submitted manuscript No.
(will be inserted by the editor)

MPH and LPH models subject to inequality constraints

Roberto Colombi[2]

Dip. Ingegneria dell'Informazione e Metodi Matematici, Università di Bergamo - Viale Marconi, 5 - 24044 Dalmine - Italy; e-mail: `colombi@unibg.it`

Received: date / Revised version: date

Abstract *Lang's (2004, 2005) MPH and HLP models include as special cases many models for contingency tables analysis that have been introduced in the effort to overcame well known limitations of the log-linear models. Here the definition of MPH and HLP models is extended to include inequality constraints. These notes are mainly intended to be a brief introduction to the Lang's (2004, 2005) papers.*

1 Multi-way contingency tables

We consider the joint probability function of q ordinal categorical variables $A_1, \ldots, A_q$, where A_j has categories in $\mathcal{A}_j = \{a_{ji_j}, i_j = 1, 2, \ldots, r_j\}$. For ordinal variables the numbering of the categories is assumed to be coherent with their order. The vector of the $c = \prod_1^q r_j$ joint probabilities will be denoted by $\boldsymbol{p}$. The set of variables that defines a given marginal distribution will be denoted by the set $\mathcal{M}$ of indices of the corresponding variables. $\mathcal{M}$ will be called marginal set and the distribution associated with it will be called $\mathcal{M}$-marginal distribution. $\mathcal{Q} = \{1, \ldots, q\}$ will refer to the joint distribution. Given a vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_q)'$ of q components $\boldsymbol{x}_{\mathcal{M}}$ will denote the vector with components $x_j : j \in \mathcal{M}$. $\mathbf{1}$ will denote a vector of ones when the dimension is clear from the context and $\mathbf{1}_{\mathcal{M}}$ a vector of ones of dimension equal to the cardinality of $\mathcal{M}$. Given a vector $\boldsymbol{x}$, $\boldsymbol{D}(\boldsymbol{x})$ will denote the diagonal matrix with the components of $\boldsymbol{x}$ on the main diagonal.

2 Basic Concepts on MPH models

It is assumed that sample frequencies come from s different strata or populations and that the sample size is fixed for some strata while it is a Poisson random variable for the remaining strata. The strata are defined by a subset of the q categorical variables A_j. The vector of the sample frequencies will be denoted by $\boldsymbol{y}$. Usually the frequencies are ordered on a population basis and within a population the categories subscripts change according the lexicographic order. The zero-one elements population matrix $\boldsymbol{Z}$ gives the vector of s sample sizes $\boldsymbol{n} = \boldsymbol{Z}'\boldsymbol{y}$ and takes into account the fact that the frequencies of $\boldsymbol{y}$ may not be arranged on a population basis. If the frequencies are ordered on a population basis and if there are c' frequencies from every stratum then $\boldsymbol{Z} = \boldsymbol{I}_s \bigotimes \boldsymbol{1}_{c'}$. The sample design matrix $\boldsymbol{Z}_F$ is formed by a subset of the columns of $\boldsymbol{Z}$ and gives the vector of fixed sample sizes: $\boldsymbol{n}_F = \boldsymbol{Z}'_F\boldsymbol{y}$. The sample design matrix $\boldsymbol{Z}_R$ is formed by a subset of the columns of $\boldsymbol{Z}$ too and gives the vector of random sample sizes: $\boldsymbol{n}_R = \boldsymbol{Z}'_R\boldsymbol{y}$. The expected value of $\boldsymbol{n}_R$ is $E(\boldsymbol{n}_R) = \boldsymbol{\delta}$. Moreover let it be $\boldsymbol{\gamma} = \boldsymbol{Q}_F\boldsymbol{n}_F + \boldsymbol{Q}_R\boldsymbol{\delta}$ the vector of the fixed sample sizes and of the expected values of the random sample sizes. $\boldsymbol{Q}_R$ and $\boldsymbol{Q}_F$ are selection matrices such that $\boldsymbol{Z}_F = \boldsymbol{Z}\boldsymbol{Q}_F$ and $\boldsymbol{Z}_R = \boldsymbol{Z}\boldsymbol{Q}_R$. Analogously we have: $\boldsymbol{n} = \boldsymbol{Q}_F\boldsymbol{n}_F + \boldsymbol{Q}_R\boldsymbol{n}_R$ for the vector of the sample sizes fixed or not. The mathematics behind MPH models relies on same important properties of the zero-one matrices $\boldsymbol{Z}$ and of $\boldsymbol{Z}_F$ presented in Lang (2004).

