




Article

Asymptotic Results for Multinomial Models

Isaac Akoto ^{1,2,*} , João T. Mexia ^{1,3,†}  and Filipe J. Marques ^{1,3} 

¹ NOVA School of Science and Technology, Campus de Caparica, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal; jtm@fct.unl.pt (J.T.M.); fjm@fct.unl.pt (F.J.M.)

² Department of Mathematics and Statistics, University of Energy and Natural Resources, Sunyani P.O. Box 214, Ghana

³ Center of Mathematics and its Applications, NOVA School of Science and Technology, Campus de Caparica, Universidade NOVA de Lisboa, 2829-516 Lisbon, Portugal

* Correspondence: i.akoto@campus.fct.unl.pt

† These authors contributed equally to this work.

Abstract: In this work, we derived new asymptotic results for multinomial models. To obtain these results, we started by studying limit distributions in models with a compact parameter space. This restriction holds since the key parameter whose components are the probabilities of the possible outcomes have non-negative components that add up to 1. Based on these results, we obtained confidence ellipsoids and simultaneous confidence intervals for models with normal limit distributions. We then studied the covariance matrices of the limit normal distributions for the multinomial models. This was a transition between the previous general results and on the inference for multinomial models in which we considered the chi-square tests, confidence regions and non-linear statistics—namely log-linear models with two numerical applications to those models. Namely, our approach overcame the hierarchical restrictions assumed to analyse the multidimensional contingency table.

Keywords: confidence ellipsoids; covariance matrices; limit distributions; classification; non-linear models



Citation: Akoto, I.; Mexia, J.T.; Marques, F.J. Asymptotic Results for Multinomial Models. *Symmetry* **2021**, *13*, 2173. <https://doi.org/10.3390/sym13112173>

Academic Editors: Jinyu Li and Sergei D. Odintsov

Received: 6 October 2021
Accepted: 4 November 2021
Published: 12 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In several fields of study such as health, business, social sciences and education, the outcomes of variables are mainly discrete, i.e., the variables only take finite or countable numbers. A discrete variable whose outcome only takes finite numbers is called a categorical variable [1]. A categorical variable consists of a set of categories that are non-overlapping [2] and the outcome could be binary (dichotomous), i.e., with just two possible levels, such as “present” or “absent” as a desired condition, or polytomous, i.e., with more than two levels, as is the case of the “Likert” scale [3]. There are two common types of polytomous variables, as can be seen in [4], which are the ordinal and nominal scale of measurement. Categorical variables, such as one’s eye colour, ethnicity and affiliations, the categories of which cannot be ordered in any way, are nominal, while categories such as the level of resistance to a drug of a patient, the level of education and economic status exhibit a natural order and are thus ordinal.

In a study in which all the observed variables are categorical, the most common way of representing the data is in a contingency table, which is a cross-tabulation of the variables [5,6]. When there are m -variables, the contingency table is an m -dimensional table—also known as a multidimensional table when the attributes are more than two. The information of a contingency table is mainly summarized through appropriate measures such as measures of association or models. Association measures, although easy in their computation and interpretation, lead to a great loss of information, as can be seen in [7]. Models are preferred in the case where a more sensitive analysis is required. A model is a “theory” or a conceptual framework about observations, and the parameters in the

model represent the “effects” that particular variables or combinations of variables have in determining the values taken by the observations.

The easiest and most common model for a contingency table is the log-linear model [8]. It is constructed by taking the natural logarithms of the cell probabilities by the analogy of the analysis of variance (ANOVA) models, as can be seen in [9–11]. Classical log-linear models are sometimes regarded in the framework of the generalized linear model (GLM). They are also important in connection with contingency matrices, as can be seen in [12]. Contemporary problems in categorical data analysis with extremely high-dimensional data with demanding computational procedures require the development of complex models. Much work has been done on the modelling of categorical data, as can be seen in [7,13]. For example, in [14], the author used regression models for modelling categorical data. In our work, we derived new asymptotic results that will enable us to obtain confidence ellipsoids and simultaneous confidence intervals, respectively, for the vector of probabilities and its components, which will enable us to overcome some inference limitations of the existing procedures.

Inferential statistical analysis requires assumptions about the probability distribution of the response variable. For categorical data, the main distribution is the multinomial distribution. Most of the time, categorical data result from n -independent and identical trials with each trial having two or more possible outcomes. When the n is identical and independent trials have the same category probabilities, then the distribution of counts in the various categories is the multinomial distribution. The binomial distribution is a special case of the multinomial distribution with just two possible outcomes for each trial. Usually, the parameters of the multinomial distribution are not known and these parameters are often estimated from the sample data by several estimation methods such as the maximum likelihood estimation (MLE), as can be seen, for instance, in [15], the minimum discrimination information (MDI) [16], weighted least squares (WLS) [17] and Bayesian estimation (BA) [18]. In a previous study [19], we wanted to minimize the average cost so we used statistical decision theory (SDT) since there were only a finite number of possible choices. We point out that we achieved consistency since the probability of selecting the choice with the least average cost tends towards 1 and when the sample size tends towards infinity.

If we have n realizations of an experiment with m possible results with probabilities p_1, \dots, p_m , we have the probability mass function, as can be seen in [20,21]:

$$\Pr \left[\bigcap_{l=1}^m (N_l = n_l) \right] = \frac{n!}{\prod_{l=1}^m n_l!} p_l^{n_l} \quad (1)$$

for the vector $\mathbf{N} = (N_1, \dots, N_m)$ of the times we obtain the different results. This probability mass function corresponds to the singular multinomial distribution $\mathbf{M}(\cdot | n, \mathbf{p})$. We name as multinomial the models for describing these sets of independent realizations of experiments with a finite number of results:

For the vector with $\mathbf{p} = (p_1, \dots, p_m)$ of probabilities, we have the vector of estimators:

$$\tilde{\mathbf{p}} = (\tilde{p}_1, \dots, \tilde{p}_m) \quad (2)$$

with:

$$\tilde{p}_l = \frac{n_l}{n}, \quad l = 1, \dots, m \quad (3)$$

Moreover, as can be seen in [22], as $n \rightarrow \infty$:

$$\sqrt{n}(\tilde{\mathbf{p}}_m - \mathbf{p}) \sim \mathcal{N}(\mathbf{0}, \mathbf{U}(\mathbf{p})) \quad (4)$$

where \sim indicates the limit distribution, in this case $\mathcal{N}(\mathbf{0}, \mathbf{U}(\mathbf{p}))$, the normal distribution with the null mean vector and covariance matrix:

$$\mathbf{U}(\mathbf{p}) = \mathcal{D}(\mathbf{p}) - \mathbf{p}\mathbf{p}^t, \tag{5}$$

where $\mathcal{D}(\mathbf{p})$ is the diagonal matrix with principal elements p_1, \dots, p_m . This result will play an important role in the asymptotic treatment of the multinomial models which is this paper’s goal.

To carry out that asymptotic treatment, we start by obtaining a convenient version of the continuous mapping theorem [23] in the next section on limit distributions. Then, we obtained confidence regions in Section 3, namely the confidence ellipsoids and simultaneous confidence intervals. Then, in Section 4, we studied the algebraic structure of the limit covariance matrix, $\mathbf{U}(\mathbf{p})$.

In Section 5, we obtained chi-square tests for hypotheses on outcome probabilities and confidence ellipsoids and simultaneous confidence intervals for them. We also considered log-linear models for which we presented a numerical application. We pointed out that our approach to these models overcame the hierarchical restriction used to analyse multidimensional contingency tables.

Our use of both the classical and the new version of the parametrized continuous mapping theorem (PCMT) enabled us to carry out statistical inference for multinomial models. This inference was similar to ANOVA and related techniques but F -tests were replaced by chi-square tests which is highly convenient since now we have an infinity of degrees of freedom for the error.

Finally, we stress the close relationship between our ANOVA-like inference using the chi-square test and the usual treatment of fixed effect models. We point out that the F -test in that treatment had interesting invariance properties that expressed the symmetry of those models, especially since those models are associated with orthogonal partitions or sub-spaces which are invariant for rotation.

2. Limit Distributions

Let \mathcal{C} be the class of continuous functions. If $I(\cdot) \in \mathcal{C}$, and the distribution F_{Y_n} , of Y_n converges to $F_Y \in \mathcal{C}$ (that is $F_{Y_n}(\mathbf{y}_\cdot) \rightsquigarrow F_Y(\mathbf{y}_\cdot)$, whenever \mathbf{y}_\cdot is a continuity point of $F_Y(\cdot)$), we have, as can be seen in [23–26]:

$$F_{I(Y_n)} \xrightarrow{n \rightarrow \infty} F_{I(Y)} \tag{6}$$

as follows from the continuous mapping theorem.

