Journal of the Nigerian Association of Mathematical Physics
***Volume 31*, (July, *2015*), pp 391 – 398**
© J. of NAMP

Application of Multi-Set to log linear models for arbitrary d-Dimensional Contingency Table and its Associated Closed-Form Formula for Maximum Likelihood Estimations

[1]O.C. Okoli, [2]N. A. Nsiegbe and [3]E.I Ezenekwe

[1,2,3]Department of Mathematics, P.M.B 02 Uli, Chukwuemeka Odumegwu Ojukwu University, Anambra State, Nigeria.

Abstract

The purpose of this research paper, is to give a classical (combinatorial) proof for the closed-form formula that evaluate the Maximum likelihood parameter estimators for certain arbitrary d -dimensional multinomial function (likelihood function) induced by an arbitrary d-dimensional Multi-index.

Keywords: Permutations, Combinations, Multi-Set, Multi-Index, Probability, contingency table.
Mathematics Subject Classification: 05A10.

1.0 Introduction

Enumerative combinatorics provides one of the basic fundamentals for discrete probability theory [1-9]. If E is subset of a finite set X ($E \subset X$) such that one chooses an element of$X$at random, the probability that the element chosen actually belong to E is given by $n(E) / n(X)$ **.** Thus, the determination of the cardinalities of these sets underlies this probability model (the so-called uniform model). However, beyond this uniform model, enumerative combinatorics have proved to be efficient whereby permeating various forms of discrete probability theory andparametric evaluation which we intend to demonstrate. Let$X = \{x_i : i = 1,2, \ldots, k\}$ $and E \subset X$, then the usual probability (p_1) of successfully selecting (picking) an element of E is given by $p_1 = n(E) / n(X)$, and then define $p_2 = 1 - p_1$ for the failure. If this selection is repeated for n-number of times, then the probability that an element of Ewill be selected exactly $j$times in n-number of trials is given by the binomial function defined by $C_j^n p_1^j p_2^{n-j}$ for two possible outcomes and k-nomial function as $C_{j_1,j_2,\ldots,j_k}^n p_1^{j_1} p_2^{j_2} \ldots p_k^{j_k}$ for k-possible outcomes. The purpose of this research, is to consider the sets $X^{(d)}$and $M(\lambda, X^{(d)})$that is more general than the set Xsuch that the index (running) variable ($\bar{\imath}_d$)=($i_1, i_2, \ldots, i_d$) is not necessarily a point, but rather a vector, where $i_r \epsilon [k_r]$, $k_r \epsilon \mathbb{N}$, $r \epsilon [d]$. Let

$$X^{(d)} = \{ x_{\bar{\imath}_d} : i_r \epsilon [k_r], k_r \epsilon \mathbb{N}, r \epsilon [d] \}$$

and then define

$$M(\lambda, X^{(d)}) = \{ {x_{\bar{\imath}_d}}^{\alpha_{\bar{\imath}_d}} : \alpha_{\bar{\imath}_d} \epsilon \alpha^{(d)} \}$$

to be the Multiset induced by$X^{(d)}$ due to the function $\lambda: X^{(d)} \to \mathbb{N}$such that $\lambda(x_{\bar{\imath}_d}) = \alpha_{\bar{\imath}_d}$. Where$\alpha^{(d)}$ is a multi-index. We then give a classical proof of the associated closed-form formula for the maximum likelihood estimator.

1.1 Multiset and Multinomial

Definition1.1.1 [10-11]: A finite multiset $M(\lambda, X)$ ($or M$) on a set X is a function $\lambda: X \to \mathbb{N}$such that

$$\sum \lambda(x) < \infty$$

If $\lambda(x) = n \forall x \in X$, then M is called an n-multiset, hence we write $n(M) = n$. Suppose $X = \{x_i : i = 1,2, \ldots, k\}$ and $\lambda: X \to \mathbb{N}$such that $\lambda(x_i) = \alpha_i$, we shall have $M = \{x_i^{\alpha_i} : i = 1,2, \ldots, k\}$, where α_i is called the multiplicity of $x_i (in M)$ and $(\alpha_1, \alpha_2, \ldots, \alpha_k)$ is called the (associated) multi-index (or weak composition), which is also a row matrix (vector). For simplicity wewrite$\alpha = (\alpha_{1i}, \alpha_{2i}, \ldots, \alpha_k)$.We quickly remark that the function $\lambda: X^{(d)} \to \mathbb{N}$ is the so-called "random variable" as often used by statisticians. To see this, given any finite Multiset M, then there exist $\lambda: X^{(d)} \to \mathbb{N}$ such that $\lambda(x_i) = \alpha_i$($i = 1,2, \ldots, k$) and $\sum \lambda(x_i) = n$ since M is finite. If we let$\lambda(x_i) = X_i$, then$X_i$is a random variable that count the occurences of outcome $x_i$in$X$ (i.e. $X_i = \alpha_i$; $i = 1,2, \ldots, k$).

Corresponding author: O.C. Okoli, E-mail: odicomatics@yahoo.com, Tel.: +234-8036941434

1.2 One Dimensional Multinomial Distribution

If for each n independent trials, there are k possible outcomes $x_1, x_2, \ldots, x_k$, with the corresponding probabilities $p_1, p_2, \ldots, p_k$ ($\sum p_i = 1$), and ifX_i (random variable) records the number of occurrances of x_i in these n trials, hence for every one-dimensional multi-index $(\alpha_{1i}, \alpha_{2i}, \ldots, \alpha_k)ofn$, then

$$p(X_i = \alpha_i;\ i = 1,2,\ldots,k) = C^n_{\alpha_1,\alpha_2,\ \ldots,\alpha_k} p_1^{\alpha_1} p_2^{\alpha_2} \ldots p_k^{\alpha_k} \qquad (1.1)$$

is the underlying probability mass function (or simply pmf) which is well known. In the sequel we shall represent the above description on$X$ that defines the$pmf$by $M(p, \alpha, X)$ ($or M(p, \lambda, X)$)and then give a formal proof for the pmf for completeness purpose. We now proceed to define certain concept and notations which will serve as a building block in this paper.

