Available online at www.sciencedirect.com

Journal of Multivariate Analysis 95 (2005) 37–49

www.elsevier.com/locate/jmva

Asymptotically optimal tests for parametric functions against ordered functional alternatives

Ming-Tien Tsai[a,*], Pranab Kumar Sen[b]

[a]Institute of Statistical Science, Academia Sinica, Taipei 11529, Taiwan, ROC
[b]University of North Carolina at Chapel Hill, USA

Received 19 May 2003
Available online 16 September 2004

Abstract

There are hypothesis testing problems for (nonlinear) functions of parameters against functional ordered alternatives for which a reduction to a conventional order-restricted hypothesis testing problem may not be feasible. While such problems can be handled in an asymptotic setup, among the available choices, it is shown that the union–intersection principle may have certain advantages over the likelihood principle or its ramifications. An application to a genomic model is also considered.

AMS 1991 subject classification: 62H15; 62H17

Keywords: Gini–Simpson index; Genomic sequence; Hamming distance; Marginal likelihood function; Nearly convex and prox-regular; Union–intersection principle

1. Introduction

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be independent and identically distributed (i.i.d.) random vectors (r.v.) having the density

$$f_{\boldsymbol{\theta}}(\mathbf{x}) = \exp\{\boldsymbol{\theta}'\mathbf{T}(\mathbf{x}) - \psi(\boldsymbol{\theta})\}\,\mathrm{d}\lambda(\mathbf{x}), \tag{1.1}$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)'$ is an unknown parametric vector, $\mathbf{T}(\mathbf{x}) = (T_1(\mathbf{x}), \ldots, T_m(\mathbf{x}))'$ is a vector of functions of $\mathbf{x}$, of given forms, and λ is a σ-finite measure on the Borel subsets of

---

[*] Corresponding author. Fax: +886-2-783-1523.

E-mail addresses: mttsai@stat.sinica.edu.tw (M.-T. Tsai), pksen@bios.unc.edu (P.K. Sen).

doi:10.1016/j.jmva.2004.07.002

R^m, m being a positive integer. Thus, $f_{\boldsymbol{\theta}}(\mathbf{x})$ is assumed to belong to the regular exponential family with the natural parameter space

$$\mathcal{N}_{\boldsymbol{\theta}} = \{\boldsymbol{\theta} : \int \mathrm{e}^{\boldsymbol{\theta}'\mathbf{T}(\mathbf{x})-\psi(\boldsymbol{\theta})}\,\mathrm{d}\lambda(\mathbf{x}) < \infty\}. \tag{1.2}$$

$\mathcal{N}_{\boldsymbol{\theta}}$ is open and $\psi(\boldsymbol{\theta})$ is continuous on $\mathcal{N}_{\boldsymbol{\theta}}$ (Brown [7]). Let $\boldsymbol{\xi}(\boldsymbol{\theta}) = (\partial/\partial\boldsymbol{\theta})\psi(\boldsymbol{\theta})$. Then $E_{\boldsymbol{\theta}}\mathbf{T}(\mathbf{X}) = \boldsymbol{\xi}(\boldsymbol{\theta})$ and $\mathbf{T}(\cdot)$ is complete, sufficient statistic for $\boldsymbol{\theta}$. As the exponential family itself is a 1-flat manifold, $\boldsymbol{\xi}(\boldsymbol{\theta})$ and the canonical parameter $\boldsymbol{\theta}$ have dual relationship to each other (Amari [1]), which has been exploited in formulating optimal tests for $\boldsymbol{\xi}(\boldsymbol{\theta})$ against global, as well as, restricted alternatives.

We consider here a set of parametric functions

$$\boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta})) = (\gamma_1(\boldsymbol{\xi}(\boldsymbol{\theta})), \ldots, \gamma_p(\boldsymbol{\xi}(\boldsymbol{\theta})))', \quad p \leqslant m \tag{1.3}$$

which need not be linear or one-to-one; it is also possible to express $\boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta}))$ as $\boldsymbol{\gamma}^*(\boldsymbol{\theta})$, though often, $\boldsymbol{\gamma}(\cdot)$ is more handy than $\boldsymbol{\gamma}^*(\cdot)$. $\boldsymbol{\gamma}(\cdot)$ is assumed to be continuously differentiable. We thus allow for $m-p$ nuisance parametric functions. Our contention is to test

$$\mathrm{H}_0 : \boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta})) = \mathbf{0} \ \text{ against } \ \mathrm{H}_1^+ : \boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta})) \geqslant \mathbf{0}. \tag{1.4}$$

A general characteristic of such parametric functions (and functional alternatives) is that even if the classical maximum likelihood estimator (MLE) $\widehat{\boldsymbol{\theta}}$ exists and can be found in a closed form, the plugged-in estimator $\boldsymbol{\gamma}(\boldsymbol{\xi}(\widehat{\boldsymbol{\theta}}))$ may not necessarily be the MLE of $\boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta}))$. This difficulty stems primarily from the fact that the $\boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta}))$ are not generally one-to-one functions of $\boldsymbol{\theta}$ or $\boldsymbol{\xi}(\boldsymbol{\theta})$, and under H_1^+, it might be even more difficult to have the MLE of $\boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta}))$, as is needed to obtain the likelihood ratio test (LRT). All the illustrative examples in Section 2 testify this feature, and raise the issue of finite sample properties of the LRT even when the densities belong to the exponential family.

Let $\Gamma = \{\boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta})) : \boldsymbol{\theta} \in \mathcal{N}_{\boldsymbol{\theta}}\}$ and $\Gamma^+ = \{\boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta})) \in \Gamma : \boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta})) \geqslant \mathbf{0}\}$. As such, it might be tempting to reduce this hypothesis testing problem to a *positive orthant alternative* problem, and thereby to use various tests available in the literature. However, we shall see in Section 3 that there are roadblocks for constructions of optimal tests based on finite sample sizes. Before that in Section 2, we consider some illustrative models that motivate our hypothesis testing problem and bring the relevance of asymptotics. Section 3 also outlines the utility of marginal likelihood functions in this context. Our main emphasis is on the role of Roy's [24] union–intersection principle (UIP) in the proposed testing problem. In Section 4, based on some basic results of Shapiro [27] asymptotic optimality properties are studied, and the likelihood principle (LP) is contrasted with the UIP. In the last Section, the proposed test procedure is applied to test for homogeneity in some genomic sequence models.

2. Some motivating illustrations

To bring out the basic differences between conventional and functional models, we consider the following. Some other reasons for illuminating these examples are also mentioned in the next section.

