ANADOLU ÜNİVERSİTESİ BİLİM VE TEKNOLOJİ DERGİSİ
ANADOLU UNIVERSITY JOURNAL OF SCIENCE AND TECHNOLOGY
Cilt/Vol.:8-Sayı/No: 2 : 475-485 (2007)

RESEARCH ARTICLE / **ARAŞTIRMA MAKALESİ**

TESTING LOG-LINEAR MODELS IN THREE-WAY CONTINGENCY TABLES USING ORDINARY AND PENALIZED POWER-DIVERGENCE TEST STATISTICS

Aylin ALIN[1], Serdar KURT[2]

ABSTRACT

Basu and Basu (1998) considered the empty cell penalty for the family of power-divergence measures, defined by Cressie and Read (1984), in multinomial models. This penalty gives the chance to change the weight put on empty cells in contingency tables by the family of power-divergence measures. Cressie and Pardo (2000) introduced the family of ordinary power-divergence test statistics to test nested log-linear models. This family includes the likelihood ratio and Pearson's chi-squared test statistics. In this study, we define the family of penalized power-divergence test statistics by using penalized power-divergence measure and penalized power-divergence estimators in the family of ordinary power-divergence test statistics. We compare these two families of test statistics for small samples in terms of size and power properties based on both asymptotic and finite sample (using a designed simulation study) results. We consider three-way contingency tables distributed according to a multinomial distribution with probabilities belonging to a log-linear model. Our results reveal penalization improves the size and power properties of ordinary power-divergence test statistics.

Keywords: Contingency table, Log-linear models, Power-divergence measure, Penalized power-Divergence measure.

ÜÇ YÖNLÜ OLUMSALLIK TABLOLARINDA LOGARİTMİK-DOĞRUSAL MODELLERİN SIRADAN VE CEZALANDIRILMIŞ GÜÇ-SAPMA TEST İSTATİSTİKLERİ İLE TEST EDİLMESİ

ÖZ

Basu ve Basu (1998) çoklu örnekleme modellerinde Cressie ve Read (1984) tarafından tanımlanan güç-sapma ölçümleri ailesi için boş hücre cezası tanımlamışlardır. Bu ceza olumsallık tablolarında boş hücrelere güç-sapma ölçümleri tarafından atanan ağırlıkları değiştirme imkanı tanımaktadır. Cressie ve Pardo (2000) iç içe geçmiş logaritmik doğrusal modellerin test edilmesi için sıradan güç-sapma test istatistikleri ailesini önermişlerdir. Bu aile olabilirlik oran testi ve Pearson'ın ki-kare test istatistatistiklerini de içermektedir. Bu çalışmada, sıradan güç-sapma test istatistikleri ailesinde cezalandırılmış güç-sapma ölçümleri ve tahmincileri kullanılarak cezalandırılmış güç-sapma test istatistikleri ailesi tanımlanmıştır. Sıradan güç-sapma test istatistikleri ailesi ile önerilen cezalandırılmış güç-sapma test istatistikleri ailesinin küçük örneklem birinci tip hata ve güç özellikleri örneklem genişliği sonsuza giderken ve sonlu örneklem (özel tasarlanmış benzetim çalışması kullanarak) durumlarına göre karşılaştırılmıştır. Benzetim çalışması için olasılıkların logaritmik-doğrusal modele uyduğu çoklu örneklem dağılımına sahip olan üç-boyutlu olumsallık tabloları baz alınmıştır. Benzetim çalışması sonuçları boş hücrelere verilen cezanın sıradan güç-sapma test istatistikleri ailesinin birinci tip hata ve güç özelliklerini geliştirdiğini göstermektedir.

Anahtar Kelimeler: Cezalandırılmış güç-sapma ölçümleri, Güç-sapma ölçümleri, Logaritmik-doğrusal modeller, Olumsallık tabloları.

[1,] Dokuz Eylul Universitesi, Fen Edebiyat Fakultesi, Istatistik Bolumu, Kaynaklar Yerleskesi, 35160, Buca, Izmir.
[2,] Dokuz Eylul Universitesi, Fen Edebiyat Fakultesi, Istatistik Bolumu, Kaynaklar Yerleskesi, 35160, Buca, Izmir.

Recieved: 29 November 2006; **Revised:** 06 March 2007; **Accepted:** 23 November 2007

1. INTRODUCTION

Log-linear models describe the association patterns among categorical variables by modeling expected cell counts in the contingency tables without distinguishing between independent and dependent variables. These models are linear in their logarithms and as stated by Christensen (1990), they have great advantages of flexibility and interpretability. The use of log-linear modeling is more appropriate in situations where applied researchers are interested in the various pairwise and higher order associations among set of independent variables. Lee and Viele (2001) used log-linear models for modeling data from train waybills particularly focusing on revealing the relationships between cargo volume and origin, destination and commodity type. Log-linear models are also used in survival analysis as accelareted failure time models (Christensen, 2000). More detailed information on log-linear models can be found in Agresti (1996), Christensen (1990), and Bishop et al. (1975).

Basu and Basu (1998) have proposed an empty cell penalty for the power-divergence measures presented by Cressie and Read (1984). This penalty gives the chance to change the weight put on empty cells in contingency tables by the family of power-divergence measures. Cressie and Pardo (2000) introduced the family of ordinary power-divergence test statistics to test nested log-linear models. In this paper, we define the family of penalized power-divergence test statistics. We compare the size and power properties of these two families of test statistics for testing log-linear models in three-way contingency tables based on both asymptotic and finite sample (using a designed simulation study) results. Our purpose is to show that the family of penalized power-divergence test statistics performs better than the family of ordinary power-divergence test statistics for small samples.

The rest of the paper is laid out as follows: After giving the descriptions of the log-linear models and power-divergence measures in next section, we will review the distributional properties of ordinary and penalized power-divergence test statistics in section three, and we will present our simulation results in section four.

