5 INCOMPLETE CONTINGENCY TABLES

An incomplete contingency table is a contingency table where information on one or more of the categorical variables is missing. It is assumed that the data are MAR and the missing data mechanism is ignorable. This chapter discusses ML estimation of cell probabilities in an incomplete contingency table by using all the observed data - including data where information on one or more of the categorical variables is missing. Lipsitz, Parzen and Molenberghs (1998) uses the Poisson generalized linear model to obtain ML estimates of cell probabilities for the saturated loglinear model whilst Little and Rubin (1987) describes and uses the EM algorithm to determine the ML estimates of cell probabilities for any loglinear model. Maximum likelihood estimation under constraints is also discussed in this chapter as a method to determine the ML estimates of cell probabilities. The advantage of this method is that it is less computational intensive compared to the more generally used EM algorithm. It also illustrates the elegance with which the method of ML estimation under constraints can be applied.

5.1 ML ESTIMATION IN INCOMPLETE CONTINGENCY TABLES

Consider an $I \times J$ contingency table with categorical variables $C_1 = \{1, 2, \ldots, I\}$ and $C_2 = \{1, 2, \ldots, J\}$. A multinomial sampling procedure is assumed. Let Y_{ij} be the count in cell (i, j), y_{ij} the observed value of Y_{ij} and $n = \sum \sum y_{ij}$ the total counts. The counts in each cell can be arranged to form the complete data vector $\mathbf{Y}' = (Y_{11}, Y_{12}, \ldots, Y_{IJ})$ with $E(\mathbf{Y}) = \boldsymbol{\mu}$, the vector of expected counts.

If information on one or both of the categories is missing the contingency table is said to be incomplete. The data to be classified in the contingency table can be split into two parts namely:
   - the fully classified cases where information on all the categories is available and,
   - the partially classified cases where information on some of the categories is missing.
It is assumed that the data are MAR and the missing data mechanism is ignorable.

In this section the saturated model is considered and the EM algorithm and ML estimation under constraints are described and illustrated as methods which uses both the fully and partially classified cases to determine the ML estimates of the cell probabilities.

5.1.1 The EM Algorithm

Multinomial Sampling
If the probability that an observation falls in cell (i, j) is π_{ij}, where $\pi_{ij} \geq 0$ and $\sum \sum \pi_{ij} = 1$ then the complete data $\mathbf{Y}$ have a multinomial distribution,

$$\mathbf{Y} \sim Mult(n; \pi_{11}, \pi_{12}, \ldots, \pi_{IJ})$$

with probability function

$$f(\mathbf{y}|\boldsymbol{\pi}) = \frac{n!}{\prod \prod y_{ij}!} \pi_{11}^{y_{11}} \pi_{12}^{y_{12}} \cdots \pi_{IJ}^{y_{IJ}} \tag{56}$$

where $\boldsymbol{\pi}' = (\pi_{11}, \pi_{12}, \ldots, \pi_{IJ})$.
The kernel of the complete data log-likelihood is

$$L(\boldsymbol{\pi}|\mathbf{y}) = y_{11} \log \pi_{11} + y_{12} \log \pi_{12} + \cdots + y_{IJ} \log \pi_{IJ}.$$

The cell counts, Y_{ij}, are the sufficient statistics and the MLE of π_{ij} is

$$\widehat{\pi}_{ij} = \frac{y_{ij}}{n}.$$

Product Multinomial Sampling

Let $Y_{i+} = \sum_j Y_{ij}$ be the total counts in row i and $\pi_{i+} = \sum_j \pi_{ij}$ the probability that an element falls in row i. If the Y_{i+} elements of row i are independent, each having a probability distribution $\frac{\pi_{ij}}{\pi_{i+}}$, $j = 1, 2, \ldots, J$, then, given the row total Y_{i+} and the vector of cell probabilities π, the elements of row i have a multinomial distribution

$$Y_{i1}, Y_{i2}, \cdots, Y_{iJ} | Y_{i+}, \pi \sim Mult\left(y_{i+}; \frac{\pi_{i1}}{\pi_{i+}}, \frac{\pi_{i2}}{\pi_{i+}}, \cdots, \frac{\pi_{iJ}}{\pi_{i+}}\right) \tag{57}$$

and $E(Y_{ij}|Y_{i+} = y_{i+}) = y_{i+}\left(\frac{\pi_{ij}}{\pi_{i+}}\right)$.

When samples from different rows are independent, the joint probability function for the entire data set is the product of I multinomial probability functions,

$$f(\mathbf{y}|\pi, y_{1+}, y_{2+}, \ldots, y_{I+}) = \prod_{i=1}^{I}\left[\frac{y_{i+}!}{y_{i1}!y_{i2}!\cdots y_{iJ}!}\left(\frac{\pi_{i1}}{\pi_{i+}}\right)^{y_{i1}}\left(\frac{\pi_{i2}}{\pi_{i+}}\right)^{y_{i2}}\cdots\left(\frac{\pi_{iJ}}{\pi_{i+}}\right)^{y_{iJ}}\right].$$

Similarly, if the column totals are fixed then the elements of column j will have a multinomial distribution

$$Y_{1j}, Y_{2j}, \cdots, Y_{Ij} | Y_{+j}, \pi \sim Mult\left(y_{+j}; \frac{\pi_{1j}}{\pi_{+j}}, \frac{\pi_{2j}}{\pi_{+j}}, \cdots, \frac{\pi_{Ij}}{\pi_{+j}}\right) \tag{58}$$

with $E(Y_{ij}|Y_{+j} = y_{+j}) = y_{+j}\left(\frac{\pi_{ij}}{\pi_{+j}}\right)$.

EM algorithm to determine the ML estimates of the cell probabilities in an incomplete $I \times J$ contingency table: data missing on both categories

If missing values occur on both C_1 and C_2, the observed data can be partitioned into three parts denoted by A, B and C respectively, where A includes units having both C_1 and C_2 observed, B includes those having only C_1 observed and C includes those where only C_2 was observed. In part A observations are fully classified and in B and C only partially. The three parts of the sample are displayed in Table 5.1. The objective is to determine the ML estimates of cell probabilities in the $I \times J$ table by using the fully and partially classified data.

TABLE 5.1 (a), (b) and (c): Classification of sample units in an incomplete $I \times J$ contingency table.