The probabilities of the vector $\boldsymbol{p}$ are called *pre-sample probabilities* while the vector of the conditional probabilities $\boldsymbol{\pi} = \boldsymbol{D}^{-1}(\boldsymbol{Z}\boldsymbol{Z}'\boldsymbol{p})\boldsymbol{p}$ is the vector of the *sample-model probabilities*.

The log-likelihood is assumed to be the product of s Multinomial probability functions and s' Poisson probability functions. Under the previous assumptions $\boldsymbol{y}$ is said to have a product Multinomial Poisson distribution.

The expected value of $\boldsymbol{y}$ is $\boldsymbol{m} = \boldsymbol{D}(\boldsymbol{Z}\boldsymbol{\gamma})\boldsymbol{\pi}$ while the expected value, given the sample sizes, is $E(\boldsymbol{y}|\boldsymbol{n}) = \boldsymbol{D}(\boldsymbol{Z}\boldsymbol{n})\boldsymbol{\pi} = \boldsymbol{N}\boldsymbol{\pi}$.

It is of interest to know under what assumptions inferences made on $\boldsymbol{\pi}$ and $\boldsymbol{\gamma}$ or $\boldsymbol{m}$ are valid inferences on $\boldsymbol{p}$ too.

The Lang's (2004) definition of $\boldsymbol{Z}$*-homogeneus functions* $\boldsymbol{h}(\cdot)$ is relevant because for these functions $\boldsymbol{h}(\boldsymbol{p}) = 0$ if and only if $\boldsymbol{h}(\boldsymbol{\pi}) = 0$ if and only if $\boldsymbol{h}(\boldsymbol{m}) = 0$. Here we remember only that $\boldsymbol{h}(\cdot)$ is $\boldsymbol{Z}$*-homogeneous* if and only if for every component $h_j(\cdot)$ of $\boldsymbol{h}(\cdot)$ it is:

$$h_j(\boldsymbol{D}(\boldsymbol{Z}\boldsymbol{\gamma})\boldsymbol{\pi}) = \gamma_{v(j)}^{p(j)} h_j(\boldsymbol{\pi})$$

where $\gamma_{v(j)}$ is a component of $\boldsymbol{\gamma}$, $v(j) \in \{1, 2, ..., s\}$, and where the constant $p(j)$ is the order of homogeneity of the $j - th$ component. An important special case is: $p(j) = 0, \forall j$, that is the case of *zero order* $\boldsymbol{Z}$*-homogeneus functions*. Proposition 5-7 of Lang (2005) describe the main consequences of these definitions.

A MPH model is a model in which i) $\boldsymbol{y}$ has a Product-Multinomial Poisson distribution with parameters $\boldsymbol{\pi}$, $\boldsymbol{\gamma}$, ii) $\boldsymbol{h}(\boldsymbol{\pi}) = \boldsymbol{0}$, iii) $\boldsymbol{h}(\boldsymbol{x})$ is a $\boldsymbol{Z}$-*homogeneus function* of class C^2 with full row rank Jacobian matrix $\frac{\partial \boldsymbol{h}(\boldsymbol{x})}{\partial \boldsymbol{x}'}$.

