

## Convergence Theorems in Multinomial Saturated and Logistic Models

### Teoremas de convergencias en los modelos saturados y logísticos multinomiales

ERICK OROZCO-ACOSTA<sup>1,a</sup>, HUMBERTO LLINÁS-SOLANO<sup>2,b</sup>,  
JAVIER FONSECA-RODRÍGUEZ<sup>3,c</sup>

<sup>1</sup>PROGRAMA DE INGENIERÍA INDUSTRIAL, FACULTAD DE INGENIERÍA, UNIVERSIDAD SIMÓN BOLÍVAR, BARRANQUILLA, COLOMBIA

<sup>2</sup>DEPARTAMENTO DE MATEMÁTICAS Y ESTADÍSTICA, DIVISIÓN DE CIENCIAS BÁSICAS, UNIVERSIDAD DE NORTE, BARRANQUILLA, COLOMBIA

<sup>3</sup>DEPARTAMENTO DE CIENCIAS BÁSICAS, CORPORACIÓN POLITÉCNICO DE LA COSTA ATLÁNTICA, BARRANQUILLA, COLOMBIA

---

#### Abstract

In this paper, we develop a theoretical study about the logistic and saturated multinomial models when the response variable takes one of  $R \geq 2$  levels. Several theorems on the existence and calculations of the maximum likelihood (ML) estimates of the parameters of both models are presented and demonstrated. Furthermore, properties are identified and, based on an asymptotic theory, convergence theorems are tested for score vectors and information matrices of both models. Finally, an application of this theory is presented and assessed using data from the R statistical program.

**Key words:** Multinomial logit model; Saturated model; Logistic regression; Maximum likelihood estimator; Score vector; Fisher information matrix.

#### Resumen

En este artículo se desarrolla un estudio teórico de los modelos logísticos y saturados multinomiales cuando la variable de respuesta toma uno de  $R \geq 2$  niveles. Se presentan y demuestran teoremas sobre la existencia y cálculos de las estimaciones de máxima verosimilitud (ML-estimaciones) de los parámetros de ambos modelos. Se encuentran sus propiedades y, usando teoría asintótica, se prueban teoremas de convergencia para los vectores de puntajes y para las matrices de información. Se presenta y analiza

---

<sup>a</sup>MSc. E-mail: [eorozco15@unisimonbolivar.edu.co](mailto:eorozco15@unisimonbolivar.edu.co)

<sup>b</sup>Dr. rer. nat. E-mail: [hllinas@uninorte.edu.co](mailto:hllinas@uninorte.edu.co)

<sup>c</sup>MSc. E-mail: [jfonsecar@pca.edu.co](mailto:jfonsecar@pca.edu.co)

una aplicación de esta teoría con datos tomados de la librería `aplore3` del programa R.

**Palabras clave:** Modelo logístico multinomial; Modelo saturado; Regresión logística; Estimador de máxima verosimilitud; Vector score; Matriz de información de Fisher.

## 1. Introduction

For classifying individuals, logistic regression models are generalized linear models used (Cox 1958), which allow covariates that have been continuously and categorically escalated to predict any outcome that has been categorically escalated (Darlington 1990). These models do not generate assumptions on explanatory variable distributions and generalize multiple regression analysis techniques to cases having a categorical dependent variable and a categorical or numerical predictor variable. Hence, regardless of study design, the logistic regression model is a direct probability model that is able to provide valid estimates (Harrell 2015). Accordingly, one of the most common model variants that is assessed is when the response variable is binary, whether nominal or ordinal, as evidenced by Hosmer & Lemeshow (2000), Agresti (2013) and Monroy, Morales & Dávila (2018). Furthermore, LLinás (2006) discussed certain theoretical details of these models.

For example, applied sciences (such as biomedical and social sciences) often deal with nominal response variables at several levels as well as a vector of explanatory variables in which certain components may be intervals while others may be nominal scale measurements. In this case and other cases in which the answers are not ordinal and the levels are not organized in a sequential order, a multinomial logistic regression may assist in assessing the relationship between nominal responses and the set of explanatory variables; moreover, its applicability is computationally accessible (Chan 2005, Long 1987).

The multinomial logit model is a generalized linear model, which can be considered as a direct extension of the binary logit model since its categories can contract and reduce it to a binary model. As per Begg & Gray (1984), this reduced model is able to properly estimate both logistic parameters and their corresponding standard errors. In fact, it is a special case of discrete choice models (or conditional logit models) introduced by McFadden (1973) who generalized binary logistic regression by allowing more than two discrete responses.

Based on our literature review, several studies have been published on multinomial logistic regression models. This technique has become widespread and its use has been very critical within social sciences, marketing applications, and demographic and educational research studies (Chuang 1997, Peng, Lee & Ingersoll 2002, Pohlman & Leitner 2003). For example, in classical works, such as Hosmer & Lemeshow (2000), Díaz & Morales (2009), Kleinbaum & Klein (2010) and Agresti (2013), these models are viewed as alternative solutions to monitor data analysis-related issues. Among their multiple applications, Anderson, Verkuilen & Peyton (2010) used a multinomial logistics model to study the

psychometric validity of a multiple response instrument. Similarly, as evidenced in Tan, Christiansen, Christensen, Kruse & Bathum (2004); Chen & Kao (2006) and McNevin, Santos, Gamez-Tato, Álvarez-Dios, Casares, Daniel, Phillips & Lareu (2013), these models have been very useful in genetics. Subsequently, to assess the reliability of cancer diagnoses, multinomial logistic regression models have been used (Lloyd & Frommer 2008). Furthermore, in epidemiology, this model has been used in an explanatory study of factors affecting malaria treatments in African pregnant women (Exavery, Mbaruku, Mbuyita, Makemba, Kinyonge & Kweka 2014). In economics, the model is implemented as part of a study that assessed occupational patterns and trends in The Netherlands (Dessens, Jansen, G. B. Jansen, Ganzeboom & Van der Heijden 2003). In demography, to determine risks associated with variables considered for each of these studies, Kim (2015) and Schnor, Vanassche & Bavel (2017) used three-level multinomial logistic regression models. Finally, Monyai, Lesaoana, Darikwa & Nyamugure (2016) applied multinomial logistic regression to educational factors derived from the 2009 General Household Survey in South Africa; however, Ekström, Esseen, Westerlund, Grafström, Jonsson & Ståhl (2018) applied logistic regression models to data collected from environmental monitoring programs.

Nevertheless, certain studies, such as Fahrmeir & Kaufmann (1985), McCullagh & Neider (2018), and Agresti (2013), failed to provide a detailed development of a general asymptotic theory for maximum likelihood (ML) estimation for independent but not identically distributed variables. However, classical Mathematical Statistics books, such as Zacks (1971) and Rao, Rao, Statistiker, Rao & Rao (1973), only mention identically distributed independent variables that are not applicable to generalized linear models. Nevertheless, many studies, such as Wedderburn (1974), Wedderburn (1976), or McCullagh (1983), discuss the more generalized concept of quasi-likelihood functions, which are important for logistical models with repeated measurements. Therefore, because of the critical role that multinomial logistic models play in several applications, this study seeks to develop a theory for independent but not identically distributed variables, i.e., a theory that is indeed outlined in the literature but not discussed in detail.

Therefore, for independent but not identically distributed variables, theoretical details must be generalized to multinomial model applications when the response variable takes any of  $R \geq 2$  levels, which is the primary contribution of this work. In fact, multinomial models in which the response variable may take one of three levels are addressed in Llinás & Carreño (2012) and Llinás, Arteta & Tilano (2016).

This paper is organized as follows. Section 2 presents the multinomial model, and Section 3, we briefly introduce the saturated model. Then, in Section 4, we discuss the results from the score vector and the information matrix for the saturated model. Subsequently, in Section 5, we develop the theory corresponding to the multinomial logistic model. Section 6 provides the results from the score vector and the information matrix for this logistic model. Then, in Section 7, we present and demonstrate a theorem on the existence of logistic parameters and Section 8 addresses an application of the previously introduced theory. The paper ends with some conclusions in the Section 9.