Example 1. *Behrens–Fisher problem*: *ordered alternatives*. Let X_{ij}, $j = 1, 2, \ldots, n_i$ be i.i.d.r.v.'s with the normal pdf mean μ_i and variance σ_i^2, $i = 1, 2, \ldots, k$, where μ_i and σ_i^2 are all unknown, and it is not apriorily assumed that the σ_i^2 are all equal. Consider then the hypothesis testing problem

$$\mathrm{H}_0 : \mu_1 = \cdots = \mu_k \ \text{ against } \ \mathrm{H}_1^{>} : \mu_1 \geqslant \cdots \geqslant \mu_k, \tag{2.1}$$

not all equal. Even for $k = 2$, the LP does not yield a similar test, and the situation is worse for $k \geqslant 3$. Actually, Linnik [14] showed that, even for $k = 2$, no similar test having certain desirable properties exists. There are some other ad hoc tests which may not be generally optimal when the σ_i^2 are all nuisance parameters.

Example 2. *Coefficient of variation problem*: *ordered alternatives*. Under the set-up of Example 1, assuming $\mu_i \neq 0, i = 1, \ldots, k$ and consider the hypothesis

$$\mathrm{H}_0 : \frac{\sigma_1}{\mu_1} = \cdots = \frac{\sigma_k}{\mu_k} = v \ \text{ against } \ \mathrm{H}_1^{>} : \frac{\sigma_1}{\mu_1} \geqslant \cdots \geqslant \frac{\sigma_k}{\mu_k}, \tag{2.2}$$

with at least one strict inequality (v is nuisance); we may refer to Berger et al. [5] who point out the difficulties in using LP for making exact inference.

Example 3. *Fieller–Creasy problem*: *ordered alternatives*. Under the set-up of Example 1, assuming that $\mu_1 \neq 0$ (but unknown), set

$$\mathrm{H}_0 : \frac{\mu_2}{\mu_1} = \cdots = \frac{\mu_k}{\mu_1} \ \text{ against } \ \mathrm{H}_1^{>} : \frac{\mu_2}{\mu_1} \geqslant \cdots \geqslant \frac{\mu_k}{\mu_1}, \tag{2.3}$$

with at least one strict inequality. Even for $k = 2$, there are some difficulties for drawing exact inference and the situation becomes worse for $k \geqslant 3$ (Glesar and Hwang [11]).

Example 4. *The ordered noncentrality problem*. Let $\mathbf{X}_{ij}$, $j = 1, 2, \ldots, n_i$ be i.i.d.r.v.'s with the multinormal distribution with mean vector $\boldsymbol{\mu}_i$ and dispersion matrix $\boldsymbol{\Sigma}_i$, $i = 1, \ldots, k$. Define the noncentrality parameters as $\Delta_i = \boldsymbol{\mu}_i' \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\mu}_i$, for $i = 1, \ldots, k$, and set

$$\mathrm{H}_0 : \Delta_1 = \cdots = \Delta_k \ \text{ vs. } \ \mathrm{H}_1^{>} : \Delta_1 \geqslant \cdots \geqslant \Delta_k, \tag{2.4}$$

with at least one strict inequality. It is difficult to find the MLE under the alternative hypothesis, and no optimal exact test may exist (Berger et al. [5]).

Example 5. *The orthant space problem*. Let $\mathbf{X}_i$, $i = 1, 2, \ldots, n$ be i.i.d.r.v.'s with $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ density, with both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ unknown. Consider the hypothesis testing problem

$$\mathrm{H}_0 : \boldsymbol{\mu} = \mathbf{0} \ \text{ vs. } \ \mathrm{H}_1^{+} : \boldsymbol{\mu} \geqslant \mathbf{0}, \ \boldsymbol{\Sigma} \text{ nuisance.} \tag{2.5}$$

Following Perlman's [18] seminal paper, lots of work have been done in this case (Wang and McDermott [29], Sen and Tsai [26], and Perlman and Wu [19]). Still, the finite sample resolutions are not final say in this matter.

Example 6. *The ordered correlation problem*. In the same set-up of Example 5, set $\boldsymbol{\Sigma} = ((\sigma_{jl}))$, $\sigma_{jl} = \sigma_j \sigma_l \rho_{jl}$ for $j, l = 1, \ldots, p$. Set then

$$\mathrm{H}_0 : \rho_{ij+1} = \cdots = \rho_{ip}, \ \forall\, i \leqslant p \ \text{ vs. } \ \mathrm{H}_1^{>} : \rho_{ij} \text{ is } \searrow \text{ in } j\ (> i), \ \forall i. \tag{2.6}$$

It is difficult to obtain MLE under H_1^+.

Example 7. *Gini–Simpson indexes*: *ordered alternatives*. Let $(n_{i1}, \ldots, n_{iJ})$ have the multinomial law $(n_i, \boldsymbol{\pi}_i)$, where $\boldsymbol{\pi}_i = (\pi_{i1}, \ldots, \pi_{iJ})'$, $\boldsymbol{\pi}_i'\mathbf{1} = 1$, for $i = 1, \ldots, k$ $(\geqslant 2)$. Define

$$\mathcal{I}(\boldsymbol{\pi}_i) = 1 - \boldsymbol{\pi}_i'\boldsymbol{\pi}_i \quad (i = 1, \ldots, k). \tag{2.7}$$

As a measure of biodiversity that was proposed by Simpson [28] while Pinheiro et al. [21] used this measure in Genomic analysis. We formulate

$$\mathrm{H}_0 : \mathcal{I}(\boldsymbol{\pi}_1) = \cdots = \mathcal{I}(\boldsymbol{\pi}_k) \text{ vs. } \mathrm{H}_1^{>} : \mathcal{I}(\boldsymbol{\pi}_1) \geqslant \cdots \geqslant \mathcal{I}(\boldsymbol{\pi}_k), \tag{2.8}$$

with at least one strict inequality. Nei [17] defined the *gene diversity* as $\boldsymbol{\pi}_i'\boldsymbol{\pi}_i$ (whose reciprocal is called the *effective* number of alleles and has good use in statistical genetics). As a result, (2.8) may also be formulated in terms of the gene identities. For further studies, we may refer to Nayak and Gastwith [16], Rao [22] and Sen [25], where other related measures have also been considered. We shall discuss them in the last section.

Example 8. *The ordered entropy problem*. For the *i*th population (in the previous example), we define the entropy function as

$$\mathcal{E}(\boldsymbol{\pi}_i) = -\sum_{j=1}^{J} \pi_{ij} \log \pi_{ij},\, i = 1, \ldots, k$$

and then formulate

$$\mathrm{H}_0 : \mathcal{E}(\boldsymbol{\pi}_1) = \cdots = \mathcal{E}(\boldsymbol{\pi}_k) \text{ against } \mathrm{H}_1^{>} : \mathcal{E}(\boldsymbol{\pi}_1) \geqslant \cdots \geqslant \mathcal{E}(\boldsymbol{\pi}_k), \tag{2.9}$$

with at least one strict inequality. Again, there are various related entropy measures, considered by Chakraborty and Rao [8], and in the last section we shall comment more on them.