A HLP model is a model in which: i) $\boldsymbol{y}$ has a Product-Multinomial Poisson distribution with parameters $\boldsymbol{\pi}$, $\boldsymbol{\gamma}$, ii) $\boldsymbol{L}(\boldsymbol{\pi}) = \boldsymbol{X}\boldsymbol{\beta}$, with $\boldsymbol{X}$ a matrix of known constants and of full column rank, iii) $\boldsymbol{L}(\boldsymbol{\pi}) = \boldsymbol{X}\boldsymbol{\beta}$ if and only if $\exists \quad \boldsymbol{\beta}^*$ such that $\boldsymbol{L}(\boldsymbol{p}) = \boldsymbol{X}\boldsymbol{\beta}^*$, iv) $\boldsymbol{L}(\boldsymbol{m}) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{a}(\gamma), \boldsymbol{a}(\gamma_1) - \boldsymbol{a}(\gamma_2) = \boldsymbol{a}(\gamma_1/\gamma_2) - \boldsymbol{a}(\mathbf{1})$, v) $\boldsymbol{L}(\boldsymbol{x})$ is of class C^2 with full row rank Jacobian matrix $\frac{\partial \boldsymbol{L}(\boldsymbol{x})}{\partial \boldsymbol{x}'}$.

HLP models are special cases of MPH models, see Lang (2005) for details. Many examples of MPH and HLP models can be found on the Lang's web page:

http://www.stat.uiowa.edu/ jblang/mph.fitting/mph.fit.documentation.htm.

The reader is strongly advised to go trough these examples carefully.

3 Inequality constrained MPH and HLP models

Inequality constrained MPH and HLP models are MPH and HLP models subject to the inequality constraints $\boldsymbol{d}(\boldsymbol{\pi}) \geq \mathbf{0}$ and where $\boldsymbol{d}(\cdot)$ is a $\boldsymbol{Z}$-*homogeneus function* of class C^2 with full rank Jacobian matrix $\frac{\partial \boldsymbol{d}(\boldsymbol{x})}{\partial \boldsymbol{x}'}$ or satisfying the Mangasarian Fromowitz constraint qualification condition. For a discussion of the relevance of the Mangasarian-Fromiwitz condition for the problem of inequalities testing see Cazzaro Colombi, 2006a. Note that from the homogeneity condition it follows that $\boldsymbol{d}(\boldsymbol{\pi}) \geq \mathbf{0}$ if and only if $\boldsymbol{d}(\boldsymbol{p}) \geq \mathbf{0}$ if and only if $\boldsymbol{d}(\boldsymbol{m}) \geq \mathbf{0}$. Finally we assume that there exists a $\boldsymbol{Z}$-*homogeneus function* $\boldsymbol{g}(\cdot)$ of class C^2 with full rank Jacobian $\frac{\partial \boldsymbol{g}(\boldsymbol{x})}{\partial \boldsymbol{x}'}$ such that $\boldsymbol{g}(\boldsymbol{x}) = \mathbf{0}$ if and only if $\boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}$, $\boldsymbol{d}(\boldsymbol{x}) = \mathbf{0}$. In the context of inequality constraints it is of interest to test the hypothesys $\boldsymbol{h}(\boldsymbol{\pi}) = \mathbf{0}, \quad \boldsymbol{d}(\boldsymbol{\pi}) = \mathbf{0}$ against $\boldsymbol{h}(\boldsymbol{\pi}) = \mathbf{0}, \quad \boldsymbol{d}(\boldsymbol{\pi}) \geq \mathbf{0}$ (test of type A according Silvapulle-Sen, 2005) and to test $\boldsymbol{h}(\boldsymbol{\pi}) = \mathbf{0}, \quad \boldsymbol{d}(\boldsymbol{\pi}) \geq \mathbf{0}$ against $\boldsymbol{h}(\boldsymbol{\pi}) = \mathbf{0}$ (test of type B). See Silvapulle-Sen (2005) for the technicalities needed to handle inequality constraints in deriving the asymptotic properties of the ML estimators and of the ML based inferential techniques.

4 Asymptotics

Let $v = \mathbf{1}'\boldsymbol{\gamma}$ and for the moment let the parameters of interest $\boldsymbol{m}$, $\boldsymbol{\pi}$, $\boldsymbol{\gamma}$ be indexed by v. That it is to say let us use the notation: $\boldsymbol{m}_v$, $\boldsymbol{\pi}_v$, $\boldsymbol{\gamma}_v$ to make explicit the dependence on v. To derive the asymptic results for $v \to \infty$ it is assumed that: $\boldsymbol{\gamma}_v/v \to \boldsymbol{w}$, $\boldsymbol{\pi}_v = \boldsymbol{\pi}$. If $\boldsymbol{W} = \boldsymbol{D}(\boldsymbol{Z}\boldsymbol{w})$, it follows that $\boldsymbol{m}_v \to \boldsymbol{W}\boldsymbol{\pi}$.

