Communications in Statistics—Theory and Methods, 39: 874–883, 2010
ISSN: 0361-0926 print/1532-415X online
DOI: 10.1080/03610920902807879

Taylor & Francis
Taylor & Francis Group

Maximum Likelihood Estimation of the Log-Binomial Model

MARTIN R. PETERSEN[1] AND JAMES A. DEDDENS[1,2]

[1]National Institute for Occupational Safety and Health, Cincinnati, Ohio, USA
[2]Department of Mathematical Sciences, University of Cincinnati, Cincinnati, Ohio, USA

Maximum likelihood estimation of prevalence ratios using the log-binomial model is problematic when the estimates are on the boundary of the parameter space. When the model is correct, maximum likelihood is often the method of choice. The authors provide a theorem, formulas, and methodology for obtaining maximum likelihood estimators of the log-binomial model and their estimated standard errors when the solution is on the boundary of the parameter space. Examples are given to illustrate the method.

Keywords Log-binomial model; Maximum likelihood; Parameter space.

Mathematics Subject Classification 62F.

1. Introduction

When a dependent variable (Y) has 2 levels (0,1) and it is reasonable to assume that values of this variable are independent of each other for any combination of independent variables (X), then one assumes that the dependent variable has a binomial distribution for each combination of the independent variables. Combining this with a functional form of a relationship between the dependent and independent variables allows one to model the relationship between them. The logistic model, $E[Y] = \frac{e^{X\beta}}{1+e^{X\beta}}$, has been used for many years as the functional form and can be fit using a generalized linear models program with a logit link and a binomial distribution. The maximum likelihood estimator (MLE) of β is easily obtained with most software packages (as long as the estimate is finite). Exponentiation of the MLE of β yields an MLE of the odds ratio (OR). Valid likelihood ratio or Wald tests of hypotheses can usually be obtained, but in cross-sectional studies, the estimated **OR** is generally not the index of interest. Instead, one is usually

Received March 5, 2008; Accepted February 9, 2009

Address correspondence to Martin R. Petersen, National Institute for Occupational Safety and Health, Mail Stop R15, 4676 Columbia Parkway, Cincinnati, OH 5226, USA; E-mail: mrp1@one.net

interested in the prevalence ratio (PR). If the probability that $Y = 1$ is less than 0.10, then the OR is approximately equal to the PR. When this probability is greater than 0.10, one can fit the log-binomial model, $E[Y] = e^{X\beta}$, using a log link and a binomial distribution (McNutt et al., 2003; Skov et al., 1998; Wacholder, 1986). Then exponentiation of the MLE of β yields an MLE of the PR, and again valid likelihood ratio or Wald tests of hypotheses can be performed.

However, there is a problem with estimating the parameters of the log-binomial model. The parameter space is the set of all β such that the probability that $Y = 1$ is in the closed interval [0, 1]. For the logistic model, the parameter space is $-\infty \le \beta \le \infty$, and the software will generally only fail for infinite estimates of β, which occur when there is a zero cell or complete separation in the sample points. For the log-binomial model, however, the parameter space is $-\infty \le X\beta \le 0$. When the MLE is on the upper boundary of this parameter space, the software may not converge because the slope of the line tangent to the likelihood at this point may not be zero. These boundary solutions sometimes occur when many independent variables are in the model, and they are also quite common even in simple models which contain quantitative independent variables (treated as continuous).

In this article, we assume that the data fit the log-binomial model, and that one has decided to obtain MLEs. A theorem is available (but not published) for obtaining exact maximum likelihood estimates when the solution is on the boundary of the parameter space, provided only one combination of values of the independent variables has an estimated probability of one (mentioned in Deddens et al., 2003). We refer to this as a single boundary point even if several observations have these same independent variable values. In this article, we extend this theorem to include any number of boundary points (at which the estimated probabilities equal one). Our theorem relates closed form estimates of some parameters with non-closed form estimates of the rest. We use some examples to illustrate the use of this theorem to obtain exact MLEs, and we discuss when it fails to obtain these estimates.

2. Log-Binomial Parameter Space Boundary Maximum Likelihood Theorem

The theorem involves finding the boundary points (assume s of them), and restricting the search for the maximum likelihood estimates to those solutions which satisfy the condition that the estimated probability is one at these boundary points. By re-writing the likelihood function to reflect this restriction, one can get a unique solution using the usual log-binomial software. That is, the estimate converges. This gives maximum likelihood estimates and their standard errors for all but the first s parameters. Closed form estimates of the first s estimates (including the intercept) are obtained algebraically using the restriction.