Definition1.2.1 [12-14] By Multi-index, we mean a k-tuple vector (a row matrix or a column matrix)α_i, where each (α_i: $[k] = \{1,2,\ldots,k\}$) is a non-negative interger. We define

The associated integer $|\alpha_i|$by

$$|\alpha| = \sum_{i=1}^{k} \alpha_i \ (1.2)$$

The associated monomial x^α by

$$x^\alpha = \prod_{i=1}^{k} x_i^{\alpha_i} \ (1.3)$$

The associated factorial $\alpha!$ by

$$\alpha! = \prod_{i=1}^{k} \alpha_i\,! \ (1.4)$$

Let $X = \{x_1, x_2, \ldots, x_k\}$ be a distinct finite set of points. If we associate to each element$x_i \in X$ with the numberα_i in αthen certainly there exist a non-empty set $M(\alpha, X)$ induced by a non-negative integer $\lambda: X \to \mathbb{N}$such that x_i has multiplicity $\alpha_i$in$M(\alpha, X) or$ ($M(\lambda, X)$), which is define by

$$M(\lambda, X) = \{x_1^{\alpha_1}, x_2^{\alpha_2}, \ldots, x_k^{\alpha_k}\}(1.5)$$

Is the multiset associated with X with respect to the non-negative integer function on X. Now consider the expansion of$(x_1 + x_2 + \cdots + x_k)^n$, observe that if $k = 2{,}3$ then we have the binomial, trinomial expansion respectively. For arbitrary but fixed positive integer k the expansion of $(\sum_{i=1}^{k} x_i)^n$is a multinomial expansion of X in one running (index) variable i, which can be referred to as one category or class of data. Observe that each $x_i$has certain number of repeatition or multiplicity in the expansion of $(\sum_{i=1}^{k} x_i)^n$. There is no loss of generality if we assume that the multiplicity of x_i in the expansion of $(\sum_{i=1}^{k} x_i)^n$isα_i; $i = 1,2,\ldots,k$provided $\sum \alpha_i = n$. Thus, this will certainly induce a multiset representation due to the multinomial expansion, as such we have $\{x_1^{\alpha_1}, x_2^{\alpha_2}, \ldots, x_k^{\alpha_k}\}$as in (1.5). Furthermore, observe that each term (string) in this multinomial (k-nomial) expansion can be given in the general form.

$$C(\alpha_1, \alpha_2,\ \ldots, \alpha_k) x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_k^{\alpha_k} \qquad (1.6)$$

Where $C(\alpha_1, \alpha_2,\ \ldots, \alpha_k)$is the associated k-nomial coefficient for each term. The following lemma gives the actual formular for $C(\alpha_1, \alpha_2,\ \ldots, \alpha_k)$ and corresponding probability mass function.

Now, we extend our description above to two-dimensional multiset and its associated two-dimensional multi-index$\alpha^{(2)} = (\alpha_{ij})$, by considering the expansion $\left(\sum_{i=1}^{k} . \sum_{j=1}^{m} x_i\right)^n$ where the multiplicity α_{ij} ($i \in [k],\ j \in [m],\ k, m \in \mathbb{N}$) for each term $x_{ij} \in X^{(2)}$($i \in [k],\ j \in [m],\ k, m \in \mathbb{N}$) induces a $k \times m$ array (vector) where each α_{ij} ($i \in [k],\ j \in [m],\ k, m \in \mathbb{N}$) is a non-negative integer. Hence for the vector

$$\alpha^{(2)} = \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1m} \\ \vdots & \ddots & \vdots \\ \alpha_{k1} & \cdots & \alpha_{km} \end{pmatrix}, or \alpha^{(2)} = (\alpha_{ij}): i \in [k], \qquad j \in [m]$$

We define

The associated integer $|\alpha^{(2)}|$ by

$$\alpha^{(2)} = \sum_{i=1}^{k} . \sum_{j=1}^{m} \alpha_{ij}\ ; where \left|\alpha_i^{(2)}\right| = \sum_{j=1}^{m} \alpha_{ij}\ ;\ i \in [k] \quad (1.7)$$

The associated monomial x^α by

$$x^{\alpha^{(2)}} = \prod_{i=1}^{k} . \prod_{j=1}^{m} x_{ij}^{\alpha_{ij}}\,; \text{where } x_i^{\alpha^{(2)}} = \prod_{j=1}^{m} x_{ij}^{\alpha_{ij}} \quad (1.8)$$

The associated factorial $\alpha!$ by

$$\alpha^{(2)}! = \prod_{i=1}^{k}.\prod_{j=1}^{m} \alpha_{ij}!; \text{ where } \alpha_i^{(2)}! = \prod_{j=1}^{m} x_{ij}^{\alpha_{ij}}! \quad (1.9)$$

Lemma1.2.1

Let $M_{(2)}(p, \alpha^{(2)}, X^{(2)})$denote a description on a finite multiset with multiplicity $\alpha_{ij}$and probability $p_{ij}$for each $\alpha_{ij} \in X^{(2)}$, then the probability that $\alpha_{ij} \in X^{(2)}$is selected exactly α_{ij}times ($i \in [k]$, $j \in [m]$, $k, m \in \mathbb{N}$) in n-trials is;

$$\mathrm{P}\big(X_{ij} = \alpha_{ij} : i \in [k], j \in [m]\big) = \frac{\mathrm{n}!}{\prod_{i=1}^{k}.\prod_{j=1}^{m} \alpha_{ij}} \prod_{i=1}^{k}.\prod_{j=1}^{m} p_{ij}^{\alpha_{ij}}$$

