \alpha_{21}, \alpha_{31}]', \end{aligned}$$

the aforementioned six logits produce the X matrix as

$$X = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \end{bmatrix}. \tag{2.165}$$

LM3. Now suppose that unlike the models (2.130) and (2.164), one uses the same level effect parameter, say $\alpha_{1\ell}$, under all response categories. Then, similar to LM1, the logits can be expressed as

$$L_{[\ell]j} = log\left[\frac{1-F_{[\ell]j}}{F_{[\ell]j}}\right] = \begin{cases} \alpha_{j0}+\alpha_{1\ell} & \text{for } j=1,\ldots,J-1;\ \ell=1,\ldots,p \\ \alpha_{j0} & \text{for } j=1,\ldots,J-1;\ \ell=p+1. \end{cases} \tag{2.166}$$

yielding six logits for the gender versus physician visit data as

$$\begin{aligned} L_{[1]1} &= \alpha_{10} + \alpha_{11} \\ L_{[1]2} &= \alpha_{20} + \alpha_{11} \\ L_{[1]3} &= \alpha_{30} + \alpha_{11}, \end{aligned}$$

and

$$\begin{aligned} L_{[2]1} &= \alpha_{10} \\ L_{[2]2} &= \alpha_{20} \\ L_{[2]3} &= \alpha_{30}. \end{aligned}$$

For

$$\begin{aligned} \alpha &= [\alpha_0', \alpha_1']' \\ &= [\alpha_{10}, \alpha_{20}, \alpha_{30}, \alpha_{11}]', \end{aligned}$$

the aforementioned six logits produce the $X : 6 \times 4$ matrix as

$$X = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \tag{2.167}$$

LM4. Suppose that we use the same regression parameters as in the model (2.166), but use the restriction on the level effect parameters as in (2.164). One may then express the logits as

$$L_{[\ell]j} = log\left[\frac{1 - F_{[\ell]j}}{F_{[\ell]j}}\right] = \begin{cases} \alpha_{j0} + \alpha_{1\ell} & \text{for } j = 1, \ldots, J-1;\ \ell = 1, \ldots, p \\ \alpha_{j0} - \sum_{\ell=1}^{p} \alpha_{1\ell} & \text{for } j = 1, \ldots, J-1;\ \ell = p+1. \end{cases} \tag{2.168}$$

yielding six logits for the gender versus physician visit data as

$$\begin{aligned} L_{[1]1} &= \alpha_{10} + \alpha_{11} \\ L_{[1]2} &= \alpha_{20} + \alpha_{11} \\ L_{[1]3} &= \alpha_{30} + \alpha_{11}, \end{aligned}$$

and

$$\begin{aligned} L_{[2]1} &= \alpha_{10} - \alpha_{11} \\ L_{[2]2} &= \alpha_{20} - \alpha_{11} \\ L_{[2]3} &= \alpha_{30} - \alpha_{11}. \end{aligned}$$

For

$$\begin{aligned} \alpha &= [\alpha_0', \alpha_1']' \\ &= [\alpha_{10}, \alpha_{20}, \alpha_{30}, \alpha_{11}]', \end{aligned}$$

the aforementioned six logits produce the $X : 6 \times 4$ matrix as

$$X = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix}. \tag{2.169}$$

Step 2. Developing notations to write $F(\pi) = Mlog\,(A\pi)$ satisfying (2.145)

Now because $J = 4$, for a given $\ell(\ell = 1, 2)$, A^* and M^* matrices by (2.143) and (2.144), are written as

$$A^* = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} : 6 \times 4, \tag{2.170}$$

and

$$M^* = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix} : 3 \times 6, \tag{2.171}$$

respectively. Note that these matrices are constructed following the definition of the logits, that is, satisfying

$$L_\ell = [L_{[\ell]1}, L_{[\ell]2}, L_{[\ell]3}]' = M^* log\ (A^* \pi_{[\ell]}),$$

as shown by (2.141)–(2.142). Thus, for the present gender versus physician visit status data, by (2.145), we write

$$M = \begin{pmatrix} M^* & 0U_{3\times 6} \\ 0U_{3\times 6} & M^* \end{pmatrix} : 6 \times 12,\ A = \begin{pmatrix} A^* & 0U_{6\times 4} \\ 0U_{6\times 4} & A^* \end{pmatrix} : 12 \times 8, \tag{2.172}$$

with $U_{3\times 6}$, for example, as the 3×6 unit matrix, satisfying $F(\pi) = Mlog\ (A\pi)$, where

$$\begin{aligned} \pi &= [\pi'_{[1]}, \pi'_{[2]}]' \\ &= [\pi_{[1]1}, \pi_{[1]2}, \pi_{[1]3}, \pi_{[1]4}, \pi_{[2]1}, \pi_{[2]2}, \pi_{[2]3}, \pi_{[2]4}]'. \end{aligned}$$

We now directly go to Step 4 and use (2.155) to compute the WLS estimate for the regression parameter vector α.