## 2. Basic Model

Assume that the variable of interest  $Y$  able to assume  $R$  levels  $0, 1, \dots, R-1$ . For each  $r = 0, 1, \dots, R-1$ , we use the notation  $p_r := P(Y = r)$  to represent the probability of  $Y$  taking the value of  $r$ . Making independent observations of  $Y$ , we obtain the sample  $Y = (Y_1, Y_2, \dots, Y_n)^T$  with  $y = (y_1, y_2, \dots, y_n)^T$ , where  $y_i \in \{0, 1, 2, \dots, R-1\}$ ,  $i = 1, \dots, n$ , is a possible value of the sampling variable  $Y_i$ . Note that the variables  $Y_i$  are independent of each other. To build the likelihood function,  $R$  independent variables will be created with values in  $\{0, 1\}$ , as follows:  $U_{ri} = 1$  si  $Y_i = r$  and  $U_{ri} = 0$ , otherwise, where  $r = 0, 1, \dots, R-1$  y  $i = 1, \dots, n$ . Note that  $U_{ri}$  has a Bernoulli distribution with parameter  $p_{ri} = P(Y_i = r)$ . In terms of  $U_{ri}$ , the sampling variables will be  $U_i = (U_{0i}, U_{1i}, \dots, U_{(R-1)i})$ , with values  $u_i = (u_{0i}, u_{1i}, \dots, u_{(R-1)i})$ , where  $\sum_{r=0}^{R-1} u_{ri} = 1$ , for a fixed  $i$  value. Therefore, we achieve the following:  $P(U_i = u_i) = \prod_{r=0}^{R-1} p_{ri}^{u_{ri}}$ ,  $i = 1, \dots, n$ . Fixing  $y$  we get the likelihood function  $L$  for parameter  $p = (p_0, p_1, \dots, p_{(R-1)})^T$ , with  $p_r := (p_{r1}, p_{r2}, \dots, p_{rn})^T$ , and using this, the logarithm of the likelihood function is obtained as follows:

$$\mathcal{L}(p) := \ln L(p) = \sum_{i=1}^n \left[ \sum_{r=0}^{R-2} u_{ri} \ln p_{ri} + \left( 1 - \sum_{r=0}^{R-2} u_{ri} \right) \ln \left( 1 - \sum_{r=0}^{R-2} p_{ri} \right) \right] \quad (1)$$

## 3. Saturated Model

The saturated model is characterized by the following assumptions:

1. The following are the basic assumptions for the model:
  - (a) There are  $K$  explanatory variables  $X_1, \dots, X_K$  (numerical or categorical) with values  $x_{1i}, \dots, x_{Ki}$ , for  $i = 1, \dots, n$  (fixed or observed according to deterministic or random variables).
  - (b) Among the  $n$  values  $(x_{1i}, \dots, x_{Ki})$  of  $X = (X_1, \dots, X_K)$ , there are  $J$  different values, thus defining  $J$  populations. Therefore,  $J \leq n$ .

For each population,  $j = 1, \dots, J$ , where:

- $n_j$  is the number of  $Y_{ij}$  observations (or  $U_{rij}$  observations in the category  $r$ ) in each  $j$  population. Then,  $n_1 + \dots + n_J = n$ .
- $Z_{rj} := \sum_{i=1}^{n_j} U_{rij}$ ,  $r = 0, 1, 2, \dots, R-1$  fixed, the random variable that represents the sum of the  $n_j$  observations  $U_{rij}$  in  $j$ . Hence,  $z_{rj} = \sum_{i=1}^{n_j} u_{rij}$ , with  $\sum_{j=1}^J z_{rj} = \sum_{i=1}^n u_{ri}$ .

For clarity, the  $j$ th  $(x_{1j}, \dots, x_{Kj})$  population will be abbreviated by the symbol  $\star$  for each fixed  $r = 0, 1, 2, \dots, R-1$ .

2. For each  $j = 1, \dots, J$  population and each  $i = 1, \dots, n$  observation in  $j$ , it is assumed that the variables  $(U_{ri} | \star)$  are independent of each other, with a Bernoulli distribution with parameter  $p_{rj} = P(U_{ri} = 1 | \star) = E(U_{ri} | \star)$ . Here  $v_{rj} := V(U_{ri} | \star) = p_{rj}(1 - p_{rj})$ . Hereinafter, the  $\star$  will be pressed. This second assumption implies that, for each  $r = 0, 1, \dots, R - 1$  and each population  $j = 1, \dots, J$ :

- (a) All  $p_{rij}$  for  $i = 1, \dots, n$  within each population  $j$  are equal. That is, the dimensional vector  $(R-1)J$  is taken as a parameter  $p = (p_1, p_2, \dots, p_J)^T$ , where  $p_j := (p_{0j}, p_{1j}, \dots, p_{(R-1)j})$ .
- (b) The  $Z_{rj}$  variables are independent among populations with a binomial distribution with parameters  $n_j$  and  $p_{rj}$ , With  $m_{rj} := E(Z_{rj}) = n_j p_{rj}$  and  $V_{rj} := V(Z_{rj}) = n_j v_{rj}$ .

In vector form,  $Z = (Z_0, Z_1, \dots, Z_{R-1})^T$  with  $Z_r := (Z_{r1}, Z_{r2}, \dots, Z_{rJ})^T$ . Values are gathered in the  $z = (z_0, z_1, \dots, z_{R-1})^T$  vector, where the  $z_r$  components are defined by  $z_r := (z_{r1}, z_{r2}, \dots, z_{rJ})^T$ . From the above:

- $m := E(Z) = (m_0, m_1, \dots, m_{(R-2)})^T$ , with  $m_r := (m_{r1}, m_{r2}, \dots, m_{rJ})^T$ .
- $V := Cov(Z)$ , which is a matrix of  $(R - 1)J \times (R - 1)J$ . The elements of this matrix can be seen in item (b) in the proof of theorem 4.

According to (1), the logarithm of the maximum likelihood function will be:

$$\mathcal{L}(p) = \sum_{j=1}^J \left[ \sum_{r=0}^{R-2} z_{rj} \ln p_{rj} + \left( n_j - \sum_{r=0}^{R-2} z_{rj} \right) \ln \left( 1 - \sum_{r=0}^{R-2} p_{rj} \right) \right]. \tag{2}$$

The following theorem evidences the ML estimator for the saturated model parameters and the possible values that the logarithm of the corresponding likelihood function can accept when assessed in a point estimate vector of the parameter vector.

**Theorem 1.** For each fixed  $r = 0, 1, \dots, R - 1$  and  $j = 1, 2, \dots, J$ , the ML estimators of  $p_{rj}$  are  $\tilde{p}_{rj} = \frac{Z_{rj}}{n_j}$  with  $\tilde{p}_{rj} = \frac{z_{rj}}{n_j}$ . Further,

$$\mathcal{L}(\tilde{p}) = \sum_{j=1}^J n_j \left[ \sum_{r=0}^{R-2} \tilde{p}_{rj} \ln \tilde{p}_{rj} + \left( 1 - \sum_{r=0}^{R-2} \tilde{p}_{rj} \right) \ln \left( 1 - \sum_{r=0}^{R-2} \tilde{p}_{rj} \right) \right] < 0. \tag{3}$$

**Proof.** This theorem is demonstrated in the Appendix hereto. □

### 4. Score and Information of the Saturated Model

In this section, we present and demonstrate some asymptotic properties both for the score vector and for the information matrix in the saturated model,

highlighting the fact that we are using independent but not identically distributed variables, thus providing the details that are not yet found in the literature. These results are important for performing comparison tests using the logistic model. In the following theorem the corresponding properties are presented.

**Theorem 2.** *In the saturated model:*

- (a) *The score vector of the sample is a column vector of size  $(R-1)J$  and is given by  $S(p) := \frac{\partial \mathcal{L}}{\partial p} = (S_0(p), S_1(p), \dots, S_{R-2}(p))^T$ , where  $S_r(p)$  is a column vector of size  $J$ , with elements  $\frac{Z_{rj} - n_j p_{rj}}{v_{rj}} = \frac{n_j(\bar{p}_{rj} - p_{rj})}{v_{rj}}$ . In addition,  $E(S(p)) = 0$ .*
- (b) *The information matrix of the sample is a square matrix of size  $(R-1)J$  and is given by*

$$\mathfrak{S}(p) := \text{Cov}(S(p)) = \begin{pmatrix} A_{00} & A_{01} & \cdots & A_{0(R-2)} \\ A_{10} & A_{11} & \cdots & A_{1(R-2)} \\ \vdots & \vdots & \ddots & \vdots \\ A_{(R-2)0} & A_{(R-2)1} & \cdots & A_{(R-2)(R-2)} \end{pmatrix},$$

where  $A_{rr'}$  are diagonal matrices of size  $J$  and whose diagonal elements are  $\frac{n_j}{v_{rj}}$  if  $r = r'$  and  $-\frac{n_j p_{rj} p_{r'j}}{v_{rj} v_{r'j}}$  if  $r \neq r'$ . For clarity,  $\tilde{\mathfrak{S}} := \mathfrak{S}(p)$ .