3. Proposed tests

From finite sample size point of views, generally we base on one of the well-known "optimal" criteria such as invariance principle (maximal invariance), unbiasedness, similar test (Neyman's structure), α-admissibility, generalized Bayes, Bayes factor, fiducial argument, integrated likelihood methods, likelihood ratio principle tests (such as the conditional, partial and marginal LRT) and score tests etc. to construct reasonable tests for various problems. However, there are some impasses to incorporate the optimal criteria mentioned above to construct finite sample (exact) optimal tests for the hypothesis problem in (1.4), and this can be easily illustrated by the motivating examples in Section 2 (for details, please see the corresponding references cited in this paper).

One of the major difficulties of the problem is that likelihood ratio based tests are hard to compute. Although, in all the examples (belong to the exponential family of densities) $\sum_{i=1}^{n} \mathbf{T}(\mathbf{x}_i)$ (or parallel statistics) is a complete sufficient statistic for $\boldsymbol{\theta}$ [and unbiased for $\boldsymbol{\xi}(\boldsymbol{\theta})$], as $\boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta}))$ may not be a one-to-one transformation of $\boldsymbol{\xi}(\boldsymbol{\theta})$, the MLE of $\boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta}))$

may not have a closed form. Secondly, as $\boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta}))$ is not generally linear in $\boldsymbol{\xi}(\boldsymbol{\theta})$, a plug-in estimator $\boldsymbol{\gamma}(\bar{\mathbf{T}}_n)$ [$\bar{\mathbf{T}}_n$ is defined in (3.2)] may not be unbiased for $\boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta}))$, and its exact distribution may be quite cumbersome to have a closed form. Therefore, Neyman's structure may not generally hold, and as a result, MLE [of $\boldsymbol{\xi}(\boldsymbol{\theta})$] based tests may not have the similar region property, nor it may be optimal in a finite sample setup. To overcome it, we, therefore, incorporate two important concepts, pivotal quantity (the normalization of minimal sufficient statistics for interesting parameters under null hypothesis) and orthogonal projection, to solve the problems. As a result, it turns out to be the same as by incorporating the efficient score statistics and UIP. Then, we proceed to appraise an asymptotic situation where the dimension of the parameter space $\mathcal{N}_{\boldsymbol{\theta}}$ is held fixed but the sample size is made to increase indefinitely. In passing, we may remark that if the dimensionality of $\mathcal{N}_{\boldsymbol{\theta}}$ is also made to increase with the sample size, an altogether different asymptotic situation may evolve, and we shall not get into that context here. With the help of geometric structure of exponential family (Amari [1]) and the asymptotic results of Shapiro [27], we can show that the proposed tests are optimal in the sense that they are asymptotically power-equivalent to those of (marginal) LRT in next section. This unified procedure is easy to implement, and has widely applications including some interesting "genomic sequences" problems which the parameter space under alternative hypotheses are no longer to be positively homogeneous cones.

We confine ourselves to (1.1)–(1.4) in the regular case, where $\boldsymbol{\gamma}(\cdot)$ is twice differentiable, and define

$$\boldsymbol{\Delta} = E_{\boldsymbol{\theta}}\{[\mathbf{T}(\mathbf{X}) - \boldsymbol{\xi}(\boldsymbol{\theta})][\mathbf{T}(\mathbf{X}) - \boldsymbol{\xi}(\boldsymbol{\theta})]'\}; \tag{3.1}$$

$\boldsymbol{\Delta}$ may, in general, depend on $\boldsymbol{\theta}$. Also, let

$$\bar{\mathbf{T}}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{T}(\mathbf{X}_i) \ \text{ and } \ \mathbf{S}_n = \frac{1}{n-1}\sum_{i=1}^{n}[\mathbf{T}(\mathbf{X}_i) - \bar{\mathbf{T}}_n][\mathbf{T}(\mathbf{X}_i) - \bar{\mathbf{T}}_n]'. \tag{3.2}$$

Then the following results hold: as $n \to \infty$,

$$\bar{\mathbf{T}}_n \to \boldsymbol{\xi}(\boldsymbol{\theta}) \ \text{ almost surely (a.s.)}, \tag{3.3}$$

$$\mathbf{S}_n \to \boldsymbol{\Delta} \ \text{ a.s.}, \tag{3.4}$$

$$n^{1/2}[\bar{\mathbf{T}}_n - \boldsymbol{\xi}(\boldsymbol{\theta})] \overset{\mathcal{D}}{\to} \mathcal{N}_m(\mathbf{0}, \boldsymbol{\Delta}). \tag{3.5}$$

Let then $\boldsymbol{\Phi} = \triangledown\boldsymbol{\gamma}(\cdot)$, $\triangledown$ being the gradient operator, be a $p \times m$ matrix (which is as well defined on $\mathcal{N}_{\boldsymbol{\theta}}$), and let

$$\boldsymbol{\Delta}^* = \boldsymbol{\Phi}\boldsymbol{\Delta}\boldsymbol{\Phi}', \tag{3.6}$$

where without loss of generality $\boldsymbol{\Delta}^*$ is assumed to be of full rank. We obtain from (3.3) to (3.6) along with the Slutzky theorem,

$$n^{1/2}[\boldsymbol{\gamma}(\bar{\mathbf{T}}_n) - \boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta}))] \overset{\mathcal{D}}{\to} \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Delta}^*), \ \text{ as } n \to \infty. \tag{3.7}$$

We note that the notion of *pivotal inference* is central to the development of Fisher's *fiducial argument*, a pivotal quantity being recognized as a generalized ancillary statistic.

Our basic contention is to project a linear pivotal quantity onto the set Γ^{+}, defined after (1.4), and it is easy to show that this coincides with the UIP approach in a natural way. To incorporate UIP, we define for each $\mathbf{b} \geqslant \mathbf{0}$,

$$\mathrm{H}_{0,\mathbf{b}} : \mathbf{b}^{t}\boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta})) = 0 \quad \text{and} \quad \mathrm{H}_{1,\mathbf{b}} : \mathbf{b}^{t}\boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta})) > 0, \tag{3.8}$$

so that

$$\mathrm{H}_0 = \bigcap\nolimits_{\{\mathbf{b} \geqslant \mathbf{0}\}} \mathrm{H}_{0,\mathbf{b}} \quad \text{and} \quad \mathrm{H}_1^{+} = \bigcup\nolimits_{\{\mathbf{b} \geqslant \mathbf{0}\}} \mathrm{H}_{1,\mathbf{b}}. \tag{3.9}$$

Also, we evaluate $\boldsymbol{\Phi}$ at $\bar{\mathbf{T}}_n$, and denote it by $\widehat{\boldsymbol{\Phi}}_n$. Let then

$$\widehat{\boldsymbol{\Delta}}_n^{*} = \widehat{\boldsymbol{\Phi}}_n \mathbf{S}_n \widehat{\boldsymbol{\Phi}}_n^{\prime}, \tag{3.10}$$

so that $\widehat{\boldsymbol{\Delta}}_n^{*}$ is a (strongly) consistent estimator of $\boldsymbol{\Delta}^{*}$.