Step 4. Computation of $\hat{\alpha}_{WLS}$ by (2.155)

Notice that V matrix in (2.155) is computed by (2.158), that is,

$$\begin{aligned} V = \text{var}[p] &= \text{var}[p'_{[1]}, p'_{[2]}]' \\ &= \begin{pmatrix} \text{var}[p_{[1]}] & \text{cov}[p_{[1]}, p'_{[2]}] \\ \text{cov}[p_{[2]}, p'_{[1]}] & \text{var}[p_{[2]}] \end{pmatrix} \\ &= \begin{pmatrix} V_1 & 0 \\ 0 & V_2 \end{pmatrix}, \end{aligned} \tag{2.173}$$

where

$$\begin{aligned} K_{[1]}V_1 &= \text{diag}[\pi_{[1]1}, \pi_{[1]2}, \pi_{[1]3}, \pi_{[1]4}] - \pi_{[1]}\pi'_{[1]} \\ K_{[2]}V_2 &= \text{diag}[\pi_{[2]1}, \pi_{[2]2}, \pi_{[2]3}, \pi_{[2]4}] - \pi_{[2]}\pi'_{[2]}. \end{aligned}$$

Table 2.25 Parameter estimates and their standard errors under selected cumulative logit models for gender versus physician visit status data

	Logit model parameters based on gender and 4 visit categories						
Logit model	Quantity	$\hat{\alpha}_{10}$	$\hat{\alpha}_{20}$	$\hat{\alpha}_{30}$	$\hat{\alpha}_{11}$	$\hat{\alpha}_{21}$	$\hat{\alpha}_{31}$
LM1	Estimate	1.893	0.537	−0.485	−1.006	−1.424	−1.461
	Standard error	0.324	0.226	0.225	0.394	0.319	0.382
LM2	Estimate	1.390	−0.175	−1.216	−0.503	−0.712	−0.750
	Standard error	0.197	0.159	0.190	0.197	0.159	0.191
LM3	Estimate	2.107	0.508	−0.524	−1.312	-	-
	Standard error	0.261	0.215	0.214	0.285	-	-
LM4	Estimate	1.451	−0.148	−1.180	−0.656	-	-
	Standard error	0.189	0.157	0.181	0.142	-	-

One however needs to use an estimate of this V matrix to compute $\hat{\alpha}_{WLS}$ by (2.155). Now because $p_{[1]}$ and $p_{[2]}$ are unbiased estimates for $\pi_{[1]}$ and $\pi_{[2]}$, respectively, V matrix may be estimated as

$$\hat{V} = \begin{pmatrix} \hat{V}_1 & 0 \\ 0 & \hat{V}_2 \end{pmatrix}, \tag{2.174}$$

where

$$K_{[1]}\hat{V}_1 = \text{diag}[p_{[1]1}, p_{[1]2}, p_{[1]3}, p_{[1]4}] - p_{[1]}p'_{[1]}$$
$$K_{[2]}\hat{V}_2 = \text{diag}[p_{[2]1}, p_{[2]2}, p_{[2]3}, p_{[2]4}] - p_{[2]}p'_{[2]},$$

with $p[1]$ and $p_{[2]}$ as given by (2.160).

Next we compute $\hat{D} = \text{diag}[Ap]$, where A is given in (2.172). Further compute $\hat{Q} = M\hat{D}^{-1}A$. Finally by using these estimates $\hat{V}, \hat{Q}$, and $F(p) = Mlog\ (Ap)$ into (2.155), we obtain $\hat{\alpha}_{WLS}$ by using X matrix from (2.163), (2.165), (2.167), and (2.169), under the models LM1, LM2, LM3, and LM4, respectively. These estimates along with their standard errors computed by (2.157) are reported in Table 2.25.

We now use the estimates from Table 2.25 and compute the logits under all four models. The observed logits are also computed using the observed proportions from Table 2.24. For interpretation convenience we display the exponent of the logits, i.e., $\exp(L_{[\ell]j})$ under all four models in Table 2.26. Notice that LM1 and LM2 produce the same logits, similarly LM3 and LM4 also produce the same logits. Thus, proper restriction on level based parameters is important but restrictions can vary. Next, it is clear from the table that LM1 (or LM2) fits the observed logits exactly, whereas the logits produced by LM3 (or LM4) are slightly different than the observed logits. This shows that level (gender) based covariates do not play the same role under all four response categories. Thus, using three different regression parameters, namely α_{1j} for $j = 1, \ldots, 3$, is more appropriate than using only one parameter, namely α_{11}.

Table 2.26 Observed and estimated logits under selected cumulative logit models for gender versus physician visit status data

		Logit estimates				
Gender	Logits	Observed	LM1	LM2	LM3	LM4
Male	$\exp(L_{[1]1})$	2.428	2.428	2.428	2.214	2.214
	$\exp(L_{[1]2})$	0.411	0.411	0.411	0.447	0.447
	$\exp(L_{[1]3})$	0.143	0.143	0.143	0.159	0.159
Female	$\exp(L_{[2]1})$	6.639	6.639	6.639	8.225	8.225
	$\exp(L_{[2]2})$	1.710	1.710	1.710	1.662	1.662
	$\exp(L_{[2]3})$	0.616	0.616	0.616	0.592	0.592

Furthermore, when logits of males are compared to those of the females, all three logits for the male group appear to be smaller than the corresponding logits for the female group, i.e.,

$$L_{[1]j} \leq L_{[2]j}, \text{ for all } j = 1, 2, 3,$$

showing that more females pay large number of visits to their physician as compared to males. These results agree with the analysis discussed in Sect. 2.2.2.4.2(a) and the results reported in Table 2.12, where it was found through direct multinomial regression fitting that females paid relatively more visits as compared to males.

