Maximum-Likelihood Analysis for Useful Classes of Multinomial-Poisson Homogeneous-Function Models

Joseph B. Lang

Department of Statistics and Actuarial Science, University of Iowa, IA 52242 USA, email: jblang@stat.uiowa.edu

Technical Report #298

November 2000

Maximum-Likelihood Analysis for Useful Classes of Multinomial-Poisson Homogeneous-Function Models

Joseph B. Lang[1]

ABSTRACT

Maximum likelihood inference for two important subclasses of multinomial-Poisson homogeneous-function (MPH) categorical data models is described. Maximum likelihood fit results, which include point estimates, goodness-of-fit statistics, and asymptotic-based approximating distributions, are described and compared for equivalent models. As an example, the effect of the sampling plan on the large-sample behavior of estimators is given in explicit form. The first subclass, the class of homogeneous linear predictor models, comprises MPH models that constrain expected counts $\mathbf{m}$ through $\mathbf{L}(\mathbf{m}) = \mathbf{X}\beta$, where the link $\mathbf{L}$ is allowed to be a many-to-one function. Generalized loglinear models, qualitative dispersion trend models, and mean response models are given as specific examples. The second subclass, the class of probability freedom models, comprises MPH models that constrain outcome probabilities π through $\pi = \mathrm{D}^{-1}(\mathbf{Z}\mathbf{Z}^T\mathrm{g}(\beta))\mathrm{g}(\beta)$, where the positive function g is sufficiently smooth and the matrix $\mathbf{Z}$ is determined by the sampling plan. This class includes the product-multinomial models that lend themselves to the multinomial-to-Poisson transformation.

Keywords: *categorical data, equivalent models, homogeneous linear predictor model, large-sample inference, multinomial-to-Poisson transformation, probability freedom model.*

1 Introduction

Lang (2000) introduced the broad class of multinomial-Poisson homogeneous-function (MPH) categorical data models and explored properties of MPH model maximum-likelihood estimators and goodness-of-fit statistics. In addition, formal definitions of model equivalence were introduced; they were based on equivalence-class partitionings of the collection of all candidate MPH models. Equivalent models were compared on the basis of their maximum likelihood fit results, which include point estimates, goodness-of-fit statistics, and asymptotic-based approximating distributions. The results were quite general. They were stated for models specified using the

[1]November 2, 2000. Joseph B. Lang is Associate Professor, Department of Statistics and Actuarial Science, University of Iowa, IA 52242, email: jblang@stat.uiowa.edu

model constraint $\mathbf{h}(\mathbf{m}) = \mathbf{0}$, where $\mathbf{m}$ is the vector of expected counts and $\mathbf{h}$ is any function in $\mathcal{H}''(\mathbf{Z})$, a broad class of $\mathbf{Z}$-homogeneous functions. The current paper explores important subclasses of $\mathbf{h}$ functions. By restricting attention to subclasses, we obtain additional useful results, and gain further insight.

This paper introduces the useful class of homogeneous linear predictor models, which have the generic form $\mathbf{L}(\mathbf{m}) = \mathbf{X}\boldsymbol{\beta}$, where the link $\mathbf{L}$ is allowed to be a many-to-one function. It follows that this class is very broad and includes models that are not of the univariate or multivariate generalized linear model form (see, e.g., McCullagh and Nelder 1989; Fahrmeir and Tutz 1994). Large-sample, maximum-likelihood inference is described and the fit results from equivalent models are compared. Generalized loglinear models (GLLMs) of the form $\mathbf{C}\log\mathbf{M}\mathbf{m} = \mathbf{X}\boldsymbol{\beta}$ (cf. Grizzle et al. 1969, Lang and Agresti 1994) are typically counted among the members of the homogeneous linear predictor models. As examples, GLLMs include standard loglinear, logit, cumulative logit, multivariate logit (Glonek and McCullagh 1995, Glonek 1996), and association-marginal models (Lang et al. 1999). Links other than $\mathbf{L}(\mathbf{m}) = \mathbf{C}\log\mathbf{M}\mathbf{m}$ are considered as well. For example, $\mathbf{L}(\mathbf{m})$ could be a vector of distribution summaries such as Gini-dispersions, mean scores, or association measures like the gamma or kappa statistic (cf. Agresti 1990, pp. 22, 366). The current paper's maximum likelihood fitting approach for these many-to-one link models serves as an attractive alternative to the more commonly used weighted least squares approach (see, e.g., Grizzle et al. 1969, Stokes et al. 1995).

This paper also considers a class of MPH "probability freedom models" that includes the product-multinomial and Poisson models of Baker (1994). In particular, Baker (1994) considered product-multinomial models that constrain outcome probabilities through $\pi_{kj} = g_{kj}(\boldsymbol{\beta})/\sum_j g_{kj}(\boldsymbol{\beta})$, where k indexes the independent multinomials and $\boldsymbol{\beta}$ is free to take on values in some unrestricted space. That paper argued that for these models, the product-multinomial estimate of $\boldsymbol{\beta}$ and the corresponding approximating variance estimate are identical to those for a particular, related Poisson model, a model that is arguably simpler to fit. Instead of restricting attention to a product-multinomial model and its Poisson relative, we more generally consider an MPH model and its *population equivalent* Poisson model (as defined below in Section 2). We also give a more detailed comparison of the maximum likelihood fit results for these population equivalent models. For example, estimates of $\boldsymbol{\pi}$, $\mathbf{m}$, and $\boldsymbol{\beta}$; their corresponding approximating variances; and goodness-of-fit statistics are compared, using results of Lang (2000).

This paper is organized as follows. Section 2 gives a brief overview of Lang (2000) and states several of the more relevant results. Section 3 introduces, and explores large-sample likelihood-based inference for, the class of *homogeneous linear predictor models.* Section 4 describes analogous results for the class of *probability freedom models.* Section 5 gives a summary and brief discussion.

2 Overview of MPH Model Results

The current paper will use many of the same notations that were used in Lang (2000). A sample of these notations follows: The symbol $\mathbf{D}^{\alpha}(\mathbf{m})$ (or $diag^{\alpha}\{m_i,\ i = 1, \ldots, c\}$) represents the α^{th} power, where α is any real number, of the diagonal matrix with the components in $\mathbf{m}$ on the diagonal. Functions that typically operate on scalars, like powers and logarithms, act on vectors in a component-wise fashion. For example, $\log \boldsymbol{\delta}$ is defined as $(\log \delta_1, \ldots, \log \delta_s)^T$, where the T represents the transpose. To denote a sum over a certain dimension of an array, a '+' sign will be used. For example, a matrix $\mathbf{Z}$ with components Z_{ik} has kth column sum equal to Z_{+k}. The symbol '$\oplus$' represents a direct sum, so for example, $\oplus_{k=1}^{K}\mathbf{A}_k$ is a block diagonal matrix with the $\mathbf{A}_k$'s making up the blocks. The indicator functional $I(\cdot)$ is defined as $I(E) = 1$ or 0 as the condition E is true or false. Finally, when referring to definitions, theorems, sections etc. from Lang (2000), the original numbering used in Lang (2000) followed by the symbol "L" will be used for ease of cross-referencing.

A vector of counts $\mathbf{Y}$ is said to be a multinomial-Poisson (MP) random vector if it comprises independent multinomial and Poisson random variables. The sampling distribution of an MP random vector is determined by a sampling plan, which in turn is characterized by a population matrix $\mathbf{Z}$, a sampling constraint matrix $\mathbf{Z}_F$ and, when $\mathbf{Z}_F \neq \mathbf{0}$, a vector of a priori known sample sizes $\mathbf{n} > 0$. Population matrix $\mathbf{Z}$ satisfies the properties (i) $Z_{ik} = I(Z_{ik} = 1)$, (ii) $Z_{i+} = 1$, and (iii) $Z_{+k} \geq 1$; it indicates the stratification scheme. Sampling constraint matrix $\mathbf{Z}_F$ comprises a subset of columns of $\mathbf{Z}$; it indicates which stratum sample sizes are a priori fixed.

The MP sampling distribution can be parameterized in terms of $E(\mathbf{Y}) = \mathbf{m}$ or $(\boldsymbol{\gamma}, \boldsymbol{\pi})$, where $\boldsymbol{\gamma} \equiv \mathbf{Z}^T\mathbf{m}$ is the vector of expected sample sizes and $\boldsymbol{\pi} \equiv \mathbf{D}^{-1}(\mathbf{Z}\mathbf{Z}^T\mathbf{m})\mathbf{m}$ is the vector of outcome probabilities; notice that $\mathbf{m} = \mathbf{D}(\mathbf{Z}\boldsymbol{\gamma})\boldsymbol{\pi}$. The notation $\mathbf{Y} \sim MP_Z(\mathbf{m}|\mathbf{Z}_F, \mathbf{n})$ means that $\mathbf{Y}$ has an MP distribution determined by the sampling plan $(\mathbf{Z}, \mathbf{Z}_F, \mathbf{n})$. The ("unrestricted") parameter space is $\omega(0|\mathbf{Z}_F, \mathbf{n}) \equiv \{\mathbf{m} : \mathbf{m} > 0, \mathbf{Z}_F^T\mathbf{m} = \mathbf{n}\}$ or, in terms of $(\boldsymbol{\gamma}, \boldsymbol{\pi})$, $\{(\boldsymbol{\gamma}, \boldsymbol{\pi}) : \boldsymbol{\gamma} =$

$\mathbf{Q}_F\mathbf{n} + \mathbf{Q}_R\boldsymbol{\delta}, \boldsymbol{\delta} > 0, \boldsymbol{\pi} > 0, \mathbf{Z}^T\boldsymbol{\pi} = \mathbf{1}\}$. The matrix $\mathbf{Q}_F$ satisfies $\mathbf{Z}_F = \mathbf{Z}\mathbf{Q}_F$ and the matrix $\mathbf{Q}_R$ is $\mathbf{Q}_F$'s orthogonal complement. These $\mathbf{Q}$ matrices are such that each column in $\mathbf{Z}$ is included in one of $\mathbf{Z}_F = \mathbf{Z}\mathbf{Q}_F$ or $\mathbf{Z}_R \equiv \mathbf{Z}\mathbf{Q}_R$, but not both. In general, the density of $\mathbf{Y} \sim MP_Z(\mathbf{m}|\mathbf{Z}_F, \mathbf{n})$ has the form

$$P(\mathbf{Y} = \mathbf{y}) = C^* \exp\{\mathbf{y}^T \log \mathbf{m} - \mathbf{1}^T\mathbf{Z}_R^T\mathbf{m}\} = C \exp\{\mathbf{y}^T \log \boldsymbol{\pi} + \mathbf{y}^T\mathbf{Z}_R \log \mathbf{Q}_R^T\boldsymbol{\gamma} - \mathbf{1}^T\mathbf{Q}_R^T\boldsymbol{\gamma}\}, \quad (1)$$

where $C^* = \mathbf{n}!\exp\{-\mathbf{n}^T \log \mathbf{n}\}/\mathbf{y}!$ and $C = \mathbf{n}!/\mathbf{y}!$ when $\mathbf{Z}_F \neq \mathbf{0}$ and $C^* = C = 1/\mathbf{y}!$ when $\mathbf{Z}_F = \mathbf{0}$. Here, the symbol $\mathbf{x}! = (x_1, x_2, \ldots, x_q)^T! \equiv x_1!x_2! \cdots x_q!$.

As an example, suppose that $\mathbf{Y} = (Y_{11}, Y_{12}, Y_{21}, Y_{22})^T \sim MP_Z(\mathbf{m}|\mathbf{Z}_F, \mathbf{n})$, where $\mathbf{Z} = \mathbf{Z}_F = \oplus_{k=1}^{2}\mathbf{1}_2$ and $\mathbf{n} = (15, 25)^T$. Then (Y_{11}, Y_{12}) and (Y_{21}, Y_{22}) are independent multinomial random vectors with sample sizes 15 and 25 and probability vectors $(\pi_{11}, \pi_{12}) = (m_{11}/m_{1+}, m_{12}/m_{1+})$ and $(\pi_{21}, \pi_{22}) = (m_{21}/m_{2+}, m_{22}/m_{2+})$. The $\mathbf{m}$ parameter space is $\{\mathbf{m} : \mathbf{m} > 0, m_{1+} = 15, m_{2+} = 25\}$ and, noting that $\mathbf{Q}_F$ and $\mathbf{Q}_R$ are the identity and zero matrix respectively, the $(\boldsymbol{\gamma}, \boldsymbol{\pi})$ parameter space is $\{(\boldsymbol{\gamma}, \boldsymbol{\pi}) : \gamma_1 = 15, \gamma_2 = 25, \boldsymbol{\pi} > 0, \pi_{1+} = \pi_{2+} = 1\}$. As another example, if $\mathbf{Z}_F = 0$, then $\mathbf{Y}$ would comprise independent Poisson components. The $\mathbf{m}$ parameter space would be $\{\mathbf{m} : \mathbf{m} > 0\}$, and noting that $\mathbf{Q}_F$ and $\mathbf{Q}_R$ are the zero and identity matrix respectively, the $(\boldsymbol{\gamma}, \boldsymbol{\pi})$ parameter space would be $\{(\boldsymbol{\gamma}, \boldsymbol{\pi}) : \gamma_1 > 0, \gamma_2 > 0, \boldsymbol{\pi} > 0, \pi_{1+} = \pi_{2+} = 1\}$. The reader is referred to Sections 2L and 3L (Lang 2000) for further discussion of population and sampling constraint matrices and for more examples of the MP distribution.