For any given $\mathbf{b}$ ($\geqslant \mathbf{0}$), consider the pivotal quantity

$$V_n(\mathbf{b}) = n^{1/2}\mathbf{b}^{\prime}\boldsymbol{\gamma}(\bar{\mathbf{T}}_n)/\{\mathbf{b}^{\prime}\widehat{\boldsymbol{\Delta}}_n^{*}\mathbf{b}\}^{1/2}. \tag{3.11}$$

From (3.9) to (3.11), we may formulate a union–intersection test (UIT) for testing H_0 vs. H_1^{+} based on the test statistic

$$U_n = \sup\,\{V_n(\mathbf{b}) : \mathbf{b} \geqslant \mathbf{0}\} \tag{3.12}$$

and the task is to find a closed expression for U_n along with its critical value, say $c_{n,\alpha}$ such that

$$P\{U_n \geqslant c_{n,\alpha} |\, \mathrm{H}_0\} \to \alpha \; (0 < \alpha < 1),$$

the desired significance level.

To obtain a closed expression for U_n, we virtually follow Sen and Tsai [26]. Let $P = \{1, 2, \ldots, p\}$, and for every $a : \emptyset \subseteq a \subseteq P$, let a' be its complement and $|a|$ its cardinality, there being 2^p subsets for which $0 \leqslant |a| \leqslant p$. For each a: $\emptyset \subseteq a \subseteq P$, we partition (following possible rearrangement) $\mathbf{Y}_n = \boldsymbol{\gamma}(\bar{\mathbf{T}}_n)$ and $\widehat{\boldsymbol{\Delta}}_n^{*}$ as

$$\mathbf{Y}_n = \begin{pmatrix} \mathbf{Y}_{na} \\ \mathbf{Y}_{na'} \end{pmatrix} \text{ and } \widehat{\boldsymbol{\Delta}}_n^{*} = \begin{pmatrix} \widehat{\boldsymbol{\Delta}}_{naa}^{*} & \widehat{\boldsymbol{\Delta}}_{naa'}^{*} \\ \widehat{\boldsymbol{\Delta}}_{na'a}^{*} & \widehat{\boldsymbol{\Delta}}_{na'a'}^{*} \end{pmatrix}, \tag{3.13}$$

and write

$$\mathbf{Y}_{na:a'} = \mathbf{Y}_{na} - \widehat{\boldsymbol{\Delta}}_{naa'}^{*}\widehat{\boldsymbol{\Delta}}_{na'a'}^{*-1}\mathbf{Y}_{na'}, \tag{3.14}$$

$$\widehat{\boldsymbol{\Delta}}_{naa:a'}^{*} = \widehat{\boldsymbol{\Delta}}_{naa}^{*} - \widehat{\boldsymbol{\Delta}}_{naa'}^{*}\widehat{\boldsymbol{\Delta}}_{na'a'}^{*-1}\widehat{\boldsymbol{\Delta}}_{na'a}^{*}, \tag{3.15}$$

$$I_{na} = I\{\mathbf{Y}_{na:a'} > \mathbf{0}, \widehat{\boldsymbol{\Delta}}_{na'a'}^{*-1}\mathbf{Y}_{na'} \leqslant \mathbf{0}\}. \tag{3.16}$$

Note that only one of the I_{na} ($\emptyset \subseteq a \subseteq P$) will be equal to 1 and the rest all 0. Then using the Kuhn–Tucker–Lagrange (KTL-) point formula theorem (Hadley [12]), we obtain that

$$U_n^2 = \sum_{\emptyset \subseteq a \subseteq P} n\{\mathbf{Y}_{na:a'}^{\prime}\widehat{\boldsymbol{\Delta}}_{naa:a'}^{*-1}\mathbf{Y}_{na:a'}\}I_{na}. \tag{3.17}$$

Using (3.7) and proceeding as in Theorem 2.1 of Sen and Tsai [26], we obtain that for every $c > 0$

$$\lim_{n\to\infty} P\{U_n^2 \geqslant c^2 | \mathrm{H}_0\} = \sum_{k=1}^{p} r(p,k;\boldsymbol{\Delta}^*) P\{\chi_k^2 \geqslant c^2\} \tag{3.18}$$

where the χ_j^2 are independent r.v.'s, with χ_j^2 having the central chisquare distribution with j ($\geqslant 0$) degrees of freedom (DF), $j = 1, \ldots, p$ and $\chi_0^2 = 0$ with probability 1. Further, for each $a : \emptyset \subseteq a \subseteq P$, for a r.v. $\mathbf{Z} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$, we define the quantities $\mathbf{Z}_{a:a'}$ etc. as in (3.14)–(3.15) with $\boldsymbol{\Delta}^*$ replaced by $\boldsymbol{\Sigma}$. Then, we have

$$r(p,k;\boldsymbol{\Sigma}) = \sum\nolimits_{\{a\subseteq P:|a|=k\}} P\{\mathbf{Z}_{a:a'} > \mathbf{0}, \boldsymbol{\Sigma}_{a'a'}^{-1}\mathbf{Z}_{a'} \leqslant \mathbf{0}\}, \tag{3.19}$$

for $k = 0, \ldots, p$. In passing, we may remark that for the global alternative $\mathrm{H}_1 : \boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta})) \neq \mathbf{0}$, we would have the supremum over all $\mathbf{b} \neq \mathbf{0}$, as a result, U_n^2 reduces to

$$U_{n0}^2 = n\mathbf{Y}_n'\widehat{\boldsymbol{\Delta}}_n^{*-1}\mathbf{Y}_n \tag{3.20}$$

(with asymptotic χ_p^2 distribution under H_0).

Let us make some comments on the scope of LP in the context of testing for restricted functional alternatives in (1.4). First, the conventional LRT; it involves the computation of the MLE of $\boldsymbol{\theta}$ [or equivalently $\boldsymbol{\xi}(\boldsymbol{\theta})$] under both H_0 and H_1^+. Under H_0, $\boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta})) = \mathbf{0}$, so that it may be more manageable than under H_1^+ where $\boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta})) \geqslant \mathbf{0}$ may create impasses for the computation of the MLE [of $\boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta}))$]. Generally, closed expression of the MLE may not be available, and on top of that the functional inequality restraints make it more difficult to obtain the restricted (R)MLE of $\boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta}))$, under H_1^+. Another possibility is to use the unrestricted (U)MLE of $\boldsymbol{\xi}(\boldsymbol{\theta})$, plug-in the $\boldsymbol{\gamma}(\cdot)$ and then obtain the approximate UMLE of $\boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta}))$. One may then use a variant of the classical Wald method, but for restricted alternatives. In this way, we end up with the same test statistic U_n^2 in (3.17). This explains the proximity of the restricted alternative Wald test and the UIT. In any case, even if we use the plug-in MLE, we are not directly using the LP totally. Thus, in the present context, we find the UIT more appealing and manageable. In the next section, we shall study some asymptotic optimality properties of the UIT, as well as, other related tests.