2.3.1.2 Binary Mapping Based Pseudo-Likelihood Estimation Approach

Based on the form of the cumulative logits from (2.130)–(2.131), in this approach we utilize the binary information at every cut point for an individual and write a likelihood function. For the purpose, for an individual i with covariate level ℓ and responding in category $h(h = 1, \ldots, J)$ [this identifies the ith individual as $i \in (\ell, h)$], we define a cut point j $(j = 1, \ldots, J-1)$ based 'working' or 'pseudo' binary variable

$$b_{i\in(\ell,h)}^{(j)} = \begin{cases} 1 \text{ for given response category } h > j \\ 0 \text{ for given response category } h \leq j, \end{cases} \tag{2.175}$$

with probabilities following (2.130)–(2.131) as

$$\begin{aligned} Pr[b_{i\in(\ell,h)}^{(j)} = 1] &= \sum_{c=j+1}^{J} \pi_{[\ell]c} = 1 - F_{[\ell]j} \\ &= \begin{cases} \frac{\exp(\alpha_{j0}+\alpha_{j\ell})}{1+\exp(\alpha_{j0}+\alpha_{j\ell})} & \text{for } \ell = 1, \ldots, p \\ \frac{\exp(\alpha_{j0})}{1+\exp(\alpha_{j0})} & \text{for } \ell = p+1. \end{cases} \end{aligned} \tag{2.176}$$

Table 2.27 Cumulative counts as responses at cut points $j = 1,\ldots,J-1$, reflecting the cumulative probabilities (2.176), under covariate level ℓ

	Binomial response		
Cut point	**Low group ($g^* = 1$)**	**High group ($g^* = 2$)**	**Total**
1	$K^*_{[\ell]1} = \sum_{c=1}^{1} K_{[\ell]c}$	$K_{[\ell]} - K^*_{[\ell]1}$	$K_{[\ell]}$
.	.	.	.
j	$K^*_{[\ell]j} = \sum_{c=1}^{j} K_{[\ell]c}$	$K_{[\ell]} - K^*_{[\ell]j}$	$K_{[\ell]}$
.	.	.	.
$J-1$	$K^*_{[\ell](J-1)} = \sum_{c=1}^{J-1} K_{[\ell]c}$	$K_{[\ell]} - K^*_{[\ell](J-1)}$	$K_{[\ell]}$

representing the probability for the binary response to be in category h beyond j; and

$$Pr[b^{(j)}_{i\in(\ell,h)} = 0] = \sum_{c=1}^{j} \pi_{[\ell]c} = F_{[\ell]j}$$
$$= = \begin{cases} \frac{1}{1+\exp(\alpha_{j0}+\alpha_{j\ell})} & \text{for } \ell = 1,\ldots,p \\ \frac{1}{1+\exp(\alpha_{j0})} & \text{for } \ell = p+1. \end{cases} \quad (2.177)$$

representing the probability for the binary response to be in a category h between 1 and j inclusive.

Now as a reflection of the cut points based cumulative probabilities (2.176)–(2.177), for convenience, we display the response counts computed from Table 2.4, at every cut points, as in Table 2.27. We use the notation $K^*_{[\ell]j} = \sum_{c=1}^{j} K_{[\ell]c}$, whereas in Table 2.4, $K_{[\ell]c}$ is the number of individuals with covariate at level ℓ those belong to category c for their responses.

Note that $K_{[\ell]} - K^*_{[\ell]j}$ follows the binomial distribution $\text{Bin}(K_{[\ell]}, 1 - F_{[\ell]j})$, where $[1 - F_{[\ell]j}] = \sum_{c=j+1}^{J} \pi_{[\ell]c} = \pi^*_{[\ell]j}$ by (2.176). Furthermore, the regression parameters in (2.176)–(2.177) may be expressed by a vector α as in (2.135), that is,

$$\alpha = \left[\alpha'_0, \alpha'_1, \ldots, \alpha'_\ell, \ldots, \alpha'_p\right]',$$

where

$$\alpha_0 = [\alpha_{10}, \ldots, \alpha_{(J-1)0}]' \text{ and } \alpha_\ell = [\alpha_{1\ell}, \ldots, \alpha_{(J-1)\ell}]',$$

for $\ell = 1,\ldots,p$. Alternatively, similar to (2.69), these parameters may be represented by

$$\alpha = [\alpha_1^{*\prime}, \ldots, \alpha_j^{*\prime}, \ldots, \alpha^{*}_{J-1}{}']', \text{ with } \alpha^*_j = [\alpha_{j0}, \alpha_{j1}, \ldots, \alpha_{j\ell}, \ldots, \alpha_{jp}]'. \quad (2.178)$$