2. LOG-LINEAR MODELS AND POWER-DIVERGENCE MEASURES

Let X, Y and Z be three categorical response variables having I, J and K levels, respectively. When we classify on these three variables, we have IJK possible combinations of classifications. $p_{ijk}(\theta) = \mathrm{P}(X = i, Y = j, Z = k)$, $i = 1,\ldots, I$; $j = 1,\ldots, J$; $k = 1,\ldots, K$, is a probability distribution of the responses (X, Y, Z) of a subject randomly chosen from a population. This distribution is assumed to be unknown, but belonging to a known family,

$$P = \{p(\theta) = (p_{111}(\theta), \ldots, p_{IJK}(\theta))^T: \theta \in \Theta\}$$

of distributions with components taking values on $\chi = \{(i, j, k) \in I \times J \times K\}$ with parameter space $\Theta \subseteq R^t$ $(t < IJK - 1)$. Hence, the true value, θ_0, of parameters vector $\theta = (\theta_1, \ldots, \theta_t)^T \in \Theta$ is assumed to be unknown.

Consider a sample $S_1, \ldots, S_n$ of size n on (X,Y, Z). The statistic $(N_{111}, \ldots, N_{IJK})^T$ is sufficient for statistical model under consideration where N_{ijk} denotes the observed frequency in ijk-th cell for $(i, j, k) \in I \times J \times K$. Poisson, multinomial and product-multinomial samplings are among the possible sampling distributions for $(N_{111}, \ldots, N_{IJK})^T$. For Poisson sampling, the total $\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} N_{ijk}$ is random. If this total is fixed, the vector $(N_{111}, \ldots, N_{IJK})^T$ becomes multinomial. If these observations come from product-multinomial sampling, certain of the margins are fixed. We will assume $(N_{111}, \ldots, N_{IJK})^T$ has multinomial distribution with probabilities belonging to a general class of log-linear models, that is ,

$$P(N_{111} = n_{111},\ldots, N_{IJK} = n_{IJK}) = \frac{n!}{n_{111}!\ldots n_{IJK}!} p_{111}(\boldsymbol{\theta})^{n_{111}} \ldots p_{IJK}(\boldsymbol{\theta})^{n_{IJK}} \tag{1}$$

for $n_{ijk} \geq 0$ such that $\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} n_{ijk} = n$ and

$$p_{ijk}(\boldsymbol{\theta}) = \frac{\exp(\boldsymbol{w}_{ijk}^T \boldsymbol{\theta})}{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} \exp(\boldsymbol{w}_{ijk}^T \boldsymbol{\theta})}$$

for $i = 1,\ldots, I$; $j = 1,\ldots, J$; $k = 1,\ldots,K$. (2)

$1 \times t$ vector $\boldsymbol{w}_{ijk}^T = \left(w_{1(ijk)},\ldots, w_{t(ijk)}\right)$ in equation (2) forms the $IJK \times t$ log-linear model matrix of explanatory variables $W = (w_{111}, \ldots, w_{IJK})^T$ which is assumed to have full column rank $t < IJK - 1$. Elements of this matrix are determined according to linear constraints on the parameter vector θ. Columns of W are linearly independent of the $IJK \times 1$ column vector $(1, \ldots, 1)^T$. Matrix form of the log-linear model given in (2) is

$$\log p(\theta^*) = \mathrm{D}\theta^* \tag{3}$$

with $IJK \times (t + 1)$ "design matrix" $\mathrm{D} = (1_{IJK \times 1}, W_{IJK \times t})$. $\theta^* = (u, \theta_1, \ldots, \theta_t)^T$ is a $(t + 1) \times 1$ column vector with

$$\mathrm{u} = -\log\left(\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} \exp(\boldsymbol{w}_{ijk}^{\mathrm{T}} \boldsymbol{\theta})\right) \tag{4}$$

which is overall mean effect parameter and calculated as the normalizing constant to insure $\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} p_{ijk}(\boldsymbol{\theta}) = 1$. Linear constraints on the parameters are as follows;

$$\sum_{i=1}^{I}\theta_{1(i)}=\sum_{j=1}^{J}\theta_{2(j)}=0 \tag{5}$$

$$\sum_{i=1}^{I}\theta_{12(ij)}=\sum_{j=1}^{J}\theta_{12(ij)}=\sum_{i=1}^{I}\theta_{13(ik)}=\sum_{k=1}^{K}\theta_{13(ik)}=\sum_{j=1}^{J}\theta_{23(jk)}=\sum_{k=1}^{K}\theta_{23(jk)}=0$$

Since the parameter values are unknown, they need to be estimated. When it comes to estimation, maximum likelihood estimator (MLE) is the mostly used estimator. MLE is known to be efficient in regular models but it is also known to be nonrobust (Menéndez et al 2001). The efficiency as well as the nonrobustness are resulting from specific properties of the logarithmic function used in the definition of the function (Pardo et al. 2001). By replacing the logarithmic function by other functions with appropriate properties, a new class of estimators such as minimum ϕ-divergence estimators can be obtained.

ϕ-divergence measure is a density based divergence and has been defined by Csiszár (1967) as follows;

$$D_{\phi}(\boldsymbol{p},\boldsymbol{q})=\sum_{j=1}^{m}q_j\phi\left(\frac{p_j}{q_j}\right)\ ;\phi\in\Phi \tag{6}$$

where $p=(p_1,\ldots,p_m)^T$ and $q=(q_1,\ldots,q_m)^T$ are discrete probability distributions. Φ^* is the class of all convex functions $\phi(x)$, $x>0$, such that at $x=1$, $\phi(1)=0$ and $\phi''(1)>0$, and at $x=0$, $0\phi(0/0)=0$ and $0\phi(p/0)=\lim_{u\to\infty}\phi(u)/u$. For every $\phi\in\Phi^*$ that is differentiable at $x=1$, the function $\psi(x)\equiv\phi(x)-\phi'(1)(x-1)$ also belongs to Φ^*. Then $\psi(x)$ and $\phi(x)$ are equivalent functions with ψ having additional property of $\psi'(1)=0$. See Lindsay (1994) and Morales et al. (1995) for the asymptotic properties of the minimum ϕ-divergence estimators,

$$\hat{\theta}_{\phi}\equiv\arg\min_{\theta\in\Theta}\mathrm{D}_{\phi}(\boldsymbol{p},\boldsymbol{q}) \tag{7}$$

