Discrete Efficient Methods 1: Testing Compound Hypotheses*

Pavel Boček, Tomáš Marek, Igor Vajda

November 2, 2006

Introduction

This report introduces into the research program DIFEM at the Department of Stochastic Informatics planned for the period 2005 – 2009. This program consist of a theoretical characterization, algorithmization, programming and experimental verification of selected class of advanced methods of statistical analysis of discrete data. It is a part of the chapter DECISION PROCESSES AND CLASSIFICATION of the Research Program DAR supported by the Ministry of Education of the Czech Republic. The aim of research program is to provide a user-friendly package of computer programs applying to concrete data the statistical methods proposed by the Department and recently published or accepted for publishing in renowned international mathematics journals. The publication in this type of journals guarantees the novelty and efficiency of these methods. The recent date of publication insures against their eventual appearance in the existing packages of statistical programs.

The general schedule of this program is as follows:

1. Theoretical research (publications and/or research reports, 2005 – 2009).
2. Algorithmization and programming (2005 – 2009).
3. Experiments with simulated data (publications and/or research reports, 2006 – 2009).
4. User-friendly finalization of programs (2008 – 2009).

A common feature of the statistical methods dealing with stochastic data and proposed by Department of Stochastic Informatics is that they are based on divergences $D(P, \widehat{P})$ between hypothetical stochastic models P from a apriori given classes $\mathcal{P}$ and empirical stochastic models $\widehat{P}$ obtained from the data. One of the internationally respected scientific achievements of the Institute are the results concerning properties of divergences of statistic models. The divergence-based statistical methods are thus a natural extension of these achievements. These methods can be classified and roughly characterized as follows:

*Supported by the MSMT grant 1M0572

a. Estimation of a model from a given class $\mathcal{P}$

$$\widetilde{P} = \arg\min_{P \in \mathcal{P}} D(P, \widehat{P})$$

b. α-size testing of a simple hypothesis $\mathcal{H} \equiv \mathcal{P}$:
Reject if

$$D(P, \widehat{P}) > D_{critical}(\alpha), 0 < \alpha < 1.$$

c. α-size testing of a composite hypothesis $\mathcal{H} \equiv \mathcal{P}$:
Reject if

$$\min_{P \in \mathcal{P}} D(P, \widehat{P}) > D_{\text{critical}}(\alpha), 0 < \alpha < 1.$$

d. Classification under an etalon class $\mathcal{P} = \{P_1, P_2, \ldots, P_n\}$:

$$\widetilde{P} = \arg\min_{P \in \mathcal{P}} D(P, \widehat{P}).$$

If $D(P, \widehat{P})$ is the classical logarithmic divergence (known also as a relative entropy of P with respect to $\widehat{P}$, or the Kullback divergence) then the estimation or classification are the classical maximum likelihood estimation or classification respectively. Similarly the testing of a simple or composite hypothesis is in this case the classical likelihood ratio or generalized likelihood ratio testing respectively. If $D(P, \widehat{P})$ is not the classical logarithmic divergence then the estimators, classifiers of the tests of hypotheses differ from the maximum likelihood classics.

Since various divergences of models P and $\widehat{P}$ are not mutually isotone (they are only "approximately isotone") the estimators, classifiers or tests obtained from various divergencies usually differ slightly. To give a hint how to choose the most efficient among these slightly different solutions we propose special methods of empirical optimization of estimators, classifiers and α-size tests based on the approaches published in the recent papers of the Department. Details are deferred to the special reports dealing with concrete divergence-depending estimators, classifiers and tests.

The activities of research program DIFEM planned for 2005 consisted of theoretic characterization, algorithmization and programming of compound statistical hypotheses based the previous papers of the Department dealing with such hypotheses. Section 1.1 introduces into the testing of hypotheses based on ϕ-divergence statistics and its robustification using more general ϕ-disparity statistics. Section 1.2 characterizes and illustrates the compound statistical hypotheses studied previously in the department. Section 1.3 summarizes a basic theory on which is based the computer program COMPOTEST the algorithm of which is given in Section 1.4. This report contains also the algorithm of the subprogram EOTEST for empirical optimization of the ϕ-disparity test (the subprogram **D** in Section 1.4).

1 Testing of compound hypotheses (program COMPOTEST)

1.1 Preliminaries

A common statistical situation is that there is given a complete system of mutually exclusive events

$$\mathcal{E} = \{E_1, E_2, \ldots, E_r\} \tag{1.1}$$

which are outcomes of a random experiment with probability distribution P. The problem is to test the hypothesis

$$\mathcal{H} : (P(E_1), P(E_2), \ldots, P(E_r)) = \boldsymbol{p} \tag{1.2}$$

for a given probability distribution

$$\boldsymbol{p} = (p_1, p_2, \ldots, p_r), \quad \prod_{j=1}^{r} p_j > 0. \tag{1.3}$$

The testing is based on the empirical evidence given by a system

$$\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_r) \tag{1.4}$$

of frequencies of the events $E_1, E_2, \ldots, E_r$ observed in

$$n = Y_1 + Y_2 + \cdots + Y_r \tag{1.5}$$

independent realizations of the experiment. Namely, the empirical probability distribution

$$\widehat{\boldsymbol{p}}(n) = (\widehat{p}_1(n), \widehat{p}_2(n), \ldots, \widehat{p}_r(n)) \triangleq \left(\frac{Y_1}{n}, \frac{Y_2}{n}, \ldots, \frac{Y_r}{n}\right) \tag{1.6}$$

is compared with the hypothetical distribution (1.3).