An MP model is characterized by a sampling plan $(\mathbf{Z}, \mathbf{Z}_F, \mathbf{n})$ and a model constraint function, $\mathbf{h}$. The notation $\mathbf{Y} \sim MP_Z(\mathbf{h}|\mathbf{Z}_F, \mathbf{n})$ means that $\mathbf{Y} \sim MP_Z(\mathbf{m}|\mathbf{Z}_F, \mathbf{n})$, where $\mathbf{m}$ falls in the parameter space

$$\omega(\mathbf{h}|\mathbf{Z}_F, \mathbf{n}) \equiv \{\mathbf{m} : \mathbf{m} > 0, \mathbf{Z}_F^T\mathbf{m} = \mathbf{n}, \mathbf{h}(\mathbf{m}) = \mathbf{0}\}.$$

The corresponding $(\boldsymbol{\gamma}, \boldsymbol{\pi})$ parameter space is $\{(\boldsymbol{\gamma}, \boldsymbol{\pi}) : \boldsymbol{\gamma} = \mathbf{Q}_F\mathbf{n} + \mathbf{Q}_R\boldsymbol{\delta}, \boldsymbol{\delta} > 0, \boldsymbol{\pi} > 0, \mathbf{Z}^T\boldsymbol{\pi} = \mathbf{1}, \mathbf{h}(\mathbf{D}(\mathbf{Z}\boldsymbol{\gamma})\boldsymbol{\pi}) = \mathbf{0}\}$. The notation $\omega(0|\mathbf{Z}_F, \mathbf{n})$ was used above for the "unrestricted" parameter space, because the "unrestricted" model can be viewed as having model constraint function $\mathbf{h}$ equal to the zero function; for this model there are only sampling constraints. Actually, when $\mathbf{Z}_F = \mathbf{0}$, there are no sampling constraints either; in this case (i.e. Poisson sampling), the notations $\mathbf{Y} \sim MP_Z(\mathbf{h}|0)$ and $\omega(\mathbf{h}|0)$ will be used.

Often the model constraint function $\mathbf{h}$ satisfies certain important properties. For example, $\mathbf{h}$ is typically smooth and the constraints $\mathbf{h}(\mathbf{m}) = \mathbf{0}$ are non-redundant. In addition, $\mathbf{h}$ is often

homogeneous relative to the sampling plan. In particular, Lang (2000) defines a **Z**-homogeneous function as follows:

Definition 3L. *Let* $\Omega = \{\mathbf{x} \in R^c : \mathbf{x} > 0\}$. *A function* $\mathbf{h} : \Omega \to R^u$ *is* **Z**-homogeneous [of order $\mathbf{p} = (p(1), \ldots, p(u))^T$] *if*

$$\mathbf{h}(\mathbf{D}(\mathbf{Z}\boldsymbol{\delta})\mathbf{x}) = \mathbf{G}(\boldsymbol{\delta})\mathbf{h}(\mathbf{x}), \quad \forall \boldsymbol{\delta} > 0, \ \forall \mathbf{x} \in \Omega,$$

where $\mathbf{G}(\boldsymbol{\delta}) = diag\{\delta_{\nu(j)}^{p(j)} : j = 1, \ldots, u\}$. *Here,* **Z** *is a* $c \times K$ *population matrix and* $\nu(j) \in \{1, \ldots, K\}$. *When the orders are not important the phrase in square brackets is omitted. The function* **h** *is* **Z**-homogeneous of order 0 *if* $\mathbf{p} = \mathbf{0}$. *In this special case,*

$$\mathbf{h}(\mathbf{D}(\mathbf{Z}\boldsymbol{\delta})\mathbf{x}) = \mathbf{h}(\mathbf{x}), \quad \forall \boldsymbol{\delta} > 0, \ \forall \mathbf{x} \in \Omega.$$

Lang (2000) gave several useful properties of **Z**-homogeneous functions. As an example, a very useful property is as follows:

Proposition 4L (Generalized Euler's Homogeneous Function Theorem). *Provided first-order derivatives exist,* **h** *is* **Z***-homogeneous of order* **p** *if and only if*

$$\mathbf{Z}^T\mathbf{D}(\mathbf{x})\mathbf{H}(\mathbf{x}) = \mathbf{A}\mathbf{D}(\mathbf{p})\mathbf{D}(\mathbf{h}(\mathbf{x})), \ \forall \mathbf{x} \in \Omega,$$

where the matrix **A** *has components that satisfy* $A_{ij} = I(A_{ij} = 1)$ *and* $A_{+j} = 1$. *Moreover,* $\mathbf{A}\mathbf{D}(\mathbf{p}) = \partial \mathbf{d}_{\mathbf{G}}(\mathbf{1})^T/\partial\boldsymbol{\delta}$, *where* $\mathbf{d}_{\mathbf{G}}(\boldsymbol{\delta})$ *is the diagonal vector of* $\mathbf{G}(\boldsymbol{\delta})$. *Here,* **G** *is the diagonal matrix satisfying* $\mathbf{h}(\mathbf{D}(\mathbf{Z}\boldsymbol{\delta})\mathbf{x}) = \mathbf{G}(\boldsymbol{\delta})\mathbf{h}(\mathbf{x})$.

Several useful classes of **Z**-homogeneous functions were defined in Lang (2000). For example, $\mathcal{H}(\mathbf{Z})$ [$\mathcal{H}_{\mathbf{p}}(\mathbf{Z})$] is the collection of all **Z**-homogeneous functions [of order **p**]. A particularly useful collection of constraints for modeling purposes is $\mathcal{H}''(\mathbf{Z})$, defined as follows:

Definition 4L. *The set* $\mathcal{H}''(\mathbf{Z})$ *contains all functions* $\mathbf{h} : \Omega \mapsto R^u$ *that satisfy the following four conditions:*

$\mathtt{H}_0$: $\omega(\mathbf{h}|\mathbf{0}) \equiv \{\mathbf{x} : \mathbf{x} > 0, \ \mathbf{h}(\mathbf{x}) = \mathbf{0}\} \neq \emptyset$.

$\mathtt{H}_1$: **h** *has continuous second-order derivatives on* Ω.

$\mathtt{H}_2$: $\mathbf{H}(\mathbf{x}) \equiv \partial \mathbf{h}^T(\mathbf{x})/\partial \mathbf{x}$ *is full column rank* u *on* Ω.

$\mathtt{H}_3$: $\mathbf{h} \in \mathcal{H}(\mathbf{Z})$.

By convention the zero function is also included in $\mathcal{H}''(\mathbf{Z})$.

An MP model with model constraint function **h** that falls in $\mathcal{H}''(\mathbf{Z})$ is called an MP homogeneous-function (MPH) model. Using the properties of **Z**-homogeneous functions, Lang (2000) argued

that the MPH model constraint function $\mathbf{h}$ satisfies the following two important properties: (i) $\mathbf{h}(\mathbf{m}) = \mathbf{0}$ if and only if $\mathbf{h}(\pi) = \mathbf{0}$ and (ii) the collection of constraints, $\mathbf{h}(\pi) = \mathbf{0}$ and $\mathbf{Z}^T\pi = \mathbf{1}$, are non-redundant. As an example of the usefulness of these properties, an MPH model has (γ, π) parameter space that can be written as a product-space, namely

$$\begin{aligned}&\{(\gamma,\pi) : \gamma = \mathbf{Q}_F\mathbf{n} + \mathbf{Q}_R\delta, \delta > 0, \pi > 0, \mathbf{Z}^T\pi = \mathbf{1}, \mathbf{h}(\mathbf{D}(\mathbf{Z}\gamma)\pi) = \mathbf{0}\}\\ &\quad = \{(\gamma,\pi) : \gamma = \mathbf{Q}_F\mathbf{n} + \mathbf{Q}_R\delta, \delta > 0, \pi > 0, \mathbf{Z}^T\pi = \mathbf{1}, \mathbf{h}(\pi) = \mathbf{0}\}\\ &\quad \equiv \mathcal{D}(\mathbf{Z}, \mathbf{Z}_F, \mathbf{n}) \times \omega(\mathbf{h}|\mathbf{Z}, \mathbf{1}),\end{aligned}$$

where $\mathcal{D}(\mathbf{Z}, \mathbf{Z}_F, \mathbf{n}) = \{\gamma : \gamma = \mathbf{Q}_F\mathbf{n} + \mathbf{Q}_R\delta, \delta > 0\}$. Moreover, by properties of functions in $\mathcal{H}''(\mathbf{Z})$, the space $\omega(\mathbf{h}|\mathbf{Z}, \mathbf{1})$ is a manifold (cf. Fleming 1977) and, therefore, topologically well-behaved. It was this manifold, product-space representation that was exploited in the derivations of many of the results in Lang (2000).

Lang (2000) gave formal definitions of model equivalence. Two equivalence-class partitions of $U(\mathbf{y})$, the collection of MPH models for data $\mathbf{y}$, were considered. The first is induced by the equivalence relation $\overset{P}{\approx}$ defined as follows. The symbols $\mathcal{P}$ and $\mathcal{S}(\mathbf{Z})$ represent the collection of population and sampling constraint matrices (relative to $\mathbf{Z}$), respectively.

Definition 5L. *Two models* $\mathcal{M}_1, \mathcal{M}_2 \in U(\mathbf{y})$ *are population equivalent, denoted* $\mathcal{M}_1 \overset{P}{\approx} \mathcal{M}_2$, *if there exists* $\mathbf{Z} \in \mathcal{P}, \mathbf{Z}_{1F}, \mathbf{Z}_{2F} \in \mathcal{S}(\mathbf{Z})$, *and* $\mathbf{h} \in \mathcal{H}''(\mathbf{Z})$ *such that* $\mathcal{M}_1 = MP_Z(\mathbf{h}|\mathbf{Z}_{1F}, \mathbf{n}_1)$ *and* $\mathcal{M}_2 = MP_Z(\mathbf{h}|\mathbf{Z}_{2F}, \mathbf{n}_2)$.

The second equivalence-class partition of $U(\mathbf{y})$ is induced by the equivalence relation $\approx$ defined as follows.

Definition 6L. *Two models* $\mathcal{M}_1, \mathcal{M}_2 \in U(\mathbf{y})$ *are equivalent, denoted* $\mathcal{M}_1 \approx \mathcal{M}_2$, *if there exists* $\mathbf{Z}_1, \mathbf{Z}_2 \in \mathcal{P}, \mathbf{Z}_{1F} \in \mathcal{S}(\mathbf{Z}_1), \mathbf{Z}_{2F} \in \mathcal{S}(\mathbf{Z}_2)$, *and* $\mathbf{h} \in \mathcal{H}''(\mathbf{Z}_1) \cap \mathcal{H}''(\mathbf{Z}_2)$ *such that* $\mathcal{M}_1 = MP_{Z_1}(\mathbf{h}|\mathbf{Z}_{1F}, \mathbf{n}_1)$ *and* $\mathcal{M}_2 = MP_{Z_2}(\mathbf{h}|\mathbf{Z}_{2F}, \mathbf{n}_2)$.

The sets $\mathcal{E}(\mathbf{h}, \mathbf{Z}, \mathbf{y})$ and $\mathcal{E}(\mathbf{h}, \mathbf{y})$ are equivalence classes induced by the relations $\overset{P}{\approx}$ and $\approx$, respectively. Any two models in $\mathcal{E}(\mathbf{h}, \mathbf{Z}, \mathbf{y})$ are population equivalent and any two models in $\mathcal{E}(\mathbf{h}, \mathbf{y})$ are equivalent. If two models are population equivalent then they are equivalent.

Equivalent models can be compared on the basis of their maximum-likelihood fit results, which include point estimates, goodness-of-fit statistics, and asymptotic-based approximating distributions.

Lang (2000) gave several numerical equivalence results for MPH models. For example, if $\mathcal{M}_1$ and $\mathcal{M}_2$ are members of $\mathcal{E}(\mathbf{h}, \mathbf{y})$, and hence equivalent, then, assuming unique existence, the maximum likelihood estimates $\hat{\mathbf{m}}_1$ and $\hat{\mathbf{m}}_2$ both solve the same set of restricted likelihood

equations, viz.

$$\begin{bmatrix} \mathbf{y} - \mathbf{m} + \mathbf{D}(\mathbf{m})\mathbf{H}(\mathbf{m})\boldsymbol{\lambda} \\ \mathbf{h}(\mathbf{m}) \end{bmatrix} = \mathbf{0}, \tag{2}$$

where $\boldsymbol{\lambda}$ is a vector of Lagrange multipliers. This leads to the numerical identities, $\hat{\mathbf{m}}_1 = \hat{\mathbf{m}}_2$, $\hat{\boldsymbol{\lambda}}_1 = \hat{\boldsymbol{\lambda}}_2$, and $\hat{\boldsymbol{\pi}}_1 = \mathbf{N}_1^{-1}\mathbf{N}_2\hat{\boldsymbol{\pi}}_2$, where $\mathbf{N}_i = \mathbf{D}(\mathbf{Z}_i\mathbf{Z}_i^T\mathbf{y}), i = 1, 2$. Also, goodness-of-fit statistics and adjusted residuals (cf. Haberman 1973) for equivalent models are shown to be numerically identical.

Lang (2000) gave several useful asymptotic results for MPH models. Among other things, it was argued that $\hat{\boldsymbol{\pi}}$ is typically a strongly consistent estimator of $\boldsymbol{\pi}$. It also gave the joint limiting normal distribution of ML estimators $(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\lambda}})$. The limiting distribution of these estimators was derived by linearly approximating equations related to the restricted likelihood equations (2). Specifically, properly normalized versions of the estimators were shown to be asymptotically equivalent to a linear function of a properly normalized version of $\mathbf{N}^{-1}\mathbf{Y}$, where $\mathbf{N} = \mathbf{D}(\mathbf{Z}\mathbf{Z}^T\mathbf{Y})$. The limiting distribution of the properly normalized sample proportions $\mathbf{N}^{-1}\mathbf{Y}$ was given and shown to depend only on $\mathbf{Z}$, not on the sampling constraint matrix $\mathbf{Z}_F$. Lang (2000) also proved that for MPH models the three estimators $\hat{\boldsymbol{\gamma}}$, $\hat{\boldsymbol{\pi}}$ and $\hat{\boldsymbol{\lambda}}$ are asymptotically mutually independent. This independence result was exploited in the derivation of the limiting distribution of properly normalized $\hat{\mathbf{m}}$ as well as other estimators.