The asymptotic properties, for $v \to \infty$, of the ML estimators and of the ML based inferential techniques relies on Lemmas 1-4 of Lang (2004) and in particular on the fact that the random variable: $v^{-1/2}(\boldsymbol{y} - \boldsymbol{N}\boldsymbol{\pi})$, has an asymptotic normal distribution with null expected value and variance matrix $\boldsymbol{\Omega} = \boldsymbol{W}\boldsymbol{D}(\boldsymbol{\pi}) - \boldsymbol{W}\boldsymbol{D}(\boldsymbol{\pi})\boldsymbol{Z}\boldsymbol{Z}'\boldsymbol{D}(\boldsymbol{\pi})$. An important role is also played by the asymptotic independence between $v^{-1/2}(\boldsymbol{y} - \boldsymbol{N}\boldsymbol{\pi})$ and $v^{-1/2}(\boldsymbol{Z}_R'\boldsymbol{y} - \boldsymbol{\delta})$. See Lang (2004, 2005) for details. We observe that, in contrast with statndard asymptotic theory, v is not a function of sample sizes but both of sample sizes and unknown parameters. The practical aspects implied by this fact are dealt with in theorem 6 of Lang (2004). Lang (2004) derived the asymptotic properties of the ML estimators of the parameters of a MPH models by the Cramer approach that is by considering these estimators as the solutions of a system of first order conditions. Here we use the more direct approach of Wald that derives the asymptotic properties of the ML estimators by looking at them as the solutions of a constrained optimization problem. This approach is more convenient in the case of inequality constraints (Andrews 1999, Silvapulle-Sen 2005) and gives a more direct result concerning the likelihood ratio test statistics.

It is convenient to reparameterize the vector $\boldsymbol{\pi}$ by the saturated log-linear model

$$\begin{aligned}\boldsymbol{\pi}(\boldsymbol{\theta}) &= \boldsymbol{D}^{-1}(\boldsymbol{Z}\boldsymbol{k}_{\boldsymbol{Z}}(\boldsymbol{\theta}))\exp\{\boldsymbol{G}\boldsymbol{\theta}\}\\ \boldsymbol{k}_{\boldsymbol{Z}}(\boldsymbol{\theta}) &= \qquad \boldsymbol{Z}'\exp\{\boldsymbol{G}\boldsymbol{\theta}\}\end{aligned}$$

where $\boldsymbol{G}$ is a design matrix that gives a saturated log-linear model within every population. We will use the following notation $\boldsymbol{H}^* = \frac{\partial \boldsymbol{h}(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}'}$ and

$$\boldsymbol{H} = \frac{\partial \boldsymbol{h}(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}'}\frac{\partial \boldsymbol{\pi}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \boldsymbol{H}^*(\boldsymbol{D}(\boldsymbol{\pi}) - \boldsymbol{D}(\boldsymbol{\pi})\boldsymbol{Z}\boldsymbol{Z}'\boldsymbol{D}(\boldsymbol{\pi}))\boldsymbol{G}.$$

An important result states that: $\boldsymbol{H} = \boldsymbol{H}^*\boldsymbol{D}(\boldsymbol{\pi})\boldsymbol{G}$ for $\boldsymbol{Z}$*-homogeneus function* of class C^2 (Lang, 2004, Proposition 6), $\forall \boldsymbol{\pi} : \boldsymbol{h}(\boldsymbol{\pi}) = 0$. The previous identity is true for all $\boldsymbol{\pi}$ for *zero order* $\boldsymbol{Z}$*-homogeneus function* of class C^2. The Jacobian matrices $\boldsymbol{D}^*$, $\boldsymbol{D}$, $\boldsymbol{G}^*$ and $\boldsymbol{G}$ are defined in a similar way with respect to $\boldsymbol{d}(\boldsymbol{\pi})$ and $\boldsymbol{g}(\boldsymbol{\pi})$.