4. Asymptotic optimality properties

The difficulty in implementing the LRT stems from the complexities involved in finding the MLE of $\boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta}))$ over the parameter space Γ^+. Let $\Theta_0 = \{\boldsymbol{\theta} : \boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta})) = \mathbf{0}\}$ and $l_n(\boldsymbol{\theta}) = n[\psi(\boldsymbol{\theta}) - \boldsymbol{\theta}'\bar{\mathbf{T}}_n]$, where the notations are introduced in the previous section. Then, to obtain the MLE under H_0 and H_1^+ respectively we need to minimize $l_n(\boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \Theta_0 \cap \mathcal{N}_{\boldsymbol{\theta}}$ and $\boldsymbol{\theta} \in \Gamma^+ \cap \mathcal{N}_{\boldsymbol{\theta}}$; let $\widehat{\boldsymbol{\theta}}_{0n}$ and $\widehat{\boldsymbol{\theta}}_n^*$ be the MLE of $\boldsymbol{\theta}$ under H_0 and H_1^+ respectively. By the KTL-point formula, they are the solutions of the following:

$$[\bigtriangledown l_n(\boldsymbol{\theta}) + \sum_{i=1}^{p} v_i \bigtriangledown \gamma_i^*(\boldsymbol{\theta})]|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_{0n}} = \mathbf{0} \tag{4.1}$$

and

$$[\bigtriangledown l_n(\boldsymbol{\theta}) + \sum_{i=1}^{p} u_i \bigtriangledown \gamma_i^*(\boldsymbol{\theta})]|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_n^*} = \mathbf{0},\ \mathbf{u} \geqslant \mathbf{0},\ u_i \gamma_i^*(\widehat{\boldsymbol{\theta}}_n^*) = 0, i = 1, \ldots, p. \tag{4.2}$$

Compare to (4.1), a closed form solution for (4.2) may not generally exist (see Example 7), or tractable. Even a closed form solution for (4.1) may not exist, though iterative methods can be used to obtain the RMLE by successive numerical iterations. We refer to McLachlan and Krishnan [15] for an extensive account of the EM algorithm and extensions which would be valuable for this iterative process; however, in view of the inequality restraints in (4.2), such algorithms may need to be modified or generalized to suit the purpose. Nevertheless, sans a closed form, finite sample optimality properties (of the MLE) may not exist, nor easy to establish even if they exist. As such, we take recourse to asymptotic analysis to establish some asymptotic optimality properties: some recent results of Geyer [10], Shapiro [27] and others help us to analyze nonlinear, non-normal asymptotic behavior of the MLE, LRT and other related tests. In that way, we may advocate the UIT for its relative computational simplicity and general asymptotic properties.

Based on $\bar{\mathbf{T}}_n$, the MLE of $\boldsymbol{\xi}(\boldsymbol{\theta})$, we consider the plug-in estimator $\mathbf{Y}_n = \boldsymbol{\gamma}(\bar{\mathbf{T}}_n)$, set an objective function $h(\mathbf{b}, \mathbf{Y}_n) = \mathbf{b}'\mathbf{Y}_n$, and aim to maximize $h(\mathbf{b}, \mathbf{Y}_n)$, subject to $\mathbf{b} \in R^{+p} = \{\mathbf{b} |\ \mathbf{b} \geqslant \mathbf{0}\}$ and $\mathbf{b}'\widehat{\boldsymbol{\Delta}}_n^*\mathbf{b} = 1$. This provides a unified method of obtaining asymptotically optimal estimator of $\boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta}))$ over the set Γ^+. For every $a \subseteq P$, if $\mathbf{u}, \mathbf{v}$ are $|a|-$ and $(p - |a|)-$ vectors respectively, we set $[\mathcal{M}_a(\mathbf{u}, \mathbf{v})]_i = u_i$ or v_i according as $i \in a$ or $i \in a'$. Then proceeding as in Section 3, the estimator of $\boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta}))$ restricted to the set Γ^+ can then be taken as

$$\widehat{\boldsymbol{\gamma}}(\boldsymbol{\xi}(\boldsymbol{\theta})) = \sum_{\emptyset \subseteq a \subseteq P} \mathcal{M}_a(\mathbf{Y}_{na:a'}, \mathbf{0}) I_{na}, \tag{4.3}$$

while the unrestricted estimator of $\boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta}))$ is $\mathbf{Y}_n$. We proceed to study properties of $\widehat{\boldsymbol{\gamma}}(\boldsymbol{\xi}(\boldsymbol{\theta}))$.

Case I: *Quasi-factorizable likelihood function*. Suppose that the likelihood function can be factorized into $g(\mathbf{x}, \boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta})))$ and $h^*(\mathbf{x}, \mathbf{v}(\boldsymbol{\xi}(\boldsymbol{\theta})))$, where $g(\cdot)$ involves only the $\boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta}))$, the parameters entering the hypothesis, while $h^*(\cdot)$ only involves the nuisance parameters $\mathbf{v}(\boldsymbol{\xi}(\boldsymbol{\theta}))$; this is the case when there is a Barndorff-cut (Bandorff-Nielsen [3]) which separates $\boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta}))$ from $\mathbf{v}(\boldsymbol{\xi}(\boldsymbol{\theta}))$. Let Θ_N be the space of nuisance parameters and Γ^+ defined as before. Then

$$\begin{aligned} &\frac{\sup_{\Gamma^+} g(\mathbf{x}, \boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta}))) \sup_{\Theta_N} h^*(\mathbf{x}, \mathbf{v}(\boldsymbol{\xi}(\boldsymbol{\theta})))}{\sup_{\Theta_0} g(\mathbf{x}, \boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta}))) \sup_{\Theta_N} h^*(\mathbf{x}, \mathbf{v}(\boldsymbol{\xi}(\boldsymbol{\theta})))} \\ &\quad = \frac{\sup_{\Gamma^+} g(\mathbf{x}, \boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta})))}{\sup_{\Theta_0} g(\mathbf{x}, \boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta})))}. \end{aligned} \tag{4.4}$$