As in Lang (2000), we will use asymptotic results to derive more practically useful approximation results. To facilitate this, Lang (2000) gave a formal definition of approximate normal distribution. Paraphrasing Definition 7L ... *Suppose that as* $0 < \alpha \to \infty$, $\alpha^s\mathbf{A}^{\mathbf{P}}(\mathbf{U}_\alpha - \boldsymbol{\mu}_\alpha) \xrightarrow{d} N(0, \boldsymbol{\Sigma})$, *where* $\boldsymbol{\mu}_\alpha$ *is a constant sequence,* $\boldsymbol{\Sigma}$ *has positive diagonal terms,* $\mathbf{A}^{\mathbf{P}} = diag\{a_i^{p_i}\}$, *and* $a_i/\alpha \to b_i > 0$. *If the sequence of matrices* $\mathbf{V}_\alpha$, *stochastic or otherwise, satisfies* $\alpha^{2s}\mathbf{A}^{\mathbf{P}}\mathbf{V}_\alpha\mathbf{A}^{\mathbf{P}} \xrightarrow{P} \boldsymbol{\Sigma}$ *then* $\mathbf{U}_\alpha$ *is said to have an approximate normal distribution with mean* $\boldsymbol{\mu}_\alpha$ *and approximating variance* $\mathbf{V}_\alpha$. *We use the notations* $\mathbf{U}_\alpha \sim \widehat{AN}(\boldsymbol{\mu}_\alpha, \mathbf{V}_\alpha)$ *and* $avar(\mathbf{U}_\alpha) = \mathbf{V}_\alpha$.

The approximate normal $\widehat{AN}$ definition is a generalization of the asymptotic normal AN definition of Serfling (1980). The important difference lies in the fact that $\mathbf{V}_\alpha$ in the definition of approximate normal is allowed to be stochastic, whereas Serfling's definition of asymptotic normal restricts attention to non-stochastic $\mathbf{V}_\alpha$. It was pointed out in Part I that if $\mathbf{U}_\alpha$ is $\widehat{AN}(\boldsymbol{\mu}_\alpha, \mathbf{V}_\alpha)$ then $P(\boldsymbol{\theta}^T\mathbf{U}_\alpha \leq q_\alpha)$ is well approximated by $P(N_\alpha \leq q_\alpha|\mathbf{V}_\alpha)$, where $N_\alpha|\mathbf{V}_\alpha \sim N(\boldsymbol{\theta}^T\boldsymbol{\mu}_\alpha, \boldsymbol{\theta}^T\mathbf{V}_\alpha\boldsymbol{\theta})$; the approximation error converges in probability to zero as α increases.

Lang (2000) gave several approximate normal results (Theorem 4L), as well as corresponding

equivalence results. Equivalence results for goodness-of-fit statistics were also given. For economy of space, we do not reproduce those results here.

Finally, Lang (2000) introduced, and explored properties of, the useful class of **Z**-homogeneous statistics. A **Z**-homogeneous statistic of order **p** has the form $\mathbf{S}(\hat{\mathbf{m}})$, where (i) $\hat{\mathbf{m}}$ is the ML estimator under $MP_Z(\mathbf{h}|\mathbf{Z}_F, \mathbf{n})$, with $\mathbf{h} \in \mathcal{H}''(\mathbf{Z})$, (ii) $\mathbf{S} \in \mathcal{H}_\mathbf{p}(\mathbf{Z})$, and (iii) **S** has continuous first order derivatives at the true $\boldsymbol{\pi}$. One useful approximation result is that $\mathbf{S}(\hat{\mathbf{m}}) \sim \widehat{AN}(\mathbf{S}(\mathbf{m}), \frac{\partial \mathbf{S}(\hat{\mathbf{m}})}{\partial \mathbf{m}^T} avar(\hat{\mathbf{m}}) \frac{\partial \mathbf{S}(\hat{\mathbf{m}})^T}{\partial \mathbf{m}})$. In words, for the class of **Z**-homogeneous statistics, the approximating variance can be derived by a formal application of the delta method to $\mathbf{S}(\hat{\mathbf{m}})$. (This formal method does not work for just any smooth function of **m**.) Another result that is useful for comparing approximating distributions across sampling plans is that, when **S** is 0-order **Z**-homogeneous, the approximating variance does not depend on the sampling constraint matrix $\mathbf{Z}_F$. As an example, this implies that the approximating variances of a 0-order **Z**-homogeneous statistic for two population equivalent models are identical.

3 Homogeneous Linear Predictor Models

This section introduces the useful class of homogeneous linear predictor models, which has members that are characterized as follows:

Definition 1. *A* **Z***-homogeneous linear predictor model is an MPH model that constrains* $\mathbf{m} = \mathbf{D}(\mathbf{Z}\boldsymbol{\gamma})\boldsymbol{\pi}$ *through* $\mathbf{L}(\mathbf{m}) = \mathbf{X}\boldsymbol{\beta}$, *where (i)* $\mathbf{L}(\mathbf{m}) = \mathbf{a}(\boldsymbol{\gamma}) + \mathbf{L}(\boldsymbol{\pi})$, *(ii)* $\mathbf{a}(\boldsymbol{\gamma}_1) - \mathbf{a}(\boldsymbol{\gamma}_2) = \mathbf{a}(\boldsymbol{\gamma}_1/\boldsymbol{\gamma}_2) - \mathbf{a}(\mathbf{1})$, *and (iii)* $\mathbf{U}^T\mathbf{L} \in \mathcal{H}''(\mathbf{Z})$ *for full column rank* **U**, *an orthogonal complement to* **X**.

Throughout this section **L** will be referred to as the "link," and **U** will denote a full column rank matrix that spans the space orthogonal to the range space of **X**. For convenience, **X** is assumed to be of full column rank.

It is important to note that, unlike univariate or multivariate generalized linear models (cf. McCullagh and Nelder 1989, Fahrmeir and Tutz 1994), the link **L** is not required to be one-to-one. This greatly broadens the class of models under consideration. Indeed the utility of the constraint specification and corresponding derivation of the approximating distributions is that it is well-suited for models with many-to-one links.

An important goal in the analysis of a homogeneous linear predictor model of the form $MP_Z(\mathbf{h}|\mathbf{Z}_F, \mathbf{n})$, where $\mathbf{h}(\mathbf{m}) = \mathbf{U}^T\mathbf{L}(\mathbf{m}) \in \mathcal{H}''(\mathbf{Z})$, is to find the approximating distribution of $\hat{\boldsymbol{\beta}}$, the ML estimator of $\boldsymbol{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{L}(\mathbf{m})$. As a first step toward this goal, we will

find the limiting distribution of properly normalized $\mathbf{L}(\hat{\mathbf{m}})$. As in Lang (2000), the subsequent limiting results are valid for a sequence of models of the form $\mathbf{Y}_\nu \sim MP_Z(\mathbf{h}|\mathbf{Z}_F, \mathbf{n}_\nu)$, where the components in the expected count vector $E(\mathbf{Y}_\nu) = \mathbf{m}_\nu = \mathbf{D}(\mathbf{Z}\boldsymbol{\gamma}_\nu)\boldsymbol{\pi}_\nu$ approach infinity in such a way that $\boldsymbol{\pi}_\nu \equiv \boldsymbol{\pi}$ is fixed and $\boldsymbol{\gamma}_\nu/\nu \to \mathbf{w} > 0$ as ν approaches infinity. For convenience, the index ν is dropped from the notation and we set $\mathbf{W} = \mathbf{D}(\mathbf{Zw})$, $\mathbf{D} \equiv \mathbf{D}(\boldsymbol{\pi})$, and $\mathbf{H} \equiv \mathbf{H}(\boldsymbol{\pi})$.

By Lemma 6L, $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\pi}}$ are asymptotically independent normal random vectors. Their limiting distributions are given in Lemma 3L and Theorem 2L, respectively. It follows that

$$\begin{aligned}\nu^{1/2}(\mathbf{L}(\hat{\mathbf{m}}) - \mathbf{L}(\mathbf{m})) &= \nu^{1/2}(\mathbf{a}(\hat{\boldsymbol{\gamma}}) + \mathbf{L}(\hat{\boldsymbol{\pi}}) - \mathbf{a}(\boldsymbol{\gamma}) + \mathbf{L}(\boldsymbol{\pi})) \\ &= \nu^{1/2}(\mathbf{a}(\hat{\boldsymbol{\gamma}}/\boldsymbol{\gamma}) - \mathbf{a}(\mathbf{1})) + \nu^{1/2}(\mathbf{L}(\hat{\boldsymbol{\pi}}) - \mathbf{L}(\boldsymbol{\pi})) \\ &\xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)\end{aligned}$$

where

$$\boldsymbol{\Sigma}_1 = \frac{\partial \mathbf{a}(\mathbf{1})}{\partial \boldsymbol{\gamma}^T}\mathbf{D}^{-1}(\mathbf{w})\mathbf{Q}_R\mathbf{Q}_R^T\frac{\partial \mathbf{a}(\mathbf{1})^T}{\partial \boldsymbol{\gamma}}, \quad \boldsymbol{\Sigma}_2 = \frac{\partial \mathbf{L}(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}^T}[\boldsymbol{\Sigma}^* - \mathbf{W}^{-1}\mathbf{D}\mathbf{Z}\mathbf{Z}^T\mathbf{D}]\frac{\partial \mathbf{L}(\boldsymbol{\pi})^T}{\partial \boldsymbol{\pi}},$$

and

$$\boldsymbol{\Sigma}^* = \mathbf{W}^{-1}\mathbf{D} - \mathbf{W}^{-1}\mathbf{D}\mathbf{H}(\mathbf{H}^T\mathbf{D}\mathbf{W}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{D}\mathbf{W}^{-1}. \tag{3}$$

Now, because $\mathbf{a}(\boldsymbol{\gamma}) = \mathbf{L}(\mathbf{m}) - \mathbf{L}(\boldsymbol{\pi})$, the chain rule gives

$$\frac{\partial \mathbf{a}(\boldsymbol{\gamma})^T}{\partial \boldsymbol{\gamma}} = \frac{\partial \mathbf{m}^T}{\partial \boldsymbol{\gamma}}\frac{\partial \mathbf{L}(\mathbf{m})^T}{\partial \mathbf{m}} = \mathbf{Z}^T\mathbf{D}\mathbf{D}^{-1}(\mathbf{Z}\boldsymbol{\gamma})\frac{\partial \mathbf{L}(\boldsymbol{\pi})^T}{\partial \boldsymbol{\pi}},$$

a quantity that, when evaluated at $\boldsymbol{\gamma} = \mathbf{1}$, gives

$$\frac{\partial \mathbf{a}(\mathbf{1})^T}{\partial \boldsymbol{\gamma}} = \mathbf{Z}^T\mathbf{D}\frac{\partial \mathbf{L}(\boldsymbol{\pi})^T}{\partial \boldsymbol{\pi}}.$$

Therefore, the asymptotic variance $\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2$ simplifies to

$$\begin{aligned}\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2 &= \frac{\partial \mathbf{L}(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}^T}[\boldsymbol{\Sigma}^* - \mathbf{W}^{-1}\mathbf{D}\mathbf{Z}\mathbf{Z}^T\mathbf{D} + \mathbf{D}\mathbf{Z}\mathbf{D}^{-1}(\mathbf{w})\mathbf{Q}_R\mathbf{Q}_R^T\mathbf{Z}^T\mathbf{D}]\frac{\partial \mathbf{L}(\boldsymbol{\pi})^T}{\partial \boldsymbol{\pi}} \\ &= \frac{\partial \mathbf{L}(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}^T}[\boldsymbol{\Sigma}^* - \mathbf{W}^{-1}\mathbf{D}\mathbf{Z}_F\mathbf{Z}_F^T\mathbf{D}]\frac{\partial \mathbf{L}(\boldsymbol{\pi})^T}{\partial \boldsymbol{\pi}}\end{aligned}$$

The next theorem, which gives the limiting distribution of properly normalized $\hat{\boldsymbol{\beta}}$, now follows immediately because ML estimator $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{L}(\hat{\mathbf{m}})$. For convenience, set $\mathbf{P}_X \equiv (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$.

Theorem 1. *Suppose that the sequence of homogeneous linear predictor models, $MP_Z(\mathbf{h}|\mathbf{Z}_F, \mathbf{n})$, with $\mathbf{Z}$-homogeneous constraint $\mathbf{h}(\mathbf{m}) = \mathbf{U}^T\mathbf{L}(\mathbf{m})$, hold. Letting $\hat{\boldsymbol{\beta}}$ be the ML estimator of $\boldsymbol{\beta}$ in the expression $\mathbf{L}(\mathbf{m}) = \mathbf{X}\boldsymbol{\beta}$, the following limiting result is obtained.*

$$\nu^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N\left(\mathbf{0},\ \mathbf{P}_X\frac{\partial \mathbf{L}(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}^T}[\boldsymbol{\Sigma}^* - \mathbf{W}^{-1}\mathbf{D}\mathbf{Z}_F\mathbf{Z}_F^T\mathbf{D}]\frac{\partial \mathbf{L}(\boldsymbol{\pi})^T}{\partial \boldsymbol{\pi}}\mathbf{P}_X^T\right),$$

where $\boldsymbol{\Sigma}^*$ is defined in (3).

It is implicit that $\boldsymbol{\beta}$ is a sequence of parameters indexed by ν. Therefore, it is generally not appropriate to make statements like $\hat{\boldsymbol{\beta}} \xrightarrow{P} \boldsymbol{\beta}$; the correct statement is $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \xrightarrow{P} 0$.

The next practically useful approximation result follows from the asymptotic result of Theorem 1 and the approximation result (Ax 4L) of Lang (2000). For convenience, set $\hat{\mathbf{D}} \equiv \mathbf{D}(\hat{\mathbf{m}})$, $\hat{\mathbf{H}} \equiv \mathbf{H}(\hat{\mathbf{m}})$, and $\mathbf{N} = \mathbf{D}(\mathbf{Z}\mathbf{Z}^T\mathbf{Y})$.

Corollary 1. *Under the conditions of Theorem 1,*

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \;\sim\; \widehat{AN}\left(\mathbf{0}, \mathbf{P}_X \frac{\partial \mathbf{L}(\hat{\mathbf{m}})}{\partial \mathbf{m}^T}[avar(\hat{\mathbf{m}})]\frac{\partial \mathbf{L}(\hat{\mathbf{m}})^T}{\partial \mathbf{m}}\mathbf{P}_X^T\right),$$

where $avar(\hat{\mathbf{m}}) = \hat{\mathbf{D}} - \hat{\mathbf{D}}\hat{\mathbf{H}}(\hat{\mathbf{H}}^T\hat{\mathbf{D}}\hat{\mathbf{H}})^{-1}\hat{\mathbf{H}}\hat{\mathbf{D}} - \mathbf{N}^{-1}\hat{\mathbf{D}}\mathbf{Z}_F\mathbf{Z}_F^T\hat{\mathbf{D}}$ *and* $\mathbf{N} = \mathbf{D}(\mathbf{Z}\mathbf{Z}^T\mathbf{Y})$.