Theorem 2.1. *Let $Y_i \sim B(1, p_i)$, and $p_i = e^{\widetilde{X}_i'\beta}$, where $\widetilde{X}_i' = (1\ X_{1i}\ X_{2i}\ \cdots\ X_{ki})$ is a row vector of independent variables for the ith observation, and $\beta = (\beta_0\ \beta_1\ \beta_2\ \ldots\ \beta_k)'$ is a column vector of unknown parameters.*

Let $\hat{\beta} = (\hat{\beta}_0\ \hat{\beta}_1\ \hat{\beta}_2\ \cdots\ \hat{\beta}_k)'$ be the MLE of β, and thus $\hat{p}_i = e^{\widetilde{X}_i'\hat{\beta}}$ is the MLE of p_i.

Let $s \le k$ be the number of distinct $\widetilde{X}_i'$ vectors for which $\hat{p}_i = 1$, $0 < s \le k$. (If $s = 0$, then the solution is not on the parameter space boundary.)

Let $t_r \in \{t_1, t_2, \ldots, t_s\}$ index these s values of $\widetilde{X}_i'$.

Let $Z_{ji1} = X_{ji} - X_{jt_1}$ and $Z_{jir} = Z_{ji(r-1)} - \left[\frac{Z_{jt_r(r-1)}}{Z_{(r-1)t_r(r-1)}}\right] Z_{(r-1)i(r-1)}$, $r = 2, 3, \ldots, s$.

Then:

(1) the maximum likelihood solution for β_s, $\beta_{s+1}, \ldots, \beta_k$ and their estimated variances and covariances can be obtained by removing all observations for which $\hat{p}_i = 1$ (redundant because their information is contained in the Z_{ijr} variables) and fitting the model $p_i = e^{\sum_{j=s}^{k} Z_{jis}\beta_j}$ by usual maximum likelihood methods. In addition,

(2) $$\hat{\beta}_0 = -\sum_{j=1}^{k} X_{jt_1}\hat{\beta}_j,$$

(3) $$\hat{\beta}_1 = -\frac{\sum_{j=2}^{k} Z_{jt_2 1}\hat{\beta}_j}{Z_{1t_2 1}}, \quad \hat{\beta}_2 = -\frac{\sum_{j=3}^{k} Z_{jt_3 2}\hat{\beta}_j}{Z_{2t_3 2}}, \ldots,$$
$$\hat{\beta}_r = -\frac{\sum_{j=r+1}^{k} Z_{jt_{(r+1)}r}\hat{\beta}_j}{Z_{rt_{(r+1)}r}}, \ldots, \quad \hat{\beta}_{s-1} = -\frac{\sum_{j=s}^{k} Z_{jt_s(s-1)}\hat{\beta}_j}{Z_{(s-1)t_s(s-1)}},$$

(4) $$\widehat{se}(\hat{\beta}_0) = \sqrt{\sum_{j=1}^{k}[\widehat{\mathrm{var}}(\hat{\beta}_j)]X_{jt_1}^2 + \sum_{j_1=1}^{k}\sum_{\substack{j_2=1\\ j_1\neq j_2}}^{k}[\widehat{\mathrm{Cov}}(\hat{\beta}_{j_1}, \hat{\beta}_{j_2})]X_{j_1t_1}X_{j_2t_1}}, \quad \text{and}$$

(5) $$\widehat{se}(\hat{\beta}_r) = \sqrt{\sum_{j=r+1}^{k}[\widehat{\mathrm{var}}(\hat{\beta}_j)]\left[\frac{Z_{jt_{(r+1)}r}}{Z_{rt_{(r+1)}r}}\right]^2 + \sum_{j_1=r+1}^{k}\sum_{\substack{j_2=r+1\\ j_1\neq j_2}}^{k}[\widehat{\mathrm{Cov}}(\hat{\beta}_{j_1}, \hat{\beta}_{j_2})]\left[\frac{Z_{j_1t_{(r+1)}r}Z_{j_2t_{(r+1)}r}}{(Z_{rt_{(r+1)}r})^2}\right]}$$
$$r = 1, 2, \ldots, s-1.$$

The proof is given in the Appendix.

Corollary 2.1 covers the one boundary point situation, which is much simpler and is usually all that is needed in practice.

Corollary 2.1. *Let $Y_i \sim B(1, p_i)$, and $p_i = e^{\widetilde{X}_i'\beta}$, where $\widetilde{X}_i' = (1\ X_{1i}\ X_{2i}\ \cdots\ X_{ki})$ is a row vector of independent variables for the ith observation, and $\beta = (\beta_0\ \beta_1\ \beta_2\ \ldots\ \beta_k)'$ is a column vector of unknown parameters. Let $\hat{\beta} = (\hat{\beta}_0\ \hat{\beta}_1\ \hat{\beta}_2\ \cdots\ \hat{\beta}_k)'$ be the MLE of β, and thus $\hat{p}_i = e^{\widetilde{X}_i'\hat{\beta}}$ is the MLE of p_i.*

Let $(1\ X_{1t_1}\ X_{2t_1}\ \cdots\ X_{kt_1})$ be the only vector for which $\hat{p}_i = 1$.