Sample part A

(a) Both variables observed

	$C_2 = 1$	$C_2 = 2$	$\cdots$	$C_2 = J$	
$C_1 = 1$	y_{11}^A	y_{12}^A	$\cdots$	y_{1J}^A	y_{1+}^A
$C_1 = 2$	y_{21}^A	y_{22}^A	$\cdots$	y_{2J}^A	y_{2+}^A
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
$C_1 = I$	y_{I1}^A	y_{I2}^A	$\cdots$	y_{IJ}^A	y_{I+}^A
	y_{+1}^A	y_{+2}^A	$\cdots$	y_{+J}^A	

Sample part B

(b) C_2 missing

$C_1 = 1$	y_{1+}^B
$C_1 = 2$	y_{2+}^B
$\vdots$	$\vdots$
$C_1 = I$	y_{I+}^B

Sample part C

(c) C_1 missing

$C_2 = 1$	$C_2 = 2$	$\cdots$	$C_2 = J$
y_{+1}^C	y_{+2}^C	$\cdots$	y_{+J}^C

Assume that the data are MAR and the missing data mechanism ignorable. Let $\mathbf{Y}^{A\prime} = \left(Y_{11}^A, Y_{12}^A, \ldots, Y_{IJ}^A\right)$, $\mathbf{Y}^{B\prime} = \left(Y_{1+}^B, Y_{2+}^B, \ldots, Y_{I+}^B\right)$ and $\mathbf{Y}^{C\prime} = \left(Y_{+1}^C, Y_{+2}^C, \ldots, Y_{+J}^C\right)$ be the random vectors with counts for sample parts A, B and C respectively. Since C_2 is missing in sample part B, the counts observed are totals across C_2. Hence, compared to sample part A, row totals are observed in sample part B and column totals in sample part C. The observed data are

$$\left\{Y_{ij}^A, Y_{i+}^B, Y_{+j}^C : i = 1, 2, \ldots, I; \; j = 1, 2, \ldots, J\right\}.$$

Let $\mathbf{Y}'_{obs} = \left(\mathbf{Y}^{A\prime}, \mathbf{Y}^{B\prime}, \mathbf{Y}^{C\prime}\right)$ be the observed data vector, $\mathbf{Y}' = (Y_{11}, Y_{12}, \ldots, Y_{IJ})$ the complete data vector and $\boldsymbol{\pi}' = (\pi_{11}, \pi_{12}, \ldots, \pi_{IJ})$ the vector of cell probabilities for which the ML estimates must be determined.

Each complete data count, Y_{ij}, can be expressed as the sum of contributions from each of the three sample parts, that is $Y_{ij} = Y_{ij}^A + Y_{ij}^B + Y_{ij}^C$. For sample part B totals across C_2 are observed, that is Y_{i+}^B, whilst the individual cell counts, Y_{ij}^B, are missing. It follows from (57) that the predictive distribution of the missing data in part B given $\mathbf{Y}_{obs}$ and $\boldsymbol{\pi}$ is a product multinomial,

$$Y_{i1}^B, Y_{i2}^B, \cdots, Y_{iJ}^B | Y_{i+}^B, \boldsymbol{\pi} \sim Mult\left(y_{i+}^B; \frac{\pi_{i1}}{\pi_{i+}}, \frac{\pi_{i2}}{\pi_{i+}}, \cdots, \frac{\pi_{iJ}}{\pi_{i+}}\right) \tag{59}$$

with $E\left(Y_{ij}^B | Y_{i+}^B = y_{i+}^B, \boldsymbol{\pi}\right) = y_{i+}^B\left(\frac{\pi_{ij}}{\pi_{i+}}\right)$.

For part C only the totals across C_1 are observed, that is Y_{+j}^C. From (58) the predictive distribution of the missing data in sample part C given $\mathbf{Y}_{obs}$ and $\boldsymbol{\pi}$ is a product multinomial given by

$$Y_{1j}^C, Y_{2j}^C, \cdots, Y_{Ij}^C | Y_{+j}^C, \boldsymbol{\pi} \sim Mult\left(y_{+j}^C; \frac{\pi_{1j}}{\pi_{+j}}, \frac{\pi_{2j}}{\pi_{+j}}, \cdots, \frac{\pi_{Ij}}{\pi_{+j}}\right) \tag{60}$$

with $E\left(Y_{ij}^C | Y_{+j}^C = y_{+j}^C, \boldsymbol{\pi}\right) = y_{+j}^C\left(\frac{\pi_{ij}}{\pi_{+j}}\right)$.

Thus, $E\left(Y_{ij}|\mathbf{Y}_{obs}, \boldsymbol{\pi}\right) = E\left(Y_{ij}^A + Y_{ij}^B + Y_{ij}^C|\mathbf{Y}_{obs}, \boldsymbol{\pi}\right) = y_{ij}^A + y_{i+}^B\left(\frac{\pi_{ij}}{\pi_{i+}}\right) + y_{+j}^C\left(\frac{\pi_{ij}}{\pi_{+j}}\right)$.

The distribution of the complete data belong to the regular exponential family with sufficient statistics the cell counts, Y_{ij}. In the E-step of the EM algorithm $E\left(Y_{ij}|\mathbf{Y}_{obs}, \boldsymbol{\pi}^{(r)}\right)$ is calculated where $\boldsymbol{\pi}^{(r)}$, $r = 0, 1, 2, \ldots$, is the rth estimate of $\widehat{\boldsymbol{\pi}}$. From (59) and (60)

$$
\begin{aligned}
E\left(Y_{ij}|\mathbf{Y}_{obs}, \boldsymbol{\pi}^{(r)}\right) &= E\left(Y_{ij}^A + Y_{ij}^B + Y_{ij}^C|\mathbf{Y}_{obs}, \boldsymbol{\pi}^{(r)}\right) \\
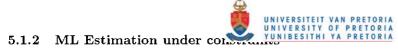&= y_{ij}^A + E\left(Y_{ij}^B|\mathbf{Y}_{obs}, \boldsymbol{\pi}^{(r)}\right) + E\left(Y_{ij}^C|\mathbf{Y}_{obs}, \boldsymbol{\pi}^{(r)}\right) \\
&= y_{ij}^A + y_{i+}^B\left(\frac{\pi_{ij}^{(r)}}{\pi_{i+}^{(r)}}\right) + y_{+j}^C\left(\frac{\pi_{ij}^{(r)}}{\pi_{+j}^{(r)}}\right).
\end{aligned} \tag{61}
$$

In the M-step $\boldsymbol{\pi}^{(r+1)}$ is calculated by substituting the results from the E-step into the expression of the MLE of $\boldsymbol{\pi}$ for the complete data. That is,

$$
\begin{aligned}
\pi_{ij}^{(r+1)} &= \frac{1}{n} E\left(Y_{ij}|\mathbf{Y}_{obs}, \boldsymbol{\pi}^{(r)}\right) \\
&= \frac{1}{n}\left[y_{ij}^A + y_{i+}^B\left(\frac{\pi_{ij}^{(r)}}{\pi_{i+}^{(r)}}\right) + y_{+j}^C\left(\frac{\pi_{ij}^{(r)}}{\pi_{+j}^{(r)}}\right)\right]
\end{aligned} \tag{62}
$$

The process iterates between (61) and (62) until convergence is attained.