An important family of ϕ-divergences is the power-divergence family defined by Cressie and Read (1984) with

$$\phi_{(\lambda)}(x)=\begin{cases}\dfrac{1}{\lambda(\lambda+1)}(x^{\lambda+1}-x) & \text{for } \lambda\neq 0,\lambda\neq -1\\ x\log x-(x-1) & \text{for } \lambda\to 0\\ -\log x+x-1 & \text{for } \lambda\to -1\end{cases} \tag{8}$$

We use $\lim_{\lambda\to 0}\dfrac{1}{\lambda(\lambda+1)}(x^{\lambda+1}-x)$ and $\lim_{\lambda\to -1}\dfrac{1}{\lambda(\lambda+1)}(x^{\lambda+1}-x)$, since it is not defined at these values. For three-way contingency tables, power-divergence measure between $\hat{\boldsymbol{p}}$ and $p(\theta)$ can be defined by the formula;

$$\mathrm{I}_{\lambda}(\hat{\boldsymbol{p}},\boldsymbol{p}(\boldsymbol{\theta}))=\frac{1}{\lambda(\lambda+1)}\sum_{i=1}^{I}\sum_{j=I}^{J}\sum_{k=1}^{K}\hat{p}_{ijk}\left[\left(\frac{\hat{p}_{ijk}}{p_{ijk}(\boldsymbol{\theta})}\right)^{\lambda}-1\right] \tag{9}$$

where $\hat{\boldsymbol{p}}=(\hat{p}_{111},\ldots,\hat{p}_{IJK})^T$ with $\hat{p}_{ijk}=\dfrac{n_{ijk}}{n}$ and $\boldsymbol{p}(\boldsymbol{\theta})=(p_{111}(\boldsymbol{\theta}),\ldots,p_{IJK}(\boldsymbol{\theta}))^T$ are observed relative frequencies (the estimator of saturated model) and true probability vectors, respectively. Minimum power-divergence estimator as the value minimizing $I_{\lambda}(\hat{\boldsymbol{p}},\boldsymbol{p}(\boldsymbol{\theta}))$ with respect to θ is defined by

$$\hat{\boldsymbol{\theta}}_{\lambda}^{I}\equiv\arg\min_{\theta\in\Theta}I_{\lambda}(\hat{\boldsymbol{p}},\boldsymbol{p}(\boldsymbol{\theta})) \tag{10}$$

For $\lambda\to 0$, (10) is equal to minimum Kullback-Leibler divergence estimator which is equivalent to the MLE of θ. There are other estimators less known than MLE and included in the power-divergence family. For example, (10) is equivalent to minimum Pearson's chi-squared estimator for $\lambda=1$, minimum Cressie-Read divergence estimator for $\lambda=2/3$, minimum Matusita distance estimator for $\lambda=-1/2$, and minimum discrimination information estimator for $\lambda\to -1$.

Harris and Basu (1997) have considered power-divergence family in the following form;

$$\mathrm{I}_{\lambda}(\hat{\boldsymbol{p}},\boldsymbol{p}(\boldsymbol{\theta}))=\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left[\frac{1}{\lambda(\lambda+1)}\hat{p}_{ijk}\left[\left(\frac{\hat{p}_{ijk}}{p_{ijk}(\boldsymbol{\theta})}\right)^{\lambda}-1\right]+\frac{p_{ijk}(\boldsymbol{\theta})-\hat{p}_{ijk}}{\lambda+1}\right] \tag{11}$$

The second component in (11) does not contribute anything to the measure and comes from using $\psi(x)\equiv\phi(x)-\phi'(1)(x-1)$ instead of $\phi(x)$. Since $\phi(x)$ and $\psi(x)$ are equivalent, equations given with (9) and (11) define the same divergence.

When it comes to small samples, MLE can perform better in terms of efficiency than many of the more robust estimators. Basu and Basu (1998) claim that this unfortunate trade off between robustness and small sample efficiency appears to be partly due to the disproportionately large weight that these divergences put on empty cells. To deal with that Basu and Basu (1998) have proposed an empty cell penalty for the minimum power-divergence estimators in multinomial models and separated equation (11) as follows;

$$I_{\lambda}(\hat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\theta})) = \sum_{i=1}^{I}\sum_{\substack{j=1 \\ \hat{p}_{ijk}\neq 0}}^{J}\sum_{k=1}^{K}\left[\frac{1}{\lambda(\lambda+1)}\hat{p}_{ijk}\left[\left(\frac{\hat{p}_{ijk}}{p_{ijk}(\theta)}\right)^{\lambda}-1\right]+\frac{p_{ijk}(\theta)-\hat{p}_{ijk}}{\lambda+1}\right]+\sum_{i=1}^{I}\sum_{\substack{j=1 \\ \hat{p}_{ijk}=0}}^{J}\sum_{k=1}^{K}\frac{p_{ijk}(\theta)}{(\lambda+1)} \quad (12)$$

Equation (12) is comprised of two parts; the part for the nonempty cells and the part for the empty cells. The second component can become very large for the values of λ closer to -1. By applying penalty on empty cells in equation (12), we get the penalized power-divergence family as given by (13);

$$P_{\lambda}^{w}(\hat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\theta})) = \sum_{i=1}^{I}\sum_{\substack{j=1 \\ \hat{p}_{ijk}\neq 0}}^{J}\sum_{k=1}^{K}\left[\frac{1}{\lambda(\lambda+1)}\hat{p}_{ijk}\left[\left(\frac{\hat{p}_{ijk}}{p_{ijk}(\boldsymbol{\theta})}\right)^{\lambda}-1\right]+\frac{p_{ijk}(\boldsymbol{\theta})-\hat{p}_{ijk}}{\lambda+1}\right]+w\sum_{i=1}^{I}\sum_{\substack{j=1 \\ \hat{p}_{ijk}=0}}^{J}\sum_{k=1}^{K}p_{ijk}(\boldsymbol{\theta}) \quad (13)$$

where w is the penalty put on empty cells. By increasing or decreasing the penalty, power-divergence measure can be made more or less sensitive to empty cells. Penalized minimum power-divergence estimator can be defined by

$$\hat{\boldsymbol{\theta}}_{\lambda}^{P^{w}} \equiv \arg\min_{\theta\in\Theta} P_{\lambda}^{w}(\hat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\theta})) \quad (14)$$