- (c)  $E\left(-\frac{\partial^2 \mathcal{L}}{\partial p^2}\right) = -\tilde{\mathfrak{S}}^*$ , where  $\tilde{\mathfrak{S}}^*$  is a square matrix of size  $(R-2)J$  given by

$$\tilde{\mathfrak{S}}^* = \begin{pmatrix} A_{00} & -A_{11} & \cdots & -A_{(R-2)(R-2)} \\ -A_{00} & A_{11} & \cdots & -A_{(R-2)(R-2)} \\ \vdots & \vdots & \ddots & \vdots \\ -A_{00} & -A_{11} & \cdots & A_{(R-2)(R-2)} \end{pmatrix},$$

where  $A_{rr}$  is the diagonal matrix described in the previous item.

- (d) Then  $Z^* = \tilde{\mathfrak{S}}^{\frac{1}{2}}(V^*)^{-1}(Z - m)$ , where  $V^* = \text{diag}\{v_0^*, v_1^*, \dots, v_{(R-2)}^*\}$ , with  $v_r^* = (v_{r1}, v_{r2}, \dots, v_{rJ})^T$ . Then,  $S(p) = \tilde{\mathfrak{S}}^{\frac{1}{2}} Z^*$  or  $Z^* = \tilde{\mathfrak{S}}^{-\frac{1}{2}} S(p)$ .

**Proof.**

- (a) We denote the random score vector of the  $i$  observation by  $S_i(p)$ . The results are obtained immediately if the following is taken into account: assumption 1 from section 3, the property  $S(p) = \sum_{i=1}^n S_i(p) = \sum_{j=1}^J \left( \sum_{i \in n_j} S_i(p) \right)$  and that  $E(Z_{rj}) = n_j p_{rj}$ .
- (b) Taking a fixed value  $j = 1, \dots, J$ . For all  $r = 0, 1, \dots, R-2$  we find that

$$V \left( \frac{Z_{rj} - n_j p_{rj}}{v_{rj}} \right) = \frac{n_j}{v_{rj}}$$

and for all  $r' = 0, 1, \dots, R - 2$  with  $r' \neq r$ :

$$Cov\left(\frac{Z_{rj} - n_j p_{rj}}{v_{rj}}, \frac{Z_{r'j} - n_j p_{r'j}}{v_{r'j}}\right) = \frac{1}{v_{rj} v_{r'j}} Cov(Z_{rj}, Z_{r'j}) = -\frac{n_j p_{rj} p_{r'j}}{v_{rj} v_{r'j}}.$$

(c) With  $j, k = 1, 2, \dots, J$  and  $r, r' = 0, 1, \dots, R - 2$ .

(1) If  $j \neq k$  and  $r \neq r'$ , then  $\frac{\partial^2 \mathcal{L}}{\partial p_{r'k} \partial p_{rj}} = 0$ . Therefore,  $E\left(\frac{\partial^2 \mathcal{L}}{\partial p_{r'k} \partial p_{rj}}\right) = 0$ .

(2) If  $j = k$  and  $r = r'$ , then:

$$\frac{\partial^2 \mathcal{L}}{\partial p_{rj}^2} = \frac{\partial}{\partial p_{rj}} \left[ \frac{Z_{rj} - n_j p_{rj}}{p_{rj}(1 - p_{rj})} \right] = - \left[ \left( \frac{Z_{rj}}{n_j} - p_{rj} \right) \left( \frac{n_j(1 - 2p_{rj})}{v_{rj}^2} \right) + \frac{n_j}{v_{rj}} \right].$$

(3) Now, for  $r \neq r'$  and  $j = k$ ,

$$\begin{aligned} \frac{\partial^2 \mathcal{L}}{\partial p_{r'k} \partial p_{rk}} &= \frac{\partial \mathcal{L}}{\partial p_{r'k}} \left[ \frac{Z_{rk}}{p_{rk}} - \frac{n_k - \sum_{r=0}^{R-2} Z_{rk}}{1 - \sum_{r=0}^{R-2} p_{rk}} \right] \\ &= \left( \frac{Z_{rk}}{n_k} - p_{rk} \right) \frac{n_k(1 - 2p_{rk})}{v_{rk}^2} + \frac{n_k}{v_{rk}}. \end{aligned}$$

That is, for  $r = r'$ ,  $E\left(\frac{\partial^2 \mathcal{L}}{\partial p_{rk}^2}\right) = -\frac{n_k}{v_{rk}}$ , and for,  $r \neq r'$ ,  $E\left(\frac{\partial^2 \mathcal{L}}{\partial p_{r'k} \partial p_{rk}}\right) = \frac{n_k}{v_{rk}}$ . The results found in these three cases therefore demonstrate the subsection.

(d) Since

$$\frac{\partial \mathcal{L}}{\partial p} = (V^*)^{-1}(Z - m) = \tilde{\mathfrak{S}}^{1/2} \left[ \tilde{\mathfrak{S}}^{-1/2} (V^*)^{-1} (Z - m) \right] = \tilde{\mathfrak{S}}^{1/2} Z^*,$$

the result is immediate

□

*Remark 1.* In this case,  $\tilde{\mathfrak{S}}$  has diagonalizable main submatrices with  $\frac{n_j}{v_{rj}} > 0$  elements and the determinant is positive, which implies that  $\tilde{\mathfrak{S}}$  positive definite and therefore non singular and the  $\tilde{\mathfrak{S}}^{-\frac{1}{2}}$  root exists (see Harville (1997)).

In the following theorem, asymptotic results are presented and demonstrated both for the score vector and the information matrix in the saturated model.

**Theorem 3.** Suppose  $\lim_{n_j \rightarrow \infty} \frac{n_j}{n v_{rj}} = \frac{1}{\sigma_{rj}^2} > 0$  for all  $j = 1, 2, \dots, J$  and all  $r = 0, 1, \dots, R - 1$ . Then, for the saturated model (when  $n \rightarrow \infty$  and  $J$  is fixed):

(a)  $\frac{1}{n}S(p) \stackrel{a}{=} \frac{1}{n}S^*(p)$ , where  $S^*(p) = \tilde{\mathfrak{S}}(\tilde{p} - p)$ , with  $\tilde{p}$  being the ML-estimator of  $p$  in the saturated model.

(b)  $\frac{1}{n} \left( -\frac{\partial^2 \mathcal{L}}{\partial p^2} \right) \stackrel{a}{=} \frac{1}{n} \tilde{\mathfrak{S}}^*$ . That is, for all  $j = 1, 2, \dots, J$  and all  $r, r' = 0, 1, \dots, R-1$ , we have

$$\frac{1}{n} \frac{\partial^2 \mathcal{L}}{\partial p_{r'j} \partial p_{rj}} - \frac{1}{n} E \left( \frac{\partial^2 \mathcal{L}}{\partial p_{r'j} \partial p_{rj}} \right) \xrightarrow{P} 0.$$

(c)  $Z^* \xrightarrow{d} \mathcal{N}_{(R-1)J} (0, I_{(R-1)J})$ .

(d)  $\frac{1}{\sqrt{n}}S(p) \xrightarrow{d} \mathcal{N}_{(R-1)J}(0, \tilde{\Xi})$ , where  $\tilde{\Xi} = \lim_{n \rightarrow \infty} \left( \frac{\tilde{\mathfrak{S}}}{n} \right)$ .

Here,  $\stackrel{a}{=}$  means asymptotic equivalence;  $\xrightarrow{P}$ , indicates convergence in probability;  $\xrightarrow{d}$ , indicates convergence in distribution;  $\mathcal{N}_{(R-1)J}$ , is the  $(R-1)J$ -dimensional normal distribution and  $I_{(R-1)J}$  is the  $(R-1)J$ -dimensional identity matrix.