As explained in Menéndez, Morales, Pardo and Vajda [4], the goodness-of-fit of $\widehat{p}(n)$ and p can be measured by the ϕ-disparity statistics

$$\widehat{T}_\phi = \sum_{j=1}^{r} p_j \phi\left(\frac{\widehat{p}_j(n)}{p_j}\right) \tag{1.7}$$

for $\phi : (0, \infty) \mapsto \mathbb{R}$ which is twice continuously differentiable in neighbourhood of 1 with $\phi(1) = 0$, $\phi''(1) > 0$ and with $\widetilde{\phi}(t) \triangleq \phi(t) - \phi'(1)(t-1)$ monotone on the intervals $(0, 1)$ and $(1, \infty)$. Since $\phi''(1) > 0$, this means that $\widetilde{\phi}(t)$ is decreasing (nonincreasing) on $(0, 1)$ and increasing (nondecreasing) on $(1, \infty)$. This also means that the limit

$$\phi(0) = \lim_{u \downarrow 0} \phi(u) \in (-\infty, \infty]$$

exists and can be used as a substitution in (1.7) when $\widehat{p}_j(n) = 0$

Let Φ be the class of the above considered functions ϕ and $\widetilde{\Phi}$ its subclass restricted to $\phi \in \Phi$ with additional property $\phi'(1) = 0$. Since the above defined $\widetilde{\phi}$ belongs to Φ if ϕ does so and (1.7) implies that

$$\widehat{T}_{\widetilde{\phi}} = \widehat{T}_{\phi},$$

we may assume without less of generality that ϕ of (1.7) belongs to $\widetilde{\Phi}$ which is the class of functions $\phi : (0, \infty) \mapsto [0, \infty)$ nonincreasing on $(0, 1)$, nondecreasing on $(1, \infty)$ and twice continuously differentiable in a neighbourhood of 1 with $\phi(1) = 0, \phi''(1) > 0$. This implies that the ϕ-disparity statistics (1.7) are nonnegative, equal zero if and only if the distributions $\widehat{\boldsymbol{p}}(n)$ and $\boldsymbol{p}$ coincide.

Well known examples of functions from $\widetilde{\Phi}$ are the strictly convex functions

$$\phi^{(a)}(t) = \frac{t^a - at + a - 1}{a(a-1)}, \quad a \in \mathbb{R},\ a \neq 0, a \neq 1 \tag{1.8}$$

and their limits

$$\phi^{(1)}(t) = t \ln t - t + 1, \quad \phi^{(0)}(t) = -\ln t + t - 1 \tag{1.9}$$

leading to the power statistics

$$\widehat{T}^{(a)} = \frac{1}{a(a-1)} \left[\sum_{j=1}^{r} \widehat{p}_j(n)^a p_j^{1-a} - 1 \right], \quad a \in \mathbb{R},\ a \neq 0,\ a \neq 1 \tag{1.10}$$

and their limits

$$\widehat{T}^{(1)} = \sum_{j=1}^{r} \widehat{p}_j(n) \ln \frac{\widehat{p}_j(n)}{p_j}, \quad \widehat{T}^{(0)} = \sum_{j=1}^{r} p_j \ln \frac{p_j}{\widehat{p}_j(n)}. \tag{1.11}$$

We see that $2n\widehat{T}^{(1)}$ is the log-likelihood ratio statistic and $2n\widehat{T}^{(0)}$ the reversed log-likelihood ratio statistic. Further,

$$2n\widehat{T}^{(2)} = n \left[\sum_{j=1}^{r} \frac{\widehat{p}_j(n)^2}{p_j} - 1 \right] = n \sum_{j=1}^{r} \frac{(\widehat{p}_j(n) - p_j)^2}{p_j} \tag{1.12}$$

is the well-known Pearson statistic and

$$2n\widehat{T}^{(0)} = n \left[\sum_{j=1}^{r} \frac{p_j^2}{\widehat{p}_j(n)} - 1 \right] = n \sum_{j=1}^{r} \frac{(\widehat{p}_j(n) - p_j)^2}{\widehat{p}_j(n)} \tag{1.13}$$

is the well-known Neyman statistic. Finally,

$$2n\widehat{T}^{(1/2)} = 8n \left[1 - \sum_{j=1}^{r} \sqrt{\widehat{p}_j(n) p_j} \right] = 4n \sum_{j=1}^{r} \left(\sqrt{\widehat{p}_j(n)} - \sqrt{p_j} \right)^2 \tag{1.14}$$

is the Freeman–Tukey statistic.

As argumented in Lindsay [2], from the point of robustness it is desirable to use the ϕ-disparity statistics with the derivatives $\phi'(t)$ bounded on $(0, \infty)$. In the above considered examples the derivative $\phi_a'(t)$ are unbounded on $(0, \infty)$ for all real a. On other hand,

$$\phi_a(t) = 1 - \exp\left\{-a(t-1)^2\right\}, \quad a > 0 \tag{1.15}$$

are examples of functions belonging to $\widetilde{\Phi}$ with the derivatives $\phi'_a(t)$ bounded on $(0, \infty)$ for all $a > 0$.

Let us now return back to general statistic $\widehat{T}_\phi$ of (1.7) which is a nonnegative measure of disparity between the empirical distribution $\widehat{p}(n)$ and the hypothetical distribution p. If this statistic exceeds certain critical value $c_\phi > 0$ then the hypothesis (1.2) is rejected. If we want to achieve the probability of the decision error (test size) equal to a given $0 < \alpha < 1$,then the critical value must depend on α, i.e. $c_\phi = c_\phi(\alpha)$. It is well-known that if $n \to \infty$ and r is fixed then the Pearson statistic $2n\widehat{T}^{(2)}$ tends in law to the χ^2-distributed random variable χ^2_{r-1} with $r-1$ degrees of freedom.

More generally (see Morales, Pardo and Vajda [5]),

$$\lim_{n\to\infty} \frac{2n}{\phi''(1)} \widehat{T}_\phi \stackrel{\mathcal{L}}{=} \chi^2_{r-1} \tag{1.16}$$

so that asymptotically (for $n \to \infty$ and r fixed) α-size test is obtained by the $(1-\alpha)$-quantile

$$c_\phi(\alpha) = \chi^2_{r-1}(1-\alpha) \tag{1.17}$$

of the standardized ϕ-disparity statistic

$$\frac{2n\widehat{T}_\phi}{\phi''(1)}. \tag{1.18}$$

Our aim is to provide a test of a given size $0 < \alpha < 1$ for several simultaneous hypotheses of the type (1.2), including the case when the number k of such hypotheses is very large.

If several simple hypotheses are considered simultaneously then we speak about compound hypotheses. Testing of compound hypotheses of the type (1.2) is described in the next section.