Now for a given form of α, we first write a pseudo-likelihood function by using the pseudo binary probabilities from (2.175)–(2.177), as

$$\begin{aligned}
L(\alpha) &= \Pi_{\ell=1}^{p+1}\Pi_{j=1}^{J-1}\Pi_{[i\in(\ell,h)|j,\ell]}^{K_{[\ell]}}\left[\{F_{[\ell]j}\}^{1-b_{i\in(\ell,h)}^{(j)}}\right]\left[\{1-F_{[\ell]j}\}^{b_{i\in(\ell,h)}^{(j)}}\right] \\
&= \Pi_{\ell=1}^{p+1}\Pi_{j=1}^{J-1}\left[\{F_{[\ell]j}\}^{\Sigma_{c=1}^{j}K_{[\ell]c}}\right]\left[\{1-F_{[\ell]j}\}^{\Sigma_{c=j+1}^{J}K_{[\ell]c}}\right] \\
&= \Pi_{\ell=1}^{p+1}\Pi_{j=1}^{J-1}\left[\{F_{[\ell]j}\}^{\Sigma_{c=1}^{j}K_{[\ell]c}}\right]\left[\{1-F_{[\ell]j}\}^{K_{[\ell]}-\Sigma_{c=1}^{j}K_{[\ell]c}}\right] \qquad (2.179) \\
&= \Pi_{j=1}^{J-1}\left[\Pi_{\ell=1}^{p}\frac{\exp\{(K_{[\ell]}-K^*_{[\ell]j})(\alpha_{j0}+\alpha_{j\ell})\}}{[1+\exp(\alpha_{j0}+\alpha_{j\ell})]^{K_{[\ell]}}}\right] \\
&\quad\times\left[\frac{\exp\{(K_{[p+1]}-K^*_{[p+1]j})(\alpha_{j0})\}}{[1+\exp(\alpha_{j0})]^{K_{[p+1]}}}\right], \qquad (2.180)
\end{aligned}$$

where $K^*_{[\ell]j}=\Sigma_{c=1}^{j}K_{[\ell]c}$ for $j=1,\ldots,J-1$, and for all $\ell=1,\ldots,p+1$.

Next, in order to write the log likelihood estimating equation in an algebraic convenient form, we use the α in the form of (2.178) and first re-express $1-F_{[\ell]j}$ and $F_{[\ell]j}$ from (2.176)–(2.177) as

$$\begin{aligned}
1-F_{[\ell]j} &= \frac{\exp(x'_{[\ell]j}\alpha)}{1+\exp(x'_{[\ell]j}\alpha)} \\
F_{[\ell]j} &= \frac{1}{1+\exp(x'_{[\ell]j}\alpha)}, \qquad (2.181)
\end{aligned}$$

where $x'_{[\ell]j}$ is the jth $(j=1,\ldots,J-1)$ row of the $(J-1)\times(J-1)(p+1)$ matrix X_ℓ, defined for ℓth level as follows:

$$\begin{aligned}
X_\ell &= \begin{pmatrix} x'_{[\ell]1} \\ x'_{[\ell]2} \\ \cdot \\ x'_{[\ell](J-1)} \end{pmatrix} \\
&= \begin{pmatrix} 1\ 01'_{\ell-1},1,01'_{p-\ell} & 0 & 01'_p & \cdot & 0 & 01'_p \\ 0 & 01'_p & 1\ 01'_{\ell-1},1,01'_{p-\ell} & \cdot & 0 & 01'_p \\ \cdot & \cdot & \cdot & \cdot\ \cdot & & \cdot \\ 0 & 01'_p & 0 & 01'_p & \cdot & 1\ 01'_{\ell-1},1,01'_{p-\ell} \end{pmatrix} \text{ for } \ell=1,\ldots,p
\end{aligned}$$

$$X_{p+1} = \begin{pmatrix} x'_{[p+1]1} \\ x'_{[p+1]2} \\ \cdot \\ x'_{[p+1](J-1)} \end{pmatrix} = \begin{pmatrix} 1\ 01'_p & 0\ 01'_p & \cdot & 0\ 01'_p \\ 0\ 01'_p & 1\ 01'_p & \cdot & 0\ 01'_p \\ \cdot\ \cdot & \cdot\ \cdot & \cdot & \cdot\ \cdot \\ 0\ 01'_p & 0\ 01'_p & \cdot & 1\ 01'_p \end{pmatrix}. \tag{2.182}$$