For MPH models the log-likelihood is a sum of a multinomial component depending only on $\boldsymbol{\theta}$ and a Poisson component depending only on $\boldsymbol{\delta}$ and the equality and inequality constraints involve only $\boldsymbol{\theta}$. Moreover the asymptotic independence between $v^{-1/2}(\boldsymbol{y} - \boldsymbol{N}\boldsymbol{\pi})$ and $v^{-1/2}(\boldsymbol{Z}'_R\boldsymbol{y} - \boldsymbol{\delta})$ (Lang, 2004) implies that the the limit, for $v \to \infty$, of the averaged Fisher matrix is block diagonal with a block corresponding to $\boldsymbol{\theta}$ and a block corresponding to $\boldsymbol{\delta}$. Thus the two components of the log-likelihood can be maximized separately and the asymptotic properties of the ML estimators of $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$ can be examined independently. The ML estimator $\hat{\boldsymbol{\theta}}_i$, $i = 0, 1, 2$, we are interested in, are the solution of the optimum problems:

$$max_{\boldsymbol{\theta}\in\Theta_i}\boldsymbol{y}'\ln(\boldsymbol{\pi}(\boldsymbol{\theta}))$$

where the parameters spaces Θ_i are so defined:

$$\begin{aligned}\Theta_0 = \{\boldsymbol{\theta} : \boldsymbol{h}(\boldsymbol{\pi}(\boldsymbol{\theta})) = \boldsymbol{0},\quad \boldsymbol{d}(\boldsymbol{\pi}(\boldsymbol{\theta})) = \boldsymbol{0}\} = \{\boldsymbol{\theta} : \boldsymbol{g}(\boldsymbol{\pi}(\boldsymbol{\theta})) = \boldsymbol{0}\},\\ \Theta_1 = \{\boldsymbol{\theta} : \boldsymbol{h}(\boldsymbol{\pi}(\boldsymbol{\theta})) = \boldsymbol{0},\quad \boldsymbol{d}(\boldsymbol{\pi}(\boldsymbol{\theta})) \geq \boldsymbol{0}\},\\ \Theta_2 = \{\boldsymbol{\theta} : \boldsymbol{h}(\boldsymbol{\pi}(\boldsymbol{\theta})) = \boldsymbol{0}\}.\end{aligned}$$

Note that in all the three previous cases the ML estimator of $\boldsymbol{\delta}$ is $\hat{\boldsymbol{\delta}} = \boldsymbol{Z}'_R\boldsymbol{y}$. In what follows $\boldsymbol{\pi}_0$ is $\boldsymbol{\pi}(\boldsymbol{\theta})$ evaluated at $\boldsymbol{\theta}_0 \in \Theta_0$ and $\boldsymbol{H}_0$ denotes $\boldsymbol{H}$ evaluated at $\boldsymbol{\theta}_0$. A similar convention is used for the other Jacobian matrices previously introduced. The following quadratic approximation of $L(\boldsymbol{\theta}) = \boldsymbol{y}'\ln(\boldsymbol{\pi}(\boldsymbol{\theta}))$ is of central importance in deriving the asymptotic properties of the ML estimators:

$$\begin{aligned} 2(L(\boldsymbol{\theta}) - L(\boldsymbol{\theta}_0)) = \\ \frac{1}{\sqrt{v}}(\boldsymbol{y} - \boldsymbol{N}\boldsymbol{\pi}_0)'\boldsymbol{G}\boldsymbol{F}^{-1}\boldsymbol{G}'(\boldsymbol{y} - \boldsymbol{N}\boldsymbol{\pi}_0) + \\ -(\sqrt{v}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) - \boldsymbol{U}_v)'\boldsymbol{F}(\sqrt{v}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) - \boldsymbol{U}_v) + R_v(\boldsymbol{\theta}). \end{aligned} \tag{1}$$