1.2 Compound multinomial hypotheses

Let us now consider the situation more realistic than that of Section 1.1 in the sense that instead of one complete system $\mathcal{E} = \{E_1, E_2, \ldots, E_r\}$ of mutually exclusive events considered in (1.1) there are k independent systems

$$\mathcal{E}_i = \{E_{i1}, E_{i2}, \ldots, E_{ir_i}\}, \quad 1 \le i \le k. \tag{2.1}$$

The problem is to test the compound hypothesis

$$\mathcal{H} : (P(E_{i1}), P(E_{i2}), \ldots, P(E_{ir_i})) = \boldsymbol{p}_i, \quad 1 \le i \le k \tag{2.2}$$

for given probability distributions

$$\boldsymbol{p}_i = (p_{i1}, p_{i2}, \ldots, p_{ir_i}), \quad 1 \le i \le k \tag{2.3}$$

with $p_{ij} > 0$

The testing is based on the empirical evidence given by systems

$$\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{ir_i}), \quad 1 \le i \le k \tag{2.4}$$

of frequencies of the events $E_{i1}, E_{i2}, \ldots, E_{ir_i}$ observed in

$$n_i = Y_{i1} + Y_{i2} + \cdots + Y_{ir_i}$$

independent realizations of the experiment characterized by a probability distribution P figuring in (2.2) with the sum

$$n = n_1 + n_2 + \cdots + n_k. \tag{2.5}$$

Namely the empirical probability distributions

$$\widehat{\boldsymbol{p}}_i(n_i) = (\widehat{p}_{i1}(n_i), \widehat{p}_{i2}(n_i), \ldots, \widehat{p}_{ir_i}(n_i)) \triangleq \left(\frac{Y_{i1}}{n_i}, \frac{Y_{i2}}{n_i}, \ldots, \frac{Y_{ir_i}}{n_i}\right) \tag{2.6}$$

for $1 \leq i \leq k$ are jointly compared with the corresponding hypothetical distributions $\boldsymbol{p}_i$ given in (2.3). The comparison is based on the compound ϕ-disparity statistics.

$$\widehat{T}_\phi = \sum_{i=1}^{k} w_i \widehat{T}_{\phi,i} \tag{2.7}$$

where

$$w_i = \frac{n_i}{n} \tag{2.8}$$

are the weights proportional to the sample sizes n_i and

$$\widehat{T}_{\phi,i} = \sum_{j=1}^{r_i} p_{ij} \phi\left(\frac{\widehat{p}_{ij}(n_i)}{p_{ij}}\right) \tag{2.9}$$

are the ϕ-disparities of distributions $\widehat{\boldsymbol{p}}_i(n_i)$ and $\boldsymbol{p}_i$ defined for the $\phi \in \Phi$ in accordance with (1.7).

To obtain the critical values $c_\phi(\alpha)$ of statistics (2.7) corresponding to the asymptotically α-size tests, we need to know the asymptotic distributions of these statistics. Free parameters of the presented model are the number of classes k, the sample sizes $n_1, n_2, \ldots, n_k$ and the distribution sizes $r_1, r_2, \ldots, r_k$. Therefore the first problem is to find the combinations of free parameters for which it is reasonable to study the asymptotics of compound ϕ-disparity statistics. The empirical approach suggests the solution of this problem based on concrete examples.

Example 2.1. Let the electoral preferences among men and women of Prague be studied by a telephonic questioning of n_1 men and n_2 women. Here $min\{n_1, n_2\}$ may be very large, of the order of 10^3, while the number of classes $k = 2$ is small. The distribution sizes r_1 and r_2 coincide and they are given by the number of elected political parties which is a relatively small number of the order of 10.

Example 2.2. Health states (e.g. diagnoses or anamnestic data) $E_{i1}, E_{i2}, \ldots$ are specified in a large ensemble of n_i individuals distinguished by a given i-th combination of symptoms. Here the range $1 \leq i \leq k$ of possible combinations of symptoms is typically large too. The states $E_{i1}, E_{i2}, \ldots, E_{ir_i}$ including their total numbers r_i, strongly depend on the combinations of symptoms $1 \leq i \leq k$. The numbers $r_1, r_2, \ldots, r_k$ themselves may be large or small – this depends on the concrete medical research. For example, we have $r_1 = r_2 = \cdots = r_k = 2$ if the medical screening is interested only in $E_{i1} =$ "normal body temperature" and $E_{i2}=$ "increased body temperature" for all combinations of symptoms $1 \leq i \leq k$. On the other hand, r_i will be large and varying with combinations of symptoms $1 \leq i \leq k$ if the medical research concentrates on anamnestic data $E_{i1}, E_{i2}, \ldots, E_{ir_i}$ detected in the class of patients with the combinations of symptoms $1 \leq i \leq k$.

Example 2.3. Relatively small numbers $n_1, n_2, \ldots, n_k$ of skeletons of hominids living in subsequent millenia $1, 2, \ldots, k$ are providing a fossilized evidence sufficient to describe presence or absence of a given 32 skeletal antropologic features. In this case $r_1 = r_2 = \cdots = r_k$ is a constant equal $5 = \log_2 32$ while k may be large, of the order of 10^3.

Motivated by these examples, we shall consider the following separate cases. Note that these cases are mutually disjoint but not exhaustive, i. e. their union does not cover all possible situations.

Case A: $\max\{k, r_1, r_2, \ldots, r_k\} \ll \min\{n_1, n_2, \ldots, n_k\}$. Mathematically, this means that k and $r_1, r_2, \ldots, r_k$ (as well as the vectors $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_k$) are fixed and

$$\min\{n_1, n_2, \ldots, n_k\} \to \infty. \tag{2.10}$$

This case is illustrated by Example 2.1.

Case B: $k \ll min\{r_1, r_2, \ldots, r_k\}, max\{r_1, r_2, \ldots, r_k\} \ll min\{n_1, n_2, \ldots, n_k\}$. Mathematically, this means that k is fixed and

$$\min\{r_1, r_2, \ldots, r_k\} \to \infty, \quad \max\left\{\frac{r_1^2}{n_1}, \ldots, \frac{r_k^2}{n_k}\right\} \to 0. \tag{2.11}$$

This case is illustrated by Example 2.2 when $max\{r_1, r_2, \ldots, r_k\}$ is much smaller than $\min\{n_1, n_2, \ldots, n_k\}$.