Let $Z_{ji} = X_{ji} - X_{jt_1}$, $j = 1, 2, \ldots, k$. Then

(1) *the maximum likelihood solution for β_1, $\beta_2, \ldots, \beta_k$ and their estimated variances, covariances, and standard errors can be obtained by removing all (redundant) observations for which $\hat{p}_i = 1$ and fitting the model $p_i = e^{\sum_{j=1}^{k} Z_{ji}\beta_j}$ by usual maximum likelihood methods. In addition,*

(2) $$\hat{\beta}_0 = -\sum_{j=1}^{k} X_{jt_1}\hat{\beta}_j,$$

(3)
$$\widehat{se}(\hat{\beta}_0) = \sqrt{\sum_{j=1}^{k}[\widehat{\mathrm{Var}}(\hat{\beta}_j)]X_{jt_1}^2 + \sum_{\substack{j_1=1 \\ j_1\neq j_2}}^{k}\sum_{j_2=1}^{k}[\widehat{\mathrm{Cov}}(\hat{\beta}_{j_1}, \hat{\beta}_{j_2})]X_{j_1t_1}X_{j_2t_1}}.$$

Proof. The result follows immediately from the theorem by setting $s = 1$ and $Z_{ji1} = Z_{ji}$.

In the simple case where $s = k = 1$, (2) and (3) reduce to $\hat{\beta}_0 = -X_{1t_1}\hat{\beta}_1$ and $\widehat{se}(\hat{\beta}_0) = \sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_1)X_{1t_1}^2}$, respectively. These were employed by Deddens et al. (2003), and a macro is available for SAS at the NIOSH web page (http://www.cdc.gov/niosh/ext-supp-mat/pr-sasmac).

3. Using the Theorem

3.1. *Example* 1

The first example comes from Deddens et al. (2003). They used the one boundary point ($s = 1$) version of this theorem, and supplied a SAS macro for the solution for one independent variable ($k = 1$). We give more details here. The data are given in Table 1. Y is the dependent variable, and X_1 is the independent variable. If we attempt to run the model $\ln[E(Y_i)] = \beta_0 + \beta_1 X_{1i}$ in SAS using PROC GENMOD, the model does not converge. Predicted values can be obtained from the model at the point where the program stopped. These range from .4749482 when $X_1 = 0$ to .9999994 when $X_1 = 10$, indicating that the boundary point occurs when $X_1 = 10$. Thus, $Z_{1i1} = X_{1i} - X_{1t_1} = X_{1i} - 10$. Removing the (redundant) data where $X_1 = 10$ and fitting the model $\ln[E(Y_i)] = \beta_1 Z_{1i1}$ yields $\hat{\beta}_1 = 0.2094$ and $\widehat{SE}(\hat{\beta}_1) = 0.1021$. Then $\hat{\beta}_0 = -10\hat{\beta}_1 = -2.0940$ and

$$\widehat{SE}(\hat{\beta}_0) = \sqrt{[\widehat{\mathrm{var}}(\hat{\beta}_1)]X_{1t_1}^2} = \sqrt{[.1021^2](10)^2} = 1.0210.$$

3.2. *Example* 2

This is an illustrative example of a quadratic relation ($k = 2$) which has 2 boundary points ($s = 2$). The data are given in Table 2. Y is the dependent variable, X_1 is the independent variable, and there are 10 values of Y at each value of X_1. Let $X_2 = X_1^2$. If we attempt to run the model $\ln[E(Y_i)] = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}$ in SAS using PROC GENMOD, the model does not converge. Predicted values are .999999 at both $(X_1\ X_2) = (1\ 1)$ and $(X_1\ X_2) = (11\ 121)$, indicating that these are the boundary points. Let $X_{1t_1} = 1, X_{2t_1} = 1, X_{1t_2} = 11$, and $X_{2t_2} = 121$. Then $Z_{1i1} = X_{1i} - X_{1t_1} =$

Table 1
Example 1 data for 1 boundary point

X_1	1	2	3	4	5	6	7	8	9	10
Y	0	0	0	0	1	0	1	1	1	1

Table 2
Example 2 data for 2 boundary points*

X_1	1	2	3	4	5	6	7	8	9	10	11
$\sum Y$	10	6	4	3	3	2	3	3	4	6	10

*There are 10 values of Y for each X_1.