It is interesting to note that this approximating variance can be obtained by formally applying the delta method directly to the function $\hat{\boldsymbol{\beta}} = \mathbf{P}_X\mathbf{L}(\hat{\mathbf{m}})$. As pointed out in Lang (2000), it is important to understand that this formal approach does *not* work for all functions.

The approximating variance $avar(\hat{\boldsymbol{\beta}})$ given in Corollary 1 can be rewritten in a form that exploits the simple form of $\mathbf{h}$ and emphasizes the role of the sampling plan on the approximating distribution. We give the result in the form of another corollary.

Corollary 2. *Assume the conditions of Theorem 1 and define* $\mathbf{B} \equiv \frac{\partial \mathbf{L}(\hat{\mathbf{m}})}{\partial \mathbf{m}^T}\hat{\mathbf{D}}\frac{\partial \mathbf{L}(\hat{\mathbf{m}})^T}{\partial \mathbf{m}}$. *Then the approximating variance of* $\hat{\boldsymbol{\beta}}$ *has the following form:* $avar(\hat{\boldsymbol{\beta}}) = \mathbf{V} - \mathbf{A}(\mathbf{Z}_F)$, *where*

$$\mathbf{V} \equiv \mathbf{P}_X[\mathbf{B} - \mathbf{B}\mathbf{U}(\mathbf{U}^T\mathbf{B}\mathbf{U})^{-1}\mathbf{U}^T\mathbf{B}]\mathbf{P}_X^T \text{ and } \mathbf{A}(\mathbf{Z}_F) \equiv \mathbf{P}_X\frac{\partial \mathbf{L}(\hat{\mathbf{m}})}{\partial \mathbf{m}^T}\hat{\mathbf{D}}\mathbf{Z}_F\mathbf{D}^{-1}(\mathbf{Z}_F^T\mathbf{Y})\mathbf{Z}_F^T\hat{\mathbf{D}}\frac{\partial \mathbf{L}(\hat{\mathbf{m}})^T}{\partial \mathbf{m}}\mathbf{P}_X^T.$$

The proof is immediate upon noting that $\mathbf{H}(\hat{\mathbf{m}}) = \frac{\partial \mathbf{L}(\hat{\mathbf{m}})^T}{\partial \mathbf{m}}\mathbf{U}$.

As we shall see, for certain classes of homogeneous linear predictor models, the matrix $\mathbf{V}$ and/or the so-called "adjustment" matrix $\mathbf{A}(\mathbf{Z}_F)$ can be written in simpler and typically more enlightening forms. For example, the adjustment $\mathbf{A}(\mathbf{Z}_F)$ reduces to the zero matrix when $\mathbf{Z}_F = \mathbf{0}$ (i.e. Poisson sampling) or, by Proposition 4L, when the link $\mathbf{L}$ is $\mathbf{Z}$-homogeneous of order 0, i.e., $\mathbf{L} \in \mathcal{H}_0(\mathbf{Z})$.

An interesting simplification of $\mathbf{V}$ can be derived when the derivative matrix $\partial \mathbf{L}(\hat{\mathbf{m}})^T/\partial \mathbf{m}$ is of full column rank. In this case, the matrix $\mathbf{B}$ of Corollary 2 is non-singular. Exploiting the relationship between full-column-rank orthogonal complements $\mathbf{U}$ and $\mathbf{X}$, a matrix algebra result pointed out in Bergsma (1997, p. 137), leads to the identity $\mathbf{V} = (\mathbf{X}^T\mathbf{B}^{-1}\mathbf{X})^{-1}$. We summarize in the form of another corollary to Theorem 1.

Corollary 3. *Assume the conditions of Theorem 1. If, in addition, the derivative matrix* $\partial \mathbf{L}(\mathbf{m})^T/\partial \mathbf{m}$ *is full column rank on* $\omega(\mathbf{h}|\mathbf{0})$*, then the approximating distribution of* $\hat{\boldsymbol{\beta}}$ *is*

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \sim \widehat{AN}\left(\mathbf{0}, \left(\mathbf{X}^T\left[\frac{\partial \mathbf{L}(\hat{\mathbf{m}})}{\partial \mathbf{m}^T}\hat{\mathbf{D}}\frac{\partial \mathbf{L}(\hat{\mathbf{m}})^T}{\partial \mathbf{m}}\right]^{-1}\mathbf{X}\right)^{-1} - \mathbf{A}(\mathbf{Z}_F)\right),$$

where $\mathbf{A}(\mathbf{Z}_F) = \mathbf{P}_X\frac{\partial \mathbf{L}(\hat{\mathbf{m}})}{\partial \mathbf{m}^T}\hat{\mathbf{D}}\mathbf{Z}_F\mathbf{D}^{-1}(\mathbf{Z}_F^T\mathbf{Y})\mathbf{Z}_F^T\hat{\mathbf{D}}\frac{\partial \mathbf{L}(\hat{\mathbf{m}})^T}{\partial \mathbf{m}}\mathbf{P}_X^T$*. In case* $\mathbf{L} \in \mathcal{H}_0(\mathbf{Z})$ *or* $\mathbf{Z}_F = \mathbf{0}$*, the adjustment* $\mathbf{A}(\mathbf{Z}_F) = \mathbf{0}$*.*

Corollaries 1, 2, and 3 afford explicit comparisons of the $\boldsymbol{\beta}$ estimators for two equivalent homogeneous linear predictor models for data $\mathbf{y}$, say $\mathcal{M}_1, \mathcal{M}_2 \in \mathcal{E}(\mathbf{h}, \mathbf{y})$. First note that maximum likelihood estimators $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$ are numerically identical because fitted values $\hat{\mathbf{m}}_1$ and $\hat{\mathbf{m}}_2$ are identical (see Lang 2000). The corollaries give the difference between the approximating variance estimates as $avar(\hat{\boldsymbol{\beta}}_1) - avar(\hat{\boldsymbol{\beta}}_2) = \mathbf{A}(\mathbf{Z}_{2F}) - \mathbf{A}(\mathbf{Z}_{1F})$. This difference is $\mathbf{0}$, i.e. the approximating variance estimates are identical, when the link $\mathbf{L}$ is 0-order $\mathbf{Z}_i$-homogeneous, $i = 1, 2$. It is also $\mathbf{0}$ when $\mathbf{Z}_{1F} = \mathbf{Z}_{2F}$.

The general results of Lang (2000) can be used to make a more comprehensive comparison of maximum likelihood fit results for two equivalent homogeneous linear predictor models $\mathcal{M}_1, \mathcal{M}_2 \in \mathcal{E}(\mathbf{h}, \mathbf{y})$. In particular, (i) point estimates $\hat{\mathbf{m}}_1$ and $\hat{\mathbf{m}}_2$ are identical and the difference between their approximating variances is $avar(\hat{\mathbf{m}}_1) - avar(\hat{\mathbf{m}}_2) = \hat{\mathbf{D}}\mathbf{Z}_{2F}\mathbf{D}^{-1}(\mathbf{Z}_{2F}^T\mathbf{y})\mathbf{Z}_{2F}^T\hat{\mathbf{D}} - \hat{\mathbf{D}}\mathbf{Z}_{1F}\mathbf{D}^{-1}(\mathbf{Z}_{1F}^T\mathbf{y})\mathbf{Z}_{1F}^T\hat{\mathbf{D}}$; (ii) point estimates of outcome probabilities are related according to $\hat{\boldsymbol{\pi}}_1 = \mathbf{N}_1^{-1}\mathbf{N}_2\hat{\boldsymbol{\pi}}_2$, where $\mathbf{N}_i = \mathbf{D}(\mathbf{Z}_i\mathbf{Z}_i^T\mathbf{y})$; (iii) the difference between their approximating variances is $\mathbf{N}_2^{-2}\hat{\mathbf{D}}\mathbf{Z}_2\mathbf{Z}_2^T\hat{\mathbf{D}}\mathbf{N}_2^{-1} - \mathbf{N}_1^{-2}\hat{\mathbf{D}}\mathbf{Z}_1\mathbf{Z}_1^T\hat{\mathbf{D}}\mathbf{N}_1^{-1}$; and (iv) goodness-of-fit statistics and residuals are identical and have the same approximating distributions.

The next two subsections explore specific classes of homogeneous linear predictor models. By restricting attention to these special classes, we can gain further insight into the roles of the link and the sampling plan on the approximation results.

3.1 Generalized Loglinear Models

Generalized loglinear models (GLLM) have the form $\mathbf{L}(\mathbf{m}) = \mathbf{C}\log\mathbf{M}\mathbf{m} = \mathbf{X}\boldsymbol{\beta}$, where $\mathbf{M}$ satisfies (i) $M_{ij} \geq 0$, (ii) $\sum_{j=1}^{c} M_{ij} > 0$, and (iii) for some $j \in \{1, \ldots, K\}$, $\sum_{h=1}^{c} M_{ih}Z_{hj} = \sum_{h=1}^{c} M_{ih}$; the matrices $\mathbf{C}^T$ and $\mathbf{X}$ are assumed to be of full column rank, without loss of generality; and the model space is assumed to be non-empty. The matrix $\mathbf{Z}$ with components Z_{ij} is a sampling constraint matrix. The structure imposed on $\mathbf{M}$ forces $\mathbf{M}\mathbf{m}$ to include linear combinations of the

m_i within, not across, the populations defined by $\mathbf{Z}$. Quite generally, the link $\mathbf{L}$ is a many-to-one function. GLLMs, which have been considered by many authors including Grizzle et al. (1969), Haber (1985), Lang and Agresti (1994), Glonek and McCullagh (1995), Bergsma (1997), and Lang et al. (1999), are becoming increasingly popular. The class of GLLMs includes loglinear, logit, cumulative logit, multivariate logistic, and association-marginal models, to name just a few.

It is often relatively simple to verify that a GLLM is a homogeneous linear predictor model. We must show that $\mathbf{L}(\mathbf{m}) = \mathbf{C}\log\mathbf{M}\mathbf{m}$ satisfies the three conditions in Definition 1.

It will prove useful to define the following matrix.

$$\mathbf{Z}(\mathbf{M}) \equiv \mathbf{D}^{-1}(\mathbf{M1})\mathbf{M}\mathbf{Z}.$$

Loosely, we speak of $\mathbf{Z}(\mathbf{M})$ as being the *population matrix induced by* $\mathbf{M}$. Technically, however, $\mathbf{Z}(\mathbf{M})$ is not generally a member of the class of population matrices as defined in Section 2L of Lang (2000), as it will only satisfy the first two defining characteristics. We point out that $\mathbf{Z}(\mathbf{M}) = \mathbf{D}^{-1}(\mathbf{M}\mathbf{m})\mathbf{M}\mathbf{D}(\mathbf{m})\mathbf{Z}$, for every positive $\mathbf{m}$, and that $\mathbf{Z}(\mathbf{I}) = \mathbf{Z}$. The matrix $\mathbf{Z}(\mathbf{M})$ satisfies the following two properties.

$$\begin{aligned}\mathbf{M}\mathbf{D}(\mathbf{Z}\delta) &= \mathbf{D}(\mathbf{Z}(\mathbf{M})\delta)\mathbf{M}\\ \log(\mathbf{Z}(\mathbf{M})\delta) &= \mathbf{Z}(\mathbf{M})\log\delta.\end{aligned}$$

These properties imply that the first condition of Definition 1 is met. This can be seen as follows:

$$\mathbf{L}(\mathbf{D}(\mathbf{Z}\gamma)\pi) = \mathbf{C}\log\mathbf{M}\mathbf{D}(\mathbf{Z}\gamma)\pi = \mathbf{C}\mathbf{Z}(\mathbf{M})\log\gamma + \mathbf{C}\log\mathbf{M}\pi. \tag{4}$$

Thus, $\mathbf{L}$ has the form $\mathbf{L}(\mathbf{m}) = \mathbf{a}(\gamma) + \mathbf{L}(\pi)$.

The second condition of Definition 1 is also met, because $\mathbf{a}(\gamma_1) - \mathbf{a}(\gamma_2) = \mathbf{C}\mathbf{Z}(\mathbf{M})\log\gamma_1/\gamma_2 = \mathbf{a}(\gamma_1/\gamma_2) - \mathbf{a}(\mathbf{1})$.

Now consider the third and final condition of Definition 1. The constraint specification of a GLLM has the form $\mathbf{h}(\mathbf{m}) = \mathbf{U}^T\mathbf{L}(\mathbf{m}) = \mathbf{U}^T\mathbf{C}\log\mathbf{M}\mathbf{m} = \mathbf{0}$. It will be assumed that $\mathbf{C}$, $\mathbf{M}$, and $\mathbf{U}$ are chosen so that the first three conditions for $\mathbf{h}$ to be in $\mathcal{H}''(\mathbf{Z})$ are satisfied. This is not a restrictive assumption, because $\mathbf{h}$ is smooth, and reasonable model specifications will lead to full column rank $\mathbf{H}$ and non-empty parameter space $\omega(\mathbf{h}|0)$. What remains to be determined is whether $\mathbf{h}$ is $\mathbf{Z}$-homogeneous. The next lemma, which follows immediately from (4), gives a simple sufficient condition.

Lemma 1. *If* $R(\mathbf{CZ(M)}) \subseteq R(\mathbf{X})$ *then* $\mathbf{h}$ *is* $\mathbf{Z}$*-homogeneous of order 0. Further, if* $\mathbf{CZ(M)} = \mathbf{0}$ *then* $\mathbf{C}\log\mathbf{Mm}$ *is* $\mathbf{Z}$*-homogeneous of order 0.*

For the special case of loglinear models, $\log\mathbf{m} = \mathbf{X}\boldsymbol{\beta}$, Lemma 1 states that $\mathbf{h(m)} = \mathbf{U}^T\log\mathbf{m}$ is $\mathbf{Z}$-homogeneous of order 0 if $R(\mathbf{Z}) \subseteq R(\mathbf{X})$, i.e., stated loosely, "if the model includes fixed-by-design parameters." Another special case is the logit model, $\mathbf{C}\log\mathbf{m} = \mathbf{X}\boldsymbol{\beta}$. Often $\mathbf{CZ} = \mathbf{0}$, because $\mathbf{C}$ has rows that create within-population contrasts of $\log m_i$'s. When this is the case, Lemma 1 states that both $\mathbf{C}\log\mathbf{m}$ and $\mathbf{h(m)} = \mathbf{U}^T\mathbf{C}\log\mathbf{m}$ are $\mathbf{Z}$-homogeneous of order 0.