Proof

Some how, our choice of selection is a little beat coupled, since for each fixed $i \in [k]$ we have to make selection from the set$\{x_{ij} : j \in [m]\}$. Let $|\alpha_i|$ be the number of times x_{ij}could be repeatedly be selected for every fixed $i \in [k]$ and let α_{ij} be the number of times x_{ij}could be repeatedly be selected given that the ith row has been selected for each$j \in [m]$. Thus, for each$i \in [k]$, the number of ways of selecting the $1^{st}, 2^{nd}, \dots, m^{th}$ element in $X^{(2)}$constituting $(\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{im})$ is given by

$$\binom{|\alpha^{(2)}| - \sum_{r=1}^{i-1} |\alpha_r^{(2)}|}{|\alpha_i^{(2)}|}\binom{|\alpha_i^{(2)}|}{\alpha_{i1}}\binom{|\alpha_i^{(2)}| - \alpha_{i1}}{\alpha_{i2}}\cdots\binom{|\alpha_i^{(2)}| - \sum_{j=1}^{m-1} \alpha_{ij}}{\alpha_{im}} \forall i \in [k]$$

$$= \binom{|\alpha^{(2)}| - \sum_{r=1}^{i-1} |\alpha_r^{(2)}|}{|\alpha_i^{(2)}|}\prod_{j=1}^{m}\binom{|\alpha_i^{(2)}| - \sum_{r=1}^{j-1} \alpha_{ir}}{\alpha_{im}} \forall i \in [k]$$

Thus, as i ranges over the set$[k]$, we have altogether

$$\prod_{i=1}^{k}\binom{|\alpha^{(2)}| - \sum_{r=1}^{i-1} |\alpha_r^{(2)}|}{|\alpha_i^{(2)}|}\prod_{j=1}^{m}\binom{|\alpha_i^{(2)}| - \sum_{r=1}^{j-1} \alpha_{ir}}{\alpha_{im}}$$

$$= \frac{\big(\sum_{i=1}^{k}.\sum_{j=1}^{m} \alpha_{ij}\big)!}{\prod_{i=1}^{k}.\prod_{j=1}^{m} \alpha_{ij}!} = \frac{\mathrm{n}!}{\prod_{i=1}^{k}.\prod_{j=1}^{m} \alpha_{ij}!}$$

Thus, consequently, we have

$$\mathrm{P}\big(X_{ij} = \alpha_{ij} : i \in [k], j \in [m]\big) = \frac{\mathrm{n}!}{\prod_{i=1}^{k}.\prod_{j=1}^{m} \alpha_{ij}} \prod_{i=1}^{k}.\prod_{j=1}^{m} p_{ij}^{\alpha_{ij}}$$

It is important to remark that in lemma 1.2.1, given the set$X^{(2)} = \{x_{ij} : i \in [k], j \in [m]\}$and $\lambda: X^{(2)} \to \mathbb{N}$**,** then that there exist a two-dimensional finite multiset $M_{(2)}\big(\alpha^{(2)}, X^{(2)}\big) = \{x_{ij}^{\alpha_{ij}} : i \in [k], j \in [m]\}$ on $X^{(2)}$ with it corresponding multi-index $\alpha^{(2)} = (\alpha_{ij}) i \in [k], j \in [m]$. By lemma 1.2.1, we wish to generalise the result for an arbitrary d-dimensional finite multiset $M_{(d)}\big(\alpha^{(d)}, X^{(d)}\big)$ with the corresponding d-dimensional multi-index $\alpha^{(d)} = \alpha_{i_1, i_2, \dots, i_d}$, by considering the expansion $\big(\sum_{i_1=1}^{k_1} \dots \sum_{i_d=1}^{k_d} x_{i_1, i_2, \dots, i_d}\big)^n$ where the multiplicity $\alpha_{i_1, i_2, \dots, i_d}$ $(i_r \epsilon [k_r], k_r \epsilon \mathbb{N}, r \epsilon [d])$ for each term $x_{i_1, i_2, \dots, i_d} \in X^{(d)} (i_r \epsilon [k_r], k_r \epsilon \mathbb{N}, r \epsilon [d])$ induces a$k_1 \times k_2 \times \cdots \times k_d$vector (array) where each$\alpha_{i_1, i_2, \dots, i_d}$ $(i_r \epsilon [k_r], k_r \epsilon \mathbb{N}, r \epsilon [d])$ is a non-negative integer. Similarly for the vector $\alpha^{(d)}$, we define

The associated integer $|\alpha^{(2)}|$ by

$$\alpha^{(d)} = \sum_{i_1=1}^{k_1} \dots \sum_{i_d=1}^{k_d} \alpha_{i_1, i_2, \dots, i_d} \quad (1.10)$$

$$where \ |\alpha_{i_1, i_2, \dots, i_u}^{(d)}| = \sum_{i_{u+1}=1}^{k_{u+1}} \dots \sum_{i_d=1}^{k_d} \alpha_{i_1, i_2, \dots, i_u i_{u+1}, \dots, i_d} ; 1 \le u < d$$

The associated monomial x^{α} by

$$x^{\alpha^{(d)}} = \prod_{i_1=1}^{k_1} \dots \prod_{j_d=1}^{k_d} x_{i_1, i_2, \dots, i_d}^{\alpha_{i_1, i_2, \dots, i_d}} \quad (1.11)$$

$$\text{where } x_{i_1, i_2, \dots, i_u}^{\alpha^{(2)}} = \prod_{i_{u+1}=1}^{k_{u+1}} \dots \prod_{i_d=1}^{k_d} x_{i_1, i_2, \dots, i_u i_{u+1}, \dots, i_d}^{\alpha_{i_1, i_2, \dots, i_u i_{u+1}, \dots, i_d}} ; 1 \le u < d$$

The associated factorial $\alpha^{(d)}!$ by

$$\alpha^{(d)}! = \prod_{i_1=1}^{k_1} \dots \prod_{j_d=1}^{k_d} \alpha_{i_1,i_2,\dots,i_d}! = |\alpha_{\bar{\imath}_d}|! \quad (1.12)$$

Where

$$\alpha^{(d)}_{i_1,i_2,\dots,i_u}! = \prod_{i_{u+1}=1}^{k_{u+1}} \dots \prod_{j_d=1}^{k_d} \alpha_{i_1,i_2,\dots,i_u i_{u+1},\dots,i_d}! \; and \; |\alpha_{\bar{\imath}_u}|! = |\alpha_{i_{u-1}1}|!\,|\alpha_{i_{u-1}2}|! \cdots |\alpha_{i_{u-1}k_u}|!$$