Case C: $k \ll \min\{r_1, r_2, \ldots, r_k\}, \{r_1, r_2, \ldots, r_k\} \approx \{n_1, n_2, \ldots, n_k\}$. Mathematically, this means that k is fixed and

$$\min\{r_1, r_2, \ldots, r_k\} \to \infty, \quad \left(\frac{r_1}{n_1}, \ldots, \frac{r_k}{n_k}\right) \to (\gamma_1, \ldots, \gamma_k) \tag{2.12}$$

where $0 < \gamma_i < \infty$ for all $1 \leq i \leq k$. This case is illustrated by Example 2.2 when r_i are comparable to n_i for all $1 \leq i \leq k$.

Case D: $\max\{r_1, r_2, \ldots, r_k, n_1, n_2, \ldots, n_k\} \ll k$. Mathematically, this means that $r_1, r_2, \ldots, r_k, n_1, n_2, \ldots, n_k$ are uniformly bounded and

$$k \to \infty. \tag{2.13}$$

This case is illustrated by Example 2.3.

1.3 Theoretical background

In this section we present results about asymptotic distributions of the compound ϕ-disparity statistics $\widehat{T}_\phi$ defined in (2.7). Unless otherwise stated, these results were proved in Morales, Pardo and Vajda [5]. The relation

$$\lim \widehat{T}_\phi \stackrel{\mathcal{L}}{=} X \tag{3.1}$$

means the convergence of $\widehat{T}_\phi$ in distribution to a random variable X. The interpretation of the convergence (3.1) in the above specified Cases $\mathbf{A}-\mathbf{D}$ are intuitively clear – for a rigorous definition we refer to pp. 339–340 in Morales, Pardo and Vajda [5].

Theorem 3.1. In the Case **A**

$$\lim \frac{2n\widehat{T}_\phi}{\phi''(1)} \stackrel{\mathcal{L}}{=} \chi^2_{r-k} \tag{3.2}$$

for n given by (2.5) and

$$r = r_1 + r_2 + \cdots + r_k. \tag{3.3}$$

Proof. See a more general variant of (3.2) in Theorem 2.2 of Morales, Pardo and Vajda [5]. That variant is established under local alternatives

$$P(E_{ij}) = p_{ij} + \frac{c_{ij}}{\sqrt{n}}, \quad 1 \le j \le r_i, \ 1 \le i \le k \tag{3.4}$$

where $c_{i_1} + c_{i_2} + \cdots + c_{ir_i} = 0$ for all $1 \le i \le k$.

Theorem 3.2. If in the Case **B**

$$\liminf \min_{1\le i\le k} r_i \min_{1\le j\le r_i} p_{ij} > 0 \tag{3.5}$$

then

$$\lim \frac{1}{\sqrt{2r}} \left[\frac{2n\widehat{T}_\phi}{\phi''} - r \right] \stackrel{\mathcal{L}}{=} N(0,1) \tag{3.6}$$

for n given by (2.5) and r given by (3.3).

Proof. This statement follows from Theorem 3.1 of Morales, Pardo and Vajda [5] proved under the local alternatives (3.4) but with the second condition of (2.11) replaced by the stronger

$$\max \left\{ \frac{r_1^2 \ln^2 n_1}{n_1}, \ldots, \frac{r_k^2 \ln^2 n_k}{n_k} \right\} \to 0.$$

The possibility to arrive at the desired end under the weaker condition considered in (2.11) is offered by replacing the central limit theorem of Inglot, Jurlewicz and Ledwina [3] used in the proof of Morales, Pardo and Vajda [5] by the central limit theorem of Györfi and Vajda [1].

Remark 3.4. The condition (3.5) is automatically satisfied under the uniform compound hypothesis

$$\mathcal{H} : \boldsymbol{p}_i = \boldsymbol{p}^{(r_i)}, \quad 1 \le i \le k \tag{3.7}$$

where $\boldsymbol{p}^{(r)}$ is the r-size uniform distribution,

$$\boldsymbol{p}^{(r)} = \left(\frac{1}{r}, \frac{1}{r}, \ldots, \frac{1}{r}\right), \quad r \ge 1. \tag{3.8}$$

In the rest of the paper

$$\mathbf{E}X, \quad var\, X \quad \text{and} \quad cov(X, Y)$$

denote the expectation and variance of a random variable X and the covariance of X and Y. Further, we put

$$\mu_i = \mathbf{E} D_\phi\left(\widehat{\boldsymbol{p}}_i(n), \boldsymbol{p}_n\right) = \mathbf{E} \sum_{j=1}^{r_i} p_{ij} \phi\left(\frac{Y_{ij}}{np_{ij}}\right) \tag{3.9}$$

and

$$\sigma_i^2 = var\, D_\phi\left(\widehat{\boldsymbol{p}}_i(n), \boldsymbol{p}_i\right) \tag{3.10}$$

for the multinomialy distributed observations

$$\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{ir_i}) \sim M_{r_i}(n_i, \boldsymbol{p}_i) \tag{3.11}$$

introduced in (2.4). If $\tilde{\boldsymbol{p}}_i(n)$ are the relative frequencies (2.6) obtained for the independently Poisson distributed

$$\tilde{\boldsymbol{Y}}_i = \left(\tilde{Y}_{i1}, \tilde{Y}_{i2}, \ldots, \tilde{Y}_{ir_i}\right) \sim \bigotimes_{j=1}^{r_i} Poisson(\lambda_{ij}) \tag{3.12}$$

with the intensities

$$\lambda_{ij} = n_i p_{ij}, \quad 1 \le j \le r_i,\ 1 \le i \le k \tag{3.13}$$

then we put

$$\tilde{\mu}_i = \mathbf{E} D_\phi\left(\tilde{\boldsymbol{p}}_i(n), \boldsymbol{p}_i\right) = \mathbf{E} \sum_{j=1}^{r_i} p_{ij} \phi\left(\frac{\tilde{Y}_{ij}}{\lambda_{ij}}\right) \tag{3.14}$$

$$\tilde{\sigma}_i^2 = var\, D_\phi\left(\tilde{\boldsymbol{p}}_i(n), \boldsymbol{p}_i\right) \tag{3.15}$$

and

$$\tilde{s}_i^2 = n\tilde{\sigma}_i^2 - \left[\sum_{i=1}^{r_i} cov\left(Y_{ij},\, p_{ij}\phi\left(\frac{\tilde{Y}_{ij}}{\lambda_{ij}}\right)\right)\right]^2. \tag{3.16}$$

The next theorem was proved in Morales, Pardo and Vajda [5].