In the previous formula

$$\boldsymbol{F} = \boldsymbol{G}'(\boldsymbol{W}\boldsymbol{D}(\boldsymbol{\pi}_0) - \boldsymbol{W}\boldsymbol{D}(\boldsymbol{\pi}_0)\boldsymbol{Z}\boldsymbol{Z}'\boldsymbol{D}(\boldsymbol{\pi}_0))\boldsymbol{G}$$

is the limit, for $v \to \infty$, of the block of the averaged Fisher matrix corresponding to $\boldsymbol{\theta}$, and the random variable $\boldsymbol{U}_v = v^{-1/2}\boldsymbol{F}^{-1}\boldsymbol{G}'(\boldsymbol{y} - \boldsymbol{N}\boldsymbol{\pi}_0)$ is the normalized score function for $\boldsymbol{\theta}$; the reminder $R_v(\boldsymbol{\theta})$ is such that:

$$sup_{\{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|<\frac{K}{\sqrt{v}}\}}\frac{R_v(\boldsymbol{\theta})}{v\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2} = o(1), \quad \forall K > 0, \quad v \to \infty.$$

For a general discussion of the assumptions under which the quadratic expansion (1) holds see Andrews (2004). Let $\sqrt{v}(\boldsymbol{\theta}_{qi} - \boldsymbol{\theta}_0)$, $i = 0, 1, 2$, be the solutions of the quadratic optimization problems:

$$min_{\boldsymbol{\theta}\in\Lambda_i}(\sqrt{v}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) - \boldsymbol{U}_v)'\boldsymbol{F}(\sqrt{v}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) - \boldsymbol{U}_v)$$

where:

$$\begin{aligned} \Lambda_0 = \{\boldsymbol{\theta} : \boldsymbol{G}_0(\boldsymbol{\theta} - \boldsymbol{\theta}_0) = \boldsymbol{0}\}, \\ \Lambda_1 = \{\boldsymbol{\theta} : \boldsymbol{H}_0(\boldsymbol{\theta} - \boldsymbol{\theta}_0) = \boldsymbol{0}, \quad \boldsymbol{D}_0(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \geq \boldsymbol{0}\}, \\ \Lambda_2 = \{\boldsymbol{\theta} : \boldsymbol{H}_0(\boldsymbol{\theta} - \boldsymbol{\theta}_0) = \boldsymbol{0}\}. \end{aligned}$$

The keys results of the Wald approach to the asymptotic properties of the ML estimators are given in the following Proposition.

Proposition 1 *If the ML estimators* $\hat{\boldsymbol{\theta}}_i$*,* $i = 0, 1, 2$ *are consistent, if* $\boldsymbol{U}_v = O_p(1)$ *and if the constraints satysfy the Mangasarian-Fromowitz condition then from the quadratic expansion (1) it follows, for* $v \to \infty$ *and* $i = 1, 2$*, that* $2(L(\hat{\boldsymbol{\theta}}_i) - L(\hat{\boldsymbol{\theta}}_{i-1}))$ *converges in distribution to*

$$\begin{aligned} min_{\boldsymbol{\theta}\in\Lambda_i}(\sqrt{v}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) - \boldsymbol{U}_v)'\boldsymbol{F}(\sqrt{v}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) - \boldsymbol{U}_v) + \\ -min_{\boldsymbol{\theta}\in\Lambda_{i-1}}(\sqrt{v}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) - \boldsymbol{U}_v)'\boldsymbol{F}(\sqrt{v}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) - \boldsymbol{U}_v). \end{aligned}$$

Moreover, for $v \to \infty$*,* $\sqrt{v}(\boldsymbol{\theta}_{qi} - \boldsymbol{\theta}_0)$*, converges in probability and in distribution to* $\sqrt{v}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_0)$*,* $i = 0, 1, 2$.

