

Article

Consistency of Decision in Finite and Numerable Multinomial Models

Isaac Akoto ^{1,2,*}  and João T. Mexia ^{1,†} 

¹ Center of Mathematics and Its Applications, NOVA School of Science and Technology, Universidade NOVA de Lisboa, Campus de Caparica, 2829-516 Caparica, Portugal

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

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

† These authors contributed equally to this work.

Abstract: The multinomial distribution is often used in modeling categorical data because it describes the probability of a random observation being assigned to one of several mutually exclusive categories. Given a finite or numerable multinomial model $\mathcal{M}(n, \mathbf{p})$ whose decision is indexed by a parameter θ and having a cost $c(\theta, \mathbf{p})$ depending on θ and on \mathbf{p} , we show that, under general conditions, the probability of taking the least cost decision tends to 1 when n tends to ∞ , i.e., we showed that the cost decision is consistent, representing a Statistical Decision Theory approach to the concept of consistency, which is not much considered in the literature. Thus, under these conditions, we have consistency in the decision making. The key result is that the estimator $\tilde{\mathbf{p}}_n$ with components $\tilde{p}_{n,i} = \frac{n_i}{n}$, $i = 1, \dots$, where n_i is the number of times we obtain the i th result when we have a sample of size n , is a consistent estimator of \mathbf{p} . This result holds both for finite and numerable models. By this result, we were able to incorporate a more general form for consistency for the cost function of a multinomial model.

Keywords: stochastic convergence; decision theory; estimators

MSC: 62F12; 62C05; 60F05



Citation: Akoto, I.; Mexia, J.T.

Consistency of Decision in Finite and Numerable Multinomial Models.

Mathematics **2023**, *11*, 2434.

<https://doi.org/10.3390/math11112434>

Academic Editor: Velizar Pavlov

Received: 14 April 2023

Revised: 20 May 2023

Accepted: 22 May 2023

Published: 24 May 2023



Copyright: © 2023 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

Classical Statistical Inference (cSI) is centered on minimizing the probabilities of errors, but in Statistical Decision Theory (SDT), minimizing the decision costs is the goal; see for instance [1,2]. In minimizing the probability of errors in cSI, or the decision costs in SDT, a desired property is to have our procedure to be consistent, that is, we would like the mass of the distribution of the sequence of the sample for the procedure to converge to the population constant as the sample size gets larger, see [3,4]. As far as we have searched prior to the writing of this manuscript, not many studies have been conducted on the consistency of decision costs for SDT as compared to that of the error probabilities in the case of cSI.

Multinomial distribution is often used in modeling categorical data because it describes the probability of a random observation being assigned to one of several mutually exclusive categories. Thus, having n independent realizations of experiments with a finite or numerable set of incompatible results with probabilities p_i , $i = 1, \dots, \infty$, the probabilities of obtaining $N_i = n_i$ times the i th result, $i = 1, \dots$ follows the multinomial distribution denoted as $\mathcal{M}(n, \mathbf{p})$ with parameters n and \mathbf{p} . The probability density function of this distribution is

$$\mathbb{P} \left[\bigcap_{i=1}^{\infty} (N_i = n_i) \mid \mathbf{p} \right] = \frac{n!}{\prod_{i=1}^{\infty} n_i!} \prod_{i=1}^{\infty} p_i^{n_i} \quad (1)$$

where $0^0 = 1$, $\mathbf{p} = (p_1, \dots, p_i, \dots)$ and $n = \sum_{i=1}^{\infty} n_i$, see [5,6]. To avoid repetitions, we give the expression for the numerable case since the particularization to the finite set is direct.

In what follows, we take a double approach in the treatment of Multinomial Models, both classical and decision theory approaches. In our classical approach, we show that $\tilde{\mathbf{p}} = (\tilde{p}_1, \dots, \tilde{p}_i, \dots)$, with $\tilde{p}_{n,i} = \frac{n_i}{n}$, $i = 1, \dots$, is a consistent estimator of $\mathbf{p} = (p_1, \dots, p_i, \dots)$. This result will play a central part in our paper. We point out that we use a finite sample to obtain a numerable family $\{\tilde{p}_{n,i}, i = 1, \dots\}$ of jointly consistent estimators. We thus have consistent results in the fold of classical statistical inference.