If \mathbf{z} is obtained superposing sub-vectors \mathbf{u} and \mathbf{v} , we put $\mathbf{z} = [\mathbf{u} \ \mathbf{v}]^t$. Then, if $\theta_n \xrightarrow{p} \theta$, with θ belonging to a compact set \mathcal{D} , and $F_{Y_n} \rightarrow F_Y$, putting $\dot{Y}_n = [Y_n^t \ \theta_n^t]^t$, $Y_n^+ = [Y_n^t \ \theta^t]^t$ and $\dot{Y} = [Y^t \ \theta^t]^t$ to show that:

$$F_{I(\dot{Y}_n)} \longrightarrow F_{I(\dot{Y})}(\cdot | \theta), \tag{7}$$

it is only needed to show that, as can be seen in [27]:

$$\sup \left\{ |F_{I(\dot{Y}_n)} - F_{I(Y_n^+)}| \right\} \xrightarrow{n \rightarrow \infty} 0 \tag{8}$$

since $F_{I(Y_n^+)}(\cdot) = F_{I(\dot{Y})}(\cdot | \theta)$

With $\xi_n(\cdot)$ and $\zeta(\cdot)$, the probability measures associated with F_{Y_n} and F_Y and representing the Cartesian product by \times , whatever $\varepsilon > 0$, there exists a parallelepiped:

$$H(\varepsilon) = \times_{i=1}^m [a(\varepsilon); b(\varepsilon)] \tag{9}$$

with $Y_n, Y \in R^n$, such that $\zeta(H(\varepsilon)) \geq 1 - \varepsilon$. Since $F_{Y_n} \rightarrow F_Y$, we have $\zeta_n(H(\varepsilon)) \xrightarrow{n \rightarrow \infty} \zeta(H(\varepsilon))$ and so there will be $n(\varepsilon)$ such that, for $n > n(\varepsilon)$:

$$\zeta_n(H(\varepsilon)) > 1 - 2\varepsilon. \tag{10}$$

Now:

$$\dot{H}(\varepsilon) = H(\varepsilon) \times \mathcal{D} \tag{11}$$

will also be a compact. Thus, if $\dot{I}(\cdot) \in \mathcal{C}$, its restriction to $\dot{H}(\varepsilon)$ will be uniformly continuous. So, whatever $\delta > 0$, there exists $\delta'(\delta) > 0$, such that, if $\dot{Y}, \dot{Y}' \in \dot{H}(\varepsilon)$ and $\|\dot{Y} - \dot{Y}'\| \leq \delta'(\delta)$, $\|\dot{I}(\dot{Y}) - \dot{I}(\dot{Y}')\| < \delta$, where $\|\cdot\|$ indicates the Euclidean norm of a vector.

Let $\dot{E}_n(\delta)$ and $\dot{E}_n(\varepsilon)$ be events that occur when $\|\dot{I}(\dot{Y}) - \dot{I}(\dot{Y}')\| < \delta$ and when $Y_n \in H(\varepsilon)$, respectively. We now establish:

Lemma 1. $Pr(\dot{E}_n(\delta)) \xrightarrow{n \rightarrow \infty} 1$.

Proof. Since the restriction on $\dot{I}(\cdot)$ to $\dot{H}(\varepsilon)$ is uniformly continuous:

$$Pr[\dot{E}_n(\delta)|\dot{E}_n(\varepsilon)] \xrightarrow{n \rightarrow \infty} 1.$$

Thus, we only have to point out that ε is arbitrary and that:

$$Pr[\dot{E}_n(\delta)|\dot{E}_n(\varepsilon)] - Pr[\dot{E}_n(\delta)] \leq 1 - Pr[\dot{E}_n(\varepsilon)] \tag{12}$$

so that:

$$Pr[\dot{E}_n(\delta)] \geq Pr[\dot{E}_n(\delta)|\dot{E}_n(\varepsilon)] - (1 - Pr[\dot{E}_n(\varepsilon)]) \tag{13}$$

where $\lim_{n \rightarrow \infty} (1 - Pr[\dot{E}_n(\varepsilon)]) \leq \varepsilon$, whatever $\varepsilon > 0$, to establish the thesis. □

Since:

$$Pr(\dot{E}_n(\delta)) \xrightarrow{n \rightarrow \infty} 1, \quad \forall \delta > 0,$$

we have:

$$\begin{cases} Pr[(\dot{I}(Y_n^+) \leq \mathfrak{z} \pm \varepsilon \mathbf{1}_m) | \dot{E}_n(\delta)] \xrightarrow{n \rightarrow \infty} Pr[\dot{I}(Y_n^+) \leq (\mathfrak{z} \pm \varepsilon \mathbf{1}_m)] = F_{\dot{I}(Y_n^+)}(\mathfrak{z} \pm \varepsilon \mathbf{1}_m) \\ Pr[(\dot{I}(\dot{Y}_n) \leq \mathfrak{z} \pm \varepsilon \mathbf{1}_m) | \dot{E}_n(\delta)] \xrightarrow{n \rightarrow \infty} Pr[\dot{I}(\dot{Y}_n) \leq (\mathfrak{z} \pm \varepsilon \mathbf{1}_m)] = F_{\dot{I}(\dot{Y}_n)}(\mathfrak{z} \pm \varepsilon \mathbf{1}_m) \end{cases} \tag{14}$$

thus, also whatever $\delta > 0$, there exists $\dot{n}(\varepsilon)$ such that, for $n \geq \dot{n}(\varepsilon)$, we have:

$$F_{\dot{I}(Y_n^+)}(\mathfrak{z}) = F_{\dot{I}(Y_n)}(\mathfrak{z}|\theta) \xrightarrow{n \rightarrow \infty} F_{\dot{I}(Y)}(\mathfrak{z}|\theta) \tag{15}$$

as well as:

$$F_{\dot{I}(Y_n^+)}(\mathfrak{z}) \xrightarrow{n \rightarrow \infty} F_{\dot{I}(Y)}(\mathfrak{z}|\theta) \tag{16}$$

whenever \mathfrak{z} is a continuity point of $F_{\dot{I}(Y)}$. Thus, we have the parametrized continuous mapping theorem (PCMT).

If $\dot{I}(\cdot) \in \mathcal{C}$, $F_{Y_n} \rightarrow F_Y \in \mathcal{C}$, and $\theta_n \xrightarrow{n \rightarrow \infty} \theta$, with $\theta \in \mathcal{D}$, a compact set:

$$F_{\dot{I}(Y_n)}(\cdot) \rightarrow F_{\dot{I}(Y)}(|\theta) \tag{17}$$

Remark 1 (Corollary to PCMT).

a With θ a parameter both for F_Y and $F_{\dot{I}(Y)}$;

b when F_Y is $\mathcal{N}(\mathbf{0}, \mathbf{V}(\boldsymbol{\theta}))$ and $\boldsymbol{\theta}_n \xrightarrow{p} \boldsymbol{\theta}$, we have:

$$F_{\mathbf{l}(\dot{Y}_n)} \rightarrow \mathcal{N}(\mathbf{0}, \mathcal{G}\mathbf{V}(\boldsymbol{\theta})\mathcal{G}^t)$$

when $\mathbf{l}(\dot{Y}_n) = [(\mathcal{G}Y_n)^t, \boldsymbol{\theta}_n^t]^t$ with \mathcal{G} a matrix and $\mathbf{V}(\boldsymbol{\theta})$ the covariance matrix of the parameter, $\boldsymbol{\theta}$.

This remark will be renamed as a corollary of PCMT (CPCMT), as can be seen in [6].

We now consider the case of a sequence of random vectors with the same limit distribution. The sequence $\{X_n\}$ of random vectors is the mean stable when all its vectors have the same mean vector $\boldsymbol{\mu}$.

Between the mean stable sequences, we may establish an equivalence relation writing $\{X_n\} \gamma \{Y_n\}$ if and only if:

$$S_n = \sup\{\|Y_n - X_n\|\} \xrightarrow{s} 0 \tag{18}$$

where \xrightarrow{s} means stochastic convergence, i.e., convergence in probability, as can be seen in [27,28]. We now establish:

Proposition 1. If $S_n \xrightarrow{s} 0$ and $F_{X_n} \rightarrow F^\circ \in \mathcal{C}$, then $F_{Y_n} \rightarrow F^\circ$, whenever $\{X_n\} \gamma \{Y_n\}$.

Proof. Let the vectors in $\{X_n\}$ and $\{Y_n\}$ have m components. Given $x \in R^m$, we consider the events $A_{n,1}(\varepsilon) = \{X_n \leq x - \varepsilon \mathbf{1}_m\}$, $A_{n,2}(\varepsilon) = \{Y_n \leq x\}$, $A_{n,3}(\varepsilon) = \{X_n \leq x + \varepsilon \mathbf{1}_m\}$ and $B_n(\varepsilon) = \{S_n \leq \varepsilon\}$.

Taking $\dot{A}_{n,i}(\varepsilon) = A_{n,i}(\varepsilon) \cap B_n(\varepsilon)$, and \Rightarrow to indicate the implication, we have $\dot{A}_{n,1}(\varepsilon) \Rightarrow \dot{A}_{n,2}(\varepsilon) \Rightarrow \dot{A}_{n,3}(\varepsilon)$. Thus, with $q_{n,i}(\varepsilon) = Pr(A_{n,i}(\varepsilon))$ and $\dot{q}_{n,i}(\varepsilon) = Pr(\dot{A}_{n,i}(\varepsilon))$, $i = 1, 2, 3$, we have $\dot{q}_{n,1} \leq \dot{q}_{n,2} \leq \dot{q}_{n,3}$.