The consistency of the ML estimators can be easily proved, see Andrews (1999). The Mangasarian-Fromovitz condition is a mild regularity condition on the constraints that allows the parametric spaces Θ_i to be locally approximable by the linear spaces or cones Λ_i (Silvapulle-Sen, 2004, Cazzaro-Colombi 2006) and $\boldsymbol{U}_v = O_p(1)$ follows from the asymptotic normality of $v^{-1/2}(\boldsymbol{y} - \boldsymbol{N}\boldsymbol{\pi})$ (Lang, 2004, Lemma 4). Under the previous assumptions the proposition follows from the general results of Andrews (1999, Theorem 3). From the previous proposition and the asymptotic normality of $\boldsymbol{U}_v$ the asymptotic chibar distribution of the likelihood ratio statistcs for tests of type A and type B follows as shown in Silvapulle-Sen (2005). The well known chi-square asymptotic distribution of the likelihood ratio statistics to test $\boldsymbol{\theta} \in \Theta_2$ against $\boldsymbol{\theta} \notin \Theta_2$ follows too. Note that when both Λ_i and Λ_{i-1} are linear spaces defined by linear constraints then the standard chisquare asymptotic distribution holds. If Λ_i is a cone (defined by equalities and inequalities) and Λ_{i-1} is a linear spaces (test type A) or if Λ_i is a linear space and Λ_{i-1} a cone (test type B) the asymptotic chibar distribution holds (Sen-Silvapulle, 2004).

4.1 The case of only equality constraints

The convergence in probability and in distribution, for $v \to \infty$, of $\sqrt{v}(\boldsymbol{\theta}_{qi} - \boldsymbol{\theta}_0)$, to $\sqrt{v}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_0)$, $i = 0, 1, 2$ is usefull to obtain the asymptotic normality of the ML estimators in the case of only equality constraints. Here we give some details of this problem for the constraints $\boldsymbol{h}(\boldsymbol{\pi}) = \boldsymbol{0}$ (the case $\boldsymbol{g}(\boldsymbol{\pi}) = \boldsymbol{0}$ is analogous) because it is an interesting alternative way of deriving the Lang (2004) results. No simple and usefull results of this kind are at at the moment known for the case of inequality constraints See Andrews (1999 section 6) for a detailed discussion of this problem. Standard results on projections onto linear spaces shows that the solution of the optimization problem

$$min_{\boldsymbol{\theta} \in \Lambda_2} (\sqrt{v}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) - \boldsymbol{U}_v)' \boldsymbol{F} (\sqrt{v}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) - \boldsymbol{U}_v)$$

is given by

$$\sqrt{v}(\boldsymbol{\theta}_{qi} - \boldsymbol{\theta}_0) = \boldsymbol{U}_v - \boldsymbol{F}^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{F}^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{U}_v.$$

From this result it follows that $\sqrt{v}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is asymptotically normal with null expected value and variance matrix:

$$\boldsymbol{\Sigma} = \boldsymbol{F}^{-1} - \boldsymbol{F}^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{F}^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{F}^{-1}.$$

Finally noting that: $\frac{\partial \boldsymbol{\pi}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = (\boldsymbol{D}(\boldsymbol{\pi}) - \boldsymbol{D}(\pi)\boldsymbol{Z}\boldsymbol{Z}'\boldsymbol{D}(\boldsymbol{\pi}))\boldsymbol{G}$ the asymptotic normal distribution of $\sqrt{v}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}_0)$ is obtained. Some algebra show that the result simplifies to the one given by Theorem 3 of Lang (2004) if $\boldsymbol{h}(\cdot)$ is a $\boldsymbol{Z}$*-homogeneus function* . Finally the asymptotic distribution of $\frac{1}{\sqrt{v}}(\hat{\boldsymbol{m}} - \boldsymbol{m})$ follows as in Theorem 4 of Lang. Note that $\hat{\boldsymbol{m}} = \boldsymbol{D}(\boldsymbol{Z}\boldsymbol{Z}'\hat{\gamma})\hat{\boldsymbol{\pi}}, \quad \hat{\gamma} = \boldsymbol{Q}_F \boldsymbol{n}_F + \boldsymbol{Q}_R \hat{\boldsymbol{\delta}}$.