Theorem 3.5. If in the Case **C** in addition to (3.5)

$$\limsup \max_{1\le i\le k} r_i \max_{1\le j\le r_i} p_{ij} < \infty \tag{3.17}$$

and $\tilde{s}_1^2, \dots, \tilde{s}_k^2$ are bounded, and bounded away from zero, then

$$\lim \frac{\sqrt{n}(\widehat{T}_\phi - \tilde{\mu})}{\tilde{s}} = N(0,1) \tag{3.18}$$

for n given by (2.5) and

$$\mu = \sum_{i=1}^{k} w_i \tilde{\mu}_i, \quad \tilde{s}^2 = \sum_{i=1}^{k} w_i \tilde{s}_i^2 \tag{3.19}$$

defined by means of the weights (2.8), and for the same $\phi \in \Phi$ as in the previous theorems satisfying the additional conditions

$$\lim_{t\downarrow 0} \phi(t) < \infty, \quad \lim_{t\to\infty} \frac{\ln \phi(t)}{t} < \infty. \tag{3.20}$$

Remark 3.6. Under the uniform compound hypotheses (3.7) the conditions (3.5) and (3.17) hold. The power functions $\phi^{(a)} \in \Phi$ defined by (1.8), (1.9) satisfy (3.20) for all $a > 0$. It is clear from the next example that for $\phi^{(2)} \in \Phi$ the parameters $\tilde{s}_1^2, \dots, \tilde{s}_k^2$ are in the Case C bounded, and bounded away from zero, as required by Theorem 3.5.

Example 3.7. For $\phi = \phi^{(2)}$ defined by (1.8) we obtain from (3.14)

$$\begin{aligned}
\tilde{\mu}_i &= \mathbf{E} \sum_{j=1}^{r_i} p_{ij}\, \phi^{(2)}\left(\frac{\tilde{Y}_{ij}}{\lambda_{ij}}\right) \\
&= \frac{1}{2n_i} \sum_{j=1}^{2} \frac{\mathbf{E}(\tilde{Y}_{ij} - \lambda_{ij})^2}{\lambda_{ij}} = r_{ij} \\
&= \frac{1}{2n_i} \sum_{j=1}^{r_i} 1 = \frac{r_i}{2n_i}.
\end{aligned}$$

Hence, by (2.12),

$$\tilde{\mu}_i = \gamma_i, \quad 1 \le i \le k. \tag{3.21}$$

Further, by (3.15),

$$\begin{aligned}
\tilde{\sigma}_i &= \frac{1}{4n_i^2} \sum_{j=1}^{r_i} \frac{var(\tilde{Y}_{ij} - \lambda_{ij})^2}{\lambda_{ij}^2} \\
&= \frac{1}{4n_i^2} \sum_{j=1}^{r_i} \frac{\lambda_{ij}(1 + 2\lambda_{ij})}{\lambda_{ij}^2} \\
&= \frac{1}{4n_i^2} \left[\sum_{j=1}^{r_i} \frac{1}{np_{ij}} + 2r_i \right].
\end{aligned}$$

Since

$$cov\left(Y_{ij}, p_{ij}\phi^2\left(\frac{\tilde{Y}_{ij}}{\lambda_{ij}}\right)\right) = \frac{1}{2n_i^2\lambda_{ij}}\,cov\left(Y_{ij},\,(Y_{ij}-\lambda_{ij})^2\right) = \frac{1}{2n_i},$$

we get from (3.16)

$$\tilde{s}_i^2 = \frac{1}{4n_i}\left[\sum_{j=1}^{r_i}\frac{1}{n_i p_{ij}} + 2r_i\right] - \frac{1}{4}\left[\sum_{i=1}^{r_i}\frac{1}{n_i}\right]^2 = \frac{r_i}{2n_i}\left(1+\frac{1}{2r_i}\sum_{j=1}^{r_i}\frac{1}{n_i p_{ij}} - \frac{r_i}{2n_i}\right).$$

Taking again into account (2.2) we get

$$\tilde{s}_i^2 = \frac{\gamma_i}{2}\left(1+\gamma_i\,D^{(2)}\left(\boldsymbol{p}^{(r_i)},\boldsymbol{p}_i\right)\right), \quad 1\le i\le k \tag{3.22}$$

where $\boldsymbol{p}^{(r_i)}$ is the r_i-valued uniform distribution defined by (3.8) and

$$D^{(2)}(\boldsymbol{p},\boldsymbol{q}) = \frac{1}{2}\sum_{i=1}^{r}\frac{(p_i-q_i)^2}{q_i} = \frac{1}{2}\left(\sum_{i=1}^{r}\frac{p_i^2}{q_i} - 1\right)$$

is the Pearson-type ϕ-disparity obtained for $\phi(t) = \phi^{(2)}(t)$ given by (1.8). Thus the parameters $\tilde{s}_i^2$ are in the Case **C** positive and finite, increasing from $\gamma_i/2$ when $\boldsymbol{p}_i$ is uniform to ∞ when at least one coordinate of $\boldsymbol{p}_i$ is approaching zero.