5.1.2   ML Estimation under constraints

The data from parts A, B and C in Table 5.1 can be considered as three independent multinomial samples. Let $n^A = \sum\sum y_{ij}^A$, $n^B = \sum y_{i+}^B$ and $n^C = \sum y_{+j}^C$ be the total counts in sample parts A, B and C respectively.

Let $\mathbf{p}^A = \dfrac{1}{n^A}\, \mathbf{y}^A$, $\mathbf{p}^B = \dfrac{1}{n^B}\, \mathbf{y}^B$ and $\mathbf{p}^C = \dfrac{1}{n^C}\, \mathbf{y}^C$ be the proportions in each sample part and $\mathbf{p}'_{obs} = \left(\mathbf{p}^{A\prime}, \mathbf{p}^{B\prime}, \mathbf{p}^{C\prime}\right)$ with $E\left(\mathbf{p}'_{obs}\right) = \boldsymbol{\pi}'_{obs} = \left(\boldsymbol{\pi}^{A\prime}, \boldsymbol{\pi}^{B\prime}, \boldsymbol{\pi}^{C\prime}\right)$.

For the saturated model the maximum likelihood estimates of $\boldsymbol{\pi}_{obs}$ can be determined under the constraints

$$\pi_{i+}^A - \pi_{i+}^B = 0 \quad \text{for } i = 1, 2, \ldots, I \tag{63}$$

and

$$\pi_{+j}^A - \pi_{+j}^C = 0 \quad \text{for } j = 1, 2, \ldots, J. \tag{64}$$

Hence, the constraint can be written as $\mathbf{A}\boldsymbol{\pi}_{obs} = \mathbf{0}$ where

$$\mathbf{A} : (I + J) \times (IJ + I + J) = \left(\begin{array}{cc} \mathbf{I}_I \otimes \mathbf{1}'_J & \\ & -\mathbf{I}_{I+J} \\ \mathbf{1}'_I \otimes \mathbf{I}_J & \end{array} \right)$$

and where $\mathbf{1}'_J$ and $\mathbf{1}'_I$ indicates $1 \times J$ and $1 \times I$ row vectors respectively with all values equal to 1.

The ML estimate of the vector of cell probabilities, under the constraint $\mathbf{A}\boldsymbol{\pi}_{obs} = \mathbf{0}$ is given by

$$\widehat{\boldsymbol{\pi}}_{obs,c} = \left(\widehat{\boldsymbol{\pi}}_c^{A\prime}, \widehat{\boldsymbol{\pi}}_c^{B\prime}, \widehat{\boldsymbol{\pi}}_c^{C\prime}\right)' = \mathbf{p}_{obs} - \left(\mathbf{A}\mathbf{V}_{\boldsymbol{\pi}_{obs}}\right)' \left(\mathbf{A}\mathbf{V}_{\boldsymbol{\pi}_{obs}}\mathbf{A}'\right)^{-1} \mathbf{A}\mathbf{p}_{obs} \tag{65}$$

where

$$\begin{aligned} \mathbf{V}_{\boldsymbol{\pi}_{obs}} &= \left(\begin{array}{ccc} \text{Cov}\left(\boldsymbol{\pi}^A\right) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \text{Cov}\left(\boldsymbol{\pi}^B\right) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \text{Cov}\left(\boldsymbol{\pi}^C\right) \end{array} \right) \\ &= \left(\begin{array}{ccc} \frac{1}{n^A}\left(D_{\boldsymbol{\pi}^A} - \boldsymbol{\pi}^A\boldsymbol{\pi}^{A\prime}\right) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{1}{n^B}\left(D_{\boldsymbol{\pi}^B} - \boldsymbol{\pi}^B\boldsymbol{\pi}^{B\prime}\right) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{1}{n^C}\left(D_{\boldsymbol{\pi}^C} - \boldsymbol{\pi}^C\boldsymbol{\pi}^{C\prime}\right) \end{array} \right). \end{aligned} \tag{66}$$

Since the constraint, $\mathbf{A}\boldsymbol{\pi}_{obs} = \mathbf{0}$, is linear in $\boldsymbol{\pi}_{obs}$ iteration is only over $\boldsymbol{\pi}_{obs}$.

The ML estimates of cell probabilities in the $I \times J$ table are given by the elements of $\widehat{\boldsymbol{\pi}}^A$ in $\widehat{\boldsymbol{\pi}}_{obs,c}$.

EXAMPLE 5.1

Determining the ML estimates of cell probabilities in an incomplete contingency table by using the EM algorithm.

Consider the data in Table 5.2 from Schafer (1997) obtained through the National Crime Survey conducted by the U.S. Bureau of the Census. Housing unit occupants were interviewed to determine whether they had been victimized by crime in the preceding six-month period. Six months later the units were visited again to determine whether the occupants had been victimized in the intervening months.

TABLE 5.2: Victimization status from the National Crime Survey.