$X_{1i} - 1$, $Z_{2i1} = X_{2i} - X_{2t_1} = X_{2i} - 1$,

$$Z_{1i2} = Z_{1i1} - \left(\frac{Z_{1t_21}}{Z_{1t_21}}\right)Z_{1i1} = Z_{1i1} - \left(\frac{11-1}{11-1}\right)Z_{1i1} = 0,$$

and

$$Z_{2i2} = Z_{2i1} - \left(\frac{Z_{2t_21}}{Z_{1t_21}}\right)Z_{1i1} = Z_{2i1} - \left(\frac{121-1}{11-1}\right)Z_{1i1} = Z_{2i1} - (12)Z_{1i1}.$$

Fitting the model $\ln[E(Y_i)] = \beta_2 Z_{2i2}$ on the data where $2 \le X_1 \le 10$ yields $\hat{\beta}_2 = 0.0561$ and $\widehat{SE}(\hat{\beta}_2) = 0.0079$. Then

$$\hat{\beta}_1 = -\frac{\sum_{j=2}^{2} Z_{jt_21}\hat{\beta}_j}{Z_{1t_21}} = -\frac{Z_{2t_21}\hat{\beta}_2}{Z_{1t_21}} = -\frac{(121-1)(0.0561)}{11-1} = -0.6732,$$

$$\begin{aligned}\hat{\beta}_0 &= -\sum_{j=1}^{2} X_{jt_1}\hat{\beta}_j = -(X_{1t_1}\hat{\beta}_1 + X_{2t_1}\hat{\beta}_2)\\ &= -[(1)(-0.6732) + (1)(0.0561)] = 0.6171,\end{aligned}$$

$$\begin{aligned}\widehat{SE}(\hat{\beta}_1) &= \sqrt{\sum_{j=2}^{2}\widehat{\mathrm{Var}}(\hat{\beta}_j)\left(\frac{Z_{jt_21}}{Z_{1t_21}}\right)^2} = \sqrt{(\widehat{\mathrm{Var}}\hat{\beta}_2)\left(\frac{Z_{2t_21}}{Z_{1t_21}}\right)^2} = \sqrt{(0.0079)^2\left(\frac{X_{2t_2}-1}{X_{1t_2}-1}\right)^2}\\ &= (0.0079)\left(\frac{121-1}{11-1}\right) = 0.0948,\end{aligned}$$

and

$$\begin{aligned}\widehat{SE}(\hat{\beta}_0) &= \sqrt{\sum_{j=1}^{2}[\widehat{\mathrm{Var}}(\hat{\beta}_j)](X_{jt_1})^2 + \sum_{\substack{j_1=1\\ j_1\neq j_2}}^{2}\sum_{j_2=1}^{2}[\widehat{\mathrm{Cov}}(\hat{\beta}_{j_1}, \hat{\beta}_{j_2})](X_{j_1t_1})(X_{j_2t_1})}\\ &= \sqrt{[\widehat{\mathrm{Var}}(\hat{\beta}_1)](X_{1t_1})^2 + [\widehat{\mathrm{Var}}(\hat{\beta}_2)](X_{2t_1})^2 + 2[\widehat{\mathrm{Cov}}(\hat{\beta}_1, \hat{\beta}_2)]X_{1t_1}X_{2t_1}}\\ &= \sqrt{(0.0948)^2(1)^2 + (0.0079)^2(1)^2 + 2[\widehat{\mathrm{Cov}}(\hat{\beta}_1, \hat{\beta}_2)](1)(1)}\\ &= \sqrt{0.00904945 + 2\widehat{\mathrm{Cov}}(\hat{\beta}_1, \hat{\beta}_2)}.\end{aligned}$$

One way to obtain this covariance is to repeat the above calculations reversing the boundary points. Thus, let $X_{1t_1} = 11$, $X_{2t_1} = 121$, $X_{1t_2} = 1$, and $X_{2t_2} = 1$. Then

$Z_{1i1} = X_{1i} - X_{1t_1} = X_{1i} - 11$, $Z_{2i1} = X_{2i} - X_{2t_1} = X_{2i} - 121$,

$$Z_{1i2} = Z_{1i1} - \left(\frac{Z_{1t_2 1}}{Z_{1t_2 1}}\right) Z_{1i1} = Z_{1i1} - \left(\frac{1-11}{1-11}\right) Z_{1i1} = 0,$$

and

$$Z_{2i2} = Z_{2i1} - \left(\frac{Z_{2t_2 1}}{Z_{1t_2 1}}\right) Z_{1i1} = Z_{2i1} - \left(\frac{1-121}{1-11}\right) Z_{1i1} = Z_{2i1} - (12) Z_{1i1}.$$