As a result, the LRT can be obtained by direct use of only the factor $g(\mathbf{x}, \boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta})))$. For (1.4), $p = m$ relates to Case I. In this no nuisance parameter problem $\boldsymbol{\gamma}(\cdot)$ can be expressed as one-to-one correspondence to $\boldsymbol{\xi}(\boldsymbol{\theta})$, and we have the so called positive orthant alternative. Robertson and Wegman [23], Dykstra and Robertson [9] and others studied the LRT and also their asymptotic distributions under the null hypotheses. Even for general $p \leqslant m$, we show that the UIT proposed in Section 3 is asymptotically power-equivalent to the LRT. For simplicity of proof, we consider the case of $p = m$. Using the general results of Shapiro [27] on nearly convex and prox-regular parameter spaces and their impact on the asymptotics of local as well as global MLE, we obtain the following results:

(i) Let $\widehat{\gamma}_l(\boldsymbol{\xi}(\boldsymbol{\theta}))$ be the local MLE of $\boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta}))$ over the set Γ^+ and $\widehat{\boldsymbol{\gamma}}(\boldsymbol{\xi}(\boldsymbol{\theta}))$ be the true MLE over the same set, then

$$n^{1/2}[\widehat{\gamma}_l(\boldsymbol{\xi}(\boldsymbol{\theta})) - \widehat{\boldsymbol{\gamma}}(\boldsymbol{\xi}(\boldsymbol{\theta}))] \overset{\mathrm{P}}{\to} \mathbf{0}, \text{ as } n \to \infty. \tag{4.5}$$

(ii) Let $\widehat{\gamma}_p(\boldsymbol{\xi}(\boldsymbol{\theta}))$ be the partial MLE of $\boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta}))$ over the set Γ^+ [based on the factor $g(\mathbf{x}, \boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta})))$]. Then if the set Γ^+ is nearly convex and prox-regular at Θ_0,

$$n^{1/2}[\widehat{\gamma}_p(\boldsymbol{\xi}(\boldsymbol{\theta})) - \widehat{\gamma}_l(\boldsymbol{\xi}(\boldsymbol{\theta}))] \overset{\mathrm{P}}{\to} \mathbf{0}, \text{ as } n \to \infty. \tag{4.6}$$

In the above formulation, for local MLE, one considers a sequence of parameter spaces $\Gamma_n^+ = \{\boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta})) : \boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta})) = \boldsymbol{\gamma}_0 + n^{-1/2}\boldsymbol{\delta},\ \boldsymbol{\delta} \in \Gamma^+\}$, and in the present context (under H_0), $\boldsymbol{\gamma}_0 = \mathbf{0}$. In the literature, these type of hypothesis are also called Pitman (or contiguous) alternatives. As a result, by (4.5) and (4.6), we obtain that if Γ^+ is nearly convex and prox-regular at Θ_0, then

$$n^{1/2}[\widehat{\gamma}_p(\boldsymbol{\xi}(\boldsymbol{\theta})) - \widehat{\boldsymbol{\gamma}}(\boldsymbol{\xi}(\boldsymbol{\theta}))] \overset{\mathrm{P}}{\to} \mathbf{0}, \text{ as } n \to \infty. \tag{4.7}$$

On the other hand, linking the local MLE's role in the UIT formulated in Section 3, and the role of $\widehat{\boldsymbol{\gamma}}(\boldsymbol{\xi}(\boldsymbol{\theta}))$ to the LRT, we conclude that if Γ^+ is nearly convex and prox-regular then the UIT and LRT are asymptotically power-equivalent. Generally, for the hypothesis testing problems, the power of any reasonable test approaches to one as the sample size is sufficiently large. As such, we usually attack the problems "*locally*" by using the notion of "contiguity". Hence, the asymptotically power-equivalent generally means "*locally*" asymptotically.

Case II: *Orthogonal parameter space*. The m-vector $\boldsymbol{\xi}(\boldsymbol{\theta})$ may be partitioned as $(\boldsymbol{\xi}'_{(1)}(\boldsymbol{\theta}), \boldsymbol{\xi}'_{(2)}(\boldsymbol{\theta}))'$ with p and $m - p$ elements, such that $\boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta})) = \boldsymbol{\gamma}(\boldsymbol{\xi}_{(1)}(\boldsymbol{\theta}))$ depends only on $\boldsymbol{\xi}_{(1)}(\boldsymbol{\theta})$, while the nuisance parameter $\mathbf{v}(\boldsymbol{\xi}(\boldsymbol{\theta})) = \mathbf{v}(\boldsymbol{\xi}_{(2)}(\boldsymbol{\theta}))$ depends only on $\boldsymbol{\xi}_{(2)}(\boldsymbol{\theta})$. Further, $\boldsymbol{\xi}_{(1)}(\boldsymbol{\theta})$ and $\boldsymbol{\xi}_{(2)}(\boldsymbol{\theta})$ are orthogonal, and as a result, $\boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta}))$ and $\mathbf{v}(\boldsymbol{\xi}(\boldsymbol{\theta}))$ are also so. Note that Case I is a special case of Case II. For regular multivariate exponential families, Barndorff–Nielsen [4] showed that $\boldsymbol{\xi}_{(1)}(\boldsymbol{\theta})$ and $\boldsymbol{\xi}_{(2)}(\boldsymbol{\theta})$ are orthogonal. Also, Hudson and Vos [13] have shown that when the orthogonalized score function is a function of $\bar{\mathbf{T}}_{n(1)}$, for drawing inference on $\boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta}))$, no Fisher information is lost in using the marginal distribution of $\bar{\mathbf{T}}_{n(1)}$ [instead of the full $\bar{\mathbf{T}}_n$ which has $m(\geqslant p)$ coordinates]. As such, under this orthogonality setup, we may work with the marginal likelihood function [for $\bar{\mathbf{T}}_{n(1)}$], and then proceed as in Case I [to draw conclusion on $\boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta}))$]. Therefore, the UIT proposed in Section 3, as modified for $\bar{\mathbf{T}}_{n(1)}$, is asymptotically power-equivalent to the corresponding marginal likelihood ratio test, though the latter may be harder to formulate. This may be illustrated with Example 5 where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are orthogonal, though the Barndorff–Nielsen characterization or the factorization of the likelihood function may not hold; we refer to Sen and Tsai [26] for the asymptotic power-equivalence of UIT and Perlman's [18] LRT for this specific problem.