Example 3.8. Let us consider the bounded disparity function

$$\phi(t) = \left(1 - t\,e^{1-t}\right)/e, \quad t>0 \tag{3.23}$$

belonging to the class Φ with the bounded derivative

$$\phi'(t) = (t-1)\,e^{-t}, \quad t>0. \tag{3.24}$$

Here

$$\widehat{T}_{\phi,i} = \frac{1}{e}\left(1-\sum_{j=1}^{r_i}\widehat{p}_{ij}(n)\exp\left\{\frac{q_{ij}-\widehat{p}_{ij}(n)}{p_{ij}}\right\}\right) \tag{3.25}$$

are the components (2.9) of the compound statistic $\widehat{T}_\phi$ given as the linear combination (2.7) with the weights (2.8). By (3.14),

$$\tilde{\mu}_i = \frac{1}{e}\left(1-\sum_{j=1}^{r_i}p_{ij}\sum_{k=1}^{\infty}\frac{k}{\lambda_{ij}}\exp\left\{1-\frac{k}{\lambda_{ij}}\right\}\frac{\lambda_{ij}^k}{k!}\,e^{-\lambda_{ij}}\right) = \frac{1}{e}\left(1-\sum_{j=1}^{r_i}p_{ij}\,\exp\left\{1-\lambda_{ij}-\frac{1}{\lambda_{ij}}\right\}\sum_{k=1}^{\infty}\frac{\tilde{\lambda}_{ij}^{k-1}}{(k-1)!}\right)$$

for

$$\tilde{\lambda}_{ij} = \lambda_{ij}\, e^{-1/\lambda_{ij}}.$$

Therefore

$$\tilde{\mu}_i = \frac{1}{e}\left(1 - \sum_{j=1}^{r_i} p_{ij} \exp\left\{\lambda_{ij}\, e^{-1/\lambda_{ij}} + 1 - \lambda_{ij} - \frac{1}{\lambda_{ij}}\right\}\right)$$

where $\lambda_{ij} = n_i\, p_{ij}$. If $\boldsymbol{p}_i = \boldsymbol{p}^{(r)}$, i. e., if $p_{ij} = 1/r_i$, then $\lambda_{ij} = 1/\gamma_i$ for γ_i given in (2.12) and all $1 \le i \le k$. Hence in this case

$$\tilde{\mu}_i = \frac{1}{e}\left(1 - \exp\left\{\frac{e^{-\gamma_i} - 1}{\gamma_i} + 1 - \frac{1}{\gamma_i}\right\}\right), \quad 1 \le i \le k. \tag{3.26}$$

Similarly we can evaluate

$$\widetilde{\sigma}_i^2 \tag{3.27}$$

and

$$\widetilde{s}_i^2 \tag{3.28}$$

In the following theorem we use for the above considered function $\phi = \phi^{(2)}$ belonging to Φ and the expectations μ_i and variances σ_i^2 defined by (3.9) and (3.10). We obtain for all $1 \le i \le k$

$$\mu_i = \frac{1}{2n_i^2} \sum_{j=1}^{r_i} \frac{\mathbf{E}(Y_j - n_i p_{ij})^2}{p_{ij}} = \frac{r_i - 1}{2n_i} \tag{3.29}$$

and

$$\begin{aligned} \sigma_i^2 &= \frac{1}{4n_i^4}\Big\{\sum_{j=1}^{r_i} \frac{\mathbf{E}(Y_j - n_i p_{ij})^4}{p_{ij}^2} + \sum_{\ell \ne j}^{r_i} \frac{\mathbf{E}[(Y_j - n_i p_{ij})\,(Y_\ell - n p_{i\ell})]}{p_{ij\ i\ell}}\Big\} - \left(\frac{r_i - 1}{2n_i}\right)^2 \\ &= \frac{r_i}{2n_i^2}\left(\frac{n_i r_i - n_i - r_i + 1}{n_i r_i} + \frac{r_i}{n_i}\, D^{(2)}(\boldsymbol{p}^{(r_i)}, \boldsymbol{p}_i)\right), \end{aligned} \tag{3.30}$$

where $D^{(2)}(\boldsymbol{p}^{(r_i)}, \boldsymbol{p}_i)$ is the same Pearson-type measure of non-uniformity of $\boldsymbol{p}_i$ as that considered in Example 3.7.

Theorem 3.9. Let in the Case **D** the probabilities p_{ij} be bounded away from 0 uniformly for all i and j and let $\boldsymbol{p}_i$ be bounded away from the uniform distribution $\boldsymbol{p}^{(r_i)}$ uniformly for all i such that $n_i = 1$. Then

$$\lim \frac{\widehat{T}^{(2)} - \overline{\mu}}{\overline{\sigma}} \stackrel{\mathcal{L}}{=} N(0,1) \tag{3.31}$$

where

$$\widehat{T}^{(2)} = \sum_{i=1}^{k} w_i\, \widehat{T}_i^{(2)} \tag{3.32}$$

for w_i given by (2.8) and the Pearson type disparities

$$\widehat{T}_i^{(2)} = \frac{n_i}{2} \sum_{j=1}^{r_i} \frac{(\widehat{p}_{ij}(n) - p_{ij})^2}{p_{ij}} \quad \text{(cf. (1.12))},$$

and

$$\overline{\mu} = \sum_{i=1}^{k} w_i \mu_i, \quad \overline{\sigma}^2 = \sum_{i=1}^{k} w_i \sigma_i^2 \tag{3.33}$$

for μ_i, σ_i^2 defined by (3.29), (3.30).

Proof. Clear from Theorem 4.1 in Morales, Pardo and Vajda [5]. □

1.4 Program COMPOTEST

Inputs parameters:

a. Natural numbers $k, r_1, r_2, \ldots, r_k, n_1, n_2, \ldots, n_k$.

b. Nonnegative integers $0 \le Y_{ij} \le n_i$ for $1 \le j \le r_i$ and $1 \le i \le k$.

c. Rational numbers $0 \le p_{ij} \le 1$ for $1 \le j \le r_i$ and $1 \le i \le k$. They define compound hypothesis.

d. Rational number $0 < \alpha < 1$. It defines the described test size.

e. A capital from set $\{A,B,C,D,E\}$. The choice from the subset $\{A,B,C,D\}$ means a manual specification of one of the cases **A**,**B**,**C**,**D** specified above. The choice E means an automatic specification of one of these cases. This specification is described in the point **E** below.

f. Subset $\mathcal{A} \subset \mathcal{A}_0 \cup \mathcal{A}_1$ where

$$\mathcal{A}_0 = \{-2, -3/2, -1, -1/2, 0, 1/2, 1, 3/2, 2, 5/2, 3\}$$

and

$$\mathcal{A}_1 = \{10, 20, 30\}$$

The set $\mathcal{A}$ specifies the class of compound statistics $\widehat{T}_\phi$ used by the program. If $a \in \mathcal{A}_0$ then $T_\phi = \widehat{T}^{(a)}$ is the compound power divergence statistic defined by (2.7) – (2.9) for $\phi = \phi^{(a)}$ given by (1.7) – (1.8). If $a = 10$ then $\widehat{T}_\phi$ is defined by (2.7) – (2.9) for ϕ given by (3.23). If $a = 20$ then $\widehat{T}_\phi$ is defined by (2.7) – (2.9) for ϕ given by a special subprogram. Finally, $a = 30$ means an empirical choice of $\widehat{T}_\phi$ from the class $a \in \mathcal{A} \cup \{10, 20\}$.