Fitting the model $\ln[E(Y_i)] = \beta_2 Z_{2i2}$ on the data where $2 \leq X_1 \leq 10$ again yields $\hat{\beta}_2 = 0.0561$ and $\widehat{SE}(\hat{\beta}_2) = 0.0079$. Again,

$$\hat{\beta}_1 = -\frac{\sum_{j=2}^{2} Z_{jt_2 1} \hat{\beta}_j}{Z_{1t_2 1}} = -\frac{Z_{2t_2 1} \hat{\beta}_2}{Z_{1t_2 1}} = -\frac{(1-121)(0.0561)}{1-11} = -0.6732,$$

$$\hat{\beta}_0 = -\sum_{j=1}^{2} X_{jt_1} \hat{\beta}_j = -(X_{1t_1} \hat{\beta}_1 + X_{2t_1} \hat{\beta}_2)$$
$$= -[(11)(-0.6732) + (121)(0.0561)] = 0.6171,$$

$$\widehat{SE}(\hat{\beta}_1) = \sqrt{\sum_{j=2}^{2} \widehat{\text{Var}}(\hat{\beta}_j) \left(\frac{Z_{jt_2 1}}{Z_{1t_2 1}}\right)^2} = \sqrt{\widehat{\text{Var}}(\hat{\beta}_2) \left(\frac{Z_{2t_2 1}}{Z_{1t_2 1}}\right)^2}$$
$$= \sqrt{(0.0079)^2 \left(\frac{X_{2t_2} - 121}{X_{1t_2} - 11}\right)^2} = (0.0079)\left(\frac{1-121}{1-11}\right) = 0.0948,$$

and

$$\widehat{SE}(\hat{\beta}_0) = \sqrt{\sum_{j=1}^{2} [\widehat{\text{Var}}(\hat{\beta}_j)](X_{jt_1})^2 + \sum_{\substack{j_1=1 \\ j_1 \neq j_2}}^{2} \sum_{j_2=1}^{2} [\widehat{\text{Cov}}(\hat{\beta}_{j_1}, \hat{\beta}_{j_2})](X_{j_1 t_1})(X_{j_2 t_1})}$$
$$= \sqrt{[\widehat{\text{Var}}(\hat{\beta}_1)](X_{1t_1})^2 + [\widehat{\text{Var}}(\hat{\beta}_2)](X_{2t_1})^2 + 2[\widehat{\text{Cov}}(\hat{\beta}_1, \hat{\beta}_2)] X_{1t_1} X_{2t_1}}$$
$$= \sqrt{(0.0948)^2 (11)^2 + (0.0079)^2 (121)^2 + 2[\widehat{\text{Cov}}(\hat{\beta}_1, \hat{\beta}_2)](11)(121)}$$
$$= \sqrt{2.00117665 + 2662 \widehat{\text{Cov}}(\hat{\beta}_1, \hat{\beta}_2)}.$$

Thus, $\sqrt{0.00904945 + 2\widehat{\text{Cov}}(\hat{\beta}_1, \hat{\beta}_2)} = \sqrt{2.00117665 + 2662\widehat{\text{Cov}}(\hat{\beta}_1, \hat{\beta}_2)}$. Solving this equation yields $\widehat{\text{Cov}}(\hat{\beta}_1, \hat{\beta}_2) = -0.00074892$. Then $\widehat{SE}(\hat{\beta}_0) = \sqrt{2.00117665 + 2662(-0.00074892)} = 0.0869$.

3.3. *Example* 3

The third example comes from a book by Allison (1999) and is readily available online (SAS, 2006). These data relate Y = death penalty (death = 1, life in prison = 0) to X_1 = defendant race (2 levels), X_2 = victim race (2 levels), and X_3 = culpability (quantitative scale). Upon running the model $\ln[E(Y_i)] = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i}$ in SAS using PROC GENMOD, the model does not converge. The 4 largest predicted values are all equal to 0.9999979 and correspond to $(X_1\ X_2\ X_3) = (1\ 1\ 5)$, indicating that this is the boundary point. Thus, $X_{1t_1} = 1$, $X_{2t_1} = 1$, $X_{3t_1} = 5$. Then $Z_{1i1} = X_{1i} - X_{1t_1} = X_{1i} - 1$, $Z_{2i1} = X_{2i} - X_{2t_1} = X_{2i} - 1$, and $Z_{3i1} = X_{3i} - X_{3t_1} = X_{3i} - 5$.