**Proof.**

(a) As per item (a) of theorem 2,  $S(p) = \left[ \text{diag}(\tilde{\mathfrak{S}}) \right] (\tilde{p} - p)$ , where  $\text{diag}(\tilde{\mathfrak{S}})$  is the diagonal matrix whose elements are the same diagonal elements of  $\tilde{\mathfrak{S}}$ . Then,  $\frac{1}{n} (S(p) - S^*(p))$  is a vector column of size  $J$  that has the following as elements

$$\sum_{r=0}^{R-2} \frac{n_j p_{rj} p_{0j}}{n v_{rj} v_{0j}} (\tilde{p}_{rj} - p_{rj}), \quad j = 1, \dots, J$$

For a fixed  $j$  and  $r$ ,

$$\lim_{n_j \rightarrow \infty} \left[ \frac{n_j p_{rj} p_{0j}}{n v_{rj} v_{0j}} (\tilde{p}_{rj} - p_{rj}) \right] = \frac{1}{\sigma_{rj}^2} \frac{p_{0j} p_{rj}}{v_{0j}} \left( \frac{Z_{rj}}{n_j} - p_{rj} \right) \xrightarrow{P} 0.$$

(b) As per item (c) of theorem 2, it is known that  $E \left( -\frac{\partial^2 \mathcal{L}}{\partial p^2} \right) = -\tilde{\mathfrak{S}}$ . Suppose  $r, r'$  and  $j$  are fix and

$$E_{rr'}(n, j) := \frac{1}{n} \left( -\frac{\partial^2 \mathcal{L}}{\partial p_{r'j} \partial p_{rj}} \right) - \frac{1}{n} E \left( -\frac{\partial^2 \mathcal{L}}{\partial p_{r'j} \partial p_{rj}} \right).$$

- If  $r = r'$ , then

$$E_{rr'}(n, j) = -\frac{1}{n} \left[ \left( \frac{Z_{rj}}{n_j} - p_{rj} \right) \frac{n_j (1 - 2p_{rj})}{v_{rj}} \right].$$

- If  $r \neq r'$ , then

$$E_{rr'}(n, j) = \frac{1}{n} \left[ \left( \frac{Z_{rj}}{n_j} - p_{rj} \right) \frac{n_j (1 - 2p_{rj} + 2v_{rj})}{v_{rj}} \right].$$



Now, it is known that  $\left(\frac{Z_{rj}}{n_j} - p_{rj}\right) \xrightarrow{p} 0$  if  $n \rightarrow \infty$ , by the weak law of the large numbers. Because

$$\frac{n_j}{nv_{rj}} \frac{(1 - 2p_{rj})}{v_{rj}} \xrightarrow{n_j \rightarrow \infty} \frac{1}{\sigma_{rj}^2} \frac{(1 - 2p_{rj})}{v_{rj}} > 0$$

and

$$\frac{n_j}{nv_{rj}} \frac{(1 - 2(p_{rj} + v_{rj}))}{v_{rj}} \xrightarrow{n_j \rightarrow \infty} \frac{1}{\sigma_{rj}^2} \frac{(1 - 2(p_{rj} + v_{rj}))}{v_{rj}} > 0,$$

then,  $\frac{1}{n} \left(-\frac{\partial^2 \mathcal{L}}{\partial p_r \partial p_r}\right) \stackrel{a}{=} \frac{1}{n} \tilde{\mathfrak{S}}^*$ .

- (c) As the variables  $Z_{rj}$  converge to the normal distribution by the multivariate central limit theorem,  $Z \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}_{(R-1)J}(m, V)$ . Then, for a fixed value of  $J$ :

$$Z^* \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}_{(R-1)J}\left(E(Z^*), \tilde{\mathfrak{S}}^{-1/2} \tilde{\mathfrak{S}} \tilde{\mathfrak{S}}^{-1/2}\right) = \mathcal{N}_{(R-1)J}(0, I).$$

- (d) Considering item (d) of theorem 2, we have:

$$\left(\frac{1}{\sqrt{n}}\right) S(p) = \frac{1}{\sqrt{n}} \tilde{\mathfrak{S}}^{1/2} \left[\tilde{\mathfrak{S}}^{-1/2} (V^*)^{-1} (Z - m)\right] = \left(\frac{\tilde{\mathfrak{S}}}{n}\right)^{1/2} Z^*.$$

Now, because

$$\frac{1}{n} \tilde{\mathfrak{S}} \xrightarrow{n \rightarrow \infty} \tilde{\Xi}, \tag{4}$$

we have  $\left(\frac{\tilde{\mathfrak{S}}}{n}\right)^{1/2} \xrightarrow{c.s} \tilde{\Xi}^{1/2}$ . Therefore,  $\left(\frac{\tilde{\mathfrak{S}}}{n}\right)^{1/2} \xrightarrow{d} \tilde{\Xi}^{1/2}$ , and with it,  $\frac{1}{\sqrt{n}} S(p) \xrightarrow{d} \mathcal{N}_{(R-1)J}(0, \tilde{\Xi})$ .

□

*Remark 2.* The assumption from theorem 3 can be interpreted as follows: “The speed” of each  $n_j \rightarrow \infty$  must be the same as that of  $n \rightarrow \infty$ . For example, in a balanced design, all the  $n_j$  are equal. In this case,  $n_j = \frac{n}{J}$ . Therefore, the amount  $\frac{1}{n} \cdot \frac{n_j}{v_j} = \frac{1}{Jv_j}$  is fixed; that is, it does not depend on  $n$ . Using the notations from the previous theorem,  $\sigma_j^2 = Jv_j$  because, in this case, the expression (4) becomes an equal value of  $\frac{1}{n} \tilde{\mathfrak{S}} = \tilde{\Xi}$ .

## 5. Multinomial Logistic Model

Assumptions 1 and 2 of section 3 include that the design matrix:

$$C = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1K} \\ 1 & x_{21} & \cdots & x_{2K} \\ \vdots & \vdots & & \vdots \\ 1 & x_{J1} & \cdots & x_{JK} \end{pmatrix}$$

has full rank with  $Rg(C) = 1 + K \leq J$ . To define the logistic model, we take one of the categories of the dependent variable  $Y$  as the reference; let us say for  $R - 1$ , with the following additional assumption:

$$g_r(x_j) = \ln \left( \frac{p_{rj}}{p_{R-1,j}} \right) = \delta_r + \beta_{r1}x_{j1} + \dots + \beta_{rK}x_{jK}, \quad r = 0, 1, \dots, R - 2,$$

where  $x_j := (1, x_{j1}, \dots, x_{jK})^T$ . With  $\beta_r := (\delta_r, \beta_{r1}, \dots, \beta_{rK})^T$ , the parameter vector  $\alpha = (\beta_0, \beta_1, \dots, \beta_{R-2})^T$  is a column vector of size  $(R-1)(1+K)$ . Note that the assumption  $Rg(C) = 1 + K$  is important for the parameter  $\alpha$  to be identifiable. For an  $x_j$  observation in population  $j$  and for each  $r$ ,

$$p_{rj} = \frac{\exp \{g_r(x_j)\}}{\sum_{s=0}^{R-1} \exp \{g_s(x_j)\}}, \quad (5)$$

where  $g_{R-1}(x_j) = 0$ . The logarithm of the likelihood function can be written as a function of  $\alpha$  as follows:

$$\mathcal{L}(\alpha) = \sum_{j=1}^J \left[ \sum_{r=0}^{R-1} z_{rj} g_r(x_j) - n_j \ln \left( \sum_{s=0}^{R-1} \exp \{g_{sj}(x_j)\} \right) \right]. \quad (6)$$

The likelihood equations are found by calculating the first derivatives of  $\mathcal{L}(\alpha)$  with respect to each of the  $(R-1)(1+K)$  unknown parameters. Therefore, for each fixed  $k = 0, 1, \dots, K$  and each  $r = 0, 1, \dots, R$ , the likelihood equations are given by

$$\frac{\partial \mathcal{L}}{\partial \beta_{rk}} = \sum_{j=1}^J x_{jk} \left( Z_{rj} - n_j \left( \frac{\exp \{g_{rj}(x_j)\}}{\sum_{s=0}^{R-1} \exp \{g_{sj}(x_j)\}} \right) \right) = \sum_{j=1}^J x_{jk} (Z_{rj} - n_j p_{rj}). \quad (7)$$

The maximum likelihood estimator is obtained by equating these equations to zero and solving the logistic parameters. The solution requires the same type of iterations that were used to obtain the estimates in binary cases and with three levels, as demonstrated in Llinás (2006) and Llinás et al. (2016), respectively. For general cases, results are presented in section 7.

## 6. Logistic Model Information and Score

The following theorem shows some properties of the score vector and the information matrix in a logistic model.

**Theorem 4.** *In a logistic model:*

- (a) *The random score of the sample is a vector column of size  $(R-1)(1+K)$  given by  $S(\alpha) := \frac{\partial \mathcal{L}}{\partial \alpha} = (I_{R-1} \otimes C^T)(Z - m)$ . Also,  $E(S(\alpha)) = 0$ . Here,  $I_{R-1}$  is the identity matrix of order  $R-1$  and the symbol  $\otimes$  means the Kronecker product.*
- (b) *The information matrix of the sample is given by  $\mathfrak{S}(\alpha) := Cov(S(\alpha)) = (I_{R-1} \otimes C^T)V(I_2 \otimes C)$ . Here,  $I_{R-1}$  and  $I_2$  are the identity matrices of order  $R-1$  and  $2$ , respectively.*
- (c)  $E(\frac{\partial^2 \mathcal{L}}{\partial \alpha^2}) = \frac{\partial^2 \mathcal{L}}{\partial \alpha^2} = -\mathfrak{S}$ .