As it is seen, $I_0(\hat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\theta})) = P_0^{1}(\hat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\theta}))$ and $I_1(\hat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\theta})) = P_1^{0.5}(\hat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\theta}))$, i.e. when $w = 1$, the penalized power-divergence puts the same weight on empty cells as Kullback-Leibler divergence does whereas when $w = 0.5$, it puts the same weight as Pearson's chi-squared does. For the values of $\lambda \leq -1$, $I_{\lambda}(\hat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\theta}))$ can not be defined even if there is only one empty cell. But, this is not the case for $P_{\lambda}^{w}(\hat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\theta}))$, since empty cells component does not depend on λ. Note that reweighting empty cells will not alter the asymptotic properties of corresponding estimator (Pardo and Pardo, 2003).

To test nested log-linear models $H_{null}: H_{l+1}$ against $H_{alt}: H_l$, with H_{null} nested in H_{alt}, Cressie and Pardo (2000, 2002) have suggested the following families of test statistics

$$T(O)_{\phi_1,\phi_2}^{(l)} = \frac{2n}{\phi_1''(1)} I_{\phi_1}(\boldsymbol{p}(\hat{\boldsymbol{\theta}}_{\phi_2}^{I(l)}), \boldsymbol{p}(\hat{\boldsymbol{\theta}}_{\phi_2}^{I(l+1)})) \quad (15)$$

$$S(O)_{\phi_1,\phi_2}^{(l)} = \frac{2n}{\phi_1''(1)}\{I_{\phi_1}(\hat{\boldsymbol{p}}, \boldsymbol{p}(\hat{\boldsymbol{\theta}}_{\phi_2}^{I(l+1)})) - I_{\phi_1}(\hat{\boldsymbol{p}}, \boldsymbol{p}(\hat{\boldsymbol{\theta}}_{\phi_2}^{I(l)}))\} \quad (16)$$

where ϕ_1 and ϕ_2 are convex functions, and $\hat{\boldsymbol{\theta}}_{\phi_2}^{I(l+1)}$ and $\hat{\boldsymbol{\theta}}_{\phi_2}^{I(l)}$ are defined by (10) under the models of H_{l+1} and H_l, respectively. By using penalized power-divergence measure and penalized minimum power-divergence estimators instead of ordinary ones in (16), we can define the following family of penalized power-divergence test statistics

$$S(P^{w})_{\phi_1,\phi_2}^{(l)} = \frac{2n}{\phi_1''(1)}\{P_{\phi_1}^{w}(\hat{\boldsymbol{p}}, \boldsymbol{p}(\hat{\boldsymbol{\theta}}_{\phi_2}^{P^{w}(l+1)})) - P_{\phi_1}^{w}(\hat{\boldsymbol{p}}, \boldsymbol{p}(\hat{\boldsymbol{\theta}}_{\phi_2}^{P^{w}(l)}))\} \quad (17)$$

where $\hat{\boldsymbol{\theta}}_{\phi_2}^{P^{w}(l+1)}$ and $\hat{\boldsymbol{\theta}}_{\phi_2}^{P^{w}(l)}$ are defined by (14) under the models of H_{l+1} and H_l, respectively.

3. DISTRIBUTIONAL PROPERTIES OF POWER DIVERGENCE TEST STATISTICS UNDER NULL AND CONTIGUOUS ALTERNATIVE HYPOTHESES

Our work has been motivated by the work of Cressie et al. (2003) which involves a nested sequence of hypotheses*,

$$H_l: \boldsymbol{p} = \boldsymbol{p}(\boldsymbol{\theta}); \quad \boldsymbol{\theta} \in \Theta_l; \quad l = 1, ..., m, \quad m \leq t < IJK\text{-}1, \quad (18)$$

where $\boldsymbol{p}(\boldsymbol{\theta}) = (p_1(\theta), \ldots, p_{IJK}(\theta))^T$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_t)^T \in \boldsymbol{\Theta_l}$. $\boldsymbol{\Theta_l}$ is the parameter space for the H_l such that $\boldsymbol{\Theta_m} \subset \boldsymbol{\Theta_{m-1}} \subset \ldots \subset \boldsymbol{\Theta_1} \equiv \boldsymbol{\Theta} \subseteq R^t$; $t < IJK$-1 and $\dim(\boldsymbol{\Theta_l}) = d_l$ with $d_m < d_{m-1} < \ldots < d_1 = t$. Cressie et al. (2003)'s strategy is to test successively,

$$H_{null}: H_{l+1} \text{ against } H_{alt}: H_l; \quad l = 1, ..., m\text{-}1, \quad (19)$$

until the first l for which H_{l+1} is rejected as a null hypothesis. Cressie et al. (2003) have used the family of test statistics given by (15) to test these nested hypotheses. When $T(O)_{\phi_1,\phi_2}^{(l)} > c$, we reject H_{null} where c is specified so that the size of the test is α;

$$P(T(O)_{\phi_1,\phi_2}^{(l)} > c \mid H_{l+1}) = \alpha\,; \ \alpha \in (0, 1) \quad (20)$$

Cressie and Pardo (2000, 2002) have shown that under multinomial sampling with probabilities belonging to a log-linear model and $H_{null}: H_{l+1}$, the test statistic $T(O)_{\phi_1,\phi_2}^{(l)}$ converges in distribution to a chi-squared distribution with $d_l - d_{l+1}$ degrees of freedom; $l = 1, ..., m$-1. Hence,

$$c = \chi_{d_l - d_{l+1}}^{2}(1-\alpha), \quad (21)$$

where $P(\chi_f^2 \leq \chi_f^2(p)) = p$. Cressie and Pardo (2000) have established that $T(O)_{\phi_1,\phi_2}^{(l)}$ and $S(O)_{\phi_1,\phi_2}^{(l)}$ are asymptotically equivalent under null hypothesis. In other words, under $H_{null}: H_{l+1}$, $S(O)_{\phi_1,\phi_2}^{(l)}$ has chi-squared distribution with $d_l - d_{l+1}$ degrees of freedom;

* Dimensions are adapted to our case.