Removing the (redundant) data where $X_1 = 1$, $X_2 = 1$, and $X_3 = 5$ and fitting the model $\ln[E(Y_i)] = \beta_1 Z_{1i1} + \beta_2 Z_{2i1} + \beta_3 Z_{3i1}$ in SAS using PROC GENMOD yielded $\hat{\beta}_1 = 0.2183$, $\widehat{SE}(\hat{\beta}_1) = 0.1290$, $\hat{\beta}_2 = 0.1128$, $\widehat{SE}(\hat{\beta}_2) = 0.1036$, $\hat{\beta}_3 = 0.4461$, and $\widehat{SE}(\hat{\beta}_3) = 0.0648$. From the estimated covariance matrix, the following were obtained: $\widehat{\mathrm{Cov}}(\hat{\beta}_1, \hat{\beta}_2) = -0.0014$, $\widehat{\mathrm{Cov}}(\hat{\beta}_1, \hat{\beta}_3) = 0.0023$, and $\widehat{\mathrm{Cov}}(\hat{\beta}_2, \hat{\beta}_3) = 0.0011$. Then $\hat{\beta}_0 = -\sum_{j=1}^{3} X_{jt_1}\hat{\beta}_j = -(X_{1t_1}\hat{\beta}_1 + X_{2t_1}\hat{\beta}_2 + X_{3t_1}\hat{\beta}_3) = -[(1)(0.2183) + (1)(0.1128) + (5)(0.4461)] = -2.6616$. Also $\sum_{j=1}^{3} [\widehat{\mathrm{Var}}(\hat{\beta}_j)](X_{jt_1})^2 = (0.1290)^2(1)^2 + (0.1036)^2(1)^2 + (0.0648)^2(5)^2 = 0.1323$ and $\sum_{\substack{j_1=1 \\ j_1 \neq j_2}}^{3} \sum_{j_2=1}^{3} [\widehat{\mathrm{Cov}}(\hat{\beta}_{j_1}, \hat{\beta}_{j_2})](X_{j_1t_1})(X_{j_2t_1}) = 2[(-0.0014)(1)(1) + (-0.0023)(1)(5) + (-0.0011)(1)(5)] = -0.0368$.

Thus, $\widehat{SE}(\hat{\beta}_0) = \sqrt{\sum_{j=1}^{3} [\widehat{\mathrm{Var}}(\hat{\beta}_j)](X_{jt_1})^2 + \sum_{\substack{j_1=1 \\ j_1 \neq j_2}}^{3} \sum_{j_2=1}^{3} [\widehat{\mathrm{Cov}}(\hat{\beta}_{j_1}, \hat{\beta}_{j_2})](X_{j_1t_1})(X_{j_2t_1})}$ $= \sqrt{0.1323 - 0.0368} = 0.3090$.

4. Practical Considerations

As can be seen from the examples given, the exact maximum likelihood is often difficult to obtain by hand in practice. In the simple linear case (e.g., Example 1), a macro exists for SAS users (Deddens et al., 2003). We were able to extend that program to work for any number of independent variables, but only one boundary point (Corollary 2.1). (Note: the intercept and its estimated standard error (−2.0936 and 1.0208, respectively for Example 1; −2.6616 and 0.3100, respectively, for Example 3) from the computer program are slightly different from those in Sec. 3 due to rounding error in the latter.) One danger is that the model on the original data may fail to converge, not because the solution is on the boundary, but because bad start values were chosen. By examining the predicted probabilities from this model, however, one can see if the solution is on the boundary or not. So, as long as software gives the predicted values for the non-convergent model (which SAS does), then this should not be a problem. It is possible, however, that if two or more predicted values were very close to one, then the point corresponding to the largest one might not be the boundary point, because the program failed to converge. This would probably have only a small effect on the solution.

Another way in which the ability to obtain the maximum likelihood solution may depend on the software is when the solution for the Z variables is very close to the boundary. For Example 3 (Allison, 1999; SAS, 2006), there is also a continuous variable for crime seriousness. The single boundary point occurs when defendant race = 1, victim race = 1, culpability = 5, and crime seriousness = 8.4. At this point, the estimated probability of the death penalty is 0.9999992 from the original non convergent model. However for the point defendant race = 1, victim race = 1, culpability = 5, and crime seriousness = 9.3, the estimated probability of the death penalty is 0.9999421. Even though this point is not exactly on the boundary, it is so close that SAS does not converge for the model containing the Z variables. Other statistical packages may do better or worse than SAS.

The other situation in which the exact solution may not be easily obtained is when there is more than one boundary point. Our Example 2 was contrived to make this happen (i.e., the observed relationship was perfectly U-shaped), so it may not happen very often in practice.

5. Discussion

We assumed throughout this article that the correct form of the log-binomial model has been fit to the data. In practice, one should make sure that this is the case. Blizzard and Hosmer (2006) discussed methods of examining goodness of fit in log-binomial models.