Now considering a decision problem, let there be a family of possible decisions. For each of these decisions, we have a cost that depends on the results of the experiment. These results will have probabilities p_i , $i = 1, \dots$. We thus have for the i th result the costs $c_i(d)$, $d \in \mathcal{D}$, $i = 1, \dots$. The average cost for decision $d \in \mathcal{D}$ will be

$$c.(d) = \sum_{i=1}^{\infty} p_i c_i(d), \quad d \in \mathcal{D}. \tag{2}$$

Assuming the $c_i(d)$, $d \in \mathcal{D}$, $i = 1, \dots$ is known, we can use the estimators

$$\tilde{p}_{n,i} = \frac{n_i}{n}, \quad i = 1, \dots, \tag{3}$$

where n_i , $i = 1, \dots$ is the number of times that in n independent realizations of the experiment, we get the i th result with cost $c_i(d)$, $d \in \mathcal{D}$, $i = 1, \dots$. We have the estimators for the costs as

$$\tilde{c}_{n.}(d) = \sum_{i=1}^{\infty} \tilde{p}_{n,i} c_i(d), \quad d \in \mathcal{D}.$$

We will show that $\tilde{p}_{n,i}$, $i = 1, \dots$, are jointly consistent even when we have a numerable set of possible results, thus the $\tilde{c}_{n.}(d)$ will also be consistent, $d \in \mathcal{D}$.

If there is a decision d° with least average cost and \tilde{d}_n° is the one with the least estimated average cost where there are n realizations of the experiment, we will show that

$$Pr(\tilde{d}_n^\circ = d) \xrightarrow{n \rightarrow \infty} 1$$

so that, see [7], we will have consistency in decision taking into the setup of experiments with finite or numerable set of incompatible results.

In the subsequent sections, we consider the multinomial model and its estimator in Section 2, where by limits distributions we show that the estimators of the multinomial model are consistent. In Section 3, we develop the cost function by statistical decision theory for multinomial models and again show that this function has the property of consistency. A further extension for the cost function is presented in Section 4.

2. Multinomial Models and Estimators

In this section, we obtain and consider estimators for the multinomial model. By limit distributions, see [8], we show that the estimators are consistent.

In \mathcal{X} , the space of vectors $\mathbf{v} = (v_1, \dots, v_i, \dots)$ with numerable sets of components such that

$$\|\mathbf{v}\|_1 = \sum_{i=1}^{\infty} |v_i| < +\infty. \tag{4}$$

We can consider $\|\mathbf{v}\|_1$ as a norm, see [9]. The sub-space Γ of \mathcal{X} constituted by the vectors \mathbf{p} with non-negative components that add up to 1 will be bounded since $\|\mathbf{p}\|_1 = 1$. Given $\mathbf{p}_1, \mathbf{p}_2 \in \Gamma$, we have

$$\|\mathbf{p}_1 - \mathbf{p}_2\|_1 \leq 2, \tag{5}$$

and if

$$\| \mathbf{p}_n - \mathbf{v} \|_1 \xrightarrow{n \rightarrow \infty} 0 \tag{6}$$

we have $\mathbf{v} \in \Gamma$ since the components of \mathbf{v} will be non-negative and add up to 1. Thus, Γ is compact since it is bounded and closed.

Let us put

$$\begin{cases} p(m) = \sum_{i=1}^m p_i \\ p^c(m) = 1 - p(m) \end{cases} \tag{7}$$

as well as

$$\mathbf{p}(m) = (p_1, \dots, p_m + p^c(m), 0, \dots) \tag{8}$$

in order to get

$$\begin{aligned} \| \mathbf{p} - \mathbf{p}(m) \|_1 &= p^c(m) + \sum_{i=m+1}^{\infty} p_i \\ &= 2p^c(m). \end{aligned} \tag{9}$$

Besides this, we have the vector

$$\tilde{\mathbf{p}}_n = (\tilde{p}_{n,1}, \dots, \tilde{p}_{n,m}, \dots) \tag{10}$$

whose components are the estimators $\tilde{p}_{n,i} = \frac{n_i}{n}$, $i = 1, \dots$. Let us put

$$\begin{cases} \tilde{p}_n(m) = \sum_{i=1}^m \tilde{p}_{n,i} \\ \tilde{p}_n^c(m) = 1 - \tilde{p}_n(m) \end{cases} \tag{11}$$

as well as

$$\tilde{\mathbf{p}}_n(m) = (\tilde{p}_{n,1}, \dots, \tilde{p}_{n,m} + \tilde{p}_n^c(m), 0, \dots) \tag{12}$$

so that

$$\| \tilde{\mathbf{p}}_n - \tilde{\mathbf{p}}_n(m) \|_1 = 2\tilde{p}_n^c(m). \tag{13}$$