Moreover, since $S_n \xrightarrow{s} 0$, whatever $\delta > 0$, there exist $n(\varepsilon, \delta)$ such that, for $n > n(\varepsilon, \delta)$, $Pr(B_n(\varepsilon)) > 1 - \delta$, so $\dot{q}_{n,i}(\varepsilon) \leq q_{n,i}(\varepsilon) \leq \dot{q}_{n,i}(\varepsilon) + \delta$, and, since δ is arbitrary:

$$\lim_{n \rightarrow \infty} \dot{q}_{n,i}(\varepsilon) = \lim_{n \rightarrow \infty} q_{n,i}(\varepsilon), \quad i = 1, 2, 3 \tag{19}$$

so:

$$\lim_{n \rightarrow \infty} q_{n,1}(\varepsilon) \leq \lim_{n \rightarrow \infty} q_{n,2}(\varepsilon) \leq \lim_{n \rightarrow \infty} q_{n,3}(\varepsilon) \tag{20}$$

since $F^\circ \in \mathcal{C}$, we have $\lim_{n \rightarrow \infty} q_{n,1}(\varepsilon) = F^\circ(x - \varepsilon \mathbf{1}_n)$ and $\lim_{n \rightarrow \infty} q_{n,3}(\varepsilon) = F^\circ(x + \varepsilon \mathbf{1}_n)$, so:

$$F^\circ(x - \varepsilon \mathbf{1}_n) \leq \lim_{n \rightarrow \infty} F_{Y_n}(x) \leq F^\circ(x + \varepsilon \mathbf{1}_n) \tag{21}$$

and given ε is arbitrary and x may be whatever, from $F^\circ \in \mathcal{C}$, we obtain:

$$F_{Y_n}(x) \xrightarrow{n \rightarrow \infty} F^\circ(x) \tag{22}$$

which completes the proof. □

We then consider normal limit distributions, starting with:

Proposition 2. Given $\mathcal{G}(\cdot) = \{g_1(\cdot), \dots, g_w(\cdot)\}$ such that its component functions have gradients $\mathbf{g}_1(\cdot), \dots, \mathbf{g}_w(\cdot)$, Hessian matrices $\mathbf{g}_1(\cdot), \dots, \mathbf{g}_w(\cdot)$, and continuous second-order partial derivatives. Whatever the mean stable sequence $\{Z_n\}$ with the invariant mean vector $\boldsymbol{\mu}$, taking $Y_n = \mathcal{G}(Z_n)$ and $X_n = \mathcal{G}(Z_n)$ with $\mathcal{G}(\cdot) = [g_1(\cdot), \dots, g_w(\cdot)]^t$, we have $\{X_n\} \gamma \{Y_n\}$, whenever $\sqrt{n}(Z_n - \boldsymbol{\mu}_n)$ converges in distribution to $F^\circ \in \mathcal{C}$.

Proof. We have $g_j(\mathbf{Z}_n) - g_j(\boldsymbol{\mu}) = \mathbf{g}_j(\boldsymbol{\mu})^t(\mathbf{Z}_n - \boldsymbol{\mu}) + \frac{1}{2}(\mathbf{Z}_n - \boldsymbol{\mu})^t \mathbf{g}_j(\boldsymbol{\theta}_{n,j})(\mathbf{Z}_n - \boldsymbol{\mu})$ with $(\boldsymbol{\theta}_{n,j})$ between \mathbf{Z}_n and $\boldsymbol{\mu}$. Since $\mathbf{Z}_n \xrightarrow{s} \boldsymbol{\mu}$, we also have $\boldsymbol{\theta}_{n,j} \xrightarrow{s} \boldsymbol{\mu}$. Then, with $\Theta_\varepsilon(\boldsymbol{\mu})$ the radius ε sphere with centre $\boldsymbol{\mu}$, $Pr(\mathbf{Z}_n, \boldsymbol{\theta}_{n,j} \in \Theta_\varepsilon(\boldsymbol{\mu})) \rightarrow 1, j = 1, \dots, m$. Now $\|\mathbf{g}_j(\boldsymbol{\mu})\|$ is a continuous function of $\boldsymbol{\mu}$ so it will have a maximum $u_{j,\varepsilon}$ in $\Theta_\varepsilon(\boldsymbol{\mu})$ that will exceed the supremum of the spectral radius of $\mathbf{g}_j(\boldsymbol{\mu})$ in $\Theta_\varepsilon(\boldsymbol{\mu})$, so:

$$\left\| \frac{1}{2}(\mathbf{Z}_n - \boldsymbol{\mu})^t \mathbf{g}_j(\boldsymbol{\theta}_{n,j})(\mathbf{Z}_n - \boldsymbol{\mu}) \right\| \leq u_{j,\varepsilon} \|\mathbf{Z}_n - \boldsymbol{\mu}\|^2, \quad j = 1, \dots, m \tag{23}$$

thus:

$$(g_j(\mathbf{Z}_n) - g_j(\boldsymbol{\mu}_n)) - \mathbf{g}_j(\boldsymbol{\mu})^t(\mathbf{Z}_n - \boldsymbol{\mu}) \xrightarrow{s} 0, \quad j = 1, \dots, m \tag{24}$$

and so:

$$\|(\mathcal{G}(\mathbf{Z}_n) - \mathcal{G}(\boldsymbol{\mu})) - \mathcal{G}(\boldsymbol{\mu})^t(\mathbf{Z}_n - \boldsymbol{\mu})\| \xrightarrow{s} 0 \tag{25}$$

and the thesis follows from Proposition 1. □

Corollary 3. *If $\sqrt{n}(\mathbf{Z}_n - \boldsymbol{\mu}) \sim \mathcal{N}(\mathbf{0}, \mathbf{V})$, under the hypothesis of Propositions 1 and 2, $\sqrt{n}[\mathcal{G}(\mathbf{Z}_n) - \mathcal{G}(\boldsymbol{\mu})] \sim \mathcal{N}[\mathbf{0}, \mathcal{G}(\boldsymbol{\mu})\mathbf{V}\mathcal{G}(\boldsymbol{\mu})^t]$.*

Proof. The thesis follows from Propositions 1 and 2 since the continuous mapping theorem, as can be seen in [23,24], implies that the limit distribution of $\sqrt{n}\mathcal{G}(\boldsymbol{\mu})(\mathbf{Z}_n - \boldsymbol{\mu})$ is $\mathcal{N}(\mathbf{0}, \mathcal{G}(\boldsymbol{\mu})\mathbf{V}\mathcal{G}(\boldsymbol{\mu})^t)$. □

3. Confidence Ellipsoids

We start by establishing the following.

Proposition 4. *If \mathbf{Y} —not necessarily normal—has a covariance matrix \mathbf{C} , with a range space $\Omega = R(\mathbf{C})$, and the mean vector $\boldsymbol{\mu}$, then:*

$$Pr(\mathbf{Y} - \boldsymbol{\mu} \in \Omega) = 1 \tag{26}$$

Proof. Let $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_m$ constitute an orthonormal basis for the orthogonal complement, Ω^\perp , of Ω . Then, $\boldsymbol{\alpha}_j^t(\mathbf{Y} - \boldsymbol{\mu})$ will have a null mean value and variance. Thus, according to the Bienaymé–Tchebycheff inequality:

$$Pr(\boldsymbol{\alpha}_j^t(\mathbf{Y} - \boldsymbol{\mu}) = 0) = 1, \quad j = 1, \dots, m.$$

Therefore, we obtain, with \mathbf{A} , the matrix with row vectors $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_m$:

$$\begin{aligned} Pr(\mathbf{Y} - \boldsymbol{\mu} \in \Omega) &= Pr[\mathbf{A}(\mathbf{Y} - \boldsymbol{\mu}) = \mathbf{0}] \\ &= Pr\left[\bigcap_{j=1}^m (\boldsymbol{\alpha}_j^t(\mathbf{Y} - \boldsymbol{\mu}) = 0)\right] \\ &= m - (m - 1) \\ &= 1, \end{aligned}$$

as follows from the Boole generalized inequalities, and so the thesis is established. □

We then have:

Lemma 2. *Given \mathbf{B} is a positive semi-definite [definite] matrix with positive eigenvalues $\varepsilon_1, \dots, \varepsilon_h$ and ε_h corresponding to eigenvectors $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_m$, we have $\mathbf{B} = \mathbf{A}^t \mathbf{D} \mathbf{A}$ and $\mathbf{B}^+ = \mathbf{A}^t \mathbf{D}^{-1} \mathbf{A}$ with $+$ indicating the Moore–Penrose inverse, \mathbf{D} the diagonal matrix with principal elements $\varepsilon_1, \dots, \varepsilon_h$ and $\mathbf{A}^t = [\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_h]$.*

Proof. It is easy to show that $\mathbf{B}\mathbf{B}^+$ and $\mathbf{B}^+\mathbf{B}$ are asymmetrical and that $\mathbf{B}\mathbf{B}^+\mathbf{B} = \mathbf{B}$ and $\mathbf{B}^+\mathbf{B}\mathbf{B}^+ = \mathbf{B}^+$ establish the thesis. \square

We can now establish the following.