We point out that sometimes $\mathbf{C}\log\mathbf{Mm} = \mathbf{X}\boldsymbol{\beta}$ can be written as $\mathbf{C}_\ell\log\mathbf{M}_\ell\mathbf{m} = \mathbf{X}_\ell\boldsymbol{\beta}_\ell, \ell = 1,\ldots,L$. In this case, $\mathbf{h(m)} = [\mathbf{h}_1(\mathbf{m})^T,\ldots,\mathbf{h}_L(\mathbf{m})^T]^T$ and so $\mathbf{h}$ is $\mathbf{Z}$-homogeneous of order 0 if $R(\mathbf{C}_\ell\mathbf{Z}(\mathbf{M}_\ell)) \subseteq R(\mathbf{X}_\ell), \ell = 1,\ldots,L$.

In summary, a GLLM with non-redundant constraints (so $\mathbf{H}$ is full column rank) and non-empty parameter space is a $\mathbf{Z}$-homogeneous linear predictor model if $R(\mathbf{CZ(M)}) \subseteq R(\mathbf{X})$.

If a GLLM is a $\mathbf{Z}$-homogeneous linear predictor model with $R(\mathbf{CZ(M)}) \subseteq R(\mathbf{X})$, Corollary 2 gives the approximating variance of $\hat{\boldsymbol{\beta}}$, which can be written in the form $avar(\hat{\boldsymbol{\beta}}) = \mathbf{V} - \mathbf{A}(\mathbf{Z}_F)$. Both $\mathbf{V}$ and the adjustment matrix $\mathbf{A}(\mathbf{Z}_F)$ can be simplified for GLLMs, as stated in the next theorem and its corollary.

Theorem 2. *Let GLLM* $\mathcal{M}$ *be a* $\mathbf{Z}$*-homogeneous linear predictor model for data* $\mathbf{y}$ *of the form* $\mathbf{C}\log\mathbf{Mm} = \mathbf{X}\boldsymbol{\beta}$*, and suppose that* $R(\mathbf{CZ(M)}) \subseteq R(\mathbf{X})$*. Then*

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \;\sim\; \widehat{AN}(\mathbf{0}, avar(\hat{\boldsymbol{\beta}}) = \mathbf{V} - \mathbf{AD}^{-1}(\mathbf{Z}_F^T\mathbf{Y})\mathbf{A}^T),$$

where $\mathbf{V}$ *is as defined in Corollary 2 and* $\mathbf{A}$ *is a full-column-rank matrix that satisfies* $\mathbf{XA} = \mathbf{CZ(M)Q}_F = \mathbf{CD}^{-1}(\mathbf{M1})\mathbf{MZ}_F$*. In case* $\mathbf{Z}_F = \mathbf{0}$ *(i.e. Poisson sampling) and/or the link* $\mathbf{C}\log\mathbf{Mm}$ *is* $\mathbf{Z}$*-homogeneous of order 0, so* $\mathbf{CZ(M)} = \mathbf{0}$*, the adjustment matrix* $\mathbf{AD}^{-1}(\mathbf{Z}_F^T\mathbf{Y})\mathbf{A}^T$ *vanishes.*

Proof of Theorem 2: The derivative $\partial\mathbf{L}(\hat{\mathbf{m}})^T/\partial\mathbf{m}$ in the adjustment matrix $\mathbf{A}(\mathbf{Z}_F)$ of Corollary 2 can be replaced by its specific form for GLLMs, namely $\mathbf{M}^T\mathbf{D}^{-1}(\mathbf{M}\hat{\mathbf{m}})\mathbf{C}^T$. Using the fact that $\mathbf{Z(M)} = \mathbf{D}^{-1}(\mathbf{M1})\mathbf{MZ} = \mathbf{D}^{-1}(\mathbf{Mm})\mathbf{MD(m)Z}$ for all positive $\mathbf{m}$, we can write

$$\mathbf{A}(\mathbf{Z}_F) = \mathbf{P}_X\mathbf{CZ(M)Q}_F\mathbf{D}^{-1}(\mathbf{Z}_F^T\mathbf{Y})\mathbf{Q}_F^T\mathbf{Z(M)}^T\mathbf{C}^T\mathbf{P}_X^T. \tag{5}$$

But, there exists a full column rank matrix $\mathbf{A}$ that satisfies $\mathbf{XA} = \mathbf{CZ(M)Q}_F = \mathbf{CD}^{-1}(\mathbf{M1})\mathbf{MZ}_F$, because $R(\mathbf{CZ(M)Q}_F) \subseteq R(\mathbf{CZ(M)}) \subseteq R(\mathbf{X})$. Replacing $\mathbf{CZ(M)Q}_F$ in (5) by $\mathbf{XA}$ gives the

initial result. The vanishing-adjustment result follows from the general form of the adjustment matrix given in Corollary 2. ■

Corollary 4. *Assume the conditions of Theorem 2. If in addition the derivative matrix* $\partial \mathbf{L}(\mathbf{m})^T/\partial \mathbf{m} = \mathbf{M}^T\mathbf{D}^{-1}(\mathbf{Mm})\mathbf{C}^T$ *is of full column rank, then*

$$avar(\hat{\boldsymbol{\beta}}) = \left(\mathbf{X}^T\left[\mathbf{C}\mathbf{D}^{-1}(\mathbf{M}\hat{\mathbf{m}})\mathbf{M}\hat{\mathbf{D}}\mathbf{M}^T\mathbf{D}^{-1}(\mathbf{M}\hat{\mathbf{m}})\mathbf{C}^T\right]^{-1}\mathbf{X}\right)^{-1} - \mathbf{A}\mathbf{D}^{-1}(\mathbf{Z}_F^T\mathbf{Y})\mathbf{A}^T, \quad (6)$$

where $\mathbf{A}$ *is full column rank and satisfies* $\mathbf{XA} = \mathbf{C}\mathbf{D}^{-1}(\mathbf{M1})\mathbf{M}\mathbf{Z}_F$. *In case* $\mathbf{Z}_F = \mathbf{0}$ *(i.e. Poisson sampling) and/or the link* $\mathbf{C}\log\mathbf{Mm}$ *is* $\mathbf{Z}$*-homogeneous of order 0, so* $\mathbf{CZ(M)} = \mathbf{0}$, *the adjustment matrix* $\mathbf{A}\mathbf{D}^{-1}(\mathbf{Z}_F^T\mathbf{Y})\mathbf{A}^T$ *vanishes.*

The proof of Corollary 4 follows immediately from Corollary 3 and Theorem 2. The next corollary, which also follows immediately from Theorem 2, gives a result that is useful for comparing inferences about $\boldsymbol{\beta}$ for equivalent models.

Corollary 5. *Assume the conditions of Theorem 2. If* β_i, *the* i^{th} *component in* $\boldsymbol{\beta}$, *corresponds to a column in* $\mathbf{X}$ *that is not needed to span the range space of* $\mathbf{C}\mathbf{D}^{-1}(\mathbf{M1})\mathbf{M}\mathbf{Z}_F$, *then* $avar(\hat{\beta}_i, \hat{\beta}_j) = V_{ij}, \forall j$, *where* V_{ij} *is the* $(i,j)^{th}$ *element in* $\mathbf{V}$. *That is, the approximating covariances do not depend on the sampling plan.*

Notice that the special class of GLLMs with links of the form $\mathbf{L}(\mathbf{m}) = \mathbf{C}\log\mathbf{m}$, with $\mathbf{C}$ of full row rank, will give

$$avar(\hat{\boldsymbol{\beta}}) = \left(\mathbf{X}^T\left[\mathbf{C}\hat{\mathbf{D}}^{-1}\mathbf{C}^T\right]^{-1}\mathbf{X}\right)^{-1} - \mathbf{A}\mathbf{D}^{-1}(\mathbf{Z}_F^T\mathbf{Y})\mathbf{A}^T,$$

provided $\mathbf{C}$ is full row rank and $R(\mathbf{X}) \supseteq R(\mathbf{CZ})$. In this case, $\mathbf{A}$ satisfies $\mathbf{XA} = \mathbf{CZ}_F$. If $\mathbf{CZ}_F = \mathbf{0}$, as it will, for example, for logit models, then $\mathbf{A} = \mathbf{0}$. Otherwise, for example for loglinear models, the adjustment is generally non-zero. In this case, Corollary 5 can be used to compare equivalent model $\boldsymbol{\beta}$ variance estimators. This generalizes the result of Palmgren (1981) and Lang (1996), where product-multinomial and Poisson loglinear models were compared.

Example 1: Marginal Stochastic Ordering. Let (B, A) be the before-intervention and after-intervention 3-level ordinal response for a randomly selected subject. To be explicit, assume that both responses take on one of the three possible values $x_1 < x_2 < x_3$. It is of interest to test the null hypothesis H_0 : "B and A have the same distribution" against the alternative H_1 : "B is stochastically larger than A;" i.e. $P(B \leq x_j) \leq P(A \leq x_j), j = 1, 2$, with at least one strict inequality. A paired-comparison experiment is conducted; the resulting data are summarized in

the following 3×3 table:

$$\begin{array}{cc|ccc|c} & & \multicolumn{3}{c|}{A} & \\ & & x_1 & x_2 & x_3 & \\ \hline & x_1 & 14 & 4 & 3 & 21 \\ B & x_2 & 10 & 15 & 7 & 32 \\ & x_3 & 8 & 9 & 24 & 41 \\ \hline & & 32 & 28 & 34 & 94 \end{array} \tag{7}$$

That is, the sufficient cross-classification counts are $\mathbf{y} = (y_{11}, y_{12}, \ldots, y_{33})^T = (14, 4, \ldots, 24)^T$, where y_{ij} is the number of $(B = x_i, A = x_j)$ events observed.

A reasonable model for the data is $\mathbf{y} \leftarrow \mathbf{Y} \sim MP_1(\mathbf{m}|\mathbf{Z}_F, n)$, where $\mathbf{Z}_F = 1_9$ or $\mathbf{Z}_F = 0$, depending on whether the sample size 94 was a priori fixed or a realization of a Poisson random variable. For this model, the stochastic ordering hypothesis H_1 is equivalent to $\log m_{1+} \leq \log m_{+1}$, $\log(m_{1+} + m_{2+}) \leq \log(m_{+1} + m_{+2})$, with at least one strict inequality. Equivalently, the hypothesis can be stated in terms of odds, viz.

$$\log \frac{m_{1+}}{m_{2+} + m_{3+}} \leq \log \frac{m_{+1}}{m_{+2} + m_{+3}}, \quad \log \frac{m_{1+} + m_{2+}}{m_{3+}} \leq \log \frac{m_{+1} + m_{+2}}{m_{+3}}.$$

A one-degree of freedom, one-sided test of stochastic ordering can be derived if it can be assumed that $odds(A \leq x_1)/odds(B \leq x_1) = odds(A \leq x_2)/odds(B \leq x_2)$. Under this assumption, a reasonable GLLM is as follows:

$$\log \frac{m_{1+}}{m_{2+} + m_{3+}} = \alpha_1, \quad \log \frac{m_{+1}}{m_{+2} + m_{+3}} = \alpha_1 + \tau, \quad \log \frac{m_{1+} + m_{2+}}{m_{3+}} = \alpha_2, \quad \log \frac{m_{+1} + m_{+2}}{m_{+3}} = \alpha_2 + \tau.$$

In matrix form, this GLLM can be written as $\mathbf{L}(\mathbf{m}) \equiv \mathbf{C} \log \mathbf{M}\mathbf{m} = \mathbf{X}\boldsymbol{\beta}$, where

$$\mathbf{C} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}, \quad \mathbf{M} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{X}\boldsymbol{\beta} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \tau \end{bmatrix}.$$

It is straightforward to see that the link $\mathbf{L}$ is 0-order $\mathbf{1}$-homogeneous and that the constraint function $\mathbf{h}(\mathbf{m}) = \mathbf{U}^T\mathbf{L}(\mathbf{m})$ is a member of $\mathcal{H}_0''(\mathbf{1})$. Hence, this MP GLLM is a homogeneous linear predictor model. Theorem 2 gives the approximating distribution of $\hat{\boldsymbol{\beta}}$. Moreover, because the link $\mathbf{L}$ has full column rank derivative and is $\mathbf{1}$-homogeneous of order 0, Corollary 4 gives

$$avar(\hat{\boldsymbol{\beta}}) = \left(\mathbf{X}^T \left[\mathbf{C}\mathbf{D}^{-1}(\mathbf{M}\hat{\mathbf{m}})\mathbf{M}\hat{\mathbf{D}}\mathbf{M}^T\mathbf{D}^{-1}(\mathbf{M}\hat{\mathbf{m}})\mathbf{C}^T \right]^{-1} \mathbf{X} \right)^{-1}.$$

Notice that inference about β does not depend on the choice of sampling constraint matrix $\mathbf{Z}_F$. That is, whether the counts are realizations of a single multinomial or four independent Poissons, large-sample inferences about β are the same.

The observed data fit the restricted GLLM reasonably well ($G^2 = 1.03$, $df = 1$). The maximum likelihood estimate of τ is $\hat{\tau} = 0.422$ ($ase \equiv \sqrt{avar} = 0.204$). For this restricted model, the parameter τ is zero or positive as H_0 or H_1 is true. Therefore, a reasonable test of H_0 versus H_1 can be obtained by referring the observed statistic $\hat{\tau}/ase(\hat{\tau}) = 0.422/0.204 = 2.07$ to the right tail of the standard normal distribution. The approximate p-value is 0.02, so there appears to be a statistically significant shift in the marginal distributions, with B stochastically larger than A.