First Visit	Second Visit		
	Crime-free	Victims	Missing
Crime-free	392	55	33
Victims	76	38	9
Missing	31	7	

Following the notation in 5.1.1,
$\mathbf{y}'_{obs} = \left(\mathbf{y}^{A\prime}, \mathbf{y}^{B\prime}, \mathbf{y}^{C\prime}\right)$ where
$\mathbf{y}^{A\prime} = \{y_{ij} : i, j = 1, 2\} = (392, 55, 76, 38)$
$\mathbf{y}^{B\prime} = \{y^B_{i+} : i = 1, 2\} = (33, 9)$
$\mathbf{y}^{C\prime} = \{y^C_{+j} : j = 1, 2\} = (31, 7)$.
The fully classified data, $\mathbf{y}^A$, were used to determine a starting value for the algorithm,
$\boldsymbol{\pi}^{(0)\prime} = \frac{1}{561}(392, 55, 76, 38) \approx (0.70, 0.10, 0.13, 0.07)$. From (62) the first estimate of $\hat{\pi}_{11}$ is

$$
\begin{aligned}
\pi^{(1)}_{11} &= \frac{1}{n}\left[y^A_{11} + y^B_{1+}\left(\frac{\pi^{(0)}_{11}}{\pi^{(0)}_{1+}}\right) + y^C_{+j}\left(\frac{\pi^{(0)}_{11}}{\pi^{(0)}_{+1}}\right)\right] \\
&= \frac{1}{641}\left[392 + 33\left(\frac{0.70}{0.80}\right) + 31\left(\frac{0.70}{0.83}\right)\right] \\
&= 0.6974.
\end{aligned}
$$

Similarly, the first estimates of $\hat{\pi}_{12}$, $\hat{\pi}_{21}$ and $\hat{\pi}_{22}$ are

$$
\pi^{(1)}_{12} = \frac{1}{641}\left[55 + 33\left(\frac{0.10}{0.80}\right) + 7\left(\frac{0.10}{0.17}\right)\right] = 0.0987
$$

$$
\pi^{(1)}_{21} = \frac{1}{641}\left[76 + 9\left(\frac{0.13}{0.20}\right) + 31\left(\frac{0.13}{0.83}\right)\right] = 0.1353
$$

$$
\pi^{(1)}_{22} = \frac{1}{641}\left[38 + 9\left(\frac{0.07}{0.20}\right) + 7\left(\frac{0.07}{0.17}\right)\right] = 0.0687.
$$

This gives $\boldsymbol{\pi}^{(1)\prime} = (0.6974, 0.0987, 0.1353, 0.0687)$ which is used to calculate the second estimate for $\hat{\boldsymbol{\pi}}$. The process continues until convergence is attained. Table 5.3 shows the values at different steps of the algorithm.

TABLE 5.3: Iterations of the EM algorithm.

r	$\pi^{(r)}_{11}$	$\pi^{(r)}_{12}$	$\pi^{(r)}_{21}$	$\pi^{(r)}_{22}$
0	0.7000	0.1000	0.1300	0.0700
1	0.6974	0.0987	0.1353	0.0687
2	0.6972	0.0986	0.1357	0.0685
3	0.6971	0.0986	0.1358	0.0685
∞	0.6971	0.0986	0.1358	0.0685

EXAMPLE 5.2

Determining the ML estimates of cell probabilities in an incomplete contingency table under constraints.

Consider the data in Example 5.1.

$\mathbf{y}'_{obs} = \left(\mathbf{y}^{A\prime}, \mathbf{y}^{B\prime}, \mathbf{y}^{C\prime}\right)$ where

$\mathbf{y}^{A\prime} = \{y_{ij} : i, j = 1, 2\} = (392, 55, 76, 38)$ and $\mathbf{p}^A = \dfrac{1}{561}\,\mathbf{y}^A$

$\mathbf{y}^{B\prime} = \{y_{i+}^{B} : i = 1, 2\} = (33, 9)$ and $\mathbf{p}^B = \dfrac{1}{42}\,\mathbf{y}^B$

$\mathbf{y}^{C\prime} = \{y_{+j}^{C} : i = 1, 2\} = (31, 7)$ and $\mathbf{p}^C = \dfrac{1}{38}\,\mathbf{y}^C$.

Let $\mathbf{p}'_{obs} = \left(\mathbf{p}^{A\prime}, \mathbf{p}^{B\prime}, \mathbf{p}^{C\prime}\right)$ with $E\left(\mathbf{p}_{obs}\right) = \boldsymbol{\pi}_{obs}$.

For the saturated model the constraint $\mathbf{A}\boldsymbol{\pi}_{obs} = \mathbf{0}$ must hold, where the elements of $\mathbf{A}$ are

$$
\begin{array}{ccccccccc}
 & \pi_{11}^A & \pi_{12}^A & \pi_{21}^A & \pi_{22}^A & \pi_{1+}^B & \pi_{2+}^B & \pi_{+1}^C & \pi_{+2}^C \\
 & 1 & 1 & 0 & 0 & -1 & 0 & 0 & 0 \\
\mathbf{A}: & 0 & 0 & 1 & 1 & 0 & -1 & 0 & 0 \\
 & 1 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\
 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & -1
\end{array}
$$

The ML estimate of $\boldsymbol{\pi}_{obs}$ under the constraint $\mathbf{A}\boldsymbol{\pi}_{obs} = \mathbf{0}$ is obtained with

$$
\widehat{\boldsymbol{\pi}}_{obs,c} = \left(\widehat{\boldsymbol{\pi}}_c^{A\prime}, \widehat{\boldsymbol{\pi}}_c^{B\prime}, \widehat{\boldsymbol{\pi}}_c^{C\prime}\right)' = \mathbf{p}_{obs} - \left(\mathbf{A}\mathbf{V}_{\pi_{obs}}\right)' \left(\mathbf{A}\mathbf{V}_{\pi_{obs}}\mathbf{A}'\right)^{-1}\mathbf{A}\mathbf{p}_{obs} \tag{67}
$$

where

$$
\mathbf{V}_{\pi_{obs}} = \begin{pmatrix}
\frac{1}{561}\left(D_{\pi^A} - \pi^A\pi^{A\prime}\right) & 0 & 0 \\
0 & \frac{1}{42}\left(D_{\pi^B} - \pi^B\pi^{B\prime}\right) & 0 \\
0 & 0 & \frac{1}{38}\left(D_{\pi^C} - \pi^C\pi^{C\prime}\right)
\end{pmatrix}.
$$

The ML estimates of the cell probabilities in the 2×2 table are given by the elements of $\widehat{\boldsymbol{\pi}}_c^A$ in $\widehat{\boldsymbol{\pi}}_{obs,c}$. This procedure gives the same values for the ML estimates as obtained with the EM algorithm in Example 5.1. Results obtained under constraints and from the Genmod procedure in SAS are shown in Table 5.4. The programs are given in the Appendix.

TABLE 5.4: ML estimates and standard errors.