The log likelihood equation for α may then be written from (2.179) as

$$\begin{aligned}
\frac{\partial Log\, L(\alpha)}{\partial \alpha} &= \sum_{\ell=1}^{p+1}\sum_{j=1}^{J-1}\Big[(K_{[\ell]} - K^*_{[\ell]j})\frac{\partial}{\partial \alpha}\{log\ (1 - F_{[\ell]j})\} \\
&+\ K^*_{[\ell]j}\frac{\partial}{\partial \alpha}\{log\ F_{[\ell]j}\}\Big] \\
&= \sum_{\ell=1}^{p+1}\sum_{j=1}^{J-1}\Big[(K_{[\ell]} - K^*_{[\ell]j})\frac{\partial}{\partial \alpha}\{log\ (\frac{\exp(x'_{[\ell]j}\alpha)}{1+\exp(x'_{[\ell]j}\alpha)})\} \\
&+\ K^*_{[\ell]j}\frac{\partial}{\partial \alpha}\{log\ \frac{1}{1+\exp(x'_{[\ell]j}\alpha)}\}\Big] \\
&= \sum_{\ell=1}^{p+1}\sum_{j=1}^{J-1}\Big[(K_{[\ell]} - K^*_{[\ell]j})\{F_{[\ell]j}x_{[\ell]j}\} \\
&-\ K^*_{[\ell]j}\{(1 - F_{[\ell]j})x_{[\ell]j}\}\}\Big] \\
&= \sum_{\ell=1}^{p+1}\sum_{j=1}^{J-1} x_{[\ell]j}\Big[K_{[\ell]}F_{[\ell]j} - K^*_{[\ell]j}\Big] \\
&= -\sum_{\ell=1}^{p+1}\sum_{j=1}^{J-1} x_{[\ell]j}\Big[K^*_{[\ell]j} - K_{[\ell]}F_{[\ell]j}\Big] \\
&= \sum_{\ell=1}^{p+1}\sum_{j=1}^{J-1} x_{[\ell]j}\Big[(K_{[\ell]} - K^*_{[\ell]j}) - K_{[\ell]}(1 - F_{[\ell]j})\Big] \qquad (2.183) \\
&= \sum_{\ell=1}^{p+1} X'_\ell\Big[y^*_{[\ell]} - K_{[\ell]}\pi^*_{[\ell]}\Big] = f(\alpha) = 0, \qquad (2.184)
\end{aligned}$$

where

$$y^*_{[\ell]} = [K_{[\ell]} - K^*_{[\ell]1}, \ldots, K_{[\ell]} - K^*_{[\ell]j}, \ldots, K_{[\ell]} - K^*_{[\ell](J-1)}]' \text{ and}$$
$$\pi^*_{[\ell]} \equiv [\pi^*_{[\ell]1}, \ldots, \pi^*_{[\ell]j}, \ldots, \pi^*_{[\ell](J-1)}]' = [1 - F_{[\ell]1}, \ldots, 1 - F_{[\ell]j}, \ldots, 1 - F_{[\ell](J-1)}]',$$

with X_ℓ matrices for $\ell = 1, \ldots, p+1$ as given in (2.182). Note that this estimating equation form in (2.184) is similar to (2.71), but they are quite different estimating equations.

2.3.1.2.1 Pseudo-Likelihood Estimates and their Asymptotic Variances

Let $\hat{\alpha}$ be the solution of $f(\alpha) = 0$ in (2.184). Assuming that $\hat{\alpha}_0$ is not a solution for $f(\alpha) = 0$ but a trial estimate, and hence $f(\hat{\alpha}_0) \neq 0$, by similar calculations as in (2.36), the iterative equation for $\hat{\alpha}*$ is obtained as

$$\hat{\alpha} = \hat{\alpha}_0 - \left[\{f'(\alpha)\}^{-1} f(\alpha)\right] |_{\alpha=\hat{\alpha}_0}. \tag{2.185}$$

Next, by similar calculations as in (2.183), one writes

$$\begin{aligned}
\frac{\partial \pi^*_{[\ell]j}}{\partial \alpha'} &= \frac{\partial(1 - F_{[\ell]j})}{\partial \alpha'} \\
&= F_{[\ell]j}(1 - F_{[\ell]j})x'_{[\ell]j} = \pi^*_{[\ell]j}(1 - \pi^*_{[\ell]j})x'_{[\ell]j},
\end{aligned} \tag{2.186}$$

yielding

$$\begin{aligned}
\frac{\partial \pi^*_{[\ell]}}{\partial \alpha'} &= \mathrm{diag}[\pi^*_{[\ell]1}(1 - \pi^*_{[\ell]1}), \ldots, \pi^*_{[\ell](J-1)}(1 - \pi^*_{[\ell](J-1)})]X_\ell \\
&= D_{\pi^*_{[\ell]}} X_\ell.
\end{aligned} \tag{2.187}$$

By (2.187), it then follows from (2.184) that

$$\begin{aligned}
f'(\alpha) &= \frac{\partial^2 Log\, L(\alpha)}{\partial \alpha \partial \alpha'} \\
&= -\sum_{\ell=1}^{p+1} K_{[\ell]} X'_\ell D_{\pi^*_{[\ell]}} X_\ell.
\end{aligned} \tag{2.188}$$

Thus, by (2.188), the iterative equation (2.185) takes the form

$$\begin{aligned}
\hat{\alpha}(r+1) = \hat{\alpha}(r) &+ \left[\sum_{\ell=1}^{p+1} K_{[\ell]} X'_\ell D_{\pi^*_{[\ell]}} X_\ell\right]^{-1} \\
&\times \left[\sum_{\ell=1}^{p+1} X'_\ell \left(y^*_{[\ell]} - K_{[\ell]} \pi^*_{[\ell]}\right)\right]_{\alpha=\hat{\alpha}(r)},
\end{aligned} \tag{2.189}$$

yielding the final estimate $\hat{\alpha}$.