Case III: *Nonorthogonal parameter space*. Under parametric restraints, the orthogonality condition in Case II may not generally hold, resulting in a harder problem of eliminating the nuisance parameters while drawing conclusion on $\boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta}))$. In such a case, the conventional likelihood principle may not suit, and usually, either Bayes methods or integrated likelihood approaches are employed. Conventional Bayes methods rest on the choice of suitable priors; in a multi-parameter case, particularly with nonlinear restraints, a choice of a subjective (or conjugate) prior may often be ambiguous, and difficult too. Computational Bayes methods are increasingly advocated for this reason. Integrated likelihood approaches may be viewed as pseudo-Bayesian, and as agreed by Barnard [2], they adhere to Fisher's fiducial methodology. Also, optimality properties of integrated likelihood methods are not that precisely known. Further, such integrated likelihood functions may not be of simple forms, so that closed solution for the integrated (I)MLE may not be tractable. Example 7 of Section 2 serves as a good illustration of all these points. Moreover, without parametric orthogonality, conditional likelihood approaches (given estimates of nuisance parameters) may result in low efficiency, even asymptotically. All these provocate the use of marginal likelihood functions for $\boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta}))$, though their exact form may no longer be simple, and the computation of marginal (M)MLE (over the set Γ^+) may still be quite difficult.

Based on all these reasons, we take recourse to the asymptotic case where, as in after (4.6), we confine ourselves to local (restricted) alternatives. Note that the MMLE of $\boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta}))$ (over the set Γ^+) may still be a problem and this leads to complications in the formulation of marginal (M)LRT. On the other hand, $\boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta}))$ is continuously differentiable. Thus, if $\bigtriangledown\boldsymbol{\gamma}(\boldsymbol{\xi}(\boldsymbol{\theta}))$ is Lipschitz continuous in a neighborhood of $\boldsymbol{\xi}_0(\boldsymbol{\theta})$, the true parameter point, then the set Γ^+ is nearly convex and prox-regular at $\boldsymbol{\xi}_0(\boldsymbol{\theta})$. This enables us to use Shapiro's [27] results, make use of the plug-in estimates $\boldsymbol{\gamma}(\widehat{\boldsymbol{\xi}}(\boldsymbol{\theta}))$, and construct UIT as in Section 3. This appears to be, at least computationally, simpler than the MLRT, and asymptotically these UIT and MLRT are power-equivalent. With respect to most of the examples in Section 2, Γ^+ is a positively homogeneous cone which is nearly convex and prox-regular, and even without this positive homogeneous cone property, in the other examples, we have the nearly convex and prox-regular condition satisfied. Thus, it seems that the suggested UIT approach combines the computational ease and the asymptotic power-equivalence to the MLRT.

5. An application in genomics

In genomic sequence analysis, we encounter a large number (say, K) of sites, where in each site, there is a purely qualitative categorical response variable (like the chemical words A, C, G, T in DNA or the amino acids in a protein sequence). Analysis of such data sets in a conventional categorical data modeling results in considerable loss of efficiency. On top of that often the null and alternative hypotheses are formulated in a way that resembles our formulation in Section 1. We refer to Pinheiro et al. [20,21] for a detailed account of statistical modeling and analyses of genomic data sets, and consider here a simpler model to illustrate our main point of interest, namely, formulation of UIT for such nonstandard restricted alternative hypotheses testing problem. We conceive of a purely categorical data model wherein a response vector $\mathbf{X} = (X_1, \ldots, X_K)'$ has K coordinates and each X_k can take on $C(\geqslant 2)$ categorical values, indexed as $1, \ldots, C$ (though there may not be any implicit ordering in these C categories). For two independent observations $\mathbf{X}_i$ and $\mathbf{X}_j$, a Hamming distance is defined as

$$d_{ij} = K^{-1} \sum_{k=1}^{K} I(X_{ik} \neq X_{jk}), \tag{5.1}$$

so that

$$\begin{aligned} \Delta &= E\{d_{ij}\} = K^{-1} \sum_{k=1}^{K} P(X_{ik} \neq X_{jk}) \\ &= K^{-1} \sum_{k=1}^{K} [1 - \sum_{c=1}^{C} (P\{X_{ik} = c\})^2] \\ &= K^{-1} \sum_{k=1}^{K} \mathcal{I}(\boldsymbol{\pi}_k), \end{aligned} \tag{5.2}$$

where $\mathcal{I}(\boldsymbol{\pi}_k)$, the Gini–Simpson index, is defined as in (2.7), and $\boldsymbol{\pi}_k = (\pi_{k1}, \ldots, \pi_{kC})'$ is the (marginal) probability (vector) for the outcome $X_{ik}, k = 1, \ldots, K$. Note that the coordinates of $\mathbf{X}$ are generally stochastically dependent (having a C^K-cell multinomial law) but the parameter Δ is a function of the K marginals $\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_K$ only. As a result, the (sample) Hamming distance for n independent $\mathbf{X}_1, \ldots, \mathbf{X}_n$ is defined as

$$U_n = \binom{n}{2}^{-1} \sum_{1 \leqslant i \leqslant j \leqslant n} d_{ij}; \tag{5.3}$$

U_n is a natural (i.e., unbiased, symmetric and consistent) estimator of Δ, and the d_{ij} can only take on the values $r/K, r = 0, 1, \ldots, K$. Further, in view of the anticipated stochastic dependence among the elements of $\mathbf{X}_i$ (or $\mathbf{X}_j$), the $I(X_{ik} \neq X_{jk})(1 \leqslant k \leqslant K)$ are not necessarily independent, so that KU_n may not have the binomial law. In fact, the exact distribution theory of U_n depends on the C^K-cell multinomial probability law, and may not be simple.

Suppose now that there are $G(\geqslant 2)$ groups of sequences where the gth group consists of n_g independent sequences having the marginal probability vectors $\boldsymbol{\pi}_1^{(g)}, \ldots, \boldsymbol{\pi}_K^{(g)}$, for $g = 1, \ldots, G$. Note that they actually have C^K-cell multinomial probability laws $\Pi^{(g)}$, whose marginal probabilities are the $\boldsymbol{\pi}_j^{(g)}, 1 \leqslant j \leqslant K; g = 1, \ldots, G$. We define the population Hamming distances as in (5.2) and denote them by $\Delta_g, g = 1, \ldots, G$. In the context of genomic studies, it has been observed that the Δ_g may vary according to the HIV positivity status of the sequences, HIV positive status may increase the Δ measure, though Δ would remain bounded in (0,1) (Pinheiro et al. [21]). Thus, it may be of interest to consider the following hypotheses:

$$\mathrm{H}_0 : \Delta_1 = \cdots = \Delta_G \ \text{ vs. } \ \mathrm{H}_1^{<} : \Delta_1 \leqslant \cdots \leqslant \Delta_G, \tag{5.4}$$

with the strict inequality holding in at least one place. Viewed from this perspective, we have a model which is a generalization of (2.8) in a genuine (discrete) multivariate case.