With

$$\tilde{m}_n = \min \{ h : 0 = \tilde{p}_{n,h+1} = \tilde{p}_{n,h+2} = \dots \}, \tag{14}$$

for $m \geq \tilde{m}_n$, we get

$$\| (\tilde{\mathbf{p}}_n - \tilde{\mathbf{p}}_n(m)) - (\mathbf{p} - \mathbf{p}(m)) \|_1 = \| \mathbf{p} - \mathbf{p}(m) \|_1 = 2p^c(m) \tag{15}$$

since $\tilde{\mathbf{p}}_n = \tilde{\mathbf{p}}_n(m)$ when $m \geq \tilde{m}_n$. We also get

$$\| \tilde{\mathbf{p}}_n - \tilde{\mathbf{p}}_n(m) \|_1 \leq 2 \| \mathbf{p} - \mathbf{p}(m) \|_1 \leq 4p^c(m). \tag{16}$$

Now, by representing stochastic convergence by $\xrightarrow[n \rightarrow \infty]{s}$, we establish Proposition 1:

Proposition 1. *The estimator $\tilde{\mathbf{p}}_n$ is said to be a consistent estimator for \mathbf{p} since*

$$\tilde{\mathbf{p}}_n \xrightarrow[n \rightarrow \infty]{s} \mathbf{p}, \tag{17}$$

that is according to the Weak Law of Large numbers, see [10,11].

Proof. Taking

$$m(\varepsilon) = \min\{m : p^c(m) \leq \varepsilon\}, \quad \forall \varepsilon > 0. \tag{18}$$

so that for $m > m(\varepsilon)$, we have

$$\mathbb{P}(p_n^c(m) \leq 2\varepsilon) \xrightarrow[n \rightarrow \infty]{s} 1, \tag{19}$$

as well as

$$\mathbb{P}\left(\|\tilde{\mathbf{p}}_n - \tilde{\mathbf{p}}_n(m)\|_1 \leq 4p^c(m)\right) \xrightarrow[n \rightarrow \infty]{s} 1, \tag{20}$$

since

$$\|\tilde{\mathbf{p}}_n - \tilde{\mathbf{p}}_n(m)\|_1 \leq 2p^c(m). \tag{21}$$

Now

$$\|\tilde{\mathbf{p}}_n(m) - \mathbf{p}(m)\|_1 \xrightarrow[n \rightarrow \infty]{s} 0, \tag{22}$$

since see [7,8,12],

$$\sqrt{n}(\tilde{\mathbf{p}}_n(m) - \mathbf{p}(m)) \xrightarrow{D} \mathcal{N}(\mathbf{0}, U(\mathbf{p}(m))) \tag{23}$$

where \xrightarrow{D} indicates convergence in distribution in this case to the normal distribution with null mean vector and covariance matrix

$$U(\mathbf{p}(m)) = \mathcal{D}(\mathbf{p}(m)) - \mathbf{p}(m)\mathbf{p}(m)^t, \tag{24}$$

where $\mathcal{D}(\mathbf{p}(m))$ is the diagonal matrix whose principal elements are the components of $\mathbf{p}(m)$. Thus

$$\mathbb{P}\left(\|\tilde{\mathbf{p}}_n - \tilde{\mathbf{p}}_n(m)\|_1 \leq \varepsilon\right) \xrightarrow[n \rightarrow \infty]{s} 1, \tag{25}$$

and so

$$\mathbb{P}\left[\left(\|\tilde{\mathbf{p}}_n - \tilde{\mathbf{p}}_n(m)\|_1 \leq 4\varepsilon\right) \cap \left(\|\tilde{\mathbf{p}}_n(m) - \mathbf{p}(m)\|_1 \leq \varepsilon\right)\right] \xrightarrow[n \rightarrow \infty]{s} 1, \tag{26}$$

as well as

$$\mathbb{P}\left(\|\tilde{\mathbf{p}}_n - \mathbf{p}(m)\|_1 \leq 5\varepsilon\right) \xrightarrow[n \rightarrow \infty]{s} 1, \tag{27}$$

whenever $m > m(\varepsilon)$.