Lemma 1.2.2

For arbitrary but fixed $d \in \mathbb{N}$ let $r \in [\,d\,]$ and $k_r \in \mathbb{N}$, , let $i_r \in [k_r]$, then the following holds;

$$\prod_{r=1}^{d} \cdot \left(\prod_{i_r=1}^{k_r} \begin{pmatrix} |\alpha_{\bar{\imath}_{r-1}}| - \sum_{j_r=1}^{i-1} |\alpha_{\bar{\imath}_{r-1} j_r}| \\ |\alpha_{\bar{\imath}_{r-1} i_r}| \end{pmatrix} \right) = \frac{n!}{\alpha^{(d)}!}$$

Where $|\alpha_{\bar{\imath}_{r-1}}| - \sum_{j_r=1}^{i-1} |\alpha_{\bar{\imath}_{r-1} j_r}| = 0 \;\forall\, r \in [\,d\,]$

Lemma 1.2.3

Let $M_{(d)}(p\,, \alpha^{(d)}, X^{(d)})$ denote a description on d-dimensional finite multiset $M_{(d)}(\alpha^{(d)}, X^{(d)})$ with corresponding multiplicity $\alpha_{i_1,i_2,\dots,i_d}$and probability $p_{i_1,i_2,\dots,i_d}$ for each $x_{i_1,i_2,\dots,i_d} \in X^{(d)}$then the probability that $x_{i_1,i_2,\dots,i_d} \in X^{(d)}$is selected exactly$\alpha_{i_1,i_2,\dots,i_d}$times $(i_r \in [k_r], k_r \in \mathbb{N}, r \in [\,d\,])inn$-trials is;

$$\mathrm{P}(X_{i_1,i_2,\dots,i_d} = \alpha_{i_1,i_2,\dots,i_d} : i_r \in [k_r], k_r \in \mathbb{N}, r \in [\,d\,]) = \frac{\mathrm{n}!}{\prod_{i_1=1}^{k_1} \dots \prod_{j_d=1}^{k_d} \alpha_{i_1,i_2,\dots,i_d}} \prod_{i_1=1}^{k_1} \dots \prod_{i_d=1}^{k_d} p_{i_1,i_2,\dots,i_d}^{\alpha_{i_1,i_2,\dots,i_d}}$$

For the next lemmas, we give the preliminary introduction as follows, for arbitrary but fixed $d \in \mathbb{N}, let\ [d] = \{1,2,\dots,d\}$ denote the set of d categorical variables. Let $\alpha^{(d)}$ be a d-dimensional array; that is $k_1 \times k_2 \times \dots \times k_d$ contingency table with cell counts (frequencies) $\alpha_{i_1,i_2,\dots,i_d}$. For any $r \in \mathbb{N}$ such that $1 < r < d$, then the subset variable $[r] \subset [d]$ generate $\alpha^{(d)}r$-dimensional sub array with cell counts (frequencies) $\alpha_{i_1,i_2,\dots,i_r} = \alpha_{\bar{\imath}_r}$ $(r \le d)$. Let $\mu_{[r](\bar{\imath}_r)}$ denote the interaction among the variables in the index subset $[r]$ of the r-dimensional sub table of $\alpha^{(d)}$ that correspond to $\bar{\imath}_r$ cell. we shall assume that $[r] = \emptyset$ if $r = 0$, so that we define $\mu_{[\emptyset](\bar{\imath}_\emptyset)} = \mu$.

Definition 1.2.2([15])

A loglinear model is said to be hierarchical if for every $r \in \mathbb{N}$ such that $1 < r < d$ $([r] \subset [d])$ for which $\mu_{[r](\bar{\imath}_r)} = 0$, then we have $\mu_{[s](\bar{\imath}_s)} = 0$ for all $s \ge r([r] \subset [s])$

Furthermore, let$S^c_{d,r}$be the set of strings of r-combinations (in increasing order) of elements of $[d]$ and $P([d]) = 2^{[d]}$denote the power set of $[d]$. Thus, for any $r \in \mathbb{N}$ such that $1 < r < d$ $([r] \subset [d])$ we define

$$P([d]: 0 \le n(\{\,\bar{\jmath}\}) \le r) := \{\{\bar{\jmath}\} \in P([d]) : 0 \le n(\{\bar{\jmath}\}) \le r\} \qquad (1.13)$$

Where $\bar{\jmath} \in S^c_{d,r}$for $r = 0,1,2,\dots,d$ $(d \le r)$with $\{\bar{\jmath}\} = \emptyset$ if $r = 0$ and $n(\{\bar{\jmath}\})$ denote the lenght of the string $\bar{\jmath} \in S^c_{d,r}$ or the cardinality of the set $\{\bar{\jmath}\} \in P([d])$. Observe that $P([d]: 0 \le n(\{\,\bar{\jmath}\}) \le r)$ is simply a subclass of $P([d])$, however, $P([d]: 0 \le n(\{\,\bar{\jmath}\}) \le r)$ is equal to the power set $P([d])$ if $d = r$. Observed that by this, is easy to see that

$$P([d]: 0 \le n(\{\bar{\jmath}\}) \le d) := \bigcup_{r=0}^{d} P([d]: n(\{\bar{\jmath}\}) = r) = \bigcup_{r=0}^{d} \{\bar{\jmath}\}_r \qquad (1.14)$$

From (1.3), notice that $P([d]: n(\{\bar{\jmath}\}) = r)$ is structurally equal to the (set) collections of elements of $S^c_{d,r}$. As a consequence of above concept and definitions, we shall rather replace the notation $\mu_{[s](\bar{\imath}_s)}$ by $\mu_{\{\bar{\jmath}\}(i_{\bar{\jmath}})}$ such that $n(\{\bar{\jmath}\}) = r$. However, these notations could be used interchangeable if need be in the course of this work, also if$n(\{\bar{\jmath}\}) = r$, then $\{\bar{\jmath}\} \in \{\bar{\jmath}\}_r$. The follwing lemma shall be useful in the sequel.