	Estimate	Std Err
π_{11}	0.6971	0.0187
π_{12}	0.0986	0.0124
π_{21}	0.1358	0.0141
π_{22}	0.0685	0.0104

EXAMPLE 5.3
Determining the ML estimates of cell probabilities in an incomplete contingency table under constraints.

Consider the data in Table 5.5 (from Lipsitz, Parzen, Molenberghs (1998)) which contains the data from the Six Cities Study, a study conducted to assess the health effects of air polution. The columns corresponds to the wheezing status (no wheeze, wheeze with cold, wheeze apart from cold) of a child at age 10. The rows represent the smoking status of the child's mother (none, medium, heavy) during that time. For some individuals the maternal smoking variable is missing, while for others the child's wheezing status is missing. The objective is to estimate the probabilities of the joint distribution of maternal smoking and respiratory illness.

TABLE 5.5: Six Cities Data: Maternal Smoking Cross-Classified by Child's Wheeze Status.

Maternal Smoking	Child's wheeze status			
	No Wheeze	Wheeze with Cold	Wheeze apart from cold	Missing
None	287	39	38	279
Moderate	18	6	4	27
Heavy	91	22	23	201
Missing	59	18	26	

Similar as in Example 5.2:
$\mathbf{y}'_{obs} = \left(\mathbf{y}^{A\prime}, \mathbf{y}^{B\prime}, \mathbf{y}^{C\prime}\right) = (287, 39, 38, 18, 6, 4, 91, 22, 23, 279, 27, 201, 59, 18, 26)$. For the constraint $\mathbf{A}\boldsymbol{\pi}_{obs} = \mathbf{0}$ the elements of $\mathbf{A}$ are given by

	π_{11}^A	π_{12}^A	π_{13}^A	π_{21}^A	π_{22}^A	π_{23}^A	π_{31}^A	π_{32}^A	π_{33}^A	π_{1+}^B	π_{2+}^B	π_{3+}^B	π_{+1}^C	π_{+2}^C	π_{+3}^C
	1	1	1	0	0	0	0	0	0	−1	0	0	0	0	0
	0	0	0	1	1	1	0	0	0	0	−1	0	0	0	0
$\mathbf{A}:$	0	0	0	0	0	0	1	1	1	0	0	−1	0	0	0
	1	0	0	1	0	0	1	0	0	0	0	0	−1	0	0
	0	1	0	0	1	0	0	1	0	0	0	0	0	−1	0
	0	0	1	0	0	1	0	0	1	0	0	0	0	0	−1

The ML estimate of $\boldsymbol{\pi}'_{obs} = \left(\boldsymbol{\pi}^{A\prime}, \boldsymbol{\pi}^{B\prime}, \boldsymbol{\pi}^{C\prime}\right)$ is obtained iteratively with

$$\widehat{\boldsymbol{\pi}}_{obs,c} = \mathbf{p}_{obs} - \left(\mathbf{A}\mathbf{V}_{\pi_{obs}}\right)' \left(\mathbf{A}\mathbf{V}_{\pi_{obs}}\mathbf{A}'\right)^{-1} \mathbf{A}\mathbf{p}_{obs}.$$

The ML estimates of the cell probabilities, given in Table 5.6, are the same as those obtained by Lipsitz, Parzen and Molenberghs (1998). Procedures give asymptotically equivalent results. Slight differences in the standard errors are indicated.
Table 5.6 also gives the ML estimates of cell probabilities when using only the 528 fully classified cases.

TABLE 5.6. ML estimates and standard errors.

	Fully Classified Cases $n = 528$		Fully and Partially Classified Cases $n = 528 + 610$	
	Estimate	Std Err	Estimate	Std Err(Genmod)
π_{11}	0.5436	0.0217	0.4747	0.0179 (0.0174)
π_{12}	0.0739	0.0114	0.0701	0.0105 (0.0102)
π_{13}	0.0720	0.0112	0.0742	0.0108 (0.0107)
π_{21}	0.0341	0.0079	0.0327	0.0065 (0.0064)
π_{22}	0.0114	0.0046	0.0120	0.0044 (0.0045)
π_{23}	0.0076	0.0038	0.0087	0.0039 (0.0041)
π_{31}	0.1723	0.0164	0.2060	0.0149 (0.0158)
π_{32}	0.0417	0.0087	0.0558	0.0094 (0.0106)
π_{33}	0.0436	0.0089	0.0658	0.0100 (0.0116)

5.2 LOGLINEAR MODELS FOR INCOMPLETE CONTINGENCY TABLES

In this section the EM algorithm and ML estimation under constraints are discussed as methods to determine the ML estimates of the cell probabilities in the complete table for any loglinear model where both the fully and partially classified cases are used.
It is assumed that the data are MAR and the missing data mechanism is ignorable.

5.2.1 The EM algorithm

The starting values used in the EM algorithm are the ML estimates of cell probabilities obtained by using only the data in the fully classified table. The process then iterates between the E-step and the M-step. In the E-step the counts in the partially classified table are distributed into the full table by using the ML estimates of the cell probabilities obtained in the M-step. In the M-step ML estimates of the cell probabilities for the filled in table are obtained and used in the E-step as the next approximation of the ML estimates of the cell probabilities in the complete table. The ML estimation procedure under constraints for loglinear models (Section 3.1) can be used in the M-step of the algorithm.

5.2.2 ML Estimation under constraints

Consider an $I \times J \times K$ contingency table with C_1, C_2 and C_3 the three categorical variables where $C_1 = \{1, 2, \ldots, I\}$, $C_2 = \{1, 2, \ldots, J\}$ and $C_3 = \{1, 2, \ldots, K\}$. Suppose that for n^A cases, information for C_1, C_2 and C_3 is known and for n^B cases C_1 is missing. The n^A cases are classified in an $I \times J \times K$ table and the n^B cases in a $J \times K$ table. The objective is to determine the ML estimates of the cell probabilities in the $I \times J \times K$ contingency table, for a specific loglinear model, by using both the n^A fully classified cases and the n^B partially classified cases. A specific loglinear model is assumed.
Suppose $I = J = K = 2$. Let $\mathbf{Y}^{A\prime} = \left(Y_{111}^A, Y_{112}^A, Y_{121}^A, Y_{122}^A, Y_{211}^A, Y_{212}^A, Y_{221}^A, Y_{222}^A\right)$ be the $IJK \times 1$ vector of cell counts for the fully classified table with $E\left(\mathbf{Y}^A\right) = \mu^A$ and let $\mathbf{Y}^{B\prime} = \left(Y_{+11}^B, Y_{+12}^B, Y_{+21}^B, Y_{+22}^B\right)$ the $JK \times 1$ vector of cell counts for the partially classified table with $E\left(\mathbf{Y}^B\right) = \mu^B$. Furthermore let $\mathbf{y}_{obs}' = \left(\mathbf{y}^{A\prime}, \mathbf{y}^{B\prime}\right)$, $\mu_{obs}' = \left(\mu^{A\prime}, \mu^{B\prime}\right)$ and $\pi_{obs}' = \left(\frac{1}{n_A}\mu^{A\prime}, \frac{1}{n_B}\mu^{B\prime}\right) = \left(\pi^{A\prime}, \pi^{B\prime}\right)$.
Two sets of constraints are imposed; the first pertains to the specific loglinear model that is fitted and the second is used to constrain the marginal probabilities in the fully and partially classified tables.