We give a theorem showing how to calculate exact maximum likelihood estimates and their standard errors for the log-binomial model. We give three examples which demonstrate the method and show that the method is difficult to implement without software (beyond that which fits the log-binomial model). Until now, the only software available that obtained an exact solution on the parameter space boundary was for a single one degree of freedom independent variable (Deddens et al., 2003). We wrote a computer program in SAS for the case where there are any number of independent variables and no more than one boundary point. It is available at the NIOSH web page: http://www.cdc.gov/niosh/ext-supp-mat/pr-sasmac

Appendix

Proof of Theorem 2.1

For this log-binomial model, the likelihood function at its maximum is $L = \prod_{\{Y_i=1\}} e^{\widetilde{X}_i'\hat{\beta}} \prod_{\{Y=0\}} (1 - e^{\widetilde{X}_i'\hat{\beta}}) = \prod_{\{Y_i=1\}} e^{\hat{\beta}_0+\sum_{j=1}^k X_{ji}\hat{\beta}_j} \prod_{\{Y=0\}} (1 - e^{\hat{\beta}_0+\sum_{j=1}^k X_{ji}\hat{\beta}_j}) = \left[\prod_{\substack{\{Y_i=1\}\\ \{X_{ji}\neq X_{jt_1}\}}} e^{\hat{\beta}_0+\sum_{j=1}^k X_{ji}\hat{\beta}_j} \prod_{\{Y=0\}} (1 - e^{\hat{\beta}_0+\sum_{j=1}^k X_{ji}\hat{\beta}_j})\right][e^{\hat{\beta}_0+\sum_{j=1}^k X_{jt_1}\hat{\beta}_j}]^{\delta_1}$, where δ_1 is the number of observations with $X_{ji} = X_{jt_1}$.

Since $e^{\hat{\beta}_0+\sum_{j=1}^k X_{jt_1}\hat{\beta}_j} = 1$, $\hat{\beta}_0 = -\sum_{j=1}^k X_{jt_1}\hat{\beta}_j$, which proves (2). Then

$$
\begin{aligned}
L &= \prod_{\substack{\{Y_i=1\}\\ \{X_{ji}\neq X_{jt_1}\}}} e^{\sum_{j=1}^k (X_{ji}-X_{jt_1})\hat{\beta}_j} \prod_{\{Y=0\}} \left(1 - e^{\sum_{j=1}^k (X_{ji}-X_{jt_1})\hat{\beta}_j}\right) \\
&= \prod_{\substack{\{Y_i=1\}\\ \{X_{ji}\neq X_{jt_1}\}}} e^{\sum_{j=1}^k Z_{ji1}\hat{\beta}_j} \prod_{\{Y=0\}} \left(1 - e^{\sum_{j=1}^k Z_{ji1}\hat{\beta}_j}\right) \\
&= \left[\prod_{\substack{\{Y_i=1\}\\ \{X_{ji}\neq X_{jt_1}\}\\ \{X_{ji}\neq X_{jt_2}\}}} e^{\sum_{j=1}^k Z_{ji1}\hat{\beta}_j} \prod_{\{Y=0\}} (1 - e^{\sum_{j=1}^k Z_{ji1}\hat{\beta}_j})\right] \left[e^{\sum_{j=1}^k Z_{jt_2 1}\hat{\beta}_j}\right]^{\delta_2}, \qquad (*)
\end{aligned}
$$

where δ_2 is the number of observations with $X_{ji} = X_{jt_2}$ (or equivalently, $Z_{ji1} = Z_{jt_2 1}$).

Since $e^{\sum_{j=1}^{k} Z_{jt_2 1}\hat{\beta}_j} = 1$, $\hat{\beta}_1 = -\frac{\sum_{j=2}^{k} Z_{jt_2 1}\hat{\beta}_j}{Z_{1t_2 1}}$.
Then,

$$
\begin{aligned}
L &= \prod_{\substack{\{Y_i=1\}\\ \{X_{ji}\neq X_{jt_1}\}\\ X_{ji}\neq X_{jt_2}}} e^{Z_{1i1}\hat{\beta}_1+\sum_{j=2}^{k} Z_{ji1}\hat{\beta}_j} \prod_{\{Y=0\}} \left(1 - e^{Z_{1i1}\hat{\beta}_1+\sum_{j=2}^{k} Z_{ji1}\hat{\beta}_j}\right) \\
&= \prod_{\substack{\{Y_i=1\}\\ \{X_{ji}\neq X_{jt_1}\}\\ \{X_{ji}\neq X_{jt_2}\}}} e^{\sum_{j=2}^{k}\left[Z_{ji1}-\frac{Z_{jt_2 1}}{Z_{1t_2 1}}Z_{1i1}\right]\hat{\beta}_j} \prod_{\{Y=0\}} \left(1 - e^{\sum_{j=2}^{k}\left[Z_{ji1}-\frac{Z_{jt_2 1}}{Z_{1t_2 1}}Z_{1i1}\right]\hat{\beta}_j}\right) \\
&= \prod_{\substack{\{Y_i=1\}\\ \{X_{ji}\neq X_{jt_1}\}\\ \{X_{ji}\neq X_{jt_2}\}}} e^{\sum_{j=2}^{k} Z_{ji2}\hat{\beta}_j} \prod_{\{Y=0\}} \left(1 - e^{\sum_{j=2}^{k} Z_{ji2}\hat{\beta}_j}\right). \qquad (**)
\end{aligned}
$$