For the gth group, we denote the sample counterpart of Δ_g by $U_n^{(g)}$, defined as in (5.3), for $g = 1, \ldots, G$. If we want to use the full likelihood function for all the $n(= n_1 + \cdots + n_G)$

sequences, using the C^K-cell multinomial law, we do not have any resolution with optimal properties. Use of partial, marginal or quasi-likelihoods may have similar problems. On the other hand, the UIT approach works out well.

We also define the entropy function $\mathcal{E}(\boldsymbol{\pi}_k^{(g)})$ as in (2.9) and let

$$\mathcal{E}_g = K^{-1} \sum_{k=1}^{K} \mathcal{E}(\boldsymbol{\pi}_k^{(g)}), \ g = 1, \ldots, G. \tag{5.5}$$

Then, parallel to (5.4), we set

$$\mathrm{H}_0 : \mathcal{E}_1 = \cdots = \mathcal{E}_G \ \text{ vs. } \ H_1^< : \mathcal{E}_1 \leqslant \cdots \leqslant \mathcal{E}_G, \tag{5.6}$$

with at least one strict inequality.

Note that Δ_g as well as $\mathcal{E}_g$ are functions of marginal probability vectors $\boldsymbol{\pi}_k^{(g)}, k = 1, \ldots, K$, and these parameters admit MMLE which we denote by $\widehat{\boldsymbol{\pi}}_k^{(g)}, k = 1, \ldots, K; g = 1, \ldots, G$. Further, if we consider the independent Poisson variables model for multidimensional contingency tables (Bishop et al. [6]), then these MMLE may also be characterized as the unrestricted (U)MLE of the $\boldsymbol{\pi}_k^{(g)}$. For both of these hypotheses testing problems, the nearly convex and prox-regular property of the associated Γ^+ can be established by standard techniques, and hence, we may directly incorporate the UIT in Section 3. The main advantage of this UIT is the computational ease of an estimated covariance matrix of the associated $\boldsymbol{\gamma}(\widehat{\boldsymbol{\pi}}^{(g)})$, using either the classical delta method or any suitable resampling scheme, such as jackknifing or bootstrapping. The computation of the RMLE and the associated information matrix would have been comparatively much more cumbersome, and hence, using the results in Section 4, we may claim that the UIT may be advocated in such nonstandard models.

Acknowledgements

We are grateful to the reviewers for their helpful comments.

References

[1] S.I. Amari, Differential-Geometrical Methods in Statistics, Lecture Notes in Statistics, vol. 28, Springer, Heidelberg, 1985.

[2] G.A. Barnard, Pivotal models and the fiducial argument, Int. Statist. Rev. 63 (1995) 309–323.

[3] O.E. Barndorff-Nielsen, Exponential Families and Conditioning, Wiley, New York, 1973.

[4] O.E. Barndorff-Nielsen, Information and Exponential Families in Statistical Theory, Wiley, New York, 1978.

[5] J.O. Berger, B. Liseo, R.L. Wolpert, Integrated likelihood methods for eliminating nuisance parameters, Statist. Sci. 14 (1999) 1–28.

[6] Y.M.M. Bishop, S.E. Fienberg, P.W. Holland, Discrete Multivariate Analysis—Theory and Practice, MIT Press, Cambridge, MA, 1975.

[7] L.D. Brown, Fundamentals of statistical exponential families, Monograph 9, Institute of Mathematical Statistics Lectures Notes, Institute of Mathematical Statistics, Hayward, CA, 1986.

[8] R. Chakraborty, C.R. Rao, Measurement of genetic variation for evolutionary studies, in: C.R. Rao, R. Chakraborty (Eds.), Handbook of Statistics Vol. 8: Statistical Methods in Biological and Medical Sciences, North-Holland, Amsterdam, 1991, pp. 271–316.

[9] R.L. Dykstra, T. Robertson, Order restricted statistical tests on multinomial and poisson parameters: the starshaped restriction, Ann. Statist. 10 (1982) 1246–1252.

[10] C.J. Geyer, On the asymptotics of constrained *M*-estimation, Ann. Statist. 22 (1994) 1993–2010.

[11] L. Gleser, J.T. Hwang, The nonexistence of $100(1-\alpha)\%$ confidence sets of finite expected diameter in errors-in-variable and related models, Ann. Statist. 15 (1987) 1351–1362.

[12] G. Hadley, Non-linear and Dynamic Programming, Addison-Wesley, Reading, MA, 1964.

[13] S. Hudson, P.W. Vos, Marginal information for expectation parameters, Canad. J. Statist. 28 (2000) 875–886.

[14] Ju.V. Linnik, Statistical Problems with Nuisance Parameters, American Mathematical Society, Providence, RI, 1968.

[15] G.J. McLachlan, T. Krishnan, The EM Algorithm and Extensions, Wiley, New York, 1997.

[16] T.K. Nayak, J.L. Gastwith, The use of diversity analysis to assess the relative influence of factors affecting the income distribution, J. Bus. Econ. Statist. 7 (1989) 453–460.

[17] M. Nei, Genetic distance between populations, Amer. Nature 106 (1972) 283–292.

[18] M.D. Perlman, One-sided problems in multivariate analysis, Ann. Math. Statist. 40 (1969) 549–567.

[19] M.D. Perlman, L. Wu, The emperor's new tests, Statist. Sci. 14 (1999) 355–381.

[20] H.P. Pinheiro, A.S. Pinheiro, P.K. Sen, Comparison of genomic sequences by the Hamming distance, J. Statist. Plann. Inference, to appear.

[21] H.P. Pinheiro, F. Seillier-Moiseiwitsch, P.K. Sen, J. Eron, Genomic sequences and quasi-multivariate CATANOVA, in Handbook of Statistics 18 (2000) 713–746.

[22] C.R. Rao, Gini-Simpson index of diversity: a characterization, generalization and applications, Util. Math. 21 (1982) 273–282.

[23] T. Robertson, E.J. Wegman, Likelihood ratio tests for order restrictions in exponential families, Ann. Statist. 6 (1978) 485–505.

[24] S.N. Roy, On a heuristic method of test construction and its use in multivariate analysis, Ann. Math. Statist. 24 (1953) 220–238.

[25] P.K. Sen, Utility-oriented Simpson-type indexes and inequality measures, Bull. Calcutta Statist. Assoc. 49 (1999) 1–22.

[26] P.K. Sen, M.T. Tsai, Two-stage likelihood ratio and union-intersection tests for one-sided alternatives: multivariate mean with nuisance dispersion matrix, J. Multivariate Anal. 68 (1999) 264–282.

[27] A. Shapiro, On the asymptotics of constrained local *M*-estimators, Ann. Statist. 28 (2000) 948–960.

[28] E.H. Simpson, The measurement of diversity, Nature 163 (1949) 688.

[29] Y. Wang, M.P. McDermott, Conditional likelihood ratio test for a nonnegative normal mean vector, J. Amer. Statist. Assoc. 93 (1998) 380–386.