**Proof.**

- (a) Considering Equation (7). Taking  $\beta_{r0} = \delta_r$ , it can be verified that  $S(\alpha) = (I_{R-1} \otimes C^T)(Z - m)$ , where  $I_{R-1}$  is the identical matrix of order  $R-1$ . With this result, we have that  $E(S(\alpha)) = (I_{R-1} \otimes C^T)(E(Z) - n_j p_{rj}) = 0$ .
- (b) For part (a),

$$\mathfrak{S}(\alpha) = (I_{R-1} \otimes C^T)Cov(Z)(I_{R-1} \otimes C^T)^T = (I_{R-1} \otimes C^T)V(I_2 \otimes C). \quad (8)$$

- (c) Considering  $m_{rj}$  as defined in assumption 2 from section 3, for  $r, r' = 0, 1, \dots, R-2$ , and  $j = 1, \dots, J, k = 1, \dots, K$  all fixed values, we have  $\frac{\partial m_{rj}}{\partial \beta_{r'k}} = n_j \frac{\partial p_{rj}}{\partial \beta_{r'k}}$ . The following two cases are found:

- If  $r = r'$ , then  $\frac{\partial p_{rj}}{\partial \beta_{rk}} = v_{rj} x_{jk}$ . That is,  $\frac{\partial m_{rj}}{\partial \beta_{rk}} = n_j v_{rj} x_{jk}$ .
- If  $r \neq r'$ ,

$$\frac{\partial p_{rj}}{\partial \beta_{r'k}} = \frac{\exp\{g_{rj}(x_j)\} \exp\{g_{r'j}(x_j)\} x_{jk}}{\left(1 + \exp\{g_{r'j}(x_j)\} + \sum_{r \neq r'} \exp\{g_{rj}(x_j)\}\right)^2} = -p_{rj} p_{r'j} x_{jk}.$$

Thus,  $\frac{\partial m_{rj}}{\partial \beta_{rk}} = -n_j p_{rj} p_{r'j} x_{jk}$ .

Considering the results obtained in both cases, we have:

$$\frac{\partial^2 \mathcal{L}}{\partial \alpha^2} = \frac{\partial}{\partial \alpha} [(I_{R-1} \otimes C^T)(Z - m)] = -(I_{R-1} \otimes C^T) \frac{\partial m}{\partial \alpha}.$$

Because,  $\frac{\partial m}{\partial \alpha} = V(I_{(R-1)} \otimes C)$ , we can conclude from Equation (8) that  $\frac{\partial^2 \mathcal{L}}{\partial \alpha^2} = -\mathfrak{S}$  and with that,  $E(\frac{\partial^2 \mathcal{L}}{\partial \alpha^2}) = E(-\mathfrak{S}) = -\mathfrak{S}$ .

□

*Remark 3.* For the particular case of non-grouped data, where  $n_j = 1, \forall j$  and  $J = n, Z = U, m = (p_0, p_1, \dots, p_{(R-2)})^T$ , with  $p_r = (p_{r1}, \dots, p_{rn})^T$  and  $C$  is the original design matrix of  $(R-1) \times (1+K)$ .

**Theorem 5.** *Considering the assumptions from sections 3 and 5:*

(a) *The matrix  $\mathfrak{S}$  has a full rank  $Rg(\mathfrak{S}) = (R-1)(1+K)$  and is positive definite.*

(b) *Suppose that  $\lim_{n_j \rightarrow \infty} \frac{n_j}{nv_{rj}} = \frac{1}{\sigma_{rj}^2} > 0$  does exist, then a  $\Xi$  square matrix, positive definite and of size  $(R-1)(1+K)$ , also exists such that:*

$$\frac{1}{\sqrt{n}}S(\alpha) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}_{(R-1)(1+K)}(0, \Xi), \text{ for a fixed value of } J.$$

Here,  $\xrightarrow{d}$  means convergence in distribution and  $\mathcal{N}_{(R-1)(1+K)}$  is the  $(R-1)(1+K)$ -dimensional normal distribution.

**Proof.**

(a) Taking into account that  $Rg(C) = 1+K$ , where  $\mathfrak{S}$  is square of  $(R-1)(1+K)$ , that  $Rg(A^T A) = Rg(A)$  and that  $Rg(I_{R-1} \otimes C) = Rg(I_{R-1})Rg(C)$ :

$$Rg(\mathfrak{S}) = Rg[(I_{R-1} \otimes C^T)V(I_{R-1} \otimes C)] = (R-1)(1+K).$$

Here,  $\otimes$  is the Kronecker product and  $I_{R-1}$ , the identity matrix of order  $R-1$ . This indicates that  $\mathfrak{S}$  has a full rank. Now, we must prove that  $\mathfrak{S}$  is positive definite. That is,  $\forall u \neq 0, u^T \mathfrak{S} u > 0$ . With  $u \neq 0$  being any column vector of  $(R-1)(1+K)$ , we have

$$\begin{aligned} u^T \mathfrak{S} u &= u^T \left[ (I_{R-1} \otimes C^T)(V^{\frac{1}{2}})^T V^{\frac{1}{2}}(I_{R-1} \otimes C) \right] u \\ &= \left( V^{\frac{1}{2}}(I_{R-1} \otimes C)u \right)^T \left( V^{\frac{1}{2}}(I_{R-1} \otimes C)u \right). \end{aligned}$$

But  $V^{1/2}(I_{R-1} \otimes C)u$  is a column vector of  $(R-1)(1+K)$ . Therefore, for all  $u \neq 0$ , it is true that  $u^T \mathfrak{S} u \geq 0$ . However,  $u^T \mathfrak{S} u = 0$  if and only if it is true that  $V^{\frac{1}{2}}(I_{R-1} \otimes C)u = 0$ . Now,

- For  $r = r'$ , the components  $n_j v_{rj} > 0, v_{rj} = p_{rj}(1 - p_{rj})$ .
- For  $r \neq r'$ , the components  $-n_j p_{rj} p_{r'j} < 0$ .

Therefore,  $u = 0$ . Then,  $\forall u \neq 0, u^T \mathfrak{S} u > 0$ . Thus,  $\mathfrak{S}$  is positive definite.

(b) With  $\lambda := (\lambda_0, \dots, \lambda_{(R-1)(1+K)})^T$  being any vector of real numbers, we want to prove that

$$\lambda^T \left( \frac{1}{\sqrt{n}}S(\alpha) \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \lambda^T S_i(\alpha)$$

has a asymptotic one-dimensional normal distribution, where  $S_i(\alpha)$  is the score vector of observation  $i$ . This is checked below:

- We have that  $E(\lambda^T S_i(\alpha)) = \lambda^T E(S_i(\alpha)) = \lambda^T(0) = 0$ .
- The information matrix  $\mathfrak{S}_i$  (for a  $Y_i$  observation in population  $j$ ) corresponding to the logistic model is given by  $\mathfrak{S}_i = Cov((S_i)(\alpha))$ . As  $\mathfrak{S}_i$  is positive definite  $\forall i$  in  $j$ ,

$$V(\lambda^T S_i(\alpha)) = \lambda^T Cov(S_i(\alpha)) \lambda = \lambda^T \mathfrak{S}_i \lambda > 0.$$

It is clear that  $V = V^* \tilde{\mathfrak{S}} V^*$ , where  $V^*$  is as per theorem 2(d). Now,  $\frac{1}{n} \mathfrak{S} = [(I_{R-1} \otimes C^T) V^*] \frac{\tilde{\mathfrak{S}}}{n} [V^*(I_{R-1} \otimes C)]$ . Therefore, as per Tilanos Theorem 3.3.1b (Tilano & Arteta 2012):  $\frac{1}{n} \mathfrak{S} \xrightarrow[n \rightarrow \infty]{} \tilde{\Xi}$ . Then,

$$\frac{1}{n} \mathfrak{S} \xrightarrow[n \rightarrow \infty]{} [(I_{R-1} \otimes C^T) V^*] \tilde{\Xi} [V^*(I_{R-1} \otimes C)] := \Xi$$

$$y \text{ Rg}(\Xi) = (R - 1)(1 + K).$$

Then, we must prove that  $\frac{1}{\sqrt{n}} S(\alpha) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}_{(R-1)(1+K)}(0, \Xi)$  for a fixed value of  $J$ . For the non-grouped case, it is true that  $\Xi$  holds  $\frac{1}{J} \mathfrak{S}(\alpha) \xrightarrow[J \rightarrow \infty]{} \Xi$ . Considering the above and knowing that  $\Xi$  is positive definite, we have

$$\frac{1}{n} \sum_{i=1}^n V \lambda^T S_i(\alpha) = \lambda^T \frac{1}{n} \mathfrak{S} \lambda \xrightarrow[n \rightarrow \infty]{} \lambda^T \Xi \lambda > 0.$$

In addition Lindberg’s condition holds, that is,

$$\forall \varepsilon > 0, \frac{1}{n} \sum_{i=1}^n E_{\lambda^T S_i(\alpha)} \left( [\lambda^T S_i(\alpha)]^2 \cdot 1_{\{[\lambda^T S_i(\alpha)]^2 > \varepsilon^2 n\}} \right) \xrightarrow[n \rightarrow \infty]{} 0.$$

Therefore, we have  $\lambda^T \left( \frac{1}{\sqrt{n}} S(\alpha) \right) \xrightarrow{d} N_1(0, \lambda^T \Xi \lambda)$ , when  $n \rightarrow \infty$ . Then, when applying the multivariate central limit theorem, it is concluded that  $\frac{1}{\sqrt{n}} S(\alpha) \xrightarrow{d} N_{2(1+K)}(0, \Xi)$ , when  $n \rightarrow \infty$  and  $J$  is fixed.