Proposition 5. If $\mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2\mathbf{B})$, with \mathbf{B}^+ the Moore–Penrose inverse of \mathbf{B} :

$$(\mathbf{Y} - \boldsymbol{\mu})^t \mathbf{B}^+ (\mathbf{Y} - \boldsymbol{\mu}) \sim \sigma^2 \chi_h^2$$

where χ_h^2 is a central chi-square distribution with $h = \text{rank}(\mathbf{B})$, degrees of freedom and σ^2 is the variance of \mathbf{Y} .

Proof. As stated in Lemma 2, we have:

$$(\mathbf{Y} - \boldsymbol{\mu})^t \mathbf{B}^+ (\mathbf{Y} - \boldsymbol{\mu}) = \ddot{\mathbf{Y}}^t \ddot{\mathbf{Y}} \sim \sigma^2 \chi_h^2$$

with $\ddot{\mathbf{Y}} = \mathcal{D}^{-\frac{1}{2}} \mathbf{A} (\mathbf{Y} - \boldsymbol{\mu})$ where $\mathcal{D}^{-\frac{1}{2}}$ is the diagonal matrix with principal elements $\varepsilon_1^{-1/2}, \dots, \varepsilon_h^{-1/2}$. We now only have to point out that $\ddot{\mathbf{Y}} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_h)$ to establish the thesis, where \mathbf{I}_h is the identity matrix. \square

We now consider confidence ellipsoids and simultaneous confidence intervals. Ellipsoids and their support planes are presented in [29]; the affine point of x belongs to the ellipsoid:

$$\zeta(\boldsymbol{\mu}, \mathbf{B}, r) = \left\{ x : (x - \boldsymbol{\mu})^t \mathbf{B}^+ (x - \boldsymbol{\mu}) \leq r \right\} \tag{27}$$

if and only if:

$$\bigcap_v \left(|v^t \boldsymbol{\mu} - v^t x| \leq \sqrt{r(v^t \mathbf{B} v)} \right), \tag{28}$$

where \bigcap_v indicates that all possible vectors v are considered. We now establish:

Proposition 6. If $\mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{B})$:

$$\Pr \left[\bigcap_v \left(|v^t \boldsymbol{\mu} - v^t \mathbf{y}| \leq \sqrt{x_{h,1-q}(v^t \mathbf{B} v)} \right) \right] = 1 - q, \tag{29}$$

with $x_{h,1-q}$, the $(1-q)$ -th quantile of χ_h^2 (the central chi-square with h degrees of freedom), when $\text{rank}(\mathbf{B}) = h$.

Proof. The proof for the case $\mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{B})$ directly follows from the previous considerations. Thus, we only have to point out that $x \in \zeta(\boldsymbol{\mu}, \mathbf{B}, r)$ is equivalent to, as can be seen in [29]:

$$\bigcap_v \left(|v^t \boldsymbol{\mu} - v^t x| \leq \sqrt{r(v^t \mathbf{B} v)} \right). \tag{30}$$

\square

Since, as we saw $(\mathbf{Y} - \boldsymbol{\mu})^t \mathbf{B}^+ (\mathbf{Y} - \boldsymbol{\mu}) \sim \sigma^2 \chi_h^2$ when $\mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2 \mathbf{B})$, we have $(\mathbf{p} - \tilde{\mathbf{p}})^t \mathbf{U}^+ (\mathbf{p}) (\mathbf{p} - \tilde{\mathbf{p}}) \sim \chi_{m-1}^2$, because, as we shall see in the next section, $\text{rank}(\mathbf{U}(\mathbf{p})) = m - 1$.

In the next section, we will obtain results on $(\mathbf{U}(\mathbf{p}))$ that will be used to obtain chi-square confidence regions for \mathbf{p} and through duality, test hypothesis on \mathbf{p} .

4. Covariance Matrices

As we saw, for $\tilde{\mathbf{p}}$, the limit covariance matrix of $\sqrt{n}(\tilde{\mathbf{p}}_m - \mathbf{p})$ is:

$$\mathbf{U}(\mathbf{p}) = \mathcal{D}(\mathbf{p}) - \mathbf{p}\mathbf{p}^t \tag{31}$$

where $\mathbf{p} = (p_1, \dots, p_m), p_j > 0, j = 1, \dots, m$ and $\sum_{j=1}^m p_j = 1$. For the rank of the covariance matrix, we have:

$$\text{rank}(\mathbf{U}(\mathbf{p})) = \text{rank}(\mathbf{D}(\mathbf{p}) - \mathbf{p}\mathbf{p}^t) \geq m - 1$$

since $\text{rank}(\mathbf{D}(\mathbf{p})) = m$ and $\text{rank}(\mathbf{p}\mathbf{p}^t) = 1$, as follows from, as can be seen in [30], page 46, that $|\text{rank}(A) - \text{rank}(B)| \leq \text{rank}(A + B)$. In addition to this, $\mathbf{U}(\mathbf{p})\mathbf{1} = \mathbf{0}$ so $\text{rank}(\mathbf{U}(\mathbf{p})) \leq m - 1$. Thus, $\text{rank}(\mathbf{U}(\mathbf{p})) = m - 1$:

Matrix $\mathbf{U}(\mathbf{p})$ is a covariance matrix which, as can be seen in, [30], is positive semi-definite. There is therefore an orthogonal matrix $\mathbf{P}(\mathbf{p})$ and a diagonal matrix $\mathbf{D}(v)$ whose principal elements are the eigenvalues v_1, \dots, v_m of $\mathbf{U}(\mathbf{p})$ such that:

$$\mathbf{U}(\mathbf{p}) = \mathbf{P}(\mathbf{p})\mathbf{D}(v)\mathbf{P}(\mathbf{p})^t \tag{32}$$

Since $\text{rank}(\mathbf{U}(\mathbf{p})) = m - 1$, we may order its eigenvalues to have $v_j > 0, j = 1, \dots, m - 1$, and $v_m = 0$. With $\mathbf{D}(v)^{\frac{1}{2}}$, the diagonal matrix with principal elements, $v_1^{1/2}, \dots, v_m^{1/2}$ and:

$$\mathbf{U}(\mathbf{p})^{\frac{1}{2}} = \mathbf{P}(\mathbf{p})\mathbf{D}(v)^{\frac{1}{2}}\mathbf{P}(\mathbf{p})^t \tag{33}$$

we will have:

$$\mathbf{U}(\mathbf{p}) = \mathbf{U}(\mathbf{p})^{\frac{1}{2}}\mathbf{U}(\mathbf{p})^{\frac{1}{2}}. \tag{34}$$

We now establish:

Lemma 3. *If the $m \times m$ matrices $\mathbf{M}_1 \cdots \mathbf{M}_w$ are such that $\mathbf{M}_j^t \mathbf{M}_{j'} = \mathbf{0}_{m \times m}$, when $j \neq j'$, we have*

$$\text{rank}\left(\sum_{j=1}^w \mathbf{M}_j\right) = \sum_{j=1}^w \text{rank}(\mathbf{M}_j).$$

Proof. With $g_j = \text{rank}(\mathbf{M}_j)$ and $\mathbf{M}_j = [\mathbf{m}_{j,1}, \dots, \mathbf{m}_{j,m}]$, $j = 1, \dots, w$, there will be g_j linearly independent column vectors $\{\dot{\mathbf{m}}_{j,l}; l \in \mathcal{D}_j\}$ of \mathbf{M}_j , $j = 1, \dots, w$. The vectors in $\bigcup_{j=1}^w \{\dot{\mathbf{m}}_{j,l}; l \in \mathcal{D}_j\}$ will be linearly independent, since, when $j \neq j'$, they are orthogonal. Thus, $\text{rank}([\mathbf{M}_1 \cdots \mathbf{M}_w]) \geq \sum_{j=1}^w g_j = \sum_{j=1}^w \text{rank}(\mathbf{M}_j)$. Moreover, if we join another column vector of $[\mathbf{M}_1 \cdots \mathbf{M}_w]$, say $\mathbf{m}'_{j',l'}$, to the set $\bigcup_{j=1}^w \{\mathbf{m}_{j,l}; l \in \mathcal{D}_j\}$, it will linearly depend on the $\{\mathbf{m}'_{j',l'}; l' \in \mathcal{C}_{j'}\}$. Thus, the vectors in the extended set will not be linearly independent. Thus, $\text{rank}([\mathbf{M}_1 \cdots \mathbf{M}_w]) = \sum_{j=1}^w g_j = \sum_{j=1}^w \text{rank}(\mathbf{M}_j)$. □

Consider that $\sum_{j=1}^w \mathbf{M}_j = [\mathbf{M}_1 \cdots \mathbf{M}_w](\mathbf{1}_w \otimes \mathbf{I}_m)$, with \otimes indicating the Kronecker matrix product, as can be seen in [31], with $\text{rank}(\mathbf{1}_w \otimes \mathbf{I}_m) = \text{rank}(\mathbf{1}_w)\text{rank}(\mathbf{I}_m) = m$. Thus, as can be seen in [30]:

$$\begin{aligned} \text{rank}\left(\sum_{j=1}^w M_j\right) &\geq \text{rank}([M_1 \cdots M_w]) + \text{rank}(\mathbf{1}_w \otimes I_m) - m \\ &= \text{rank}([M_1, \cdots, M_w]) \\ &= \sum_{j=1}^w \text{rank}(M_j). \end{aligned}$$

and:

$$\text{rank}\left(\sum_{j=1}^w M_j\right) \leq \text{rank}([M_1, \cdots, M_w]) = \sum_{j=1}^w \text{rank}(M_j) \quad (35)$$

so:

$$\text{rank}\left(\sum_{j=1}^w M_j\right) = \sum_{j=1}^w \text{rank}(M_j), \quad (36)$$

as we wished to establish.