$l = 1, ..., m\text{-}1$. The family of $S(O)^{(l)}_{\phi_1,\phi_2}$ yields the likelihood ratio test statistic when $\phi_1 \equiv \phi_2 \equiv \phi_{(0)}$. Moreover, when $\phi_1 \equiv \phi_{(1)}$ and $\phi_2 \equiv \phi_{(0)}$, the family yields a statistics based on the difference of Pearson chi-squared statistics with maximum likelihood estimation used to obtain expected frequencies (e.g., Agresti (1996, p. 197)). However, the nonnegativity of $S(O)^{(l)}_{\phi_1,\phi_2}$ does not hold when $\phi_1 \neq \phi_2$ (Cressie and Pardo, 2000). Thus, for the case considered by Agresti, the difference of Pearson chi-squared statistics is not necessarily nonnegative.

On the contrary to H_{null}, theoretical results for $T(O)^{(l)}_{\phi_1,\phi_2}$ and $S(O)^{(l)}_{\phi_1,\phi_2}$ under alternative hypotheses are not easy to obtain except for an appropriately specified sequence of contiguous alternatives (also known as Pitman alternatives). Regarding these alternatives, Haberman (1974) was the first to study them. He has proved that the asymptotic distribution of the likelihood ratio and Pearson's test statistics under the sequence of contiguous alternatives is non-centrally chi-squared with d_l - d_{l+1} degrees of freedom. Oler (1985) and Fenech and Westfall (1988) have presented studies regarding these contiguous alternative hypotheses in multinomial populations with probabilities belonging to log-linear models. Later, Cressie et al. (2003) have defined the sequence of contiguous alternative hypotheses;

$$H_{l+1,n} : \boldsymbol{p} = \boldsymbol{p}_n(\boldsymbol{\theta}) = \boldsymbol{p}(\boldsymbol{\theta}) + \boldsymbol{s}/\sqrt{n}\,,\ \boldsymbol{\theta} \in \boldsymbol{\Theta}_{l+1},\ n \geq n_0 > 0, \tag{22}$$

which is contiguous to the null hypothesis H_{l+1} where $\boldsymbol{s} \equiv (s_{111},...,s_{IJK})^T$ is a fixed IJK x 1 vector such that $\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} s_{ijk} = 0$ and n is the total count parameter of the multinomial distribution. For multinomial sampling with probabilities belonging to a log-linear model, they have proved that under the sequence of contiguous alternative hypotheses given by (22), $T(O)^{(l)}_{\phi_1,\phi_2}$ has non-central chi-squared distribution with d_l - d_{l+1} degrees of freedom; $l = 1, ..., m\text{-}1$. Non-centrality parameter for this distribution is,

$$\mu = \boldsymbol{s}^T \mathrm{diag}(\boldsymbol{p}(\boldsymbol{\theta})^{-1/2})(\boldsymbol{A}_{(l)} - \boldsymbol{A}_{(l+1)})\mathrm{diag}(\boldsymbol{p}(\boldsymbol{\theta})^{-1/2})\boldsymbol{s}\,, \tag{23}$$

where

$\boldsymbol{A}_{(i)} = \mathrm{diag}(\boldsymbol{p}(\boldsymbol{\theta})^{-1/2})\boldsymbol{\Sigma}_{\boldsymbol{p}(\boldsymbol{\theta})}\boldsymbol{W}_{(i)}(\boldsymbol{W}_{(i)}^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{p}(\boldsymbol{\theta})}\boldsymbol{W}_{(i)})^{-1}\boldsymbol{W}_{(i)}^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{p}(\boldsymbol{\theta})}\mathrm{diag}(\boldsymbol{p}(\boldsymbol{\theta})^{-1/2})$; $i = l, l+1$.
$\boldsymbol{W}_{(i)} = (w_{111}, \ldots, w_{IJK})^T$ is IJK x t log-linear model matrix of explanatory variables under the model H_i, $i = l$, $l+1$, and $\boldsymbol{\Sigma}_{\boldsymbol{P}(\boldsymbol{\theta})} = \mathrm{diag}\ \boldsymbol{p}(\boldsymbol{\theta}) - \boldsymbol{p}(\boldsymbol{\theta})\boldsymbol{p}(\boldsymbol{\theta})^{\mathrm{T}}$. So, as $n \to \infty$,

$$\pi_n \to P(\chi^2_{d_l - d_{l+1},\mu} > c)\,, \tag{24}$$

where $\pi_n \equiv P(T(O)^{(l)}_{\phi_1,\phi_2} > c \,|\, H_{l+1,n})$

Based on the results established by Cressie and Pardo (2000), Haberman (1974), Oler (1985) and Fenech and Westfall (1988), there is no doubt about that $T(O)^{(l)}_{\phi_1,\phi_2}$ and $S(O)^{(l)}_{\phi_1,\phi_2}$ are asymptotically equivalent under the sequence of contiguous alternative hypotheses given by (22). In this paper, we compare $S(O)^{(l)}_{\phi_1,\phi_2}$ and $S(P^w)^{(l)}_{\phi_1,\phi_2}$ in terms of size and power properties for testing nested log-linear models in three-way contingency tables under the assumption of multinomial sampling. As mentioned by Basu and Basu (1998), the family of penalized power-divergence test statistics has the same asymptotic distribution with the family of ordinary power-divergence measures since they differ only in empty cells. Hence, we only focus on small sample properties of these test statistics. Our comparison is based on both asymptotic and finite sample (using a designed simulation study) results. We describe these results in following section.