Lemma 1.2.4

For arbitrary but fixed $d \in \mathbb{N}$ and let $\mu_{\{\bar{\jmath}\}(i_{\bar{\jmath}})}$ be as define above such that $\bar{\jmath} \in S^c_{d,r}(\{\bar{\jmath}\} \in P([d]: n(\{\bar{\jmath}\}) = p)$ then$\{\bar{\jmath}\} \in \mathrm{P}([d]: 0 \le n(\{\bar{\jmath}\}) \le d)$

$$\sum_{q=0}^{d} \cdot \sum_{1 \le j_1 < j_2 < \dots < j_u \le d} \mu_{\{\bar{\jmath}\}(i_{\bar{\jmath}})} = \sum_{\{\bar{\jmath}\} \in \mathrm{P}([d]: 0 \le n(\{\bar{\jmath}\}) \le d)} \mu_{\{\bar{\jmath}\}(i_{\bar{\jmath}})}$$

Lemma 1.2.5
Let $\alpha^{(d)}$ be a d-dimensional $k_1 \times k_2 \times \cdots \times k_d$ contingency table with cell counts (frequencies)$\alpha_{i_1,i_2,\ldots,i_d}(i_r \in [k_r], k_r \in \mathbb{N}, r \in [d])$, then the saturated loglinear model for the d-way table of variables indexed in $[d]$ is given by

$$\eta_{\bar{\jmath}_d} = \sum_{\{\bar{\jmath}\}\in P([d]:0\le n(\{\bar{\jmath}\})\le d)} \mu_{\{\bar{\jmath}\}(i_{\bar{\jmath}})}$$

Where $\log_e \mu_{\bar{\jmath}_d} = \eta_{\bar{\jmath}_d}$

1.3 Maximum Likelihood Function

Now, Observe that the probability mass function (pmf) associated with$M_{(d)}(p, \alpha^{(d)}, X^{(d)})$ in lemma 1.2.2 is a product-multinomial distribution function. Hence, we wish to determine in a close form, the maximum likelihood estimator for the log-likehood function, which is a function of the parameter $\mu_{\{\bar{\jmath}\}(i_{\bar{\jmath}})}$.

Definition1.3.1 [16]
For a given set of points $\{x_i : i = 1,2,\ldots,n\}$, the function $\mathcal{L}$ defined by

$$\mathcal{L}(\theta|\cdot) = L(\theta_1,\ldots,\theta_k|x_1,\ldots,x_k) = \prod_{i=1}^{n} f(\theta_1,\ldots,\theta_k|x_i) \qquad (1.15)$$

Is called the likelihood function for the distribution function $f(\theta_1,\ldots,\theta_k|x_i)$,where $x_1,\ldots,x_k$ are the independent observations that are identically distributed. The maximum likelihood estimators $\widehat{\theta_i}$are the values of the parameters $\theta_1,\ldots,\theta_k$ that optimizes $\mathcal{L}(\overline{\theta})$or equivalently, $\log \mathcal{L}(\overline{\theta})$.

Let

$$\varphi(r) = \binom{|\alpha_{\bar{\imath}_{r-1}}| - \sum_{j_r=1}^{i-1}|\alpha_{\bar{\jmath}_r}|}{|\alpha_{\bar{\imath}_r}|} \qquad (1.16)$$

For $r \in [d]$ define

$$f(\cdot|x_{i_r}) := \prod_{i_r=1}^{k_r} \varphi(r)\, p_{i_r}^{x_{i_r}}$$

Then $f(\cdot|x_{i_r})$ is a multinomial distribution for the independent observation $x_{i_r},\ldots,x_{i_d}$ which are identically distributed. The function $\mathcal{L}$ defined by

$$\mathcal{L}(\cdot|x_{i_1,i_2,\ldots,i_r}) = \prod_{i=1}^{d} f(\cdot|x_{i_r})$$

is a likelihood function for the multinomial distribution$f(\cdot|x_{i_r})$.
For every $r \in [d]$, if we let$\lambda(x_{i_r}) = X_{i_r}$, then X_{i_r} is a random variable that count the occurences of event $x_{i_r} \in X$ (i.e.$x_{i_r}; i = 1,\ldots,k$), then

$$\mathcal{L}(\cdot|x_{i_1,i_2,\ldots,i_r}) = \prod_{i=1}^{d} f(\cdot|x_{i_r})$$

$$\prod_{r=1}^{d}\cdot\left(\prod_{i_r=1}^{k_r}\binom{|\alpha_{\bar{\imath}_{r-1}}| - \sum_{j_r=1}^{i-1}|\alpha_{\bar{\jmath}_r}|}{|\alpha_{\bar{\imath}_r}|} p_{i_r}^{x_{i_r}}\right) =$$

$$\prod_{i_1=1}^{k_1}\binom{|\alpha| - \sum_{j_1=1}^{i_1-1}|\alpha_{j_1}|}{|\alpha_{i_1}|}\prod_{i_2=1}^{k_2}\binom{|\alpha_{i_1}| - \sum_{j_2=1}^{i_2-1}|\alpha_{i_1j_2}|}{|\alpha_{i_1i_2}|}$$

$$\cdots$$

$$\prod_{i_u=1}^{k_u}\binom{|\alpha_{\bar{\imath}_{u-1}}| - \sum_{j_u=1}^{i_u-1}|\alpha_{\bar{\imath}_{u-1}j_u}|}{|\alpha_{\bar{\imath}_{u-1}i_u}|}\prod_{i_{u+1}=1}^{k_{u+1}}\binom{|\alpha_{\bar{\imath}_u}| - \sum_{j_{u+1}=1}^{i_{u+1}-1}|\alpha_{\bar{\imath}_uj_{u+1}}|}{|\alpha_{\bar{\imath}_ui_{u+1}}|}$$