Let Q_1, \cdots, Q_w now be pairwise orthogonal orthogonal projection matrices (POOPM) with $Q_1 = \frac{1}{m} \mathbf{1}_m \mathbf{1}_m^t$. Now, $\mathbf{U}(\mathbf{p})$ is $m \times m$ with rank $m - 1$. Thus, its nullity space, $N(\mathbf{p})$ will have a dimension 1 and since $\alpha_1 = \frac{1}{\sqrt{m}} \mathbf{1}_m \in N(\mathbf{p})$, $\alpha_1 \alpha_1^t$ will be the orthogonal projection matrix on $N(\mathbf{p})$, since $\mathbf{U}(\mathbf{p})$ is symmetrical so its range space $R(\mathbf{p})$ will be the orthogonal complement $N(\mathbf{p})^\perp$ of $N(\mathbf{p})$. The orthogonal projection matrix $T(\mathbf{p})$ on $R(\mathbf{p})$ will then be:

$$T(\mathbf{p}) = I_m - Q_1. \quad (37)$$

Thus, if $\sum_{j=1}^w Q_j = I_m$, we will have $T(\mathbf{p}) = \sum_{j=2}^w Q_j$ as well as

$$\mathbf{U}(\mathbf{p}) = T(\mathbf{p})\mathbf{U}(\mathbf{p}) = \sum_{j=2}^w Q_j \mathbf{U}(\mathbf{p}) \quad (38)$$

and, according to Lemma 3:

$$m - 1 = \text{rank}(\mathbf{U}(\mathbf{p})) = \sum_{j=1}^w \text{rank}(Q_j \mathbf{U}(\mathbf{p})). \quad (39)$$

Now:

$$\text{rank}(Q_j \mathbf{U}(\mathbf{p})) \leq \text{rank}(Q_j), \quad j = 2, \cdots, w$$

and $m - 1 = \sum_{j=1}^w \text{rank}(Q_j)$. Thus, we must have:

$$\text{rank}(Q_j \mathbf{U}(\mathbf{p})) = \text{rank}(Q_j), \quad j = 2, \cdots, w \quad (40)$$

We now highlight that $\mathbf{U}(\mathbf{p})$ and $\mathbf{U}(\mathbf{p})^{\frac{1}{2}}$ have the same eigenvectors associated with positive eigenvalues. These eigenvectors constitute an orthonormal basis for $R(\mathbf{U}(\mathbf{p})) = R(\mathbf{U}(\mathbf{p})^{\frac{1}{2}})$. Thus:

$$T(\mathbf{p})\mathbf{U}(\mathbf{p})^{\frac{1}{2}} = \mathbf{U}(\mathbf{p})^{\frac{1}{2}} \quad (41)$$

and, reasoning as above, we obtain:

$$\text{rank}\left(Q_j \mathbf{U}(\mathbf{p})^{\frac{1}{2}}\right) = \text{rank}(Q_j), \quad j = 2, \cdots, w \quad (42)$$

Matrices Q_1, \cdots, Q_w naturally appear, as can be seen in [32], when there are factors that cross or groups of nested factors that cross. The sums of squares of effects and

interactions of these factors are the $\|A_j Y\|^2, j = 2, \dots, w$, and $\|A_1 Y\|^2$ can be associated with the general mean.

5. Inference

5.1. Chi-Square Tests

According to the PCMT, when:

$$H_0(A) : A p = 0 \tag{43}$$

holds, the limit distribution of:

$$L_n(A) = n(A \tilde{p}_n)^t (A U(\tilde{p}_n) A^t)^+ (A \tilde{p}_n) \tag{44}$$

will be that of χ_r^2 with $r = \text{rank}(A U(\tilde{p}_n) A^t)$, since when $H_0(A)$ holds:

$$\sqrt{n} A \tilde{p}_n \sim \mathcal{N}(0, A U(p_n) A^t) \tag{45}$$

and we also have $\tilde{p}_n \xrightarrow{p} p$. We thus have for $H_0(A)$ a q limit level test with statistic $L_n(A)$ and a critical value $x_{r,1-q}$, the $(1 - q)$ -th is a quantile of χ_r^2 .

Moreover, under any alternative to $H_0(A)$:

$$H_1(A) : A p = q \tag{46}$$

we have, whatever $K > 0$:

$$Pr(L_n(A) > K) \xrightarrow[n \rightarrow \infty]{} 1 \tag{47}$$

so the chi-square tests will be strongly consistent.

Let us assume that the probabilities p_1, \dots, p_m correspond to the treatments of a fixed effects model in which d factors, with h_1, \dots, h_d levels, cross (for instance, probabilities of cures for different treatments). We then have:

$$m = \prod_{l=1}^d h_l \tag{48}$$

which, as can be seen in [32], tests the effects and interactions of the factors. These effects and interactions correspond to subsets of $\bar{d} = \{1, \dots, d\}$. Thus, the null set, ϕ , will correspond to the general mean value, if the set has one element, it will be associated with the effects of the factor whose index belongs to the set. Otherwise, if the set has more than one element, it will be associated with the interaction between the levels of the factors with those indices. The sets can be ordered by the indexes:

$$j(\phi) = 1 + \sum_{l \in \phi} 2^{l-1} \tag{49}$$

Putting ϕ_j to indicate the j -th set, $j = 1, \dots, 2^d$, we have, as can be seen in [30], the matrices:

$$A(\phi_j) = \bigotimes_{j=1}^d A_l(\phi_j), \quad j = 1, \dots, 2^d \tag{50}$$

where:

$$\begin{cases} A_l(\phi_j) = \frac{1}{\sqrt{h_l}} \mathbf{1}_{h_l}^t, & l \notin \phi_j \quad j = 1, \dots, 2^d \\ A_l(\phi_j) = \mathbf{T}_{h_l}, & l \in \phi_j \quad j = 1, \dots, 2^d, \end{cases} \tag{51}$$

we then have, with $A_j = A(\varphi_j)$, $j = 1, \dots, 2^d$:

$$g_j = \text{rank}(A_j) = \prod_{j=\varphi_j} (h_j - 1) \quad j = 1, \dots, 2^d \tag{52}$$

Thus, for testing the:

$$H_{0,j} = H_0(A_j) : j = 1, \dots, 2^d \tag{53}$$

we have the statistic $L_n(A_j)$, with g_j degrees of freedom, $j = 1, \dots, 2^d$.

Another interesting case is that of cross-nesting factors. The factors in the h -th group have $a_{h,1}, \dots, a_{h,f_h}$ levels, $h = 1, \dots, d$. There are then $b_{h,v} = \prod_{v'=1}^v a_{h,v'}$, $v = 1, \dots, f_h$, combinations of levels of the first v factors in group h , and we also put $b_{h,0} = 1, h = 1, \dots, d$. Each of the combinations contains $c_{h,v} = \prod_{v'=v+1}^{f_h} a_{h,v'}$, $0 \leq v \leq f_h$ or $c_{h,f_h} = 1$, for combinations of levels of the remaining factors, and we have the matrices:

$$\begin{cases} A_{h,0} = \frac{1}{\sqrt{c_{h,0}}} \mathbf{1}_{c_{h,0}}^t, \quad h = 1, \dots, d \\ A_{h,j} = I_{b_{h,j-1}} \otimes T_{a_{h,j}} \otimes \frac{1}{\sqrt{c_{h,j}}} \mathbf{1}_{c_{h,j}}^t \quad j = 1, \dots, f_h, \quad h = 1, \dots, d, \end{cases} \tag{54}$$

where T_r is obtained by deleting the first row equal to $\frac{1}{\sqrt{r}} \mathbf{1}_r^t$ of a $r \times r$ orthogonal matrix and \otimes indicates the Kronecker matrix product. These matrices have ranks:

$$\begin{cases} g_{h,0} = 1, \quad h = 1, \dots, d \\ g_{h,j} = b_{h,j} - b_{h,j-1} \quad h = 1, \dots, d, \end{cases} \tag{55}$$

where $b_{h,0} = 1, h = 1, \dots, d$.

The effects and interactions in this cross-nesting are associated with the vectors $\mathbf{j} = (j_1, \dots, j_d)$ with $j_h = 0, \dots, f_h, h = 1, \dots, d$. So \mathbf{j} is associated with the matrices:

$$A_{\mathbf{j}} = \bigotimes_{h=1}^d A_{h,j_h}; \quad \mathbf{j} \in \Gamma \tag{56}$$

with ranks:

$$g_{\mathbf{j}} = \prod_{h=1}^d g_{h,j_h}; \quad \mathbf{j} \in \Gamma \tag{57}$$

and where $\Gamma = \{\mathbf{j}; 0 \leq j_h \leq f_h, h = 1, \dots, d\}$.