□

*Remark 4.* For the case of non-grouped data, where  $J = n$ , the assumption given in (b) makes no sense because  $J$  is not fixed. Then, it is assumed immediately that  $\frac{1}{J} \mathfrak{S}$  has a positive definite limit  $\Xi$  and  $\frac{1}{\sqrt{J}} S(\alpha) \xrightarrow{d} \mathcal{N}_{(R-1)(1+K)}(0, \Xi)$  when  $J \rightarrow \infty$ .

## 7. Existence and Calculations of the Logistic Parameters

**Theorem 6** (Existence theorem). *The ML-estimations  $\hat{\alpha}$  of  $\alpha$  exist, are unique and calculated according to the following recursion formula:*

$$\begin{aligned}\hat{\alpha}^{(0)} &= 0 \\ \hat{\alpha}^{(t+1)} &= \hat{\alpha}^{(t)} + [(I_{R-1} \otimes C^T) \hat{V}^{(t)} (I_{R-1} \otimes C)]^{-1} (I_{R-1} \otimes C^T)(Z - \hat{m}^{(t)}),\end{aligned}$$

where  $\hat{V}$  and  $\hat{m}$  are the estimated matrix of covariances and the expected vector of  $Z$ , respectively, (hence,  $\hat{V}$  and  $\hat{m}$ ), defined in section 3 assumption 2. In addition, asymptotically we have

$$\begin{aligned}\sqrt{n}(\hat{\alpha} - \alpha) &\stackrel{a}{=} Cov^{-1} \left( \frac{1}{\sqrt{n}} \frac{\partial \mathcal{L}}{\partial \alpha} \right) \cdot \left( \frac{1}{\sqrt{n}} \frac{\partial \mathcal{L}}{\partial \alpha} \right) \\ &= \sqrt{n} [(I_{R-1} \otimes C^T) V (I_{R-1} \otimes C)]^{-1} (I_{R-1} \otimes C^T)(Z - m).\end{aligned}$$

**Proof.** By theorem 4 (c)  $\frac{\partial^2 \mathcal{L}}{\partial \alpha^2} = -\mathfrak{S}$  is a full rank matrix  $(R-1)(K+1)$ . Therefore, only ML-estimations  $\hat{\alpha}$  may be used as solutions of the  $(R-1)(K+1)$  equations

$$S(\alpha) := \frac{\partial \mathcal{L}(\alpha)}{\partial \alpha} = 0. \quad (9)$$

Alternatively, by theorem 4 (a),  $(I_{R-1} \otimes C^T)(Z - m) = 0$ . The following must be true:  $\frac{\partial \mathcal{L}(\hat{\alpha})}{\partial \alpha} = 0$ . Using the Taylor approximation, if  $\alpha_1$  is a point between  $\alpha$  and  $\hat{\alpha}$ , then

$$\frac{\partial \mathcal{L}(\alpha)}{\partial \alpha} = \frac{\partial \mathcal{L}(\hat{\alpha})}{\partial \alpha} + \frac{\partial^2 \mathcal{L}(\alpha)}{\partial \alpha^2}(\alpha - \hat{\alpha}).$$

Considering Equation (9), this expression can be rewritten by theorem 4 as  $\hat{\alpha} - \alpha = [(I_{R-1} \otimes C^T) V (I_{R-1} \otimes C)]^{-1} (I_{R-1} \otimes C^T)(Z - m)$ , where  $V_1 = V(\alpha_1)$ . As  $\alpha_1$  is a point on the line segment that joins  $\alpha$  and  $\hat{\alpha}$ ,  $\alpha_1 = t\hat{\alpha} + (1-t)\alpha$ , for all  $t \in [0, 1]$ . Under the assumption that  $\hat{\alpha}$  is strongly consistent for  $\alpha$ , that is,  $\hat{\alpha} \xrightarrow{c.s.} \alpha$ ,  $n \rightarrow \infty$ , by components  $\alpha_1 \xrightarrow{a.s.} \alpha \Rightarrow \alpha_1 \xrightarrow{P} \alpha$ . This implies that  $\sqrt{n}(\hat{\alpha} - \alpha) = [(I_{R-1} \otimes C^T) \frac{1}{n} V_1 (I_{R-1} \otimes C)]^{-1} \frac{1}{\sqrt{n}} (I_{R-1} \otimes C^T)(Z - m)$ . As  $\alpha_1 \xrightarrow{P} \alpha$ , when  $n \rightarrow \infty$ , we have

$$\begin{aligned}\underbrace{\left[ (I_{R-1} \otimes C^T) \frac{1}{n} V_1 (I_{R-1} \otimes C) \right]^{-1} \frac{1}{\sqrt{n}} (I_{R-1} \otimes C^T)(Z - m)}_{\sqrt{n}(\hat{\alpha} - \alpha)} &\stackrel{a}{=} \\ \underbrace{\left[ (I_{R-1} \otimes C^T) \frac{1}{n} V (I_{R-1} \otimes C) \right]^{-1} \frac{1}{\sqrt{n}} (I_{R-1} \otimes C^T)(Z - m)}_{\sqrt{n}[(I_{R-1} \otimes C^T) V (I_{R-1} \otimes C)]^{-1} (I_{R-1} \otimes C^T)(Z - m)}.\end{aligned}$$

We have  $\hat{\alpha} \cong \alpha + [(I_{R-1} \otimes C^T)V(I_{R-1} \otimes C)]^{-1} (I_{R-1} \otimes C^T)(Z - m)$ . Replacing  $\alpha$  on the right side by the  $t$ -th approximation  $\hat{\alpha}^{(t)}$  of  $\alpha$ , we obtain the recursion formula that provides the  $(t + 1)$  - *th* approximation  $\hat{\alpha}^{(t+1)}$  of  $\hat{\alpha}$ .  $\square$

## 8. Example

Fontanella, Early & Phillips (2008) present results from a study that examines the influence of both clinical and non-clinical factors on level of aftercare decisions. The corresponding data set (named APS data) were modified to protect confidentiality. It can be downloaded from the `aplore3` library in the R program. For the application, we selected only two variables, which are described in Table 1.

TABLE 1: Code sheet for the variables in the Study.

Variable	Description	Code/Values	Name
1	Placement combined	0 = Residential, 1 = Intermediate residential, 2 = Day treatment or Outpatient	PLACE3 (Y)
2	History of violence	0=No, 1= Yes	VIOL (X)

To apply the multinomial logistics model, we have used *Day treatment or Outpatient* (2) as the reference outcome value. When these data were entered in the R statistical package and grouped into populations. First, the Table 2 shows the cross-classification of PLACE3 versus history of violence (VIOL). We observe that the dependent variable  $Y$  takes one of the possible  $R = 3$  values and the explanatory variable  $X$  is dichotomic.

TABLE 2: Cross-Classification of Placement (PLACE3) by History of Violence (VIOL).