Subprograms:

A. Evaluation of $\widehat{T}^{(a)}$ for $a \in \mathcal{A}_0$. Here the compound statistics $\widehat{T}^{(a)}$ are evaluated for the input parameters $a \in \mathcal{A}$ using the formulas (2.7) – (2.9) with $\phi = \phi^{(a)}$ of (1.8), (1.9) inserted in (2.9). In other words, these statistics are evaluated by the formula

$$\widehat{T}^{(a)} = \frac{1}{n} \sum_{i=1}^{k} n_i \widehat{T}_i^{(a)} \tag{4.1}$$

where n and n_i are the input parameters and $\widehat{T}_i^{(a)}$ are the power divergence statistics given by (1.10), (1.11) with p_j replaced by the input parameters p_{ij} and the empirical probabilities $\widehat{p}_j(n)$ by the rations

$$\widehat{p}_{ij}(n_i) = \frac{Y_{ij}}{n_i} \tag{4.2}$$

of the input parameters Y_{ij} and n_i as it is required by (2.6).

B. Evaluation of $\widehat{T}^{(a)}$ for $a = 10$. Here the compound ϕ-divergence statistic $\widehat{T}_\phi$ is evaluated by the formulas (2.7) – (2.9) for ϕ given by (3.23). In other words,

$$\widehat{T}_\phi = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{r_i} n_i p_{ij} \phi \left(\frac{\widehat{p}_{ij}(n_i)}{p_{ij}} \right) \tag{4.3}$$

where n, n_i and p_{ij} are the input parameters, $\widehat{p}_{ij}(n_i)$ are given by (4.2) for the input parameters Y_{ij} and n_i and

$$\phi(t) = \frac{1 - te^{1-t}}{e}, \quad t \geq 0. \tag{4.4}$$

C. Evaluation of $\widehat{T}^{(a)}$ for $a = 20$. Here the compound ϕ-divergence statistic $\widehat{T}_\phi$ is evaluated by the formulas (4.3) for ϕ given by a special input subprogram. The function $\phi(t)$ must satisfy the assumptions presented in Section 1.1.

D. Evaluation of the empirically optimal $\widehat{T}_\phi$ when $a = 30$. In this case the compound ϕ-divergence statistic $\widehat{T}_\phi$ is selected from the class $\{\widehat{T}^{(a)} : a \in \mathcal{A}\}$ where the set $\mathcal{A}$ is the input parameter. In this class $\widehat{T}^{(a)}$ denotes the power divergence statistic evaluated by the subbprogram **A** if $a \in \mathcal{A}_0$, $\widehat{T}^{(10)}$ denotes the statistic evaluated by the subprogram **B** if $a = 10$ belongs to $\mathcal{A}$ and, similarly, $\widehat{T}^{(20)}$ denotes the statistic evaluated by the subprogram **C** if $a = 20$ belongs to $\mathcal{A}$. The selection is done by a program EOTEST (Empirically Optimized Test) described in the next step.

The EOTEST program has as input parameters natural number $M > 1$ (typically a multiple of the above mentioned input parameter n) N (typically $N = 10^4$) and a sequence $\xi_1, \xi_2, \ldots$ of independent uniformly distributed binary digits. For every $1 \leq m \leq M$ and $1 \leq i \leq k$ the program proceeds as follows.

(a) Using random digits $\xi_1, \xi_2, \ldots \xi_{n_i}$ which do not coincide (i. e. $\xi_{j_1} \neq \xi_{j_2}$ for at least one pair $1 \leq j_1 < j_2 \leq n_i$) the support set $S_i = \{1, 2, \ldots, n_i\}$ is randomly splitted in two subsupports $S_{i,A}$ and $S_{i,B}$. Put

$$p_{i,A} = \sum_{j \in S_{i,A}} p_{ij}, \quad p_{i,B} = \sum_{j \in S_{i,B}} p_{ij} \tag{4.5}$$

and

$$\beta_{i,A} = \min \left\{ 1, \frac{1 - p_{i,A}}{p_{i,A}} \right\}, \quad \beta_{i,B} = \min \left\{ 1, \frac{1 - p_{i,B}}{p_{i,B}} \right\}. \tag{4.6}$$

Since p_{ij} is assumed to be positive for all $1 \le j \le n_i$, it holds

$$0 < p_{i,A}\beta_{i,A} = (1 - p_{i,A})\beta_{i,B} < 1. \tag{4.7}$$

Therefore the numbers

$$q_{ij} = \{ \begin{array}{ll} p_{ij}(1+\beta_{i,A}) & \text{for} \quad j \in S_{i,A} \\ p_{ij}(1-\beta_{i,B}) & \text{for} \quad j \in S_{i,B} \end{array} \tag{4.8}$$

are nonnegative. We shall prove that if

$$q_{i,A} = \sum_{j \in S_{i,A}} q_{ij}, \quad q_{i,B} = \sum_{j \in S_{i,B}} q_{ij}$$

then

$$q_{i,A} + q_{i,B} = 1 \tag{4.9}$$

i.e. that $\boldsymbol{q}_i = (q_{ij} : 1 \le j \le r_i)$ is a probability distribution. By (4.5), (4.8) and (4.9),

$$\begin{aligned} q_{i,A} + q_{i,B} &= p_{i,A}(1+\beta_{i,A}) + (1-p_{i,A})(1-\beta_{i,B}) \\ &= 1 + p_{i,A}\beta_{i,A} - (1-p_{i,A})\beta_{i,B} \\ &= 1 + min\{p_{i,A}, 1-p_{i,A}\} - min\{p_{i,A}, 1-p_{i,A}\} \\ &= 1 \end{aligned}$$