To test the hypothesis:

$$H_{0,\mathbf{j}} = H_0(A_{\mathbf{j}}); \mathbf{j} \in \Gamma \tag{58}$$

we have the statistic $L_n(A_{\mathbf{j}})$ with $g_{\mathbf{j}}$ degrees of freedom, $\mathbf{j} \in \Gamma$.

5.2. Confidence Regions

According to the continuity of the Moore–Penrose inverses, as can be seen in [30], pp. 221–224, we have:

$$\mathbf{U}(\tilde{\mathbf{p}}_n)^+ \xrightarrow{n \rightarrow \infty} \mathbf{U}(\mathbf{p})^+,$$

so the PCMT gives:

$$n(\mathbf{p} - \tilde{\mathbf{p}}_n)^t \mathbf{U}(\tilde{\mathbf{p}}_n)^+ (\mathbf{p} - \tilde{\mathbf{p}}_n) \sim \chi_{m-1}^2 \tag{59}$$

as well as:

$$n(\mathbf{A}_j \mathbf{p} - \mathbf{A}_j \tilde{\mathbf{p}}_n)^t (\mathbf{A}_j \mathbf{U}(\tilde{\mathbf{p}}_n) \mathbf{A}_j^t)^+ (\mathbf{A}_j \mathbf{p} - \mathbf{A}_j \tilde{\mathbf{p}}_n) \sim \chi_{g_j}^2, \quad j \in \Gamma \tag{60}$$

for models with factors crossing and cross-nesting.

Thus:

$$\left\{ \begin{aligned} &Pr \left[n(\mathbf{p} - \tilde{\mathbf{p}}_n)^t \mathbf{U}(\tilde{\mathbf{p}}_n)^+ (\mathbf{p} - \tilde{\mathbf{p}}_n) \leq x_{m-1,1-q} \right] \xrightarrow{n \rightarrow \infty} 1 - q \\ &Pr \left[n(\mathbf{A}_j \mathbf{p} - \mathbf{A}_j \tilde{\mathbf{p}}_n)^t (\mathbf{A}_j \mathbf{U}(\tilde{\mathbf{p}}_n) \mathbf{A}_j^t)^+ (\mathbf{A}_j \mathbf{p} - \mathbf{A}_j \tilde{\mathbf{p}}_n) \leq x_{r(\mathbf{A}_j),1-q} \right] \xrightarrow{n \rightarrow \infty} 1 - q, \quad j = 2, \dots, w \\ &Pr \left[n(\mathbf{A}_j \mathbf{p} - \mathbf{A}_j \tilde{\mathbf{p}}_n)^t (\mathbf{A}_j \mathbf{U}(\tilde{\mathbf{p}}_n) \mathbf{A}_j^t)^+ (\mathbf{A}_j \mathbf{p} - \mathbf{A}_j \tilde{\mathbf{p}}_n) \leq x_{r(\mathbf{A}_j),1-q} \right] \xrightarrow{n \rightarrow \infty} 1 - q, \quad j \in \Gamma \end{aligned} \right. \tag{61}$$

We can now apply Proposition 6 to obtain:

$$\left\{ \begin{aligned} &Pr \left[\bigcap_v \left(|v^t \mathbf{p} - v^t \tilde{\mathbf{p}}_n| \leq \sqrt{x_{n-1,1-q}(v^t \mathbf{U}(\tilde{\mathbf{p}}_n) v)} \right) \right] \xrightarrow{n \rightarrow \infty} 1 - q \\ &Pr \left[\bigcap_v \left(|v^t \mathbf{A}_j \mathbf{p} - v^t \mathbf{A}_j \tilde{\mathbf{p}}_n| \leq \sqrt{x_{g_j,1-q}(v^t \mathbf{A}_j \mathbf{U}(\tilde{\mathbf{p}}_n) \mathbf{A}_j^t v)} \right) \right] \xrightarrow{n \rightarrow \infty} 1 - q, \quad j = 2, \dots, w \\ &Pr \left[\bigcap_v \left(|v^t \mathbf{A}_j \mathbf{p} - v^t \mathbf{A}_j \tilde{\mathbf{p}}_n| \leq \sqrt{x_{g_j,1-q}(v^t \mathbf{A}_j \mathbf{U}(\tilde{\mathbf{p}}_n) \mathbf{A}_j^t v)} \right) \right] \xrightarrow{n \rightarrow \infty} 1 - q, \quad j \in \Gamma \end{aligned} \right. \tag{62}$$

5.3. Non-Linear Statistics

We assume that the component functions $g_l(\cdot)$, $l = 1, \dots, w$ of $G(\cdot)$ have continuous partial derivatives of the second order to apply Proposition 2 and its corollary in showing that:

$$\sqrt{n}(G(\tilde{\mathbf{p}}_n) - G(\mathbf{p})) \sim \mathcal{N}(\mathbf{0}, G(\mathbf{p}) \mathbf{U}(\mathbf{p}) G(\mathbf{p})^t). \tag{63}$$

An interesting application of this result is that to log-linear models, as can be seen in [12], in which we use:

$$\begin{cases} G(\tilde{\mathbf{p}}_n) = \mathbf{log}(\tilde{\mathbf{p}}_n) = (\log \tilde{p}_{n,1}, \dots, \log \tilde{p}_{n,m}) \\ G(\mathbf{p}) = \mathbf{log}(\mathbf{p}) = (\log p_1, \dots, \log p_m) \end{cases} \tag{64}$$

We now have:

$$G(\mathbf{p}) = \mathcal{D} \left(\frac{1}{p_1}, \dots, \frac{1}{p_m} \right) = \mathcal{D}(\mathbf{p})^{-1}, \tag{65}$$

so:

$$G(\mathbf{p}) \mathbf{U}(\mathbf{p}) G(\mathbf{p})^t = \mathcal{D}(\mathbf{p})^{-1} - \mathbf{1}_m \mathbf{1}_m^t = \mathbf{W}(\mathbf{p}), \tag{66}$$

since $\mathcal{D}(\mathbf{p})^{-1} \mathbf{p} = \mathbf{1}$.

Moreover, since $\mathcal{D}(\mathbf{p})^{-1}$ is invertible, taking:

$$\mathbf{U}^0(\mathbf{p}) = \mathcal{D}(\mathbf{p})^{-1} \mathbf{U}(\mathbf{p}) \mathcal{D}(\mathbf{p})^{-1} \tag{67}$$

we have:

$$\text{rank}(\mathbf{U}^0(\mathbf{p})) = \text{rank}(\mathbf{U}(\mathbf{p})) = m - 1. \tag{68}$$

Thus:

$$\mathbf{U}^0(\mathbf{p}) \left(\frac{\mathbf{1}}{\|\mathbf{p}\|} \mathbf{p} \right) = \mathbf{0} \tag{69}$$

so $\frac{1}{\|\mathbf{p}\|}\mathbf{p}$ belonging to the nullity space $N^0(\mathbf{p})$ of $\mathbf{U}^0(\mathbf{p})$ constitutes an orthonormal basis for that space. Since $\mathbf{U}^0(\mathbf{p})$ is symmetrical, its range space $R^0(\mathbf{p})$ will be $N^0(\mathbf{p})^\perp$, and the orthogonal projection matrix on $R^0(\mathbf{p})$ will be:

$$\mathbf{T}^0(\mathbf{p}) = \mathbf{I}_m - \frac{1}{\|\mathbf{p}\|}\mathbf{p}\mathbf{p}^t \quad (70)$$

Let \mathbf{A}^0 have row vectors that constitute an orthonormal basis for a sub-space ∇^0 . Putting:

$$\begin{cases} \mathbf{l}^0(\mathbf{p}) = \mathbf{log}(\mathbf{p}) \\ \mathbf{l}^0(\tilde{\mathbf{p}}_n) = \mathbf{log}(\tilde{\mathbf{p}}_n) \end{cases} \quad (71)$$

We have, according to the PCMT:

$$\sqrt{n}[\mathbf{A}\mathbf{l}^0(\tilde{\mathbf{p}}_n) - \mathbf{A}\mathbf{l}^0(\mathbf{p})] \sim \mathcal{N}(\mathbf{0}, \mathbf{A}\mathbf{U}^0(\mathbf{p})\mathbf{A}^t).$$

Thus, to test:

$$H_0(\mathbf{A}) : \mathbf{A}\mathbf{l}^0(\mathbf{p}) = \mathbf{0} \quad (72)$$

we have the limit q level chi-square test with the statistic $n\mathbf{l}^0(\tilde{\mathbf{p}}_n)^t(\mathbf{A}\mathbf{U}^0(\mathbf{p})\mathbf{A}^t)^+\mathbf{l}^0(\tilde{\mathbf{p}}_n)$ and the critical value $x_{r^0(\mathbf{A}),1-q}$ with $r^0(\mathbf{A}) = \text{rank}(\mathbf{A}\mathbf{U}^0(\mathbf{p})\mathbf{A}^t)$. These tests will be strongly consistent as follows from $\mathbf{l}^0(\tilde{\mathbf{p}}_n) \xrightarrow[n \rightarrow \infty]{p} \mathbf{l}^0(\mathbf{p})$.