$$\cdots$$

$$\prod_{i_{d-1}=1}^{k_{d-1}}\binom{|\alpha_{\bar{\imath}_{d-2}}|-\sum_{j_{d-1}=1}^{i_{d-1}-1}|\alpha_{\bar{\imath}_{d-2}j_{d-1}}|}{|\alpha_{\bar{\imath}_{d-2}i_{d-1}}|}\prod_{i_d=1}^{k_d}\binom{|\alpha_{\bar{\imath}_{d-1}}|-\sum_{j_d=1}^{i_d-1}|\alpha_{\bar{\imath}_{d-1}j_d}|}{|\alpha_{\bar{\imath}_{d-1}i_d}|}\left(\prod_{i_1=1}^{k_1}\cdots\prod_{i_d=1}^{k_d}p_{\bar{\imath}_d}^{\alpha_{\bar{\imath}_d}}\right)$$
$$= \mathrm{P}\left(X_{i_1,i_2,\dots,i_d}=\alpha_{i_1,i_2,\dots,i_d} : i_r \in [k_r], k_r \in \mathbb{N}, r \in [\,d\,]\right)$$

We shall state the result that follows in the next theorem and then give a formal prove of it.

2.0 Main Results

Theorem 2.1

Let $\mathcal{L}(\cdot\,|x_{i_1,i_2,\dots,i_r})$ be the product-multinomial probability distribution function associated with the d-dimensional contingency table, $\alpha^{(d)}$ in$M_{(d)}(p\,,\alpha^{(d)},\,X^{(d)})$, then the maximum likeihood estimators is given by

$$\hat{\mu}_{\{\bar{\jmath}\}(i_{\bar{\jmath}})} = \begin{cases} \log_e\left(\dfrac{\sum_{\bar{\imath}_d}\alpha_{\bar{\imath}_d}}{I_{\bar{\imath}_d}}\right);\, if\ r=0 \\ \log_e\left(\dfrac{\alpha_{(i_{\bar{\jmath}_r},+)}}{I_{i_{\bar{\jmath}_d}\setminus i_{\bar{\jmath}_r}}}\right)-\sum_{q=0}^{r-1}.\sum_{\{\bar{\jmath}_q\}\in \mathrm{P}(\{\bar{\jmath}_r\}:n(\{\bar{\jmath}_q\})=q)}\hat{\mu}_{\{\bar{\jmath}_q\}(i_{\bar{\jmath}_q})}; \\ \{\bar{\jmath}_r\}\in\left(P([d]:n(\{\bar{\jmath}_r\})=r)\right) \text{ if } 1\le \mathrm{r} \le \mathrm{d} \end{cases}$$

Where $\alpha_{(i_{\bar{\jmath}_r},+)} = \sum_{i_{\bar{\jmath}_d}\setminus i_{\bar{\jmath}_r}}\alpha_{\bar{\imath}_d}$

Proof

By definition 1.3.1, observe that;

$$\mathcal{L}(\cdot\,|x_{i_1,i_2,\dots,i_r}) = \binom{n}{(\alpha_{\bar{\imath}_d})}\prod_{\bar{\imath}_d}p_{\bar{\imath}_d}^{\alpha_{\bar{\imath}_d}}$$

For arbitrary but fixed $d\in\mathbb{N}$, let $\mu_{\bar{\imath}_d}$ be the expected cell count with the corresponding(expected) probability $\pi_{\bar{\imath}_d}$, then we have $\mu_{\bar{\imath}_d} = n\pi_{\bar{\imath}_d}$,

where $n\ = \sum_{\bar{\imath}_d}\alpha_{\bar{\imath}_d}$.

Then we seek to solve the problem of

$$maximize\ \ \mathcal{L}(\cdot\,|x_{i_1,i_2,\dots,i_r}) = \binom{n}{(\alpha_{\bar{\imath}_d})}\prod_{\bar{\imath}_d}p_{\bar{\imath}_d}^{\alpha_{\bar{\imath}_d}}$$
$$subject\ to:\ \mu_{\bar{\imath}_d} = n\pi_{\bar{\imath}_d}$$

Equivalently, we consider;

$$maximize\ \ln\mathcal{L}(\cdot\,|x_{i_1,i_2,\dots,i_r}) = \ln\binom{n}{(\alpha_{\bar{\imath}_d})}\prod_{\bar{\imath}_d}p_{\bar{\imath}_d}^{\alpha_{\bar{\imath}_d}}$$
$$subject\ to:\ \mu_{\bar{\imath}_d} = n\pi_{\bar{\imath}_d}$$

Now, observe that;

$$\ln\mathcal{L}(\cdot\,|x_{i_1,i_2,\dots,i_r}) = \ln\binom{n}{(\alpha_{\bar{\imath}_d})} + \ln\left(\prod_{\bar{\imath}_d}p_{\bar{\imath}_d}^{\alpha_{\bar{\imath}_d}}\right) = \ln\binom{n}{(\alpha_{\bar{\imath}_d})} + \sum_{\bar{\imath}_d}\alpha_{\bar{\imath}_d}\ln\pi_{\bar{\imath}_d}$$
$$= \ln\binom{n}{(\alpha_{\bar{\imath}_d})} + \sum_{\bar{\imath}_d}\alpha_{\bar{\imath}_d}\ln\left(\frac{\mu_{\bar{\imath}_d}}{n}\right) = \ln\binom{n}{(\alpha_{\bar{\imath}_d})} + \sum_{\bar{\imath}_d}\alpha_{\bar{\imath}_d}\left(\ln\mu_{\bar{\imath}_d}-\ln n\right) = \ln\binom{n}{(\alpha_{\bar{\imath}_d})} + \sum_{\bar{\imath}_d}\alpha_{\bar{\imath}_d}\ln\mu_{\bar{\imath}_d} - \ln n\sum_{\bar{\imath}_d}\alpha_{\bar{\imath}_d}$$
$$= \ln\binom{n}{(\alpha_{\bar{\imath}_d})} + \sum_{\bar{\imath}_d}\alpha_{\bar{\imath}_d}\ln\mu_{\bar{\imath}_d} - n\ln n$$