4. SIMULATION STUDY

In simulation study, we consider the case of 2 x 2 x 2 contingency tables, so we have eight possible combinations, i.e. $IJK = 8$. Among the nested models, we consider the following two hypotheses for brevity.

H_1: $p_{ijk}(\boldsymbol{\theta}) = \exp\,[\mathrm{u} + \theta_{1(i)} + \theta_{2\,(j)} + \theta_{3\,(k)} + \theta_{12(ij)}]$; $i, j, k = 1, 2$

H_2: $p_{ijk}(\boldsymbol{\theta}) = \exp\,[u + \theta_{1(i)} + \theta_{2\,(j)} + \theta_{3\,(k)}]$; $i, j, k = 1, 2$

Following Oler (1985), we chose moderate main effects $\exp\,[\theta_{1(1)}] = \exp\,[\theta_{2(1)}] = \exp\,[\theta_{2(1)}] = 5/6$ and big interaction effect $\exp\,[\theta_{12(11)}] = 3/4$ with the linear constraints given by (5). Overall mean effect parameter u is calculated as given by (4).

Let $\boldsymbol{p_0} \in H_{null}$ and $\boldsymbol{p}_{1,\,n} \in H_{alt}$. Here, $\boldsymbol{p}_{1,\,n}$ is subscripted with n, since its entries may depend on n. By this simulation study, we want to obtain the following exact probabilities:

$$\alpha_n^{S(O)^{(1)}} \equiv P(S(O)^{(1)}_{\phi_1,\phi_2} > c \,|\, \boldsymbol{p_0}),\ \pi_n^{S(O)^{(1)}} \equiv P(S(O)^{(1)}_{\phi_1,\phi_2} > c \,|\, \boldsymbol{p_{1,n}}) \tag{25}$$

$$\alpha_n^{S(P^w)^{(1)}} \equiv P(S(P^w)^{(1)}_{\phi_1,\phi_2} > c \,|\, \boldsymbol{p_0}),\ \pi_n^{S(P^w)^{(1)}} \equiv P(S(P^w)^{(1)}_{\phi_1,\phi_2} > c \,|\, \boldsymbol{p_{1,n}}) \tag{26}$$

Penalization values (w) are chosen as 0.5 and 1. As mentioned above, the nonnegativity of $S(O)^{(l)}_{\phi_1,\phi_2}$ does not hold when $\phi_1 \neq \phi_2$. Thus, we only consider the combinations with $\phi_1 \equiv \phi_2 \equiv \phi_{(\lambda)}$ with λ = -0.9, -0.8,

-0.7, -0.6, -0.5, 0, 2/3, 1. The other important reason to chose $\phi_1 \equiv \phi_2 \equiv \phi_{(\lambda)}$ other than getting nonnegative test statistics is that it seems quite reasonable to match ϕ_1 and ϕ_2 as mentioned by Read and Cressie (1988, p.32).

Our design for the simulation study is similar to the one carried out by Cressie et. al (2003). For a given $\boldsymbol{p_0}$ which is defined by (27), the various choices of n and $\boldsymbol{p_{1,n}}$ represent the design.

$$\boldsymbol{p_0} \in H_2: p_{ijk}(\boldsymbol{\theta}) = \exp\left[u+\theta_{1(i)}+\theta_{2(j)}+\theta_{3(k)}\right];\ i,j,k=1,2. \quad (27)$$

Sample sizes considered for this simulation study are n = 15, 25, 35. To show that $S(O)^{(l)}_{\phi_1,\phi_2}$ and $S(P^w)^{(l)}_{\phi_1,\phi_2}$ are asymptotically chi-squared with d_l - d_{l+1} degrees of freedom under H_2, we have also calculated $\alpha_n^{S(O)^{(1)}}$ and $\alpha_n^{S(P^w)^{(1)}}$ for n = 45, 75. $\boldsymbol{p_{1,n}}$ is chosen as both a contiguous and a fixed alternative which are defined by (28) and (29), respectively.

$$\boldsymbol{p}^*_{\mathbf{1},n} = \boldsymbol{p_0} + \sqrt{\frac{25}{n}}(\boldsymbol{p_1} - \boldsymbol{p_0}), \quad (28)$$

$$\boldsymbol{p_1} \in H_1: p_{ijk}(\boldsymbol{\theta}) = \exp\left[u+\theta_{1(i)}+\theta_{2(j)}+\theta_{3(k)}+\theta_{12(ij)}\right];\ i,j,k=1,2 \quad (29)$$

As noticed, $\boldsymbol{p}_1$ is also used in the definition of contiguous alternative. As n increases, $\boldsymbol{p}^*_{\mathbf{1},n}$ converges to $\boldsymbol{p_0}$ at the rate of $n^{-1/2}$; that is $\{\boldsymbol{p}^*_{\mathbf{1},n}\}$ is a sequence of contiguous alternatives. For n < 25, the contiguous alternatives are further from H_{null} than are fixed alternatives and $\boldsymbol{p}^*_{\mathbf{1,25}} = \boldsymbol{p_1}$.

Exact probability estimations given with (25) and (26) are obtained using 1000 simulations from the multinomial distributions with (n, $\boldsymbol{p_0}$), (n, $\boldsymbol{p}_1$) and (n, $\boldsymbol{p}^*_{\mathbf{1},n}$). All calculations have been done using Mathematica 5.0 except the asymptotic power values which are obtained by Matlab 5.3. As Cressie et. al. (2003), we use two basic criteria for a good performance to compare test statistics: 1) Good exact power and size, 2) Good agreement of exact and asymptotic probabilities. For the first criteria, we consider H_{null} : $\boldsymbol{p_0} \in H_2$ against H_{alt}: $\boldsymbol{p_1} \in H_1$. For the second, we use H_{null} : $\boldsymbol{p_0} \in H_2$ against H_{alt} : $\boldsymbol{p}^*_{\mathbf{1},n}$.