Moreover, we have the limit level $1 - q$ confidence ellipsoids given by

$$\begin{cases} \Pr\left[\left(\mathbf{l}^0(\mathbf{p}) - \mathbf{l}^0(\tilde{\mathbf{p}}_n)\right)^t \mathbf{U}^0(\tilde{\mathbf{p}}_n)^+ \left(\mathbf{l}^0(\mathbf{p}) - \mathbf{l}^0(\tilde{\mathbf{p}}_n)\right) \leq x_{m-1,1-q}\right] \xrightarrow[n \rightarrow \infty]{} 1 - q \\ \Pr\left[\left(\mathbf{A}\mathbf{l}^0(\mathbf{p}) - \mathbf{A}\mathbf{l}^0(\tilde{\mathbf{p}}_n)\right)^t \mathbf{A}\mathbf{U}^0(\tilde{\mathbf{p}}_n\mathbf{A}^t)^+ \left(\mathbf{A}\mathbf{l}^0(\mathbf{p}) - \mathbf{A}\mathbf{l}^0(\tilde{\mathbf{p}}_n)\right) \leq x_{r^0(\mathbf{A}),1-q}\right] \xrightarrow[n \rightarrow \infty]{} 1 - q \end{cases} \quad (73)$$

We can now apply Proposition 6 to obtain:

$$\begin{cases} \Pr\left[\bigcap_v \left(|\mathbf{v}^t \mathbf{l}^0(\mathbf{p}) - \mathbf{v}^t \mathbf{l}^0(\tilde{\mathbf{p}}_n)| \leq \sqrt{x_{m-1,1-q}(\mathbf{v}^t \mathbf{U}^0(\tilde{\mathbf{p}}_n) \mathbf{v})}\right)\right] \xrightarrow[n \rightarrow \infty]{} 1 - q \\ \Pr\left[\bigcap_v \left(|\mathbf{v}^t \mathbf{A}\mathbf{l}^0(\mathbf{p}) - \mathbf{v}^t \mathbf{A}\mathbf{l}^0(\tilde{\mathbf{p}}_n)| \leq \sqrt{x_{r^0(\mathbf{A}),1-q}(\mathbf{v}^t \mathbf{A}\mathbf{U}^0(\tilde{\mathbf{p}}_n) \mathbf{A}^t \mathbf{v})}\right)\right] \xrightarrow[n \rightarrow \infty]{} 1 - q, \quad j = 2, \dots, w \end{cases} \quad (74)$$

5.4. Numerical Example

In this section, we apply our results, for non-linear statistic, to a dataset on coronary heart disease analysed by [12,33], with the log-linear model, a non-linear statistic. In all, a total of 1330 sick patients were categorized with respect to three variables, namely blood pressure, serum cholesterol and whether they had a coronary heart disease or not. The Blood pressure, which was the first variable had four categorical levels. The second variable, which was, Serum cholesterol, also had four categorical levels while the third variable, which indicated the presence of coronary heart disease, had two levels. So in all, the data had 32 classes with a total of 1330 sick patients. Kindly refer to Section 5.6 of [12] for details on the variables and their levels as well as a cross-classification of the frequencies for each category.

To proceed with the analysis by the application of our method, as can be seen in Equation (63), we firstly estimate the probabilities of each of the 32 classes. In order to apply the non-linear statistics, we calculated the logarithms of those estimates. As in Equation (63), the composite functions of the estimates were normally distributed with a null mean vector and a certain covariance matrix. We proceeded by evaluating our covariance matrix—firstly by determining the Jacobian matrix of the gradient of our composite function, as in Equation (65) and by Equation (66), and then determining the covariance matrix for our composite function.

To test the hypothesis of the absence of effects and interactions, we started by obtaining matrices A using orthogonal matrices.

The first orthogonal matrix we considered is:

$$P_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \quad (75)$$

Since we have 32 classes, we build the P matrices up to P_{32} with Kronecker matrix products:

$$P_4 = P_2 \otimes P_2$$

$$P_8 = P_2 \otimes P_4$$

$$P_{16} = P_2 \otimes P_8$$

$$P_{32} = P_2 \otimes P_{16}, \quad \text{where } \otimes \text{ is the Kronecker product.}$$

After getting the orthogonal matrix P_{32} , we determined our A matrices. Since we have three factors, we have eight A matrices defined as follows.

Let the set containing the factors indexes be $\varphi = \{1, 2, 3\}$, then the subsets of φ and the corresponding factor effects and interactions are:

- $j = 1, \quad \varphi = \emptyset$: overall or general mean effect
- $j = 2, \quad \varphi = \{1\}$: effects of the first factor
- $j = 3, \quad \varphi = \{2\}$: effects of the second factor
- $j = 4, \quad \varphi = \{1, 2\}$: interactions between the first and second factors
- $j = 5, \quad \varphi = \{3\}$: effects of the third factor
- $j = 6, \quad \varphi = \{1, 3\}$: interactions between the first and third factors
- $j = 7, \quad \varphi = \{2, 3\}$: interactions between the second and third factors
- $j = 8, \quad \varphi = \{1, 2, 3\}$: interactions between all factors

Now, the A_j , $j = 1, \dots, 8$ are obtained from P_{32} as follows: A_1 is the first row; A_2 is the second to fourth rows; A_3 is the fifth, ninth and thirteenth rows; A_4 is the sixth, seventh, eighth, tenth, eleventh, twelfth, fourteenth, fifteenth and sixteenth rows. Similarly, A_5 is the seventeenth row; A_6 is the eighteenth, nineteenth and twentieth rows; A_7 is the twenty-first, twenty-fifth and twenty-ninth rows; and A_8 is the twenty-second, twenty-third, twenty-fourth, twenty-sixth, twenty-seventh, twenty-eighth, thirtieth, thirty-first and thirty-second rows.

Now, according to Equation (44), we have the covariance matrices:

$$V_j(\mathbf{p}) = A_j^t W(\mathbf{p}) A_j, \quad j = 1, \dots, 8, \quad (76)$$

where $W(\mathbf{p})$ is defined in Equation (66), for the:

$$Z_j(\mathbf{p}) = A_j G(\mathbf{p}), \quad j = 1, \dots, 8, \quad (77)$$

where $G(\mathbf{p})$ is given by Equation (64). The sum of squares (SS) for the statistic $Z_j(\mathbf{p})$, $j = 1, \dots, 8$ are given by

$$SS_j = n Z_j(\mathbf{p})^t V_j(\mathbf{p})^+ Z_j(\mathbf{p}), \quad j = 1, \dots, 8 \quad (78)$$

where n is the total number of our observations, and $+$ indicates the Moore–Penrose inverses of the $V_j(\mathbf{p})$, $j = 1, \dots, 8$. The SS_j , $j = 1, \dots, 8$ are chi-squares with degrees of freedom the $rank(A_j)$, $j = 1, \dots, 8$.

Table 1 is an ANOVA-like table that presents the results of our analysis of the coronary heart disease data. This gives the sources of variation of the general mean, the main effects and the interaction effects. It also presents the degrees of freedom and the sum of squares for these effects. The significance level of these effects are indicated in *. From Table 1, we see that the general mean is highly significantly different from 0. Moreover, the factors blood pressure, serum cholesterol and coronary heart disease as well as both interactions between the first and second factors with the third factor were highly significant. The interactions in which the first and second factors did partake were not significant. We pointed out that we were able to consider the interaction between the three factors by using our approach. In the classical analysis of the data, as can be seen in [12] (Section 5.6), this would not be possible.

Table 1. ANOVA-like table for the coronary heart disease data.

Sources of Variation	Degree of Freedom	Sum of Squares
General mean, μ	1	4752.504 ***
1	3	30.289 ***
2	3	20.915 ***
1 \times 2	9	5.623
3	1	365.871 ***
1 \times 3	3	24.204 ***
2 \times 3	3	22.284 ***
1 \times 2 \times 3	9	4.568

*** indicates "significance" at 0.005 levels. Source: author's own calculations.

Again, we applied our method to analyse the General Social Survey (GSS 2008). The General Social Survey (GSS) is a survey which conducts basic scientific research on the structure and development of American society with a data-collection program designed to both monitor societal change within the United States and to compare the United States with other nations [34]. We considered three categorical variables, "Education", "Political party affiliation" and "Gender". The "Education" variable had 5 categorical levels, while "Political party affiliation" and "Gender" has 7 and 2 categorical levels respectively, as analyzed by [7] in Section 3.2.4.

Just as in Tables 1 and 2, an ANOVA-like table, presents the results of our analysis. It gives the sources of variation of the general mean, the main effects and the interaction effects for our GSS (2008) data. It also presents the degrees of freedom and the sum of square for these effects. The significance level of these effects is indicated in *. From the table, both factors 1 and 2 have significant effects as well as a significant interaction. The interactions between factors 2 and 3 is also significant. Factor 3 had neither significant effects nor interactions with the exception of its interaction with factor 2.

We can thus conclude that the core of significance was in factors 1 and 2 (education and political party affiliation) and that both genders behave similarly.

Table 2. ANOVA-like table for the GSS 2008 data.