PLACE3 (Y)	History of violence (X)		Total
	No (0)	Yes (1)	
Intermediate residential (0)	26	104	130
Residential (1)	15	104	119
Day treatment or Outpatient (2)	80	179	259
Total	121	387	508

We found that  $J = 2$ , with  $n_1 = 121$  (Group for  $X = 0$ ) and  $n_2 = 387$  (Group for  $X = 1$ ). In this case,  $Z = (15, 104, 26, 104)^T$ . By applying theorem 1, the vectors of estimated parameters were  $\tilde{p} = (0.124, 0.269, 0.215, 0.269)^T$  and  $\mathcal{L}(\tilde{p}) = -515.7323$ . The matrix  $\hat{V} := \widehat{Cov}(Z)$  is of  $4 \times 4$  and has the form:

$$\hat{V} = \begin{pmatrix} 13.140 & 0.000 & -3.223 & 0.000 \\ 0.000 & 76.052 & 0.000 & -27.948 \\ -3.223 & 0.000 & 20.413 & 0.000 \\ 0.000 & -27.948 & 0.000 & 76.052 \end{pmatrix}$$

By applying theorem 2, we could easily verify that  $E(S(\tilde{p})) = 0$  and that the information matrix in the saturate model is provided by

$$\widehat{\mathfrak{S}}(\tilde{p}) := \widehat{Cov}(S(\tilde{p})) = \begin{pmatrix} 1114.189 & 0.000 & -175.925 & 0.000 \\ 0.000 & 1969.306 & 0.000 & -723.702 \\ -175.925 & 0.000 & 717.231 & 0.000 \\ 0.000 & -723.702 & 0.000 & 1969.306 \end{pmatrix}$$

Since  $J = 1 + K$  ( $J = 2$  and  $K = 1$ ) then  $C$  would be a non-singular or invertible matrix. Therefore,  $\hat{\alpha} = (\hat{\beta}_0, \hat{\beta}_1)^T$  with  $\hat{\beta}_r = C^{-1}\hat{g}_r$  for  $r = 0, 1$ . Here

$$\hat{\beta}_r := (\hat{\delta}_r, \hat{\beta}_{r1})^T, \quad \hat{g}_r := \left( \ln \left( \frac{\hat{p}_{r1}}{\hat{p}_{21}} \right), \ln \left( \frac{\hat{p}_{r2}}{\hat{p}_{22}} \right) \right)^T$$

That is, there is a one-to-one relationship between the parameters of the saturated model and those of the logistic model. That is, the models express the same thing, where  $\hat{p}_{rj} = \tilde{p}_{rj}$  for each  $j = 1, 2$ . Taking into account that

$$\hat{g}_0 = \begin{pmatrix} -1.674 \\ -0.543 \end{pmatrix}, \quad \hat{g}_1 = \begin{pmatrix} -1.124 \\ -0.543 \end{pmatrix}, \quad C^{-1} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$$

we found that  $\hat{\alpha} = (-1.674, 1.131, -1.124, 0.581)^T$ . Applying the theorem 4, the estimation of the information matrix in logistic model is given by

$$\widehat{\mathfrak{S}}(\hat{\alpha}) := \widehat{Cov}(S(\hat{\alpha})) = \begin{pmatrix} 89.192 & 76.052 & -31.171 & -27.948 \\ 76.052 & 76.052 & -27.948 & -27.948 \\ -31.171 & -27.948 & 96.465 & 76.052 \\ -27.948 & -27.948 & 76.052 & 76.052 \end{pmatrix}$$

As the estimator of the covariance matrix of the maximum likelihood estimator is the inverse of the observed information matrix, then the estimated covariance matrix for the fitted model is

$$\widehat{V}(\hat{\alpha}) = [\widehat{\mathfrak{S}}(\hat{\alpha})]^{-1} = \begin{pmatrix} 0.079 & -0.079 & 0.013 & -0.013 \\ -0.079 & 0.094 & -0.013 & 0.018 \\ 0.013 & -0.013 & 0.051 & -0.051 \\ -0.013 & 0.018 & -0.051 & 0.066 \end{pmatrix}$$

When we apply the package *nnet* in R, we get the same results as above. The results of fitting the three-category logistic regression model, using the *multinom* function, to these data are presented in Table 3. In this table, appear the estimated coefficients, the estimated standard error of the coefficients, the values of the estimated odds ratio ( $\widehat{OR}$ ) and the 95% confidence interval for the odds ratio for  $PLACE3 = 0$  versus  $PLACE3 = 2$  and for  $PLACE3 = 1$  versus  $PLACE3 = 2$ .

From the Table, we see that statistically the VIOL variable is significantly associated with adolescent placement. Table 4 shows the results obtained when performing the comparison test of the null model with the logistic model.



TABLE 3: Results of Fitting the Logistic Regression Model to the Data.

Logit	Coefficient	Estimation	Std. Error	$\widehat{OR}$	95% CI
0	Constant	-1.674	0.2814		
	VIOL	1.131	0.3072	3.10	(1.70, 5.66)
1	Constant	-1.124	0.2257		
	VIOL	0.581	0.2572	1.79	(1.08, 2.96)

TABLE 4: Comparison test of the null model with the logistic model.

Model	DF	$-2\mathcal{L}(\widehat{\theta})$	$\Delta DF$	$\Delta(-2\mathcal{L}(\widehat{\theta}))$	P-value
Null	1014	1048.742			
Residual	1012	1031.465	2	17.2774	0.0002

From this Table, the log-likelihood for the constant only model is  $\mathcal{L}(\widehat{\delta}_0) = 524.37093$  and the loglikelihood of the fitted model is  $\mathcal{L}(\widehat{\alpha}) = 515.73225$ . The value of the statistic is

$$2[\mathcal{L}(\widehat{\alpha}) - \mathcal{L}(\widehat{\delta}_0)] = 17.2774,$$

which yields a p-value of 0.0002. In conclusion, having a history of violence is a significant factor for being placed in some type of residential facility.

## 9. Conclusions

Recent studies, such as Zacks (1971), Rao et al. (1973), Fahrmeir & Kaufmann (1985), McCullagh & Neider (2018), and Agresti (2013), fail to provide a detailed development of a general asymptotic theory for ML estimation for independent but not identically distributed variables of generalized linear models. Furthermore, other works, such as Wedderburn (1974), Wedderburn (1976), or McCullagh (1983), only limit themselves to discussing the more general concept of quasi-likelihood functions, which are important for logistic models with repeated measurements. Based on these gaps identified in the literature, for independent but not identically distributed variables, theoretical details must be generalized to multinomial model applications when the response variable consider any  $R \geq 2$  levels. This is the primary contribution of this work. In fact, multinomial models where the response variable may take one of three levels are addressed in LLinás & Carreño (2012) and LLinás et al. (2016)). In this study, we extended these last two works to cover the cases where  $R > 3$ .

We assessed multinomial logistic and saturated models where the response variable takes one of  $R \geq 2$  values, emphasizing the fact that we used independent but not identically distributed variables, thus providing details that are not yet found in the literature. For this purpose, we demonstrated the properties of the score vector and the information matrices for these models. Furthermore, based on an asymptotic theory, we proved the convergence theorems for the score

vector and information matrices of both models and emphasized on the fact that these vectors exhibit normal multivariate distributions. Moreover, we validated a theorem about the existence and calculation of ML estimates for the parameters of the multinomial saturated model. Similarly, we presented and demonstrated a theorem for the existence of ML estimates for logistic parameters, thus briefly explaining the iteration method used for its calculation, i.e., the NewtonRaphson method.

Based on the results from this work, we will be able to compare the logistic model against its corresponding saturated model, which will allow us to reduce the number of observations and perform faster computer-based assessments. For future studies, the results yielded by this study may be used to construct test statistics, as well as their corresponding asymptotic distributions.

## Appendix

*Proof of theorem 1.* We fix  $r$  and  $j$  as constant values. For  $0 < p_{rj} < 1$ , of (2) we have that  $\frac{\partial \mathcal{L}}{\partial p_{rj}} = 0$  if and only if

$$z_{rj} - z_{rj} \sum_{r=0}^{R-2} p_{rj} - n_j p_{rj} + p_{rj} \sum_{r=0}^{R-2} z_{rj} = 0.$$

That is,  $p_{rj} = \frac{z_{rj}}{n_j}$ . Now:

$$\frac{\partial^2 \mathcal{L}}{\partial p_{rj}^2} = - \left[ \frac{z_{rj}}{p_{rj}^2} + \frac{\left( n_j - \sum_{r=0}^{R-2} z_{rj} \right)}{\left( 1 - \sum_{r=0}^{R-2} p_{rj} \right)^2} \right].$$

However, since  $0 < \sum_{r=0}^{R-2} z_{rj} < n_j$ , we have:

$$\frac{\partial^2 \mathcal{L}}{\partial p_{rj}^2} \Big|_{p_{rj} = \frac{z_{rj}}{n_j}} = - \left[ \frac{n_j^2}{z_{rj}} + \frac{n_j}{\left( n_j - \sum_{r=0}^{R-2} z_{rj} \right)} \right] < 0.$$

That is,  $\tilde{p}_{rj} = \frac{z_{rj}}{n_j}$ . Now, the following extreme cases must be assessed:

- If  $\sum_{r=0}^{R-2} z_{rj} = 0$ , then  $\frac{\partial^2 \mathcal{L}}{\partial p_{rj}^2} \Big|_{p_{rj} = \tilde{p}_{rj}} = - \frac{n_j}{1 - \sum_{r=0}^{R-2} \tilde{p}_{rj}} < 0$ . In this case,  $\mathcal{L}$

decreases in  $\tilde{p}_{rj}$ . That is,  $\mathcal{L}(\tilde{p})$  assumes a maximum value when  $\tilde{p}_{rj} = 0$ .