so that (4.9) holds. The sets A, B and consequently, the parameters $\beta_{i,A}$ and $\beta_{i,B}$ depend on the n-th block of the random variables $\xi_1, \xi_2, \ldots, \xi_{n_i}$, i.e. they depend on $1 \le m \le M$. The program evaluates for every $1 \le m \le M$ the local alternatives to the distributions $\boldsymbol{p}_i$ defined by

$$\boldsymbol{p}_i^{(m)} = \left(1 - \frac{1}{\sqrt{n_i}}\right)\boldsymbol{p}_i + \frac{1}{\sqrt{n_i}}\boldsymbol{q}_i^{(m)} \tag{4.10}$$

where $\boldsymbol{q}_i^{(m)}$ are the distributions defined by (4.8) for $A = A^{(m)}$ and $B = B^{(m)}$. Instead of (4.10) we can use the equivalent formula

$$p_{ij}^{(m)} = \{ \begin{array}{ll} p_{ij}(1+\frac{\beta_{i,A}}{\sqrt{n_i}}) & \text{for} \quad j \in S_{i,A} \\ p_{ij}(1-\frac{\beta_{i,B}}{\sqrt{n_i}}) & \text{for} \quad j \in S_{i,B} \end{array} \tag{4.11}$$

From here we see that in the random subsupport $A = A^{(m)}$ the local alternatives slightly increase the values p_{ij} and in the random complement $B = B^{(m)}$ they slightly decrease these values. At the same time we see that the local alternatives are random.

(b) For every $1 \le m \le M$, the program simulates N independent random realizations of the input data Y_{ij} generated by the models $\{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_k\}$ and $\{\boldsymbol{p}_1^{(m)}, \ldots, \boldsymbol{p}_k^{(m)}\}$ and evaluates the corresponding relative frequencies $\pi(\widehat{T}^{(a)}, \alpha)$ and $\pi_m(\widehat{T}^{(a)}, \alpha)$ of rejections of the compound hypothesis $\mathcal{H}$ given in (2.2) by the α-size test using the compound statistics $\widehat{T}^{(a)}$ for given input parameters $a \in \mathcal{A}$ and $0 < \alpha < 1$.

(c) The program evaluates the weighted absolute distances of the local alternatives $\boldsymbol{p}_i^{(m)}$ from the hypotheses $\boldsymbol{p}_i$,

$$
\begin{aligned}
L_{1,m} &= \sum_{i=1}^{k} \frac{n_i}{n} \sum_{j=1}^{n_i} |p_{ij} - p_{ij}^{(m)}| \\
&= \sum_{i=1}^{k} \frac{n_i}{n} \frac{1}{\sqrt{n_i}} \left[p_{i,A}^{(m)} \beta_{i,A}^{(m)} + p_{i,B}^{(m)} \beta_{i,B}^{(m)} \right] \\
&= \frac{2}{n} \sum_{i=1}^{k} \sqrt{n_i}\ \min\{p_{i,A}^{(m)} (1 - p_{i,B}^{(m)})\} \qquad (4.12)
\end{aligned}
$$

(d) The program evaluates empirically optimized statistic

$$
\widehat{T}_\phi \triangleq \operatorname{argmax} \frac{1}{M} \sum_{m=1}^{M} \frac{\pi(\widehat{T}_\phi^{(a)}, \alpha) - \pi_m(\widehat{T}_\phi^{(a)}, \alpha)}{L_{1,M}} \qquad (4.13)
$$

where the maximization extends over the input class of statistics $\{\widehat{T}^{(a)} : a \in \mathcal{A}\}$ and $\pi(\widehat{T}^{(a)}, \alpha)$, $\pi_m(\widehat{T}^{(a)}, \alpha)$ defined in (b). The actual size of the $\widehat{T}_\phi$-based test of the compound hypothesis $\mathcal{H}$ of (2.2) can be estimated by

$$
\widehat{\alpha} = \pi(\widehat{T}_\phi, \alpha) \qquad (4.14)
$$

and the power of this test under local alternatives of type

$$
\mathcal{A}_m : (P(E_{i1}), P(E_{i2}), \ldots, P(E_{ir_i})) = \boldsymbol{p}_i^{(m)}, \quad 1 \le i \le k \qquad (4.15)
$$

can be estimated by

$$
\widehat{\pi} = \frac{1}{M} \sum_{m=1}^{M} \pi_m(\widehat{T}_\phi, \alpha). \qquad (4.16)
$$

(e) The program prints $a_{opt} \in \mathcal{A}$ defined by

$$
\widehat{T}^{(a_{opt})} = \widehat{T}_\phi
$$

for $\widehat{T}_\phi$ given by (4.13), i. e. a_{opt} indicates the empirically optimal statistic in the input set $\{\widehat{T}^{(a)} : a \in \mathcal{A}\}$. It also prints the estimates $\widehat{\alpha}$ and $\widehat{\pi}$ of the actual size and power of the empirically optimized test corresponding to this statistics.

E. Automatic specification of the case. This subprogram will be proposed later after collecting experience with the programs **A**-**D**.

F. Control of the statistics which are at the disposal for the given case (similarly as the subprogram **E**, this subprogram will be proposed later).

G. Evaluation of the quantiles $\chi_m^2(1-\alpha)$ for $1 \le m \le r_1 + r_2 + \ldots r_k$.

H. Evaluation of the standard normal quantiles $F(1-\alpha)$.

References

[1] L. Györfi and I. Vajda (2002): *"Asymptotic distributions for goodness-of-fit statistics in sequence of multinomial models."*, Statistics and Probability Letters **56**, 57-67.

[2] B.G. Lindsay (1994): *"Efficiency versus robustness."* Annals of Statistics **22**, 1081-1114.

[3] T. Inglot, T. Jurlevwicz and T. Ledwina (1990): *"Asymptotics for multinomial goodnes-of-fit for simple hypothesis."*, Theory of Probability and Applications **35**, 771-777.

[4] M.L. Menéndez, D. Morales, L. Pardo and I. Vajda (1998): *"Two approaches to grouping of data and related disparity statistics."*, Communications in Statistics - Theory and Methods **27**, 609-633.

[5] D. Morales, L. Pardo and I. Vajda (2003): *"Asymptotic laws for disparity statistics in product multinomial models."*, Journal of Multivariate Analysis **85**, 335-360.