Constraint 1
The saturated loglinear model for the fully classified data is

$$\log \mu^A = \mathbf{X}\boldsymbol{\beta} \tag{68}$$

where μ^A is the vector with expected cell frequencies, $\mathbf{X} : IJK \times IJK$ is the design matrix and $\boldsymbol{\beta} : IJK \times 1$ is the parameter vector for the saturated model.
The unsaturated model can be written as

$$\log \mu^A = \mathbf{X}_u \boldsymbol{\beta}_u \tag{69}$$

where $\mathbf{X}_u$ is a submatrix of $\mathbf{X}$ given in (68) and $\boldsymbol{\beta}_u$ is the parameter vector of the model.
Let $\mathbf{P} = \mathbf{I} - \mathbf{X}_u \left(\mathbf{X}_u' \mathbf{X}_u\right)^{-1} \mathbf{X}_u'$. The constraint for the model in (69) is

$$\mathbf{g}_1\left(\mu^A\right) = \mathbf{P} \log \mu^A = \mathbf{P}\mathbf{X}_u\boldsymbol{\beta}_u = \mathbf{0}. \tag{70}$$

Constraint 2
The sum of the expected cell probabilities in the $I \times J \times K$ fully classified table over category C_1, gives the expected marginal cell probabilities,

$$\pi_{+jk}^A = \sum_{i=1}^{I} \pi_{ijk}^A, \text{ for } j = 1, 2, \ldots, J \text{ and } k = 1, 2, \ldots, K.$$

The constraint which must hold between the two classified tables is

$$\pi^A_{+jk} = \pi^B_{+jk}, \text{ for all } j, k.$$

Hence, the second constraint can be written as

$$g_2\left(\boldsymbol{\mu}_{obs}\right) = \left(\begin{array}{cc} \dfrac{1}{n^A}\mathbf{1}'_I \otimes \mathbf{I}_{JK} & -\dfrac{1}{n^B}\mathbf{I}_{JK} \end{array} \right) \boldsymbol{\mu}_{obs} = \mathbf{0}. \tag{71}$$

Combining (70) and (71) gives

$$\mathbf{g}\left(\boldsymbol{\mu}_{obs}\right) = \left(\begin{array}{c} \mathbf{g}_1\left(\boldsymbol{\mu}^A\right) \\ \mathbf{g}_2\left(\boldsymbol{\mu}_{obs}\right) \end{array} \right) = \left(\begin{array}{c} \mathbf{0} \\ \mathbf{0} \end{array} \right)$$

The ML estimate of $\boldsymbol{\mu}_{obs}$ subject to $\mathbf{g}\left(\boldsymbol{\mu}_{obs}\right) = \mathbf{0}$ is determined iteratively with

$$\widehat{\boldsymbol{\mu}}_{obs,c} = \mathbf{y}_{obs} - \left(\mathbf{G}_{\boldsymbol{\mu}_{obs}}\mathbf{V}_{\boldsymbol{\mu}_{obs}}\right)' \left(\mathbf{G}_{\mathbf{y}_{obs}}\mathbf{V}_{\boldsymbol{\mu}_{obs}}\mathbf{G}'_{\boldsymbol{\mu}_{obs}}\right)^{-1} \mathbf{g}\left(\mathbf{y}_{obs}\right) + o\left(\|\mathbf{y}_{obs} - \boldsymbol{\mu}_{obs}\|\right) \tag{72}$$

where

$$\mathbf{G}_{\boldsymbol{\mu}_{obs}} = \frac{\partial \mathbf{g}\left(\boldsymbol{\mu}_{obs}\right)}{\partial \boldsymbol{\mu}_{obs}} = \left(\begin{array}{c} \dfrac{\partial \mathbf{g}_1\left(\boldsymbol{\mu}^A\right)}{\partial \boldsymbol{\mu}_{obs}} \\ \dfrac{\partial \mathbf{g}_2\left(\boldsymbol{\mu}_{obs}\right)}{\partial \boldsymbol{\mu}_{obs}} \end{array} \right) = \left(\begin{array}{cc} \mathbf{P}\mathbf{D}_{\boldsymbol{\mu}^A}^{-1} & \mathbf{0}_{IJK \times JK} \\ \dfrac{1}{n^A}\mathbf{1}'_I \otimes \mathbf{I}_{JK} & -\dfrac{1}{n^B}\mathbf{I}_{JK} \end{array} \right),$$

$$\mathbf{G}_{\mathbf{y}_{obs}} = \frac{\partial \mathbf{g}\left(\boldsymbol{\mu}_{obs}\right)}{\partial \boldsymbol{\mu}_{obs}}|_{\boldsymbol{\mu}_{obs}=\mathbf{y}_{obs}} \text{ and}$$

$$\mathbf{V}_{\boldsymbol{\mu}_{obs}} = \mathbf{D}_{\boldsymbol{\mu}_{obs}} - \frac{1}{n^A + n^B}\boldsymbol{\mu}_{obs}\boldsymbol{\mu}'_{obs}.$$

The ML estimates of the cell probabilities in the $I \times J \times K$ table is

$$\widehat{\boldsymbol{\pi}}^A = \frac{\widehat{\boldsymbol{\mu}}_c^A}{n^A}.$$

EXAMPLE 5.4

Determining the maximum likelihood estimates of cell probabilities in an incomplete contingency table for any loglinear model.