Using Lemma 1.2.4 and Lemma 1.2.5 we have;

$$\ln\mathcal{L}(\cdot\,|x_{i_1,i_2,\dots,i_r}) = \ln\binom{n}{(\alpha_{\bar{\imath}_d})} + \sum_{\bar{\imath}_d}\alpha_{\bar{\imath}_d}\left(\sum_{\{\bar{\jmath}\}\in \mathrm{P}([d]:0\le n(\{\bar{\jmath}\})\le d)}\mu_{\{\bar{\jmath}\}(i_{\bar{\jmath}})}\right) - n\ln n$$
$$= \ln\binom{n}{(\alpha_{\bar{\imath}_d})} + \sum_{\bar{\imath}_d}\alpha_{\bar{\imath}_d}\left(\sum_{q=0}^{d}.\sum_{\bar{\jmath}\in S_{d,r}^{c}}\mu_{\{\bar{\jmath}\}(i_{\bar{\jmath}})}\right) - n\ln n \quad (2.1)$$

The constrain condition implies that $\sum_{\bar{\imath}_d} \alpha_{\bar{\imath}_d} = n$. Now we construct appropriate Largragian function $\mathcal{G}$for the maximization problem as such;

$$\mathcal{G}\left(\mu_{\{\bar{\jmath}\}(i_{\bar{\jmath}})};\ \lambda\right) = \ln\left(\mu_{\{\bar{\jmath}\}(i_{\bar{\jmath}})}|\alpha_{i_1,i_2,\dots,i_d}\right) + \lambda\left(n - \sum_{\bar{\imath}_d}\mu_{\bar{\imath}_d}\right)$$

$$= \ln\binom{n}{(\alpha_{\bar{\imath}_d})} + \sum_{\bar{\imath}_d}\alpha_{\bar{\imath}_d}\left(\sum_{\{\bar{\jmath}\}\in \mathrm{P}([d]:0\le n(\{\bar{\jmath}\})\le d)}\mu_{\{\bar{\jmath}\}(i_{\bar{\jmath}})}\right) - n\ln n + \lambda\left(n - \sum_{\bar{\imath}_d}\mu_{\bar{\imath}_d}\right)$$

$$= \ln\binom{n}{(\alpha_{\bar{\imath}_d})} + \sum_{\bar{\imath}_d}\alpha_{\bar{\imath}_d}\left(\sum_{\{\bar{\jmath}\}\in \mathrm{P}([d]:0\le n(\{\bar{\jmath}\})\le d)}\mu_{\{\bar{\jmath}\}(i_{\bar{\jmath}})}\right) - n\ln n + \lambda\left(n - \sum_{\bar{\imath}_d} Exp\left(\sum_{\{\bar{\jmath}\}\in \mathrm{P}([d]:0\le n(\{\bar{\jmath}\})\le d)}\mu_{\{\bar{\jmath}\}(i_{\bar{\jmath}})}\right)\right)$$

By optimality condition, we have that for $1 \le \mathrm{r} \le \mathrm{d}$ and$\bar{\jmath} \in S_{d,r}^c$

$$\frac{\partial\mathcal{G}}{\partial\mu_{\{\bar{\jmath}\}(i_{\bar{\jmath}})}} = \sum_{\bar{\imath}_d\backslash\bar{\imath}_r}\alpha_{\bar{\imath}_d} - \lambda\sum_{\bar{\imath}_d\backslash\bar{\imath}_r} Exp\left(\sum_{\{\bar{\jmath}\}\in \mathrm{P}([d]:0\le n(\{\bar{\jmath}\})\le d)}\mu_{\{\bar{\jmath}\}(i_{\bar{\jmath}})}\right) = 0 \quad (2.2)$$

and

$$\frac{\partial\mathcal{G}}{\partial\lambda} = n - \sum_{\bar{\imath}_d} Exp\left(\sum_{\{\bar{\jmath}\}\in \mathrm{P}([d]:0\le n(\{\bar{\jmath}\})\le d)}\mu_{\{\bar{\jmath}\}(i_{\bar{\jmath}})}\right) = 0 \quad (2.3)$$

Solving equation (2.2) and (2.3) we have that$\lambda = 1$, so that

$$\sum_{\bar{\imath}_d\backslash\bar{\imath}_r} Exp\left(\sum_{\{\bar{\jmath}\}\in \mathrm{P}([d]:0\le n(\{\bar{\jmath}\})\le d)}\hat{\mu}_{\{\bar{\jmath}\}(i_{\bar{\jmath}})}\right) = \sum_{\bar{\imath}_d\backslash\bar{\imath}_r}\alpha_{\bar{\imath}_d}$$

$$\Rightarrow \sum_{\bar{\imath}_d\backslash\bar{\imath}_r} Exp\left(\sum_{\{\bar{\jmath}\}\in \mathrm{P}([d]:0\le n(\{\bar{\jmath}\})\le r-1)}\hat{\mu}_{\{\bar{\jmath}\}(i_{\bar{\jmath}})}\right) Exp\left(\sum_{\{\bar{\jmath}\}\in \mathrm{P}([d]:n(\{\bar{\jmath}\})=r)}\hat{\mu}_{\{\bar{\jmath}\}(i_{\bar{\jmath}})}\right)$$
$$\times Exp\left(\sum_{\{\bar{\jmath}\}\in \mathrm{P}([d]:r+1\le n(\{\bar{\jmath}\})\le d)}\hat{\mu}_{\{\bar{\jmath}\}(i_{\bar{\jmath}})}\right) = \sum_{\bar{\imath}_d\backslash\bar{\imath}_r}\alpha_{\bar{\imath}_d}$$