Next because

$$\begin{aligned}\mathrm{var}[y^*_{[\ell]j} - K_{[\ell]}\pi^*_{[\ell]j}] &= \mathrm{var}[K^*_{[\ell]j} - K_{[\ell]}F_{[\ell]j}] \\ &= \mathrm{var}[\sum_{c=1}^{j} K_{[\ell]c}], \end{aligned} \tag{2.190}$$

and $K^*_{[\ell]j}$ follows the binomial distribution with parameters $K_{[\ell]}$ and $\pi^*_{[\ell]j} = [1 - F_{[\ell]j}]$, one writes

$$\mathrm{var}[y^*_{[\ell]j} - K_{[\ell]}\pi^*_{[\ell]j}] = K_{[\ell]}F_{[\ell]j}[1 - F_{[\ell]j}] = K_{[\ell]}\pi^*_{[\ell]j}[1 - \pi^*_{[\ell]j}]. \tag{2.191}$$

It then follows from (2.189) that $\mathrm{var}(\hat{\alpha})$ has the formula given by

$$\mathrm{var}(\hat{\alpha}) = \left[\sum_{\ell=1}^{p+1} K_{[\ell]} X'_\ell D_{\pi^*_{[\ell]}} X_\ell\right]^{-1}. \tag{2.192}$$

2.3.1.3 Binary Mapping Based GQL Estimation Approach

By Table 2.27, consider the response vector

$$y^*_{[\ell]} = [K_{[\ell]} - K^*_{[\ell]1}, \ldots, K_{[\ell]} - K^*_{[\ell]j}, \ldots, K_{[\ell]} - K^*_{[\ell](J-1)}]'$$

[see also (2.184)], where

$$y^*[\ell]j = [K_{[\ell]} - K^*_{[\ell]j}] \sim \mathrm{Bin}(K_{[\ell]}, \pi^*_{[\ell]j}),$$

with

$$\pi^*_{[\ell]j} = 1 - F_{[\ell]j} = \frac{\exp(x'_{[\ell]j}\alpha)}{1 + \exp(x'_{[\ell]j}\alpha)}$$

by (2.181). By following Sutradhar (2003, Section 3), one may then write a GQL estimating equation for α as

$$\sum_{\ell=1}^{p+1} \frac{\partial [K_{[\ell]}\pi^{*\prime}_{[\ell]}]}{\partial \alpha} \left[\mathrm{cov}(Y^*_{[\ell]})\right]^{-1} [y^*_{[\ell]} - K_{[\ell]}\pi^*_{[\ell]}] = 0, \tag{2.193}$$

where

$$\frac{\partial \pi^{*\prime}_{[\ell]}}{\partial \alpha} = X'_\ell D_{\pi^*_{[\ell]}}$$

by (2.187), and

$$\mathrm{cov}[Y^*_{[\ell]}] = K_{[\ell]} D_{\pi^*_{[\ell]}}$$

by (2.191). The GQL estimating equation (2.193) then reduces to

$$\sum_{\ell=1}^{p+1} X'_\ell [y^*_{[\ell]} - K_{[\ell]} \pi^*_{[\ell]}] = 0,$$

which is the same as the pseudo-likelihood estimating equation given by (2.184). Hence the GQL estimate of α is the same as the likelihood estimate found by (2.189), and its asymptotic covariance matrix is the same as that of the likelihood estimates given by (2.192).

2.3.1.4 Some Remarks on GQL Estimation for Fitting the Multinomial Model (3.63) Subject to Category Order Restriction

Notice from (2.184) that

$$\begin{aligned} y^*_{[\ell]} &= [K_{[\ell]} - K^*_{[\ell]1}, \ldots, K_{[\ell]} - K^*_{[\ell]j}, \ldots, K_{[\ell]} - K^*_{[\ell](J-1)}]' \\ &= [y^*_{[\ell]1}, \ldots, y^*_{[\ell]j}, \ldots, y^*_{[\ell](J-1)}]', \end{aligned} \tag{2.194}$$

is a cumulative response vector with its expectation

$$\begin{aligned} E[y^*_{[\ell]}] &= K_{[\ell]} \pi^*_{[\ell]} \\ &\equiv K_{[\ell]} [\pi^*_{[\ell]1}, \ldots, \pi^*_{[\ell]j}, \ldots, \pi^*_{[\ell](J-1)}]' \\ &= K_{[\ell]} [1 - F_{[\ell]1}, \ldots, 1 - F_{[\ell]j}, \ldots, 1 - F_{[\ell](J-1)}]', \end{aligned} \tag{2.195}$$

with

$$\pi^*_{[\ell]j} = 1 - F_{[\ell]j} = \sum_{c=j+1}^{J} \pi_{[\ell]c}, \tag{2.196}$$