To compare the test statistics, we calculate $g_1(\lambda,\lambda) \equiv \left|AP_{C,n}(\lambda,\lambda) - SEP_{C,n}(\lambda,\lambda)\right|$ and $g_2(\lambda,\lambda) \equiv (SEP_{F,n}(\lambda,\lambda) - STS_{F,n}(\lambda,\lambda))^{-1}$ $AP_{C,n}(\lambda,\lambda)$ is the asymptotic power under contiguous alternative, $SEP_{C,n}(\lambda,\lambda)$ is the simulated exact power under contiguous alternative, $SEP_{F,n}(\lambda,\lambda)$ is the simulated exact power under fixed alternative and $STS_{F,n}(\lambda,\lambda)$ is the simulated test size of the test statistic $S(i)^{(1)}_{\phi(\lambda),\phi(\lambda)}$ for i = O, P^w . According to Fenech and Westfall (1988), and Cressie et al. (2003), approximation to the asymptotic power of $S(i)^{(1)}_{\phi(\lambda),\phi(\lambda)}$ can be obtained by defining $\boldsymbol{p}_n(\boldsymbol{\theta}^{(1)}) \equiv \boldsymbol{p}(\boldsymbol{\theta}^{(2)}) + \frac{1}{\sqrt{n}}\boldsymbol{s}$ where $\boldsymbol{s} = \sqrt{n}(\boldsymbol{p}(\boldsymbol{\theta}^{(1)}) - \boldsymbol{p}(\boldsymbol{\theta}^{(2)}))$ and then substituting into the definition of μ, noncentrality parameter, and then finally μ into the right side of (24). $\boldsymbol{p}(\boldsymbol{\theta}^{(1)})$ and $\boldsymbol{p}(\boldsymbol{\theta}^{(2)})$ are true probability vectors under the model H_1 and H_2, respectively. We consider the statistic $S(i)^{(1)}_{\phi(\lambda),\phi(\lambda)}$ to be better than the other statistic $S(i)^{(1)}_{\phi(\lambda^*),\phi(\lambda^*)}$ iff

$$g_1(\lambda,\lambda) < g_1(\lambda^*,\lambda^*) \text{ and } g_2(\lambda,\lambda) < g_2(\lambda^*,\lambda^*) \quad (30)$$

For these comparisons, only the statistics that satisfy the following inequality are considered.

$$\left|\operatorname{logit}(1-\alpha_n^{S(i)^{(1)}}) - \operatorname{logit}(1-\alpha)\right| \le e, \quad (31)$$

where logit(p) ≡ ln(p / (1- p)), i = (O) ordinary, (P^w) penalized. This inequality has been proposed by Dale (1986) and measures the closeness of nominal size and exact size obtained from the simulation. The two probabilities are considered fairly close if they satisfy (31) with e = 0.7. For α = 0.05, which is nominal size of the power-divergence test statistics in this study, e = 0.7 corresponds to

$$\alpha_n^{S(i)^{(1)}} \in [0.0254, 0.0959] \quad (32)$$

The simulation results are given in Tables 1-9 at the end of the manuscript under the heading of Tables. Table 1-3, Table 4-6 and Table 7-9 show the exact size ($\alpha_n^{T(i)^{(1)}}$), (exact power-asymptotic power) and (exact power - exact size) values of $S(i)^{(1)}_{\phi(\lambda_1),\phi(\lambda_2)}$, respectivly.Since $I_0(\hat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\theta})) = P_0^1(\hat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\theta}))$ and $I_1(\hat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\theta})) = P_1^{0.5}(\hat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\theta}))$, we get $S(O)^{(1)}_{\phi(0),\phi(0)} = S(P^1)^{(1)}_{\phi(0),\phi(0)}$ and $S(O)^{(1)}_{\phi(1),\phi(1)} = S(P^{0.5})^{(1)}_{\phi(1),\phi(1)}$. Thus, the values in the tables corresponding to these estimators are equal.

For n = 15, the statistics that satisfy (32) correspond to the $S(O)^{(1)}_{\phi(\lambda),\phi(\lambda)}$ for λ = 1, 2/3, $S(P^1)^{(1)}_{\phi(\lambda),\phi(\lambda)}$ for λ = -0.5, -0.6, -0.7, -0.8, -0.9 and $S(P^{0.5})^{(1)}_{\phi(\lambda),\phi(\lambda)}$

for all of the values of λ. For negative values of λ, especially large negative values, exact sizes of $S(O)^{(1)}_{\phi(\lambda),\phi(\lambda)}$ are far behind satisfying (32). However, penalization improves the exact sizes of the test statistics especially for large negative values of λ. Increasing or decreasing the penalty value affects the sensitivity of the test statistics to empty cells. This causes the decreament of the exact size performances of some statistics as expected. For instance, as seen in Table 2 exact sizes of $S(P^1)^{(1)}_{\phi(1),\phi(1)}$ and $S(P^1)^{(1)}_{\phi(2/3),\phi(2/3)}$ have increased since the weight put on empty cells in these test statistics increase with penalization value of w = 1. The improvement in the performance of the test statistics $S(O)^{(1)}_{\phi(\lambda),\phi(\lambda)}$ for large negative values of λ due to penalty for n = 15 can clearly be seen at the histograms of the exact null distributions of $S(O)^{(1)}_{\phi(\lambda),\phi(\lambda)}$ and $S(P^{0.5})^{(1)}_{\phi(\lambda),\phi(\lambda)}$. Since the conclusions are same for $S(O)^{(1)}_{\phi(-0.9),\phi(-0.9)}$ and $S(P^{0.5})^{(1)}_{\phi(-0.9),\phi(-0.9)}$, we only present in here the histograms of the $S(O)^{(1)}_{\phi(-0.8),\phi(-0.8)}$ and $S(P^{0.5})^{(1)}_{\phi(-0.8),\phi(-0.8)}$ for brevity. Our interest is in the right hand tail area of the histograms and how well χ^2_1 density approximates it.