Example 2: Marginal Homogeneity. Consider the setting and data of the previous example. Suppose, however, that it is of interest to test H_0 : "B and A have the same distribution" versus the broader alternative H_1^* : "B and A have different distributions." Again, the data are modeled as $\mathbf{y} \leftarrow \mathbf{Y} \sim MP_1(\mathbf{m}|\mathbf{Z}_F, n)$, where $\mathbf{Z}_F = \mathbf{1}_9$ or $\mathbf{Z}_F = 0$ depending on whether the sample size 94 was a priori fixed or a realization of a Poisson random variable. For this model, H_1^* is equivalent to $\log m_{1+} \neq \log m_{+1}$, and/or $\log(m_{1+} + m_{2+}) \neq \log(m_{+1} + m_{+2})$. Therefore, it is reasonable to consider the following GLLM:

$$\log m_{1+} = \alpha_1, \quad \log m_{+1} = \alpha_1 + \tau_1, \quad \log(m_{1+} + m_{2+}) = \alpha_2, \quad \log(m_{+1} + m_{+2}) = \alpha_2 + \tau_2.$$

which has the matrix form $\mathbf{L}(\mathbf{m}) \equiv \log \mathbf{M}\mathbf{m} = \mathbf{X}\beta$, where

$$\mathbf{M} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 \end{bmatrix}, \quad \mathbf{X}\beta = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \tau_1 \\ \tau_2 \end{bmatrix}.$$

This MP GLLM imposes no model constraints, i.e. h is the zero function. Therefore, the model is a homogeneous linear predictor model. Moreover, because $\mathbf{M}$ is full row rank, the link $\mathbf{L}$ has full column rank derivative. It follows that $\hat{\beta}$ has an approximate normal distribution with approximate mean β and, using Corollary 4, approximating variance

$$avar(\hat{\beta}) = \left(\mathbf{X}^T \left[\mathbf{D}^{-1}(\mathbf{M}\hat{\mathbf{m}})\mathbf{M}\hat{\mathbf{D}}\mathbf{M}^T\mathbf{D}^{-1}(\mathbf{M}\hat{\mathbf{m}})\right]^{-1}\mathbf{X}\right)^{-1} - \mathbf{A}\mathbf{D}^{-1}(\mathbf{Z}_F^T\mathbf{Y})\mathbf{A}^T.$$

Here, the matrix $\mathbf{A}$ satisfies $\mathbf{X}\mathbf{A} = \mathbf{D}^{-1}(\mathbf{M}1)\mathbf{M}1Q_F = Q_F 1_4$, where Q_F equals 1 or 0 depending on whether the sample size 94 was fixed a priori or not. In either case, the first two columns of

$\mathbf{X}$ span $Q_F\mathbf{1}_4$, so Corollary 5 implies that inference about the two τ parameters does not depend on the choice of sampling constraint matrix $\mathbf{Z}_F$. That is, whether the counts are realizations of a single multinomial or four independent Poissons, large-sample inferences about $\boldsymbol{\tau}$ are the same.

We fit this unrestricted GLLM and obtained the following maximum likelihood fit results: $\hat{\tau}_1 = 0.421$, $\hat{\tau}_2 = 0.124$,

$$avar(\hat{\boldsymbol{\tau}}) = \begin{bmatrix} 0.0372 & 0.0071 \\ 0.0071 & 0.0085 \end{bmatrix}.$$

For this GLLM, the marginal homogeneity hypothesis H_0 is equivalent to $\boldsymbol{\tau} = \mathbf{0}$. Therefore, a two degree of freedom test of H_0 versus H_1^* could be obtained by referring the Wald statistic $\hat{\boldsymbol{\tau}}^T[avar(\hat{\boldsymbol{\tau}})]^{-1}\hat{\boldsymbol{\tau}} = 5.03$ to the right-hand tail of the central $\chi^2(2)$ distribution. This gives an approximate p-value of 0.08. Alternatively, the likelihood ratio test of H_0 versus H_1^* could be used. This gives $G^2 = 5.21$, $df = 2$ for an approximate p-value of 0.07. Evidently, there is insufficient evidence at the 0.05 level to reject the null hypothesis of marginal homogeneity. This is in contrast to the conclusion of Example 1, where a restricted alternative (i.e. stochastic ordering) to marginal homogeneity was used.

3.2 Homogeneous Linear Predictor Models with 0-order Links

This section restricts attention to MP models that constrain $\mathbf{m}$ through $\mathbf{L}(\mathbf{m}) = \mathbf{X}\boldsymbol{\beta}$, where the link $\mathbf{L}$ is 0-order $\mathbf{Z}$-homogeneous and $\mathbf{U}^T\mathbf{L}$ is in $\mathcal{H}_0''(\mathbf{Z})$. That these models are in fact homogeneous linear predictor models, is straightforward to see—simply set $\mathbf{a}(\cdot)$ equal to the zero function in Definition 1.

Corollary 2 implies that $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \sim \widehat{AN}(\mathbf{0}, \mathbf{V})$ and, hence, the sampling plan plays no role in this approximating distribution. Specifically, because the link function $\mathbf{L}$ is 0-order $\mathbf{Z}$-homogeneous, Proposition 4L implies that adjustment matrix $\mathbf{A}(\mathbf{Z}_F)$ is zero. Evidently, any two equivalent homogeneous linear predictor models with 0-order link give rise to numerically identical $\boldsymbol{\beta}$ estimators that have identical approximating variances.

As noted above, the matrix $\mathbf{V}$ simplifies when, in addition, the derivative matrix $\frac{\partial \mathbf{L}(\hat{\mathbf{m}})^T}{\partial \mathbf{m}}$ is full column rank. For this class of 0-order link models, Corollary 3 states that the approximating variance can be re-expressed simply as

$$avar(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^T\mathbf{B}^{-1}\mathbf{X})^{-1}, \tag{8}$$

where $\mathbf{B}$ is defined in Corollary 2.

The next two examples of 0-order link models make use of these results. They also serve to illustrate just how broad and how useful this class of models is.

Example 1: Trend Model for Qualitative Dispersion. Consider the following table of counts taken from Lloyd (1999, p. 72).

Table 1. Marital Status of Danes Data

Age	Single	Married	Divorced	Sample Gini
17-21	17	1	0	0.105
21-25	16	8	0	0.444
25-30	8	17	1	0.476
30-40	6	22	4	0.477
40-50	5	21	6	0.510
50-60	3	17	8	0.538
60-70	2	8	6	0.594
70+	1	3	5	0.568

The sample Gini-dispersion measure is given for each population, where populations are defined by the eight age groups. Let S and A represent the marital status and age of a randomly selected subject. The true Gini-dispersion measure for population k is given by $G_k = 1 - \sum_{j=1}^{3}[P(S = j|A = k)]^2$, and can be interpreted as follows. Suppose that two subjects are independently and randomly selected from population k; the probability that their outcomes are different is G_k. Of interest is whether there is a positive linear trend in the Gini-dispersion measures over age groups. We restrict attention to the linear trend model $G_k = \beta_0 + \beta_1 x_k,\ k = 1, \ldots, 8,$, where $\{x_k\}$ is a collection of scores assigned to the age groups.

Initially, we'll assume that the data $\mathbf{y} = (y_{11}, y_{12}, y_{13}, y_{21}, \ldots, y_{83})^T = (17, 1, 0, 16, \ldots, 5)^T$ are the result of sampling plan $(\mathbf{Z}_1, \mathbf{Z}_{1F}, \mathbf{n}_1)$, where $\mathbf{Z}_1 = \oplus_{k=1}^{8} \mathbf{1}_3$; the sampling constraint matrix $\mathbf{Z}_{1F}$ is left unspecified. That is, a stratified sample of subjects from each of the 8 age groups is taken; whether a sample size is fixed a priori or is random is left unspecified. For this MP model, say $MP_{Z_1}(\mathbf{m}|\mathbf{Z}_{1F}, \mathbf{n}_1)$, the probabilities are defined as

$$\pi_{kj} = P(S = j|A = k) = \left(\mathbf{D}^{-1}(\mathbf{Z}_1\mathbf{Z}_1^T\mathbf{m})\mathbf{m}\right)_{kj} = \frac{m_{kj}}{m_{k+}}.$$

Thus, $G_k = 1 - \sum_{j=1}^{3} \pi_{kj}^2 = 1 - \sum_{j=1}^{3}(m_{kj}/m_{k+})^2$, and the trend model of interest, say $\mathcal{M}_1 = MP_{Z_1}(\mathbf{h}|\mathbf{Z}_{1F}, \mathbf{n}_1)$, can be specified as

$$\mathbf{L}(\mathbf{m}) = \begin{bmatrix} 1 - \sum_{j=1}^{3}(m_{1j}/m_{1+})^2 \\ 1 - \sum_{j=1}^{3}(m_{2j}/m_{2+})^2 \\ \vdots \\ 1 - \sum_{j=1}^{3}(m_{8j}/m_{8+})^2 \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_8 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \mathbf{X}\beta, \tag{9}$$

or, equivalently, $\mathbf{h}(\mathbf{m}) \equiv \mathbf{U}^T\mathbf{L}(\mathbf{m}) = \mathbf{0}$, where $\mathbf{U}$ is an 8×6 full-column-rank matrix that spans the space orthogonal to the range space of $\mathbf{X}$. This MP model is a homogeneous linear predictor model with 0-order link.

It follows that ML estimator $\hat{\beta} \sim \widehat{AN}(\beta, \mathbf{V})$, or, because $\mathbf{B}$ is non-singular in this setting, $\hat{\beta} \sim \widehat{AN}(\beta, (\mathbf{X}^T\mathbf{B}^{-1}\mathbf{X})^{-1})$. As noted above, this approximate normal distribution does not depend on the sampling constraint matrix $\mathbf{Z}_{1F}$, so the approximation is valid for stratified Poisson or multinomial sampling.

Now suppose hypothetically that instead of a stratified random sample, a cross-sectional random sample was taken, so the population matrix is $\mathbf{Z}_2 = \mathbf{1}_{24}$. Whether the total sample size is fixed a priori or random is left unspecified; that is, the sampling constraint matrix $\mathbf{Z}_{2F}$ is left unspecified. For this MP model, say $MP_{Z_2}(\mathbf{m}|\mathbf{Z}_{2F}, \mathbf{n}_2)$, the probabilities are defined as

$$\pi_{kj} = P(A = k, S = j) = \left(\mathbf{D}^{-1}(\mathbf{Z}_2\mathbf{Z}_2^T\mathbf{m})\mathbf{m}\right)_{kj} = \frac{m_{kj}}{m_{++}},$$

so that $P(S = j|A = k) = \pi_{kj}/\pi_{k+} = m_{kj}/m_{k+}$. This implies that $G_k = 1 - \sum_{j=1}^{3}(m_{kj}/m_{k+})^2$, and the trend model, say $\mathcal{M}_2 = MP_{Z_2}(\mathbf{h}|\mathbf{Z}_{2F}, \mathbf{n}_2)$ can be specified using the same link and design matrix (or same constraint function $\mathbf{h}$) as in (9) for model $\mathcal{M}_1$. Obviously, $\mathbf{L}$ is $\mathbf{Z}_2$-homogeneous of order 0 and the model $\mathcal{M}_2$ is also a homogeneous linear predictor model.

In fact, the two models $\mathcal{M}_1$ and $\mathcal{M}_2$ are equivalent. By results herein, large-sample inferences about β, the Gini dispersion values G_k's, and the goodness of fit of the trend model will be identical for the two models.

The linear trend model $\mathcal{M}_1$ was fitted, with x_k scores $(1.9, 2.3, 2.7, 3.5, 4.5, 5.5, 6.5, 8)$ corresponding to mid-points of the age intervals. The likelihood-ratio statistic for testing the goodness of fit of the linear trend model did not indicate overall lack of fit ($G^2 = 6.62$, $df = 6$). However, the first sample dispersion value 0.105 was much smaller than the fitted value 0.400, indicating the possible need for a more general model. For illustrative purposes, we do not pursue better fitting models here. Instead, we address our objective of testing whether the linear trend term β_1 is positive. To this purpose, we refer the observed Wald statistic $\hat{\beta}_1/\sqrt{avar(\hat{\beta}_1)} = 0.0363/0.0140 = 2.59$ to the standard normal distribution, and obtain an approximate p-value of 0.005. There is a statistically significant positive linear trend in the Gini-dispersion measures.

Example 2: Mean response models.

Mean response MP models can be specified as $\mathbf{Y} \sim MP_Z(\mathbf{m}|\mathbf{Z}_F, \mathbf{n})$, where $\mathbf{L}(\mathbf{m}) = \mathbf{M}\pi(\mathbf{m}) = \mathbf{X}\beta$, and $\pi(\mathbf{m}) = \mathbf{D}^{-1}(\mathbf{Z}\mathbf{Z}^T\mathbf{m})\mathbf{m}$ is the vector of probabilities. The link vector $\mathbf{M}\pi(\mathbf{m})$ comprises

a collection of linear combinations of probabilities, where $\mathbf{M}$ is a full row rank matrix of scores. Generally, $\mathbf{M}$ is *not* full column rank so $\mathbf{L}$ is a many-to-one link. Because of this, commonly-used statistical software fits these models using weighted least squares (see, e.g., Stokes et al. 1995, pp. 386-392) rather than maximum likelihood.

The maximum likelihood estimator $\hat{\boldsymbol{\beta}}$ has approximating variance $avar(\hat{\boldsymbol{\beta}}) = \mathbf{V}$ that, for these mean response models, has a particularly simple form. Because $\partial \mathbf{L}(\mathbf{m})^T/\partial \mathbf{m} = \mathbf{D}^{-1}(\mathbf{m})[\mathbf{D}(\boldsymbol{\pi}(\mathbf{m})) - \mathbf{D}(\boldsymbol{\pi}(\mathbf{m}))\mathbf{Z}\mathbf{Z}^T\mathbf{D}(\boldsymbol{\pi}(\mathbf{m}))]\mathbf{M}^T$, a little algebra shows that $\mathbf{B} = \mathbf{M}[avar(\mathbf{N}^{-1}\mathbf{Y})]\mathbf{M}^T$, where $avar(\mathbf{N}^{-1}\mathbf{Y}) = \mathbf{N}^{-1}(\hat{\mathbf{D}} - \mathbf{N}^{-1}\hat{\mathbf{D}}\mathbf{Z}\mathbf{Z}^T\hat{\mathbf{D}})\mathbf{N}^{-1}$ is the approximating variance of the vector of sample proportions. Furthermore, provided the derivative matrix $\partial \mathbf{L}(\mathbf{m})^T/\partial \mathbf{m}$ is of full column rank, as it typically will be, we can give a very simple form for the approximating variance, namely,

$$avar(\hat{\boldsymbol{\beta}}) = \left(\mathbf{X}^T\left[\mathbf{M}[avar(\mathbf{N}^{-1}\mathbf{Y})]\mathbf{M}^T\right]^{-1}\mathbf{X}\right)^{-1}.$$

As before, any model in the same equivalence class $\mathcal{E}(\mathbf{h}, \mathbf{y})$, where $\mathbf{h}(\mathbf{m}) = \mathbf{U}^T\mathbf{M}\boldsymbol{\pi}(\mathbf{m})$, will give numerically identical ML estimates of $\boldsymbol{\beta}$ and numerically identical approximating variances. Similarly, equivalent models will give identical goodness of fit statistics and adjusted residuals.