Now, we also have

$$\|\mathbf{p} - \mathbf{p}(m)\|_1 = 2p_n^c \leq 2\varepsilon, \tag{28}$$

if $m > m(\varepsilon)$, then

$$\mathbb{P}\left(\|\tilde{\mathbf{p}}_n - \mathbf{p}\|_1 \leq 7\varepsilon\right) \xrightarrow[n \rightarrow \infty]{s} 1, \tag{29}$$

which establishes the thesis. \square

Corollary 2. If $w(\mathbf{p})$ is a continuous function of \mathbf{p} , we have

$$w(\tilde{\mathbf{p}}_n) \xrightarrow[n \rightarrow \infty]{s} w(\mathbf{p}). \tag{30}$$

The thesis follows from Proposition 1 and the Slutsky theorem, see [13,14].

Corollary 3. With $\mathbf{d}_h(\mathbf{v})$ the vector of the indexes of the h largest components of $\mathbf{v}_n \in \mathcal{X}$, we have see [10]

$$\mathbf{d}_h(\tilde{\mathbf{p}}_n) \xrightarrow[n \rightarrow \infty]{s} \mathbf{d}_h(\mathbf{p}) \tag{31}$$

if the $h + 1$ largest components of \mathbf{p} are distinct.

Proof. When the $h + 1$ largest components of \mathbf{p} are distinct, \mathbf{p} will be a continuity point of $\mathbf{d}_h(\cdot)$ and the thesis follows from Corollary 3. \square

The results in this section belong to the study of consistency in classical Statistical Inference (cSI). These classical inferences are, in most instances, made without regard to the use to which they are to be put. In the next section, we go into consistency for Statistical Decision Theory.

3. Cost Function for Multinomial Models

In Statistical Decision Theory (SDT), the goal is to incorporate more than just sample data in order to arrive at the optimal decision, unlike cSI. The knowledge of the possible consequences of a decision is much incorporated and this knowledge is quantified as the cost incurred for each possible decision that is taken. According to [15], Abraham Wald was the first person to thoroughly examine the inclusion of a cost function in statistical analysis.

The cost function represents the costs associated with taking a particular decision. It is a function that maps every possible decision and outcome to a real-valued cost. The cost function is used to evaluate the performance of various decision rules in terms of their expected cost. The goal of statistical decision theory is to identify the decision rule that minimizes the expected cost, see [16].

Now, we go back to the decision problem we presented in the Introduction and consider the cost function for multinomial models.

Let $c(\mathbf{p}, \tilde{\mathbf{p}}_n, d)$, $d \in \mathcal{D}$, be the cost for decision $d \in \mathcal{D}$, where \mathbf{p} is the vector of probabilities and $\tilde{\mathbf{p}}_n$ are the estimated probabilities of the n results. We will assume that this cost is the sum of two components, both non-negative, $c_0(\mathbf{p}, d)$, that in a given decision $d \in \mathcal{D}$, depends only on \mathbf{p} , the vector of probabilities, and $c_1(\mathbf{p}, \tilde{\mathbf{p}}_n, d)$ that depends on the estimation errors. Namely, we take

$$c(\mathbf{p}, \tilde{\mathbf{p}}_n, d) = c_0(\mathbf{p}, d) + k_d \left(\|\tilde{\mathbf{p}}_n - \mathbf{p}\|_1 \right), \quad d \in \mathcal{D} \tag{32}$$

with $k_d > 0$, $d \in \mathcal{D}$, so

$$c_0(\mathbf{p}, d) = \min\{c(\mathbf{p}, \tilde{\mathbf{p}}_n, d)\}, \quad d \in \mathcal{D}$$

since, as we saw

$$\tilde{\mathbf{p}}_n \xrightarrow[n \rightarrow \infty]{s} \mathbf{p}$$

we will have

$$c(\mathbf{p}, \tilde{\mathbf{p}}_n, d) \xrightarrow[n \rightarrow \infty]{s} c_0(\mathbf{p}, d), \quad d \in \mathcal{D}. \tag{33}$$

Thus, for every $d \in \mathcal{D}$, the limit cost will be $c_0(\mathbf{p}, d)$. It is now easy to see that if there is $d(\mathbf{p})$ such that

$$c_0(\mathbf{p}, d^0(\mathbf{p})) \leq c_0(\mathbf{p}, d), \quad d \in \mathcal{D},$$

with $\tilde{d}_n^0(\mathbf{p})$ as the decision with the least estimated cost. We have

$$Pr(\tilde{d}_n^0(\mathbf{p}) = d^0(\mathbf{p})) \xrightarrow[n \rightarrow \infty]{s} 1$$

Proposition 4. We have consistency for the cost function given by Equation (32) whenever Equation (33) holds.