where, by (2.63) and (2.64), the multinomial probabilities are defined as

$$\pi_{[\ell]c} = \begin{cases} \frac{\exp(\beta_{c0}+\beta_{c\ell})}{1+\sum_{g=1}^{J-1}\exp(\beta_{g0}+\beta_{g\ell})} & \text{for } c = 1, \ldots, J-1 \\ \frac{1}{1+\sum_{g=1}^{J-1}\exp(\beta_{g0}+\beta_{g\ell})} & \text{for } c = J, \end{cases} \tag{2.197}$$

for $\ell = 1, \ldots, p$, whereas for $\ell = p+1$, these probabilities are given as

$$\pi_{[p+1]c} = \begin{cases} \frac{\exp(\beta_{c0})}{1+\sum_{g=1}^{J-1}\exp(\beta_{g0})} & \text{for } c = 1,\ldots,J-1 \\ \frac{1}{1+\sum_{g=1}^{J-1}\exp(\beta_{g0})} & \text{for } c = J. \end{cases} \quad (2.198)$$

Use $x_{[\ell]J} = 01_{(J-1)(p+1)}$ for all $\ell = 1,\ldots,p+1$, along with $x_{[\ell]c}$ from (2.182) for $c = 1,\ldots,J-1$; and $\ell = 1,\ldots,p+1$, and re-express all $\pi_{[\ell]c}$ in (2.196)–(2.197) as

$$\pi_{[\ell]c} = \frac{\exp(x'_{[\ell]c}\beta)}{\sum_{g=1}^{J}\exp(x'_{[\ell]g}\beta)}, \quad (2.199)$$

where, similar to (2.178),

$$\beta = [\beta_1^{*\prime},\ldots,\beta_j^{*\prime},\ldots,\beta_{J-1}^{*\prime}]', \text{ with } \beta_j^* = [\beta_{j0},\beta_{j1},\ldots,\beta_{j\ell},\ldots,\beta_{jp}]'. \quad (2.200)$$

Note that α parameters in (2.178) and β parameters in (2.198) are different, even though they have some implicit connection. Here, one is interested to estimate β for the purpose of comparing $\pi^*_{[\ell]j} = \sum_{c=1}^{J}\pi_{[\ell]c}$ with $1-\pi^*_{[\ell]j} = \sum_{c=1}^{j}\pi_{[\ell]c}$. We construct a GQL estimating equation (Sutradhar 2004, 2011) for β as follows.

2.3.1.4.1 GQL Estimating Equation for β

2.3.1.4.1 (a) Computation of $\text{cov}(y^*_{[\ell]}) = \Gamma_{[\ell]} = (\gamma_{[\ell]jh}) : (J-1)\times(J-1)$

The elements of the Γ matrix are computed as follows.

$$\begin{aligned} \gamma_{[\ell]jj} &= \text{var}[y^*_{[\ell]j} - K_{[\ell]}\pi^*_{[\ell]j}] \\ &= \text{var}[K^*_{[\ell]j} - K_{[\ell]}F_{[\ell]j}] \\ &= \text{var}[\sum_{c=1}^{j} K_{[\ell]c}] \\ &= K_{[\ell]}\left[\sum_{c=1}^{j}\pi_{[\ell]c}(1-\pi_{[\ell]c}) - \sum_{c\neq c'}^{j}\pi_{[\ell]c}\pi_{[\ell]c'}\right], \text{ for } j = 1,\ldots,J-1. \end{aligned} \quad (2.201)$$

Next, for $j < h, j,h = 1,\ldots,J-1$,

$$\begin{aligned} \gamma_{[\ell]jh} &= \text{cov}[y^*_{[\ell]j} - K_{[\ell]}\pi^*_{[\ell]j}, y^*_{[\ell]h} - K_{[\ell]}\pi^*_{[\ell]h}] \\ &= \text{cov}[\sum_{c=1}^{j} K_{[\ell]c}, \sum_{c=1}^{h} K_{[\ell]c}] \end{aligned}$$

$$= K_{[\ell]}\left[\sum_{c=1}^{j}\pi_{[\ell]c}(1-\pi_{[\ell]c}) - \sum_{c\neq c'}^{j}\pi_{[\ell]c}\pi_{[\ell]c'} - \sum_{c=1}^{j}\sum_{c'=j+1}^{h}\pi_{[\ell]c}\pi_{[\ell]c'}\right] \quad (2.202)$$

Also it follows that $\gamma_{[\ell]jh} = \gamma_{[\ell]hj}$.

2.3.1.4.1 (b) Computation of $\frac{\partial \pi^{*\prime}_{[\ell]}}{\partial \beta} : (J-1)(p+1)\times(J-1)$