- If  $\sum_{r=0}^{R-2} z_{rj} = n_j$  then  $\frac{\partial^2 \mathcal{L}}{\partial p_{rj}^2} \Big|_{p_{rj}=\tilde{p}_{rj}} = \frac{n_j}{\tilde{p}_{rj}} > 0$ . In this case,  $\mathcal{L}$  increases in  $\tilde{p}_{rj}$ .  
That is,  $\mathcal{L}(\tilde{p})$  assumes a maximum when  $\tilde{p}_{rj} = 1$ .

□

[Recibido: octubre de 2019 — Aceptado: mayo de 2020]

## References

- Agresti, A. (2013), *Categorical Data Analysis*, Wiley Series in Probability and Statistics, 3 edn, John Wiley and Sons.
- Anderson, C., Verkuilen, J. & Peyton, B. (2010), ‘Modeling polytomous item responses using simultaneously estimated multinomial logistic regression models’, *Journal of Educational and Behavioral Statistics* **35**(4), 422–452.
- Begg, C. B. & Gray, R. (1984), ‘Calculation of polychotomous logistic regression parameters using individualized regressions’, *Biometrika* **71**(1), 11–18.
- Chan, Y. H. (2005), ‘Biostatistics 305. multinomial logistic regression’, *Singapore medical journal* **46**(6), 259–269.
- Chen, Y.-H. & Kao, J.-T. (2006), ‘Multinomial logistic regression approach to haplotype association analysis in population-based case-control studies’, *BMC Genetics* **7**(1), 1–12.
- Chuang, H.-L. (1997), ‘High school youths’ dropout and re-enrollment behavior’, *Economics of Education Review* **16**(2), 171–186.
- Cox, D. R. (1958), ‘The regression analysis of binary sequences’, *Journal of the Royal Statistical Society: Series B (Methodological)* **20**(2), 215–232.
- Darlington, R. B. (1990), *Regression and linear models*, McGraw-Hill College.
- Dessens, J., Jansen, W., G. B. Jansen, W., Ganzeboom, B. & Van der Heijden, P. (2003), ‘Patterns and trends in occupational attainment of first jobs in the netherlands, 1930-1995: Ordinary least squares regression versus conditional multinomial logistic regression’, *Journal of the Royal Statistical Society. Series A (Statistics in Society)* **166**(1), 63–84.
- Díaz, L. & Morales, M. (2009), *Análisis estadístico de datos categóricos*, Editorial Universidad Nacional de Colombia.
- Ekström, M., Esseen, P.-A., Westerlund, B., Grafström, A., Jonsson, B. & Ståhl, G. (2018), ‘Logistic regression for clustered data from environmental monitoring programs’, *Ecological Informatics* **43**, 165 – 173.

- Exavery, A., Mbaruku, G., Mbuyita, S., Makemba, A., Kinyonge, I. & Kweka, H. (2014), 'Factors affecting uptake of optimal doses of sulphadoxine-pyrimethamine for intermittent preventive treatment of malaria in pregnancy in six districts of tanzania', *Malaria Journal* **13**(1), 2–9.
- Fahrmeir, L. & Kaufmann, H. (1985), 'Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models', *Annals of Statistic* **13**(1), 342–368.
- Fontanella, C. A., Early, T. J. & Phillips, G. (2008), 'Need or availability? modeling aftercare decisions for psychiatrically hospitalized adolescents', *Children and Youth Services Review* **30**(7), 758 – 773.
- Harrell, F. (2015), *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*, Springer Series in Statistics, Springer International Publishing.
- Harville, D. (1997), *Matrix algebra from a statistician's perspective*, Springer-Verlag.
- Hosmer, D. & Lemeshow, S. (2000), *Applied logistic regression*, Wiley Series in Probability and Statistics, second edition edn, John Wiley & Sons Inc.
- Kim, J. (2015), 'School socioeconomic composition and adolescent sexual initiation in malawi', *Studies in Family Planning* **46**(3), 263–279.
- Kleinbaum, D. & Klein, M. (2010), *Logistic Regression. A Self-Learning Text*, Statistics for Biology and Health, third edition edn, Springer-Verlag New York.
- LLinás, H. (2006), 'Precisiones en la teoría de los modelos logísticos', *Revista Colombiana de Estadística* **29**(2), 239–265.
- LLinás, H., Arteta, M. & Tilano, J. (2016), 'El modelo de regresión logística para el caso en que la variable de respuesta puede asumir uno de tres niveles: estimaciones, pruebas de hipótesis y selección de modelos', *Revista de Matemática: Teoría y Aplicaciones* **23**(1), 173–197.
- LLinás, H. & Carreño, C. (2012), 'The multinomial logistic model for the case in which the response variable can assume one of three levels and related models', *Revista Colombiana de Estadística* **35**(1), 131–138.
- Lloyd, C. J. & Frommer, D. J. (2008), 'An application of multinomial logistic regression to estimating performance of a multiple-screening test with incomplete verification', *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **57**(1), 89–102.
- Long, J. S. (1987), 'A graphical method for the interpretation of multinomial logit analysis', *Sociological Methods & Research* **15**(4), 420–446.
- McCullagh, P. (1983), 'Quasi-likelihood functions', *The Annals of Statistics* **11**(1), 59–67.

- McCullagh, P. & Nelder, J. (2018), *Generalized Linear Models*, CRC Press.
- McFadden, D. (1973), *Conditional Logit Analysis of Qualitative Choice Behavior*, BART impact studies final report series: Traveler behavior studies, Institute of Urban and Regional Development, University of California.
- McNevin, D., Santos, C., Gamez-Tato, A., Álvarez-Dios, J., Casares, M., Daniel, R., Phillips, C. & Lareu, M. (2013), 'An assessment of bayesian and multinomial logistic regression classification systems to analyse admixed individuals', *Forensic Science International: Genetics Supplement Series* **4**(1), 63 – 64.
- Monroy, L., Morales, M. & Dávila, L. (2018), *Análisis estadístico de datos categóricos*, Universidad Nacional de Colombia.
- Monyai, S., Lesaoana, M., Darikwa, T. & Nyamugure, P. (2016), 'Application of multinomial logistic regression to educational factors of the 2009 general household survey in south africa', *Journal of Applied Statistics* **43**(1), 128–139.
- Peng, C.-Y. J., Lee, K. L. & Ingersoll, G. M. (2002), 'An introduction to logistic regression analysis and reporting', *The journal of educational research* **96**(1), 3–14.
- Pohlman, J. & Leitner, D. (2003), 'A comparison of ordinary least squares and logistic regression', *The Ohio journal of science* **103**, 118–125.
- Rao, C. R., Rao, C. R., Statistiker, M., Rao, C. R. & Rao, C. R. (1973), *Linear statistical inference and its applications*, Vol. 2, Wiley New York.
- Schnor, C., Vanassche, S. & Bavel, J. (2017), 'Stepfather or biological father? education-specific pathways of postdivorce fatherhood', *Demographic Research* **37**, 1659–1694.
- Tan, Q., Christiansen, L., Christensen, K., Kruse, T. A. & Bathum, L. (2004), 'Apolipoprotein e genotype frequency patterns in aged danes as revealed by logistic regression models', *European Journal of Epidemiology* **19**(7), 651–656.  
\*<http://www.jstor.org/stable/3582754>
- Tilano, J. & Arteta, M. (2012), Modelos logísticos multinomiales: estimaciones, pruebas y selección de modelos, Master's thesis, Universidad del Norte.
- Wedderburn, R. W. M. (1974), 'Quasi-likelihood functions, generalized linear models, and the gauss-newton method', *Biometrika* **61**(3), 439–447.
- Wedderburn, R. W. M. (1976), 'On the existence and uniqueness of the maximum likelihood estimates for certain generalized linear models', *Biometrika* **63**(1), 27–32.
- Zacks, S. (1971), *The Theory of Statistical Inference*, Wiley series in probability and mathematical statistics, John Wiley.