If the $h + 1$ largest components of \mathbf{p} are obtained, as an alternative to Equation (32), we may take

$$c(\mathbf{p}, \tilde{\mathbf{p}}_n) = c_0(\mathbf{p}) + c_1 \left(\|\mathbf{d}_h(\tilde{\mathbf{p}}_n) - \mathbf{d}_h(\mathbf{p})\|_1 \right), \tag{34}$$

reobtaining Proposition 4, since

$$\left\| \mathbf{d}_h(\tilde{\mathbf{p}}_n) - \mathbf{d}_h(\mathbf{p}) \right\|_1 \xrightarrow[n \rightarrow \infty]{s} 0 \tag{35}$$

and so we continue to have

$$c(\mathbf{p}, \tilde{\mathbf{p}}_n) \xrightarrow[n \rightarrow \infty]{s} c_0(\mathbf{p}). \tag{36}$$

We may also take

$$c_1(\tilde{\mathbf{p}}_n - \mathbf{p}) = (\tilde{\mathbf{p}}_n - \mathbf{p})^t \mathbf{M}(\tilde{\mathbf{p}}_n - \mathbf{p}) \tag{37}$$

with \mathbf{M} a positive definite matrix, see [8,9], or

$$c_1(\tilde{\mathbf{p}}_n - \mathbf{p}) = (\mathbf{d}_h(\tilde{\mathbf{p}}_n) - \mathbf{d}_h(\mathbf{p}))^t \mathbf{M}(\mathbf{d}_h(\tilde{\mathbf{p}}_n) - \mathbf{d}_h(\mathbf{p})). \tag{38}$$

Thus, there is a wide range of possible cost functions, namely, with $\mathbf{g}(\cdot)$, a continuous function

$$\tilde{\boldsymbol{\gamma}}_n = \mathbf{g}(\tilde{\mathbf{p}}_n) \tag{39}$$

will, according to the Slutsky theorem, be a consistent estimator of $\boldsymbol{\gamma} = \mathbf{g}(\mathbf{p})$. Moreover, if we have a cost function

$$c(\boldsymbol{\gamma}, \tilde{\boldsymbol{\gamma}}_n) = c_0(\boldsymbol{\gamma}) + c_1(\|\tilde{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}\|_1), \tag{40}$$

where $c_0(\boldsymbol{\gamma})$ is the cost that depends on $\boldsymbol{\gamma}$ and $c_1(\cdot)$ is continuous and such that $c_1(0) = 0$, we can use again the Slutsky theorem, to get

$$c(\boldsymbol{\gamma}, \tilde{\boldsymbol{\gamma}}_n) \xrightarrow[n \rightarrow \infty]{s} c_0(\boldsymbol{\gamma}). \tag{41}$$

We thus extended our previous results on \mathbf{p} to any parameter given by a continuous function of \mathbf{p} , such as the cost function.

4. Extension on Cost Functions

In this section, we develop an extension for the previous Section 3 to incorporate a more general form of our results on consistency for the cost function for multinomial models.

For instance, we consider

$$\gamma(\varphi) = \sum_{j \in \varphi} p_j \tag{42}$$

the sum of the probabilities of the results with indexes in φ . These results may be of interest and so we are led to consider their joint probability. A direct extension of this case is given by

$$\gamma(\varphi_1, \dots, \varphi_k) = \sum_{l=1}^k a_l \sum_{j \in \varphi_l} p_j \tag{43}$$

where we consider k sets of results. The coefficients a_1, \dots, a_k value the relevances of the corresponding sets of results.