In Table 5.7, from Little and Rubin (1987), the survival of infants are related according to the amount of prenatal care received by the mothers and the clinic they attended. For data in Table 5.7(a) information on survival, prenatal care and clinic attended were recorded but in Table 5.7(b) information of the clinic attended is missing.

TABLE 5.7. A 2^3 contingency table with partially classified observations.

Clinic (C)	Prenatal Care (P)	Survival (S) Died	Survived	
		(a) *Fully Classified Cases*		
A	Less	3	176	
	More	4	293	
B	Less	17	197	
	More	2	23	715 cases
		(b) *Partially Classified Cases (Clinic missing)*		
	Less	10	150	
	More	5	90	255 cases

The ML estimates of cell probabilities for different loglinear models are given in Table 5.8. The cell probabilities are given in the form $100\hat{\pi}_{CPS}$ where

$C = 1$ if Clinic = "A" and $C = 2$ if Clinic = "B";

$P = 1$ if Care = "Less" and $P = 2$ if Care = "More";

$S = 1$ if Survival = "Died" and $S = 2$ if Survival = "Survived".

The saturated model $\{CPS\}$ was fitted to the incomplete data as explained in section 5.1.2 and the models $\{PS, CS, CP\}$, $\{CS, CP\}$ and $\{PS, CS\}$ were fitted by using the EM algorithm and the ML procedure under constraints.

TABLE 5.8: ML estimates of cell probabilities for different loglinear models.

	$\{CPS\}$	$\{PS, CS, CP\}$	$\{PS, CS\}$	$\{CS, CP\}$
$100\hat{\pi}_{111}$	0.4639	0.4350	0.8327	0.4963
$100\hat{\pi}_{112}$	25.4410	25.4680	36.7015	25.4203
$100\hat{\pi}_{121}$	0.7560	0.7913	0.3053	0.7579
$100\hat{\pi}_{122}$	38.8092	38.7845	28.4910	38.8208
$100\hat{\pi}_{211}$	2.6289	2.6578	2.2601	2.6787
$100\hat{\pi}_{212}$	28.4765	28.4495	17.2160	28.4150
$100\hat{\pi}_{221}$	0.3780	0.3427	0.8287	0.2939
$100\hat{\pi}_{222}$	3.0465	3.0712	13.3647	3.1172

Only the $\{CS, CP\}$ loglinear model is discussed in more detail. The programs are given in the Appendix.

The EM algorithm

The observed frequency vector for the 715 fully classified cases is $\mathbf{y}^{A\prime}=(3,176,4,293,17,197,2,23)$ and for the 255 partially classified cases $\mathbf{y}^{B\prime}=(10,150,5,90)$.

The EM algorithm is used to determine $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\pi}}$, the ML estimates of the cell frequencies and cell probabilities in the 2^3 table.

The steps for the EM algorithm are as follows:

Step 1: Starting value for the EM algorithm

The starting value of the EM algorithm are the ML estimates obtained by using only the data in the fully classified table.

From section 3.1.3, $\boldsymbol{\mu}^{(0)}$, the first approximation of $\hat{\boldsymbol{\mu}}$, is determined iteratively with

$$\boldsymbol{\mu}^{(0)} = \mathbf{y}^A - \mathbf{A}_C \left(\mathbf{A}_C' \mathbf{D}_{\mathbf{y}^A}^{-1} \mathbf{A}_C \right)^{-1} \mathbf{g}\left(\mathbf{y}^A\right) + o\left(\left\|\mathbf{y}^A - \boldsymbol{\mu}\right\|\right) \tag{73}$$

and from this, $\boldsymbol{\pi}^{(0)} = \dfrac{\boldsymbol{\mu}^{(0)}}{715}$.

Step 2: E-Step

In the E-step $\boldsymbol{\pi}^{(r)}$, $r = 0, 1, \ldots$ is used to distribute the 255 partially classified counts into the full table. The filled in frequency vector at the rth step of the EM algorithm is

$$\mathbf{y}^{(r)} = \mathbf{y}^A + \frac{\boldsymbol{\pi}^{(r)}}{\left[\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \otimes \mathbf{I}_{JK} \right] \boldsymbol{\pi}^{(r)}} \# \left[\begin{pmatrix} 1 \\ 1 \end{pmatrix} \otimes \mathbf{y}^B \right]$$

where the division and multiplication indicated with "#" in the last term is elementwise.

Step 3: M-Step

In the M-step $\mathbf{y}^{(r)}$ is used to obtain the next approximation of the ML estimate of $\boldsymbol{\mu}$,

$$\boldsymbol{\mu}^{(r+1)} = \mathbf{y}^{(r)} - \mathbf{A}_C \left(\mathbf{A}_C' \mathbf{D}_{\mathbf{y}^{(r)}}^{-1} \mathbf{A}_C \right)^{-1} \mathbf{g}\left(\mathbf{y}^{(r)}\right) + o\left(\left\|\mathbf{y}^{(r)} - \boldsymbol{\mu}\right\|\right).$$

The next approximation of $\hat{\boldsymbol{\pi}}$ is $\boldsymbol{\pi}^{(r+1)} = \dfrac{\boldsymbol{\mu}^{(r+1)}}{970}$, $r = 0, 1, 2, \ldots$.

The EM algorithm iterates between Step 2 and Step 3 until covergence is attained.

Table 5.9 gives values at different steps of the algorithm.

TABLE 5.9: Values at different steps of the EM algorithm for the $\{CS, CP\}$ model.

Cell	$\mathbf{y}^A$	$r=0$ M-Step $100\boldsymbol{\pi}^{(0)}$	$r=0$ E-Step $\mathbf{y}^{(0)}$	$r=1$ M-Step $100\boldsymbol{\pi}^{(1)}$	$r=1$ E-Step $\mathbf{y}^{(1)}$	$r=2$ M-Step $100\boldsymbol{\pi}^{(2)}$	$r=2$ E-Step $\mathbf{y}^{(2)}$	$r=10$ M-Step $100\boldsymbol{\pi}^{(10)}$	$r=10$ E-Step $\mathbf{y}^{(10)}$
111	3	0.3682	4.3400	0.4802	4.5030	0.4919	4.5468	0.4963	4.5632
112	176	24.6668	246.8579	25.4165	246.8436	25.4201	246.8343	25.4203	246.8280
121	4	0.6109	7.4363	0.7338	7.5561	0.7513	7.5894	0.7579	7.6031
122	293	40.9276	376.4384	38.8409	376.3140	38.8230	376.3048	38.8208	376.3106
211	17	2.3794	25.6600	2.7148	25.4970	2.6883	25.4532	2.6787	25.4368
212	197	27.5507	276.1421	28.3988	276.1564	28.4099	276.1657	28.4150	276.1720
221	2	0.2780	3.5637	0.2980	3.4439	0.2953	3.4106	0.2939	3.3969
222	23	3.2185	29.5616	3.1170	29.6860	3.1202	29.6952	3.1171	29.6894