Sources of Variation	Degree of Freedom	Sum of Squares
General mean, μ	1	49,387.730 ***
1	4	860.829 ***
2	6	84.692 ***
1 \times 2	24	89.636 ***
3	1	1.943
1 \times 3	4	1.624
2 \times 3	6	17.466 **
1 \times 2 \times 3	24	27.902 ***

*** and ** indicate "significance" at 0.005 and 0.01 levels respectively. Source: author's own calculations.

Comparing Procedures

Our procedure is based on the asymptotic distribution given by

$$\sqrt{n}(\tilde{\mathbf{p}}_n - \mathbf{p}) \sim \mathcal{N}(\mathbf{0}, \mathbf{U}(\mathbf{p})) \quad (79)$$

and by

$$\sqrt{n}(\mathbf{l}(\tilde{\mathbf{p}}_n) - \mathbf{l}(\tilde{\mathbf{p}})) \sim \mathcal{N}(\mathbf{0}, \mathbf{J}(\mathbf{p})\mathbf{U}(\mathbf{p})\mathbf{J}(\mathbf{p})^t) \quad (80)$$

where $\mathbf{J}(\cdot)$ is the Jacobian matrix of the gradients of $\mathbf{l}(\cdot)$ (the row vectors of $\mathbf{J}(\cdot)$ are the gradients of the components of $\mathbf{l}(\cdot)$):

This enabled us to, given an orthogonal partition:

$$\mathbf{R}^m = \boxplus_{j=1}^d \nabla_j, \quad (81)$$

where m is the number of probabilities, test the hypotheses:

$$H_{0,j} : \mathbf{l}(\mathbf{p}) \in \nabla_j^\perp. \quad (82)$$

In our numerical example, we have:

$\nabla(\emptyset)$: general mean

$\nabla(\{1\})$: null effect of the first factor

$\nabla(\{2\})$: null effect of the second factor

$\nabla(\{3\})$: null effect of the third factor

$\nabla(\{1,2\})$: null interaction of the first and second factors

$\nabla(\{1,3\})$: null interaction of the first and third factors

$\nabla(\{2,3\})$: null interaction of the second and third factors

$\nabla(\{1,2,3\})$: null interaction of the three factors

(83)

This way, we overcame the requirement of using hierarchical models.

6. Conclusions

Multinomial models having their limit distribution given by $\sqrt{n}(\tilde{\mathbf{p}}_n - \mathbf{p}) \sim \mathcal{N}(\mathbf{0}, \mathbf{U}(\mathbf{p}))$ with $\text{rank}(\mathbf{U}(\mathbf{p})) = m - 1$ and \mathbf{p} belonging to a compact set led us to establish the parametrized continuous mapping theorem (PCMT) to carry out inference for them. Namely, we obtained the confidence ellipsoids and confidence intervals. Moreover, we obtained chi-square tests for the hypothesis:

$$H_0(A) : \mathbf{A}\mathbf{p} = \mathbf{0}$$

which led to ANOVA-like inference for both multinomial and log-linear models. We pointed out that now hierarchical assumptions made on log-linear models are no longer necessary and all effects and interactions can be tested without any restrictions. Actually, this was used in the two numerical examples we presented. In addition to this, the replacement of F -tests by chi-square ones increases the power of our inference by replacing a finite number of degrees of freedom for the error by an infinity of them.

Author Contributions: This research article was written while preparing my Ph.D. thesis. Thus, I performed most of the efforts while interacting with my supervisors who are also the co-authors of this article. Conceptualization, I.A. and J.T.M.; Data curation, I.A.; Formal analysis, I.A. and

J.T.M.; Methodology, I.A., J.T.M. and F.J.M.; Project administration, F.J.M.; Software, I.A. and F.J.M.; Supervision, J.T.M. and F.J.M.; Validation, J.T.M.; Visualization, I.A. and F.J.M.; Writing—original draft, I.A.; Writing—review & editing, J.T.M. and F.J.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) through project UIDB/00297/2020.

Data Availability Statement: The data on coronary heart disease used in this article were obtained from [12,33].

Acknowledgments: I would like to especially acknowledge the much appreciated and extensive support from my supervisors, J.T.P.N. Mexia and Filipe Marques for their supervision, resources and in-depth contributions to my Ph.D. I am very grateful to them.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Bishop, Y.M.; Fienberg, S.E.; Holl, P.W. *Discrete Multivariate Analysis: Theory and Practice*; Springer Science & Business Media: New York, NY, USA, 2007.
- Agresti, A. *Categorical Data Analysis*, 3rd ed.; Wiley & Sons: Hoboken, NJ, USA, 2013.
- Agresti, A. *Analysis of Ordinal Categorical Data*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2010.
- Tang, W.; He, H.; Tu, X.M. *Applied Categorical and Count Data Analysis*; CRC Press: Boca Raton, FL, USA, 2012.
- Fagerl, M.W.; Lydersen, S.; Laake, P. *Statistical Analysis of Contingency Tables*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2017.
- Lloyd, C.J. *Statistical Analysis of Categorical Data*; Wiley: New York, NY, USA, 1999.
- Kateri, M. *Contingency Table Analysis*, 1st ed.; Methods and Implementation Using R; Editorial Advisory Board: Aachen, Germany, 2014.
- Kateri, M.; Balakrishnan, N. Statistical evidence in contingency tables analysis. *J. Stat. Plan. Inference* **2008**, *138*, 873–887. [[CrossRef](#)]
- Birch, M.W. Maximum likelihood in three-way contingency tables. *J. R. Stat. Soc. Ser. (Methodol.)* **1963**, *25*, 220–233. [[CrossRef](#)]
- Goodman, L.A. Interactions in multidimensional contingency tables. *Ann. Math. Stat.* **1964**, *35*, 632–646. [[CrossRef](#)]
- Fienberg, S.E.; Rinaldo, A. Three centuries of categorical data analysis: Log-linear models and maximum likelihood estimation. *J. Stat. Plan. Inference* **2007**, *137*, 3430–3445. [[CrossRef](#)]
- Everitt, B.S. *The Analysis of Contingency Tables*; Chapman and Hall/CRC: New York, NY, USA, 2019; pp. 94–99.
- Andersen, E.B. *Introduction to the Statistical Analysis of Categorical Data*; Springer Science & Business Media: New York, NY, USA, 2012.
- Tutz, G. *Regression for Categorical Data*; Cambridge University Press: Cambridge, UK, 2011.
- Fienberg, S.E.; Rinaldo, A. Maximum likelihood estimation in log-linear models. *Ann. Stat.* **2012**, *40*, 996–1023. [[CrossRef](#)]
- Irel, C.T.; Kullback, S. Minimum discrimination information estimation. *Biometrics* **1968**, *24*, 707–713.
- Grizzle, J.E.; Starmer, C.F.; Koch, G.G. Analysis of categorical data by linear models. *Biometrics* **1969**, *25*, 489–504. [[CrossRef](#)]
- Fienberg, S.E. When did Bayesian inference become “Bayesian”? *Bayesian Anal. Int. Soc. Bayesian Anal.* **2006**, *1*, 1–40. [[CrossRef](#)]
- Akoto, I.; Mexia, J.T.; Guerreiro, G.R. Discriminant analysis in discrete models: An application to HIV Treatment status. *Symmetry* **2021**, submitted.
- Rohatgi, V.K.; Saleh, A.K.M. *An Introduction to Probability and Statistics*; John Wiley & Sons: Hoboken, NJ, USA, 2015; pp. 189–192.
- Taboga, M. *Lectures on Probability Theory and Mathematical Statistics*; CreateSpace Independent Publishing Platform: North Charleston, SC, USA, 2012; pp. 431–438.
- Wilks, S.S. *Mathematical Statistics*, 2nd ed.; John Wiley & Sons: New York, NY, USA, 1962; p. 262.
- Kallenberg, O. *Foundations of Modern Probability*, 2nd ed.; Springer: New York, NY, USA, 2010; p. 76.
- Rao, J.N.K. On two simple schemes of unequal probability sampling without replacement. *J. Indian Stat. Assoc.* **1965**, *3*, 80–173.
- Mukhopadhyay, P. *Complex Surveys: Analysis of Categorical Data*; Springer: New York, NY, USA, 2016; pp. 223–240.
- DasGupta, A. *Probability for Statistics and Machine Learning: Fundamentals and Advanced Topics*; Springer Science & Business Media: New York, NY, USA, 2011; pp. 268–282.
- Van der Vaart, A.W. *Asymptotic Statistics*; Cambridge University Press: Cambridge, UK, 2000; Volume 3.
- Loeve, M.M. *Probability Theory I*; Springer: Berlin/Heidelberg, Germany; New York, NY, USA, 1977; p. 151.
- Scheffé, H. *The Analysis of Variance*; John Wiley & Sons: New York, NY, USA, 1959; pp. 406–411.
- Schott, J.R. *Matrix Analysis for Statistics*, 3rd ed.; John Wiley & Sons: New York, NY, USA, 2016; p. 46.
- Silvey, S.D. *Statistical Inference*; Reprinted; Chapman & Hall: New York, NY, USA, 1975.
- Fonseca, M.; Mexia, J.; Zmysłony, R. Estimators and tests for variance components in cross nested orthogonal designs. *Discuss. Math. Probab. Stat.* **2003**, *23*, 175–201.
- Ku, H.H.; Kullback, S. Log-linear models in contingency table analysis. *Am. Stat.* **1974**, *28*, 115–122.
- National Opinion Research Center. *General Social Survey*; Inter-University Consortium for Political and Social Research: Ann Arbor, MI, USA, 2016. [[CrossRef](#)]