As a concrete illustration, consider the paired-comparison experiment and resulting data (7) of Example 1, Section 3.1. Assume that it makes sense to assign the scores $x_1 < x_2 < x_3$ to the three levels of both responses. A research hypothesis of interest is $E(A) \neq E(B)$, where $E(A) = \sum_{i=1}^3 x_i P(A = x_i)$ and $E(B) = \sum_{i=1}^3 x_i P(B = x_i)$. That is, the mean after-intervention response is hypothesized to be different than the mean before-intervention response.

As stated previously, a reasonable data model has random component $\mathbf{y} = (y_{11}, y_{12}, \ldots, y_{33})^T = (14, 4, \ldots, 24)^T \leftarrow \mathbf{Y} \sim MP_Z(\mathbf{m}|\mathbf{Z}_F, n)$, where $\mathbf{Z} = \mathbf{1}_9$ and $\mathbf{Z}_F$ is either 0 (Poisson sampling) or 1_9 (full-multinomial sampling with $n = 94$ fixed a priori). For this MP model, $\pi_{ij} = m_{ij}/m_{++} = P(B = x_i, A = x_j)$. A relevant systematic component has the form $\mathbf{L}(\mathbf{m}) = \mathbf{M}\boldsymbol{\pi}(\mathbf{m}) = \mathbf{X}\boldsymbol{\beta}$, where

$$\boldsymbol{\pi}(\mathbf{m}) = \mathbf{m}/m_{++}, \quad \mathbf{M} = \begin{bmatrix} x_1 & x_1 & x_1 & x_2 & x_2 & x_2 & x_3 & x_3 & x_3 \\ x_1 & x_2 & x_3 & x_1 & x_2 & x_3 & x_1 & x_2 & x_3 \end{bmatrix}, \quad \mathbf{X}\boldsymbol{\beta} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} \alpha \\ \beta \end{bmatrix}.$$

Note that $\mathbf{L}(\mathbf{m}) = [E(B), E(A)]^T$ and $\mathbf{L} \in \mathcal{H}_0(\mathbf{Z})$. In fact, this MP model is a $\mathbf{Z}$-homogeneous linear predictor model with 0-order link.

We fitted the model with scores $\{x_i = i\}$ and, for convenience, $\mathbf{Z}_F = 0$. The ML estimates of α and β are $\hat{\alpha} = 2.213$ and $\hat{\beta} = -0.191$; the approximating standard errors are $ase(\hat{\alpha}) \equiv$

$\sqrt{avar(\hat{\alpha})} = 0.081$ and $ase(\hat{\beta}) = 0.089$. These estimates and standard errors would not change if the multinomial model with $\mathbf{Z}_F = \mathbf{1}_9$ were fitted. Now, because $\beta = E(A) - E(B)$, we can test whether $E(A) \neq E(B)$ by referring the observed Wald statistic $\left(\hat{\beta}/ase(\hat{\beta})\right)^2 = 4.592$ to the $\chi^2(1)$ distribution. The approximate p-value is 0.032. Alternatively, a test of $E(A) = E(B)$ vs. $E(A) \neq E(B)$ could be based on the goodness of fit of the reduced model with $\mathbf{X} = [1,\ 1]^T$. For this reduced model, the observed likelihood-ratio statistic is $G^2 = 4.486$ ($df = 1$); referring this value to the $\chi^2(1)$ distribution gives an approximate p-value of 0.034. Note that the value of G^2 and its $\chi^2(1)$ approximating distribution would not change if, instead of the Poisson model, the multinomial model with $\mathbf{Z}_F = \mathbf{1}_9$ were fitted. In sum, because the p-values are relatively small (0.032 and 0.034), we conclude that there is statistical evidence that the before and after mean responses are different. As an aside, it is interesting to note that if the row and column marginal counts of (7) were incorrectly treated as independent multinomial realizations, one would conclude that there is insufficient evidence to reject the null hypothesis $E(A) = E(B)$—the null model goodness-of-fit statistic value is $G^2 = 2.596$ ($df = 1$), which gives an approximate p-value equal to 0.107.

Before closing this section, we make two comments. First, although we gave just two examples, the 0-order homogeneous linear predictor models include many other important many-to-one link models. Example 1 used the Gini-dispersion and Example 2 used a mean score, but any distribution summary measure could be used. Moreover, these measures need not summarize univariate distributions; one could compare bivariate association measures, like the kappa (cf. Cohen, 1960) or gamma (cf. Goodman and Kruskal, 1979) statistics, across populations (see, e.g., Carr et al. 1989 or Stokes 1999). It is also important to note that we do not need independent estimates of the distributions that are to be summarized and compared. As illustrated by the marginal mean response model example above, the margins of multivariate distributions can be summarized and compared as well.

The second comment regards the fitting method. The form of the approximating variance of equation (8) hints at a possible relationship between the maximum likelihood and weighted least squares (Grizzle et al. 1969) fit results. Indeed, for 0-order homogeneous linear predictor models that satisfy the conditions of Corollary 3, it can be shown that the approximating variance of the ML estimator $\hat{\boldsymbol{\beta}}$ given in (8) also serves as an approximating variance of the weighted least squares (WLS) estimator $\hat{\boldsymbol{\beta}}_w$. In practice, the ML approximating variance is not used in the

WLS analysis for the obvious reason—it requires computation of the ML estimates. Instead, an asymptotically equivalent estimate based on the empirical proportions is used. We point out that in sparse data settings this empirical-proportion-based estimate can have poorer finite-sample properties than the ML estimate. In a subsequent paper, the ML-WLS relationship is more fully explored.

4 Probability Freedom Models

Baker (1994) considered product-multinomial and Poisson models that, using the current notation, constrain the probabilities through $\boldsymbol{\pi} = \mathbf{D}^{-1}(\mathbf{Z}\mathbf{Z}^T\mathbf{g}(\boldsymbol{\beta}))\mathbf{g}(\boldsymbol{\beta})$, where $\boldsymbol{\beta}$ is a $p \times 1$ "freedom" parameter; i.e. allowable values of the vector $\boldsymbol{\beta}$ comprise a p-dimensional set that without loss of generality can be taken to be R^p. That paper argued that for these models, the product-multinomial estimate of $\boldsymbol{\beta}$ and the corresponding approximating variance estimate are identical to those for a particular, related Poisson model, a model that is arguably simpler to fit. Baker (1994) described the multinomial-to-Poisson transformation, which transforms a product-multinomial model into the appropriate Poisson model. As a simple example, consider a 2×2 table where the row counts make up two independent multinomials. The product-multinomial model of row homogeneity can be specified as $\boldsymbol{\pi} = (\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22})^T = (e^{\beta}/(1+e^{\beta}),\ 1/(1+e^{\beta}),\ e^{\beta}/(1+e^{\beta}),\ 1/(1+e^{\beta}))^T$. The multinomial-to-Poisson transform gives the corresponding Poisson loglinear model as $\log \mathbf{m} = (\phi_1 + \beta, \phi_1, \phi_2 + \beta, \phi_2)^T$. Baker (1994) argues that inferences about β are identical for the product-multinomial and Poisson models. Baker gives several other useful examples.

Instead of restricting attention to a product-multinomial model and its Poisson relative, we more generally consider an MP model and its *population equivalent* Poisson model (as defined in Section 2). The equivalence results of Lang (2000), some of which are included in this paper, can be used to compare the MP and Poisson model maximum likelihood fit results, including point estimates of $\boldsymbol{\pi}$, $\mathbf{m}$, and $\boldsymbol{\beta}$; the corresponding approximating variances; and goodness-of-fit statistics.

Recalling that $\boldsymbol{\pi} = \mathbf{D}^{-1}(\mathbf{Z}\mathbf{Z}^T\mathbf{m})\mathbf{m}$ in an MP model, the MP analogue of the product-multinomial freedom model of Baker (1994) can be written as $\mathcal{M}_1 : \mathbf{y} \leftarrow \mathbf{Y} \sim MP_Z(\mathbf{m}|\mathbf{Z}_F, \mathbf{n})$, where $\mathbf{m}$ falls in the parameter space

$$\omega_1 \equiv \{\mathbf{m} : \mathbf{m} > 0, \mathbf{Z}_F^T\mathbf{m} = \mathbf{n},\ \mathbf{D}^{-1}(\mathbf{Z}\mathbf{Z}^T\mathbf{m})\mathbf{m} = \mathbf{D}^{-1}(\mathbf{Z}\mathbf{Z}^T\mathbf{g}(\boldsymbol{\beta}))\mathbf{g}(\boldsymbol{\beta}),\ \boldsymbol{\beta} \in R^p\}.$$

Here, as in Baker (1994), $\mathbf{g}$ is a positive function and the sufficiently smooth function $\mathbf{f}(\boldsymbol{\beta}) \equiv \mathbf{D}^{-1}(\mathbf{Z}\mathbf{Z}^T\mathbf{g}(\boldsymbol{\beta}))\mathbf{g}(\boldsymbol{\beta})$ is one-to-one on R^p. For definiteness, suppose that $\mathbf{Z}$ is $c \times K$.

Define $\mathbf{f}_e^{-1} : \omega(0|\mathbf{Z}, \mathbf{1}) \mapsto R^p$ as the extension of the one-to-one inverse function $\mathbf{f}^{-1} : \mathbf{f}(R^p) \mapsto R^p$. Also, define the link function $\mathbf{L}$ as $\mathbf{L}(\mathbf{m}) = \mathbf{D}^{-1}(\mathbf{Z}\mathbf{Z}^T\mathbf{m})\mathbf{m}$. It follows that

$$\begin{aligned}\omega_1 &= \{\mathbf{m} : \mathbf{m} > 0, \mathbf{Z}_F^T\mathbf{m} = \mathbf{n}, \mathbf{L}(\mathbf{m}) = \mathbf{f}(\boldsymbol{\beta}), \boldsymbol{\beta} \in R^p\} \\ &= \{\mathbf{m} : \mathbf{m} > 0, \mathbf{Z}_F^T\mathbf{m} = \mathbf{n}, \mathbf{L}(\mathbf{m}) - \mathbf{f}(\mathbf{f}_e^{-1}(\mathbf{L}(\mathbf{m}))) = \mathbf{0}\} \\ &= \{\mathbf{m} : \mathbf{m} > 0, \mathbf{Z}_F^T\mathbf{m} = \mathbf{n}, \mathbf{h}(\mathbf{m}) = \mathbf{0}\} \\ &= \omega(\mathbf{h}|\mathbf{Z}_F, \mathbf{n}),\end{aligned}$$

where $\mathbf{h}(\mathbf{m}) = \mathbf{0}$ comprises the $u \equiv c - K - p$ non-redundant constraints in $\mathbf{L}(\mathbf{m}) - \mathbf{f}(\mathbf{f}_e^{-1}(\mathbf{L}(\mathbf{m}))) = \mathbf{0}$. Note that for sufficiently smooth $\mathbf{f}$, the constraint function $\mathbf{h}$, which only depends on $\mathbf{m}$ through $\mathbf{L}(\mathbf{m})$, will fall in $\mathcal{H}_0''(\mathbf{Z})$. Thus, the MP model $\mathcal{M}_1 = MP_Z(\mathbf{h}|\mathbf{Z}_F, \mathbf{n})$ is homogeneous and, hence, a member of the equivalence class $\mathcal{E}(\mathbf{h}, \mathbf{Z}, \mathbf{y})$.

Consider the population equivalent Poisson model $\mathcal{M}_2 = MP_Z(\mathbf{h}|0) \in \mathcal{E}(\mathbf{h}, \mathbf{Z}, \mathbf{y})$, with parameter space

$$\omega(\mathbf{h}|0) = \{\mathbf{m} : \mathbf{m} > 0, \mathbf{h}(\mathbf{m}) = \mathbf{0}\} = \{\mathbf{m} : \mathbf{m} > 0, \mathbf{L}(\mathbf{m}) = \mathbf{f}(\boldsymbol{\beta}), \boldsymbol{\beta} \in R^p\}.$$

This parameter space can be re-written in a useful way, namely,

$$\omega(\mathbf{h}|0) = \{\mathbf{m} : \mathbf{m} = \mathbf{D}(\mathbf{Z}\boldsymbol{\gamma})\mathbf{g}(\boldsymbol{\beta}), \boldsymbol{\gamma} > 0, \boldsymbol{\beta} \in R^p\} \equiv \omega_2.$$

To see this, note that $\mathbf{m} \in \omega(\mathbf{h}|0)$ implies that

$$\begin{aligned}\mathbf{m} &= \mathbf{D}(\mathbf{Z}\mathbf{Z}^T\mathbf{m})\mathbf{D}^{-1}(\mathbf{Z}\mathbf{Z}^T\mathbf{g}(\boldsymbol{\beta}))\mathbf{g}(\boldsymbol{\beta}) \\ &= \mathbf{D}(\mathbf{Z}(\mathbf{Z}^T\mathbf{m}/\mathbf{Z}^T\mathbf{g}(\boldsymbol{\beta})))\mathbf{g}(\boldsymbol{\beta}) = \mathbf{D}(\mathbf{Z}\boldsymbol{\gamma})\mathbf{g}(\boldsymbol{\beta}),\end{aligned}$$

where $\boldsymbol{\gamma} = \mathbf{Z}^T\mathbf{m}/\mathbf{Z}^T\mathbf{g}(\boldsymbol{\beta}) > 0$. Thus, $\mathbf{m} \in \omega_2$. If $\mathbf{m} \in \omega_2$ then, using properties of population matrix $\mathbf{Z}$,

$$\begin{aligned}\mathbf{D}^{-1}(\mathbf{Z}\mathbf{Z}^T\mathbf{m})\mathbf{m} &= \mathbf{D}^{-1}(\mathbf{Z}\mathbf{Z}^T\mathbf{D}(\mathbf{Z}\boldsymbol{\gamma})\mathbf{g}(\boldsymbol{\beta}))\mathbf{D}(\mathbf{Z}\boldsymbol{\gamma})\mathbf{g}(\boldsymbol{\beta}) \\ &= \mathbf{D}^{-1}(\mathbf{Z}\boldsymbol{\gamma})\mathbf{D}^{-1}(\mathbf{Z}\mathbf{Z}^T\mathbf{g}(\boldsymbol{\beta}))\mathbf{D}(\mathbf{Z}\boldsymbol{\gamma})\mathbf{g}(\boldsymbol{\beta}) \\ &= \mathbf{D}^{-1}(\mathbf{Z}\mathbf{Z}^T\mathbf{g}(\boldsymbol{\beta}))\mathbf{g}(\boldsymbol{\beta})\end{aligned}$$

and $\mathbf{m} = \mathbf{D}(\mathbf{Z}\boldsymbol{\gamma})\mathbf{g}(\boldsymbol{\beta}) > 0$. Thus, $\mathbf{m} \in \omega(\mathbf{h}|0)$.