Maximum likelihood estimation under constraints

Let $\mathbf{y}^{A\prime} = (3, 176, 4, 293, 17, 197, 2, 23)$ and $\mathbf{y}^{B\prime} = (10, 150, 5, 90)$ be the observed frequency vectors for the 715 fully and 255 partially classified cases respectively with $E\left(\mathbf{Y}^A\right) = \boldsymbol{\mu}^A$ and $E\left(\mathbf{Y}^B\right) = \boldsymbol{\mu}^B$. Furthermore $\mathbf{y}'_{obs} = \left(\mathbf{y}^{A\prime}, \mathbf{y}^{B\prime}\right)$ and $\boldsymbol{\mu}'_{obs} = \left(\boldsymbol{\mu}^{A\prime}, \boldsymbol{\mu}^{B\prime}\right)$. Assume a multinomial sampling scheme.

From Section 5.2.2 the ML estimate of $\boldsymbol{\mu}'_{obs} = \left(\boldsymbol{\mu}^{A\prime}, \boldsymbol{\mu}^{B\prime}\right)$ subject to $\mathbf{g}\left(\boldsymbol{\mu}_{obs}\right) = \mathbf{0}$ is determined iteratively with

$$\widehat{\boldsymbol{\mu}}_{obs,c} = \mathbf{y}_{obs} - \left(\mathbf{G}_{\boldsymbol{\mu}_{obs}} \mathbf{V}_{\boldsymbol{\mu}_{obs}}\right)' \left(\mathbf{G}_{\mathbf{y}_{obs}} \mathbf{V}_{\boldsymbol{\mu}_{obs}} \mathbf{G}'_{\boldsymbol{\mu}_{obs}}\right)^{-1} \mathbf{g}\left(\mathbf{y}_{obs}\right) + o\left(\|\mathbf{y}_{obs} - \boldsymbol{\mu}_{obs}\|\right) \tag{74}$$

where

$$\mathbf{g}\left(\boldsymbol{\mu}_{obs}\right) = \begin{pmatrix} \mathbf{g}_1\left(\boldsymbol{\mu}^A\right) \\ \mathbf{g}_2\left(\boldsymbol{\mu}_{obs}\right) \end{pmatrix} = \begin{pmatrix} \mathbf{P}\log\boldsymbol{\mu}^A \\ \frac{1}{715}(1,1)\otimes\mathbf{I}_4\boldsymbol{\mu}_{obs} \quad -\frac{1}{255}\mathbf{I}_4\boldsymbol{\mu}_{obs} \end{pmatrix}$$

$$\mathbf{G}_{\boldsymbol{\mu}_{obs}} = \frac{\partial\mathbf{g}\left(\boldsymbol{\mu}_{obs}\right)}{\partial\boldsymbol{\mu}_{obs}} = \begin{pmatrix} \dfrac{\partial\mathbf{g}_1\left(\boldsymbol{\mu}^A\right)}{\partial\boldsymbol{\mu}_{obs}} \\ \dfrac{\partial\mathbf{g}_2\left(\boldsymbol{\mu}_{obs}\right)}{\partial\boldsymbol{\mu}_{obs}} \end{pmatrix} = \begin{pmatrix} \mathbf{PD}_{\boldsymbol{\mu}^A}^{-1} & \mathbf{0}_{8\times 4} \\ \frac{1}{715}\mathbf{1}'_2\otimes\mathbf{I}_4 & -\frac{1}{255}\mathbf{I}_4 \end{pmatrix},$$

$$\mathbf{G}_{\mathbf{y}_{obs}} = \frac{\partial\mathbf{g}\left(\boldsymbol{\mu}_{obs}\right)}{\partial\boldsymbol{\mu}_{obs}}\Big|_{\boldsymbol{\mu}_{obs}=\mathbf{y}_{obs}} \text{ and}$$

$$\mathbf{V}_{\boldsymbol{\mu}_{obs}} = \mathbf{D}_{\boldsymbol{\mu}_{obs}} - \frac{1}{970}\boldsymbol{\mu}_{obs}\boldsymbol{\mu}'_{obs}.$$

The ML estimates of the cell probabilities in the incomplete contingency table are the elements of $\dfrac{\widehat{\boldsymbol{\mu}}_c^A}{n^A}$ and are the same as those obtained with the EM algorithm.

Table 5.10 gives the ML estimates of cell probabilities obtained under constraints when using only the 715 fully classified cases and when using all 970 counts. The standard errors are also given.

TABLE 5.10: ML estimates obtained under constraints for the $\{CS, CP\}$ model.

	$n = 715$		$n = 970$	
	Estimate	Std Error	Estimate	Std Error
π_{111}	0.0037	0.0014	0.0050	0.0014
π_{112}	0.2467	0.0160	0.2542	0.0153
π_{121}	0.0061	0.0023	0.0076	0.0022
π_{122}	0.4093	0.0183	0.3882	0.0159
π_{211}	0.0238	0.0054	0.0268	0.0050
π_{212}	0.2755	0.0166	0.2842	0.0158
π_{221}	0.0028	0.0008	0.0029	0.0008
π_{222}	0.0322	0.0064	0.0312	0.0063

5.3 CONCLUSION

This dissertation has illustrated maximum likelihood estimation procedures for a number of generalized linear models for categorical data. The results obtained with the method under constraints are the same as those obtained with the more generally used Newton-Raphson, Fisher scoring and EM algorithms. The advantage of the method under constraints is that it is computationally less intensive and also more flexible to incorporate different models.

In this chapter the method was further developed to determine maximum likelihood estimates for loglinear models when the contingency table is incomplete and the missing data mechanism is ignorable. This illustrates the elegance with which the method under constraints can be applied.

This opens up new opportunities for the study of maximum likelihood estimation. This includes models for incomplete data when the missing data mechanism is ignorable, such as logistic regression and analysis of variance. Furthermore the same models for incomlete data can be studied when the missing data mechanism is not ignorable.