$$\Rightarrow Exp\left(\sum_{\{\bar{\jmath}\}\in \mathrm{P}([d]:n(\{\bar{\jmath}\})=r)}\hat{\mu}_{\{\bar{\jmath}\}(i_{\bar{\jmath}})}\right) =$$
$$\frac{\sum_{\bar{\imath}_d\backslash\bar{\imath}_r}\alpha_{\bar{\imath}_d}}{\sum_{\bar{\imath}_d\backslash\bar{\imath}_r} Exp\left(\sum_{\{\bar{\jmath}\}\in \mathrm{P}([d]:r+1\le n(\{\bar{\jmath}\})\le d)}\hat{\mu}_{\{\bar{\jmath}\}(i_{\bar{\jmath}})}\right) Exp\left(\sum_{\{\bar{\jmath}\}\in \mathrm{P}([d]:0\le n(\{\bar{\jmath}\})\le r-1)}\hat{\mu}_{\{\bar{\jmath}\}(i_{\bar{\jmath}})}\right)}$$
$$= \frac{\sum_{\bar{\imath}_d\backslash\bar{\imath}_r}\alpha_{\bar{\imath}_d}}{\sum_{\bar{\imath}_d\backslash\bar{\imath}_r} Exp\left(\sum_{\{\bar{\jmath}\}_q:r+1\le q\le d}\hat{\mu}_{\{\bar{\jmath}\}(i_{\bar{\jmath}})}\right) Exp\left(\sum_{\{\bar{\jmath}\}_q:0\le q\le r-1}\hat{\mu}_{\{\bar{\jmath}\}(i_{\bar{\jmath}})}\right)}$$
$$= \frac{\sum_{\bar{\imath}_d\backslash\bar{\imath}_r}\alpha_{\bar{\imath}_d}}{\sum_{\bar{\imath}_d\backslash\bar{\imath}_r} Exp\left(\sum_{\{\bar{\jmath}\}_q:0\le q\le d\backslash\{\bar{\jmath}\}_q:0\le q\le r}\hat{\mu}_{\{\bar{\jmath}\}(i_{\bar{\jmath}})}\right) Exp\left(\sum_{\{\bar{\jmath}\}_q:0\le q\le r-1}\hat{\mu}_{\{\bar{\jmath}\}(i_{\bar{\jmath}})}\right)}$$

Hence we have

$$\hat{\mu}_{\{\bar{\jmath}\}(i_{\bar{\jmath}})}|_{(\{\bar{\jmath}\}\in \mathrm{P}([d]:n(\{\bar{\jmath}\})=r))}$$
$$\log_e\left(\frac{\sum_{\bar{\imath}_d\backslash\bar{\imath}_r}\alpha_{\bar{\imath}_d}}{\sum_{\bar{\imath}_d\backslash\bar{\imath}_r} Exp\left(\sum_{\{\bar{\jmath}\}_q:0\le q\le d\backslash\{\bar{\jmath}\}_q:0\le q\le r}\hat{\mu}_{\{\bar{\jmath}\}(i_{\bar{\jmath}})}\right)}\right) - \left(\sum_{\{\bar{\jmath}\}_q:0\le q\le r-1}\hat{\mu}_{\{\bar{\jmath}\}(i_{\bar{\jmath}})}\right)$$

Since, $n(\{\bar{\jmath}\}) = r$, we write;

$$\hat{\mu}_{\{\bar{\jmath}\}(i_{\bar{\jmath}})} = \begin{cases} \log_e\left(\frac{\sum_{\bar{\imath}_d}\alpha_{\bar{\imath}_d}}{I_{\bar{\imath}_d}}\right); if\ r=0 \\ \log_e\left(\frac{\alpha_{(i_{\bar{\jmath}_r},+)}}{I_{i_{\bar{\jmath}_d}\setminus i_{\bar{\jmath}_r}}}\right) - \sum_{q=0}^{r-1} \sum_{\{\bar{\jmath}_q\}\in P(\{\bar{\jmath}_r\}:n(\{\bar{\jmath}_q\})=q)} \hat{\mu}_{\{\bar{\jmath}_q\}}(i_{\bar{\jmath}_q}); \\ \{\bar{\jmath}_r\} \in \left(P([d]:n(\{\bar{\jmath}_r\})=r)\right) \text{ if } 1 \le r \le d \end{cases}$$

3.0 Conclusion

The results obtained in this paper solve certain maximum likelihood parameter estimation problem in the generalised sense. Researchers in multivariate analysis will find the result obtained in this paper useful when solving for value(s) that maximize certain arbitrary$d$-dimensional multinomial function (likelihood function). We are not aware of the existence of the results obtained in this paper in literature.

4.0 References

[1] C. Berge, Principles of Combinatorics, vol. 72 in Mathematics in Science and Engineering a series of monographs and textbooks, Academic press New York and London, 1971.
[2] L. Babai, Discrete. Mathematics Lecture Notes. Oct. 24, 2003.
[3] G. Christel and G. Stefan, Introduction to probability theory. University of Jyvskyl_a, 2009.
[4] R. Johnsonbauch, Discrete Mathematics , New York. Macmillian,(1984),pg.43-65.
[5] R. H. Kenneth. , Discrete Mathematics and its Applications, Singapore. McGraw UHill (1991), pg. 232-296.
[6] B. P. Kenneth, Combinatorics through guided discovery. Nov. 6, 2004.
[7] Seymour Lipschutz, Theory and Problems of Discrete mathematics, Schaum'sSeries,McGraw-Hill,Inc,1976.
[8] L. Mejibro, Discrete Distributions, Leif Mejibro and Ventus publishing ApS,2009, pp 56-71.
[9] Leif Mejibro, Introduction to Probability, Ventus Pulishing ApS, 2009.
[10] E. A. Bender and S. G. Williamson, Foundations of combinatorics withapplications, 2005.
[11] S.E. Payne, Applied combinatorics - MTH6409, Student version-fall, 2003. pg.19.
[12] Chidume C. E., Applicable functional analysis. ICTP publishing section,Italy,(2003).
[13] Fucik S. and Kufner A., Nonlinear differential equations. New York, Elsevier scienctific publishing company, 1980.
[14] Muhammed A., an Introdution to Sobolev spaces and applications (Diploma Project) ICTP, Trieste, Italy, 1999.
[15] J.E. Johndrow and et al.,Tensor rank of log-linear models, imsart-generic ver.2011/11/15 file: BPTD structure-arxiv.tex date: April 3, 2014.
[16] O. Ingram and et'al., Probability models and Applications, Macmillan publishing Co. Inc