In general, let us have a succession $\{\mathbf{Y}_n\}$ of observation vectors whose distributions depend on a parameter $\boldsymbol{\theta}$ for which we have a consistent estimator $\tilde{\boldsymbol{\theta}}_n$. Then, if we have a cost function

$$c(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}_n) = c_0(\boldsymbol{\theta}) + c_1(\|\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\|_1) \tag{44}$$

with $c_0(\boldsymbol{\theta})$ a continuous function of $\boldsymbol{\theta}$ and $c_1(\cdot)$ also continuous, and such that

$$\min\{c_1(\|z\|_1)\} = c_1(0) = 0, \tag{45}$$

we have

$$c(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}_n) \xrightarrow[n \rightarrow \infty]{s} c_0(\boldsymbol{\theta}) \tag{46}$$

which implies consistency for the cost functions. Thus, the two consistent features display the relation we had already found for multinomial models. Namely, we obtain the following result:

Proposition 5. *If we have a consistent estimator $\tilde{\theta}_n$ for a parameter θ , we have consistency for cost functions*

$$c(\theta, \tilde{\theta}_n) = c_0(\theta) + c_1(\|\tilde{\theta}_n - \theta\|_1)$$

where $c_1(\cdot)$ is continuous with minimum $c_1(0)$.

The extension behind this proposition, and getting consistent estimators for a numerable set of parameters, the components of \mathbf{p} , from a finite sample are maybe the most interesting features of our discussion.

5. Final Remark

In this study, based on limit distribution, by considering the vector of probabilities for the multinomial model, we showed, using classical Statistical Inference, that the estimators for the vector of probabilities are consistent. Due to the limitation of classical Statistical Inference but not incorporating the knowledge of the possible consequences of a decision, we used a Statistical Decision Theory approach to quantify the cost incurred for each possible decision by obtaining a cost function for the vector of probabilities. We showed that the estimators of cost function are consistent.

Our results on having consistency for the estimator of probabilities leads to consistency of decision function; in this, we hope to have opened an interesting line of work on multinomial and other models using Statistical Decision Theory.

Author Contributions: Conceptualization, I.A. and J.T.M.; methodology, I.A. and J.T.M.; software, I.A.; formal analysis, J.T.M.; writing—original draft, I.A.; writing—review & editing, I.A. and J.T.M.; supervision, J.T.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: All relevant data are within the manuscript.

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

References

1. Le Cam, L. *Asymptotic Methods in Statistical Decision Theory*, 2nd ed.; Springer: New York, NY, USA, 2012.
2. Liese, F.; Miescke, K.-J. *Statistical Decision Theory: Estimation, Testing, and Selection*; Springer Science and Business Media: New York, NY, USA, 2008; pp. 1–52.
3. Berger, C.; Casella, G.T. *Statistical Inference*, 2nd ed.; Cengage Learning: Duxbury, CA, USA, 2005; pp. 232–233.
4. Hogg, R.V.; McKean, J.W.; Craig, A.T. *Introduction to Mathematical Statistics*, 6th ed.; Pearson Education, Inc.: Upper Saddle River, NJ, USA, 2005; pp. 204–205.
5. Rohatgi, V.K.; Saleh, A.K.M.E. *An Introduction to Probability and Statistics*, 3rd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2015; pp. 189–191.
6. Evans, M.; Hastings, N.; Peacock, B.; Forbes, C. *Statistical Distributions*, 4th ed.; John Wiley & Sons: Hoboken, NJ, USA, 2011; pp. 135–136.
7. Wilks, S.S. *Mathematical Statistics*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 1962; p. 261.
8. Akoto, I.; Mexia, J.T.; Marques, F.J. Asymptotic results for multinomial models. *Symmetry* **2021**, *13*, 2173. [[CrossRef](#)]
9. Schott, R.J. *Matrix Analysis for Statistics*, 3rd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2016; p. 12.
10. Akoto, I. Asymptotic Treatment for Multinomial Models and Applications. Ph.D. Thesis, NOVA University Lisbon, School of Science and Technology, Caparica, Portugal, 2022.
11. Resnick, S. *A Probability Path*, 2005th ed.; Springer Science and Business Media: New York, NY, USA, 2019; pp. 204–208.
12. Van der Vaart, A.W. *Asymptotic Statistics*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2000; pp. 25–34.
13. Kallenberg, O. *Foundations of Modern Probability*, 2nd ed.; Springer: Jersey City, NJ, USA, 1997; p. 571.
14. Slutsky, E.E. Qualche proposizione relative alla teoria delle funzioni aleatorie. *Giorn. Dell'Istituto Ital. Degli Attuari* **1937**, *8*, 183–199.

15. Wald, A. Statistical decision functions. *Ann. Math. Stat.* **1949**, *20*, 165–205. [[CrossRef](#)]
16. Berger, J.O. *Statistical Decision Theory and Bayesian Analysis*, 2nd ed.; Springer Science and Business Media: New York, NY, USA, 1985; pp. 1–2.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.