5 Hierarchical Multinomial Marginal Models (HMMM) and MPH models

The vector $\boldsymbol{\eta}$ of the parameters of an HMMM model (Bartolucci-Colombi-Forcina, 2007, Colombi Cazzaro, 2006) can be explicitly written in matrix form as

$$\boldsymbol{\eta} = \boldsymbol{C} \log(\boldsymbol{M}\boldsymbol{\pi}), \qquad (2)$$

where the rows of $\boldsymbol{C}$ are contrasts and $\boldsymbol{M}$ is a matrix of zeros and ones which sums the probabilities to obtain the necessary marginal probabilities. Assuming without loss of generality that the probabilities of $\boldsymbol{\pi}$ are arranged on a population basis it is, as we previously noted, $\boldsymbol{Z} = \boldsymbol{I}_s \otimes \boldsymbol{1}'_c$. Moreover, as the same generalized marginal interactions are defined within every strata, it follows that:

$$\boldsymbol{M} = \boldsymbol{I}_s \otimes \boldsymbol{M}^*$$
$$\boldsymbol{C} = \boldsymbol{I}_s \otimes \boldsymbol{C}^*$$

where the matrix of contrasts $\boldsymbol{C}^*$ and the zero-one matrix $\boldsymbol{M}^*$ depend on the generalized marginal interactions that are used. The definitions of these matrices were given by Colombi and Forcina (2001) and have been extended by Cazzaro and Colombi (2006b) in order to include recursive type interactions.

Simple algebra shows that:

$$\boldsymbol{D}^{-1}(\boldsymbol{M}\boldsymbol{1})\boldsymbol{M}\boldsymbol{Z} = \boldsymbol{0}. \qquad (3)$$

From (3) and theorem 1 of Lang (2005) it follows that constraints on the vector $\boldsymbol{\eta}$ of the type:

$$\begin{aligned} \boldsymbol{\eta} &= \boldsymbol{X}\boldsymbol{\beta} \\ \boldsymbol{U}'\boldsymbol{\eta} &= \boldsymbol{0} \\ \boldsymbol{K}\boldsymbol{\eta} &\geq 0 \end{aligned}$$

define special cases of MPH or HLP models.

Many examples of HMMM models are presented in the examples and demos included in the R-package hmmm developed by Colombi and Cazzaro (2006). See also the paper *Hierarchical Multinomial Marginal Models and the R-package hmmm: a brief introduction* included in the documentation of the package.

References

1. Andrews, D.,W.,K., Estimation when a parameter is on a boundary, Econometrica, **67**, (1999) 1341-1383
2. Bartolucci, F., Colombi, R., and Forcina, A., An extended class of marginal link functions for modelling contingency tables by equality and inequality constraints, Statistica Sinica, (2007) to appear.

3. Bergsma, W. P. and Rudas, T., Marginal models for categorical data, Annals of Statistics **30**, (2002) 140-159.
4. Cazzaro, M., Colombi, R., Maximum likelihood inference for log-linear models subject to constraints of double monotone dependence, Statistical Methods and Applications **15**, (2006a) 177-190.
5. Cazzaro, M., Colombi, R., A new family of interactions for modelling contingency tables, (2006b) submitted.
6. Cazzaro, M., Colombi, R., Modelling two way contingency tables with recursive logits and odds ratios, (2007) submitted.
7. Colombi, R., Cazzaro M., *hmmm: hierarchical multinomial marginal models*, R package version 1.0.21, (2006)
8. Colombi, R., Forcina, A., Marginal regression models for the analysis of positive association of ordinal response variables, Biometrika **88**, (2001) 1007-1019.
9. Dardanoni, V., Forcina, A., A unified approach to likelihood inference on stochastic orderings in a nonparametric context, Journal of the American Statistical Association **93**, (1998) 1112-1123.
10. R Development Core Team (2006), *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.
11. Glonek, G., F., V., McCullagh, P., Multivariate logistic models for Contingency Tables. Journal of the Royal Statistical Society, B, **57**, (1995) 533-546.
12. Lang, J., B., Multinomial Poisson Homogeneous Models for Contingency Tables. The Annnals of Statistics **32**, (2004) 340-383.
13. Lang, J., B., Homogeneous Linear Predictor Models for Contingency Tables. Journal of the American Statistical Association **100**, (2005) 121-134.
14. Silvapulle, M.J., Sen, P.K., *Constrained statistical inference* (Wiley, New Jersey 2005).