Summarizing, we have that the MP model $\mathcal{M}_1$ and the Poisson model $\mathcal{M}_2$ are population equivalent homogeneous models. Model $\mathcal{M}_1$ has parameter space $\omega_1 \equiv \{\mathbf{m} : \mathbf{m} > 0, \mathbf{Z}_F^T\mathbf{m} = \mathbf{n}, \boldsymbol{\pi} = \mathbf{D}^{-1}(\mathbf{Z}\mathbf{Z}^T\mathbf{g}(\boldsymbol{\beta}))\mathbf{g}(\boldsymbol{\beta}),\ \boldsymbol{\beta} \in R^p\}$ and model $\mathcal{M}_2$ has parameter space $\omega_2 \equiv \{\mathbf{m} : \mathbf{m} = \mathbf{D}(\mathbf{Z}\boldsymbol{\gamma})\mathbf{g}(\boldsymbol{\beta}), \boldsymbol{\gamma} > 0, \boldsymbol{\beta} \in R^p\}$.

Using the equivalence results of Lang (2000), we make several useful observations regarding the comparison of the fit results for the two models $\mathcal{M}_1$ and $\mathcal{M}_2$. The first of these observations leads to a comparison result that was noted by Baker (1994) for the special-case product-multinomial setting; the remaining observations give a more complete description of the comparisons. (1) The ML estimator of β under $\mathcal{M}_i$, say $\hat{\beta}_i$, is equal to $\mathbf{f}^{-1}(\mathbf{L}(\hat{\mathbf{m}}_i))$, and hence is a **Z**-homogeneous statistic of order 0 (see Section 2). The **Z**-homogeneous statistics results of Section 7L (Lang 2000), imply that not only are $\hat{\beta}_1$ and $\hat{\beta}_2$ numerically identical, they also have identical approximating variance estimates. (2) Goodness-of-fit statistics and adjusted residuals are identical. (3) Fitted values are numerically identical, i.e. $\hat{\mathbf{m}}_1 = \hat{\mathbf{m}}_2$. (4) Their approximating variance estimates are related through $avar(\hat{\mathbf{m}}_1) = avar(\hat{\mathbf{m}}_2) - \mathbf{N}^{-1}\hat{\mathbf{D}}\mathbf{Z}_F\mathbf{Z}_F^T\hat{\mathbf{D}}$, where $\mathbf{N} = \mathbf{D}(\mathbf{Z}\mathbf{Z}^T\mathbf{y})$ and $\hat{\mathbf{D}} = \mathbf{D}(\hat{\mathbf{m}}_2)$. (5) Probability estimates are numerically identical; i.e. $\hat{\boldsymbol{\pi}}_1 = \hat{\boldsymbol{\pi}}_2$. (6) Their approximating variance estimates are identical and can be computed as $avar(\hat{\boldsymbol{\pi}}_1) = avar(\hat{\boldsymbol{\pi}}_2) = \mathbf{N}^{-1}avar(\hat{\mathbf{m}}_2)\mathbf{N}^{-1} - \mathbf{N}^{-2}\hat{\mathbf{D}}\mathbf{Z}\mathbf{Z}^T\hat{\mathbf{D}}\mathbf{N}^{-1}$.

Example. Consider the MP model $\mathbf{Y} \sim MP_Z(\mathbf{m}|\mathbf{Z}_F, \mathbf{n})$ for data $\mathbf{y} = (y_{11}, y_{12}, y_{21}, y_{22})^T$, where $\mathbf{Z} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}^T$ and $\mathbf{m}$ is constrained through

$$\mathbf{L}(\mathbf{m}) \equiv \mathbf{D}^{-1}(\mathbf{Z}\mathbf{Z}^T\mathbf{m})\mathbf{m} = \boldsymbol{\pi} = \mathbf{D}^{-1}(\mathbf{Z}\mathbf{Z}^T\mathbf{g}(\beta))\mathbf{g}(\beta) \equiv \mathbf{f}(\beta),$$

where $\mathbf{g}(\beta) = (\mathrm{e}^{\beta}, 1, \mathrm{e}^{\beta}, 1)^T$. This is the model of row homogeneity. For this model $f^{-1} : \mathrm{f}(R) \mapsto R$ can be defined as $f^{-1}(\mathbf{x}) = \log(1 - x_2) - \log(x_2)$. The extension f_e^{-1} has the same definition, but its domain is $\{\mathbf{x} : \mathbf{x} > 0, \mathbf{Z}^T\mathbf{x} = \mathbf{1}\}$. It follows that $\mathbf{L}(\mathbf{m}) - \mathrm{f}(f_e^{-1}(\mathbf{L}(\mathbf{m}))) = (0, 0, m_{21}/m_{2+} - m_{11}/m_{1+}, m_{22}/m_{2+} - m_{12}/m_{1+})^T$, and the one $(c - K - p = 4 - 2 - 1 = 1)$ non-redundant constraint is $h(\mathbf{m}) \equiv m_{21}/m_{2+} - m_{11}/m_{1+} = 0$.

The population equivalent Poisson model $MP_Z(h|0)$ has parameter space $\omega(\mathbf{h}|0) = \{\mathbf{m} : \mathbf{m} = \mathrm{D}(\mathbf{Z}\boldsymbol{\gamma})\mathbf{g}(\beta), \boldsymbol{\gamma} > 0, \beta \in R\}$, which can be written as

$$\omega(\mathbf{h}|0) = \{\mathbf{m} : m_{k1} = \gamma_k \mathrm{e}^{\beta}, m_{k2} = \gamma_k, \gamma_k > 0, k = 1, 2,\ \beta \in R\}.$$

This Poisson model has the simple loglinear form $\log \mathbf{m} = (\phi_1 + \beta, \phi_1, \phi_2 + \beta, \phi_2)^T$, where $\phi_k = \log(\gamma_k)$. By the results of this section, this Poisson loglinear model can be fit, and the fit results explicitly modified so that they coincide with the fit results for the population equivalent MP freedom model.

5 Discussion

This paper set out to explore properties of maximum likelihood fit results for two important subclasses of MPH models. Both subclasses, homogeneous linear predictor models and probability freedom models, have the generic form $\mathbf{L}(\mathbf{m}) = \mathbf{f}(\boldsymbol{\beta})$, where $\mathbf{m}$ is the vector of expected counts and $\boldsymbol{\beta}$ is a parameter vector of interest. Exploiting the special structure of these models, the large-sample behaviors of maximum likelihood estimators such as $\hat{\boldsymbol{\beta}}$ under equivalent MPH models were described and compared. These results use asymptotic arguments that are valid provided the numbers of constraints and populations are fixed and all of the expected counts approach infinity at the same rate. This precludes the regression setting where the numbers of constraints and populations grow concomitantly.

As the examples of Section 3 illustrate, the class of homogeneous linear predictor models is very rich, because the link $\mathbf{L}$ in $\mathbf{L}(\mathbf{m}) = \mathbf{X}\boldsymbol{\beta}$ is allowed to be many-to-one. When $\mathbf{L}$ is many-to-one, the likelihood cannot be reparameterized in terms of $\boldsymbol{\beta}$ alone, so standard methods for obtaining maximum likelihood fit results are not applicable. For this reason, these many-to-one link models are typically fitted using non-likelihood methods such as weighted least squares (cf. Grizzle et al. 1969, Stokes et al. 1995). This paper uses the less standard approach of Aitchison and Silvey (e.g., 1958, 1960) and rewrites the models in terms of constraint functions. Using the constraint approach, maximum likelihood estimation is straightforward, and describing the large-sample behavior of estimators is relatively simple. Moreover, the constraint approach is enlightening in that the effect of the sampling plan on the large-sample behavior of maximum likelihood estimators and goodness of fit statistics can be seen in very explicit form.

The class of probability freedom models of Section 4 contains the class of models considered in Baker (1994). The class herein is broader because many more sampling plans are considered– we do not restrict attention to Poisson and product-multinomial counts; instead, counts are allowed to be realizations of random vectors that have a multinomial-Poisson (MP) distribution. Baker (1994) gives several examples to illustrate the usefulness of the multinomial-to-Poisson transformation. In particular, for each example, Baker argues that inferences about $\boldsymbol{\beta}$ in $\boldsymbol{\pi} = \mathbf{f}(\boldsymbol{\beta})$ are identical for both the original product-multinomial and the transformed Poisson model. The current paper expands on this not only by considering the broader collection of MP models, but also by giving a more complete comparison of maximum likelihood fit results for the MP model

and the population equivalent Poisson model.

Finally, as mentioned previously, maximum likelihood fitting for homogeneous linear predictor models is relatively straightforward. In fact, maximum likelihood fitting is straightforward for any MPH model (including probability freedom models) that can be explicitly written in constraint form. The author has written a computer program that produces maximum likelihood fit results for any MPH model. The program uses a modified Newton-Raphson algorithm that is similar in spirit to the algorithms of Aitchison and Silvey (1958) and Lang and Agresti (1994) to solve the restricted likelihood equations (2) described in Section 2.

6 References

Agresti, A. (1990), *Categorical Data Analysis,* New York: John Wiley & Sons.

Aitchison, J. and Silvey, S.D. (1958), "Maximum-Likelihood Estimation of Parameters Subject to Restraints," *Ann. Math. Statist.*, **29**, 813-828.

Aitchison, J. and Silvey, S.D. (1960), "Maximum-Likelihood Estimation Procedures and Associated Tests of Significance," *J. Royal Statist. Soc. - B*, **1**, 154-171.

Baker, S.G. (1994), "The Multinomial-Poisson Transformation," *The Statistician,* **43**, No. 4, 495-504.

Bergsma, W.P. (1997), *Marginal Models for Categorical Data*, Tilburg: Tilburg University Press.

Carr, G.J., Hafner, K.B., and Koch, G.G. (1989), "Analysis of Rank Measures of Association for Ordinal Data from Longitudinal Studies," *J. Amer. Statist. Assoc.*, **84**, No. 407, 797-804.

Cohen, J. (1960), "A Coefficient of Agreement for Nominal Scales," *Educ. Psychol. Meas.*, **20**, 37-46.

Fahrmeir, L. and Tutz, G. (1994), *Multivariate Statistical Modelling Based on Generalized Linear Models*, New York: Springer-Verlag.

Fleming, W. (1977), *Functions of Several Variables*, 2nd edition, New York: Springer.

Glonek, G.F.V. (1996), "A Class of Regression Models for Multivariate Categorical Responses," *Biometrika*, **83**, No. 1, 15-28.

Glonek, G.F.V. and McCullagh, P. (1995), "Multivariate Logistic Models," *J. Royal Statist. Soc. - B*, **57**, 533-546.

Goodman, L.A. and Kruskal, W.H. (1979), *Measures of Association for Cross Classifications*, New York: Springer-Verlag.

Grizzle, J.E., Starmer, C.F., and Koch, G.G. (1969), "Analysis of Categorical Data by Linear Models," *Biometrics*, **25**, 489-504.

Haberman, S.J. (1973), "The Analysis of Residuals in Cross-Classification Tables," *Biometrics*, **29**, 205-220.

Lang, J.B. (1996), "On the Comparison of Multinomial and Poisson Loglinear Models," *J. Royal Statist. Soc.- B*, **58**, No. 1, 253-266.

Lang, J.B. (2000), "Multinomial-Poisson Homogeneous-Function Models for Categorical Data," Technical Report #297, University of Iowa, Department of Statistics and Actuarial Science.

Lang, J.B. and Agresti, A. (1994), "Simultaneously Modeling Joint and Marginal Distributions of Multivariate Categorical Responses," *J. Amer. Statist. Assoc.*, **89**, 625-632.

Lang, J.B., McDonald, J.W., and Smith, P.W.S. (1999), "Association-Marginal Modeling of Multivariate Categorical Responses: A Maximum Likelihood Approach," *J. Amer. Statist. Assoc.*, **94**, 1161-71.

Lloyd, C.J. (1999), *Statistical Analysis of Categorical Data*, New York: John Wiley & Sons.

McCullagh, P. and Nelder, J.A. (1989), *Generalized Linear Models*, 2nd edn., London: Chapman and Hall.

Palmgren, J. (1981), "The Fisher Information Matrix for Log Linear Models Arguing Conditionally on Observed Explanatory Variables," *Biometrika*, **68**, 2, 563-566.

Serfling, R.J. (1980), *Approximation Theorems of Mathematical Statistics*, New York: John Wiley & Sons.

Stokes, M.E. (1999), "Recent Advances in Categorical Data Analysis," in *Proceedings for the 24th Annual SAS User Groups International Conference*, Cary, NC: SAS Institute Inc.

Stokes, M.E., Davis, C.S., and Koch, G.G. (1995), *Categorical Data Analysis Using the SAS System*, Cary, NC: SAS Institute Inc.