Repeating the analogous steps from (*) to (**) for the remainder of the boundary points yields $\hat{\beta}_r = -\frac{\sum_{j=r+1}^{k} Z_{jt_{(r+1)}r}\hat{\beta}_j}{Z_{rt_{(r+1)}r}}$, $2 \le r \le s-1$, which, together with the derivation of $\hat{\beta}_1$ above, proves (3).

The likelihood will then be $\prod_{\substack{\{Y_i=1\}\\ \{X_{ji}\neq X_{jt_1}\}\\ \{X_{ji}\neq X_{jt_2}\}\\ \vdots \\ \{X_{ji}\neq X_{jt_s}\}}} e^{\sum_{j=s}^{k} Z_{jis}\beta_j} \prod_{\{Y=0\}} (1 - e^{\sum_{j=s}^{k} Z_{jis}\beta_j})$. Since all boundary points have been eliminated, this can be solved for $\beta_s, \beta_{s+1}, \ldots, \beta_k$ and their estimated variances and covariances using standard methods. This proves (1).

$\widehat{se}(\hat{\beta}_0) = \sqrt{\sum_{j=1}^{k} \widehat{\mathrm{var}}(\hat{\beta}_j)X_{jt_1}^2 + \sum_{j_1=1}^{k}\sum_{\substack{j_2=1\\ j_1\neq j_2}}^{k}[\widehat{\mathrm{Cov}}(\hat{\beta}_{j_1}, \hat{\beta}_{j_2})]X_{j_1 t_1}X_{j_2 t_1}}$ follows trivially from $\hat{\beta}_0 = -\sum_{j=1}^{k} X_{jt_1}\hat{\beta}_j$, which proves (4).

$$
\widehat{se}(\hat{\beta}_r) = \sqrt{\sum_{j=r+1}^{k}[\widehat{\mathrm{var}}(\hat{\beta}_j)]\left[\frac{Z_{jt_{(r+1)}r}}{Z_{rt_{(r+1)}r}}\right]^2 + \sum_{j_1=r+1}^{k}\sum_{\substack{j_2=r+1\\ j_1\neq j_2}}^{k}[\widehat{\mathrm{Cov}}(\hat{\beta}_{j_1}, \hat{\beta}_{j_2})]\left[\frac{Z_{j_1 t_{(r+1)}r}Z_{j_2 t_{(r+1)}r}}{(Z_{rt_{(r+1)}r})^2}\right]}
$$

$r = 1, 2, \ldots, s-1$ follows trivially from $\hat{\beta}_r = -\frac{\sum_{j=r+1}^{k}\hat{\beta}_j Z_{jt_{(r+1)}r}}{Z_{rt_{(r+1)}r}}$, $1 \le r \le s-1$, which proves (5).

Disclaimer

The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the National Institute for Occupational Safety and Health.

References

Allison, P. D. (1999). *Logistic Regression using SAS: Theory and Application.* Cary, NC: SAS Institute Inc., Ch. 2.

Blizzard, L., Hosmer, D. (2006). Parameter estimation and goodness-of-fit in log binomial regression. *Biometr. J.* 48:5–22.

Deddens, J. A., Petersen, M. R., Lei, X. (2003). Estimation of prevalence ratios when PROC GENMOD does not converge. *Proc. 28th Ann. SAS Users Group Int. Conf.* March 30 – April 2; http://www2.sas.com/proceedings/sugi28/270-28.pdf

McNutt, L. A., Wu, C., Xue, X., Hafner, J. P. (2003). Estimating the relative risk in cohort studies and clinical trials of common outcomes. *Amer. J. Epidemiol.* 157:940–943.

SAS Institute Inc. (2006). Logistic regression using the SAS system: Theory and application. SAS Publishing, Cary, NC: SAS Institute Inc; http://ftp.sas.com/ samples/A55770

Skov, T., Deddens, J., Petersen, M. R., Endahl, L. (1998). Prevalence proportion ratios: estimation and hypothesis testing. *Int. J. Epidemiol.* 27:91–95.

Wacholder, S. (1986). Binomial regression in GLIM, estimating risk ratios and risk differences. *Amer. J. Epidemiol.* 123:174–184.