It is sufficient to compute the derivative of a general element, say $\pi^*_{[\ell]j}$ with respect to β. That is,

$$\begin{aligned}\frac{\partial \pi^*_{[\ell]j}}{\partial \beta} &= \sum_{c=j+1}^{J}\frac{\partial \pi_{[\ell]c}}{\partial \beta} \\ &= \sum_{c=j+1}^{J}\left[\pi_{[\ell]c}\left\{x_{[\ell]c} - \sum_{u=1}^{J}\pi_{[\ell]u}x_{[\ell]u}\right\}\right] \\ &= \sum_{c=j+1}^{J}\left[\pi_{[\ell]c}\left\{x_{[\ell]c} - X'_{\ell}\pi_{[\ell]}\right\}\right] \\ &= A^*_{[\ell]j}(x,\beta):\ (J-1)(p+1)\times 1,\ \text{(say)}, \end{aligned} \quad (2.203)$$

yielding

$$\begin{aligned}\frac{\partial \pi^{*\prime}_{[\ell]}}{\partial \beta} &= \left(A^*_{[\ell]1}(x,\beta)\ \ldots\ A^*_{[\ell]j}(x,\beta)\ \ldots\ A^*_{[\ell](J-1)}(x,\beta)\right)\ (J-1)(p+1)\times(J-1) \\ &= A^*_{[\ell]}(x,\beta),\ \text{(say)}. \end{aligned} \quad (2.204)$$

Next, by following Sutradhar (2004), and using (2.200)–(2.201), and (2.203), we can write a GQL estimating equation for β as

$$\begin{aligned}&\sum_{\ell=1}^{p+1}K_{[\ell]}\frac{\partial \pi^{*\prime}_{[\ell]}}{\partial \beta}\Gamma^{-1}_{[\ell]}\left(y^*_{[\ell]} - K_{[\ell]}\pi^*_{[\ell]}\right) \\ &= \sum_{\ell=1}^{p+1}K_{[\ell]}A^*_{[\ell]}(x,\beta)\Gamma^{-1}_{[\ell]}\left(y^*_{[\ell]} - K_{[\ell]}\pi^*_{[\ell]}\right) = 0 \end{aligned} \quad (2.205)$$

The solution of this GQL estimating equation (2.204) for β may be obtained iteratively by using the iterative formula

$$\hat{\beta}(r+1) = \hat{\beta}(r) + \left[\left\{ \sum_{\ell=1}^{p+1} K_{[\ell]}^2 A_{[\ell]}^*(x,\beta) \Gamma_{[\ell]}^{-1} A_{[\ell]}^{*\prime}(x,\beta) \right\}^{-1} \right.$$
$$\left. \times \sum_{\ell=1}^{p+1} K_{[\ell]} A_{[\ell]}^*(x,\beta) \Gamma_{[\ell]}^{-1} \left(y_{[\ell]}^* - K_{[\ell]} \pi_{[\ell]}^* \right) \right]_{\beta=\hat{\beta}(r)}, \quad (2.206)$$

yielding the final GQL estimate $\hat{\beta}_{GQL}$, along with its asymptotic (as $\min_{1 \le \ell \le p+1} K_{[\ell]} \to \infty$) covariance matrix

$$\mathrm{cov}[\hat{\beta}_{GQL}] = \left[\sum_{\ell=1}^{p+1} K_{[\ell]}^2 A_{[\ell]}^*(x,\beta) \Gamma_{[\ell]}^{-1} A_{[\ell]}^{*\prime}(x,\beta) \right]^{-1}. \quad (2.207)$$

References

Agresti, A. (1984). *Analysis of ordinal categorical data.* New York: Wiley.

Agresti, A. (1990). *Categorical data analysis* (1st ed.). New York: Wiley.

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York: Wiley.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis: Theory and practice*. Cambridge: MIT Press.

Grizzle, J. E., Starmer, C. F., & Koch, G. G. (1969). Analysis of categorical data by linear models. *Biometrics, 25*, 489–504.

Haberman, S. J. (1978). *Analysis of qualitative data. Introductory topics* (Vol. 1). New York: Academic Press.

Koch, G. G., Singer, J. M., & Stokes, M. E. (1992). Some aspects of weighted least squares analysis for longitudinal categorical data. In J. H. Dwyer, M. Feinleib, P. Lippert, & H. Hoffmeister (Eds.), *Statistical Models for Longitudinal Studies of Health, Monographs in Epidemiology and Biostatistics* (Vol. 16, pp. 215–258). New York: Oxford University Press. ISBN: 0-19-505473-3

Rao, C. R. (1973). *Linear statistical inference and its applications.* New York: Wiley.

Semenya, K., & Koch, G. G. (1980). *Compound function and linear model methods for the multivariate analysis of ordinal categorical data. Mimeo Series* (Vol. 1323). Chapel Hill: University of North Carolina Institute of Statistics.

Semenya, K., Koch, G. G., Stokes, M. E., & Forthofer, R. N. (1983). Linear models methods for some rank function analyses of ordinal categorical data. *Communication in Statistics Series A, 12*, 1277–1298.

Sutradhar, B. C. (2003). An overview on regression models for discrete longitudinal responses. *Statistical Science, 18*, 377–393.

Sutradhar, B. C. (2004). On exact quasilikelihood inference in generalized linear mixed models. *Sankhya B: The Indian Journal of Statistics, 66*, 261–289.

Sutradhar, B. C. (2011). *Dynamic mixed models for familial longitudinal data.* New York: Springer.

http://www.springer.com/978-1-4939-2136-2

Longitudinal Categorical Data Analysis
Sutradhar, B.C.
2014, XVIII, 369 p., Hardcover
ISBN: 978-1-4939-2136-2