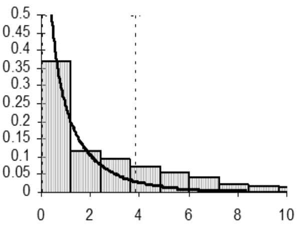


Fig. 1a.

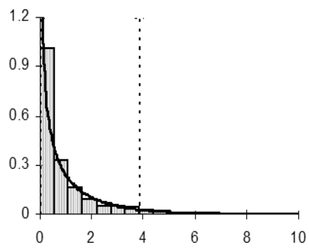


Fig. 1b.

Fig. 1a. Histogram of the null distribution of $S(O)^{(1)}_{\phi(-0.8),\phi(-0.8)}$ for n = 15 with density function of χ^2_1 distribution superimposed

Fig. 1b. Histogram of the null distribution of $S(P^{0.5})^{(1)}_{\phi(-0.8),\phi(-0.8)}$ for n = 15 with density function of χ^2_1 distribution superimposed.

Fig. 1a shows the histogram of $S(O)^{(1)}_{\phi(-0.8),\phi(-0.8)}$, where the poor approximation to the χ^2_1 density is evident from the relatively heavy tail of the statistic. On the other hand, the right tail of the histogram $S(P^{0.5})^{(1)}_{\phi(-0.8),\phi(-0.8)}$ is very well approximated by the χ^2_1 density. The vertical dashed lines on the histograms correspond to the 5% critical point of χ^2_1 (i.e. $c = \chi^2_1(0.95) = 3.84$ as defined by (21)). The usual likelihood ratio ($S(O)^{(1)}_{\phi(0),\phi(0)}$) does not satisfy (32), either. However, as seen from Table 3, penalization improves this statistic, too. For n = 25, all of the penalized power-divergence test statistics and $S(O)^{(1)}_{\phi(\lambda),\phi(\lambda)}$ for λ = 0, 1, 2/3 satisfy (32). This is also the same for n = 35 with two more statistics ($S(O)^{(1)}_{\phi(-0.5),\phi(-0.5)}$ and $S(O)^{(1)}_{\phi(-0.6),\phi(-0.6)}$) satisfying (32). Since all of the ordinary power-divergence test statistics are all asymptotically equivalent, ordinary power-divergence test statistics with large negative values of λ get exact sizes very close to nominal size of 0.05 as n gets larger even though they behave entirely opposite for small samples. To illustrate this behavior, we plot the departures of the exact sizes of $S(O)^{(1)}_{\phi(\lambda),\phi(\lambda)}$ from the nominal size of $\alpha = 0.05$ against the sample sizes of 15, 25, 35, 45, and 75. This plot is given with Fig. 2. In general, as n gets larger, all of the power-divergence test statistics satisfy Eq.(32) by getting closer to the nominal size $\alpha = 0.05$.

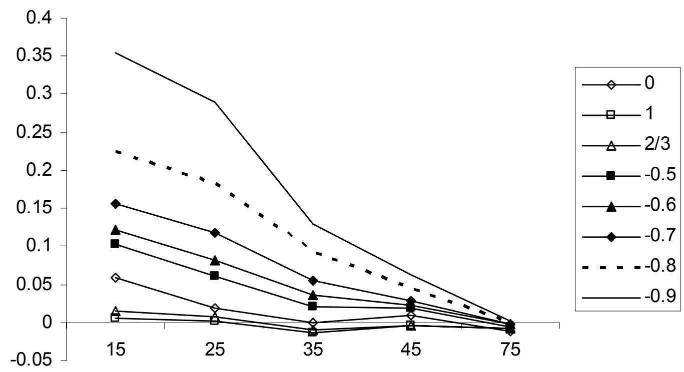


Fig.2 (Exact size - Nominal size of 0.05) for $S(O)^{(1)}_{\phi(\lambda),\phi(\lambda)}$

Regarding the departures of exact power of the ordinary test statistics from their asymptotic powers and from their exact sizes penalization improves them especially for large negative values of λ. Among all the test statistics that satisfy (32) for n = 15, the test statistics which satisfy (30) are $S(P^{0.5})^{(1)}_{\phi(\lambda),\phi(\lambda)}$ for λ = -0.8, -0.9. For n = 25 and 35, $S(P^1)^{(1)}_{\phi(\lambda),\phi(\lambda)}$ with large negative values of λ perform better among the test statistics satisfying (32).

5. CONCLUSION AND DISCUSSION

In this paper, we have defined the family of penalized power-divergence test statistics by using penalized power-divergence measure and penalized power-divergence estimators in the family of ordinary power-divergence test statistics. Our aim was to compare the size and power properties of ordinary and penalized power-divergence test statistics for testing log-linear models in three-way contingency tables for small samples. In general penalization improves the size and power properties of the ordinary power-divergence statistics in 2x2x2 contingency tables especially for large negative values of λ. It also appears that penalized power-divergence test statistics with large negative values of λ perform better than all of the ordinary ones, including the likelihood ratio test statistic. It seems that penalization value of w = 0.5 is better for n=15, whereas w=1 is a better choice for n=25 and n=35.

Standard log-linear models do not allow incorporation of ordered categories which can be encountered in lots of applied studies. One method of incorporating such information is to specify ordered factor scores and fit model which will no longer be log-linear (see, Christensen, 1990, Ch. V). The problem of considering models in which non-linear terms have been added was considered for the first time by Tukey (1949) who solved the problem of testing interaction in two-way ANOVA with one observation per cell. There have been several extensions of this test to the models with different functions of interactions. Alin and Kurt (2006) give detailed review of these methods. These are called two-stage tests procedure in which parameters are first estimated using an additive model, and then the estimates are treated as known constants for the second stage of the test. Pardo and Pardo (2005) have presented the families based on ϕ-divergences and studied their size and power properties for testing non-additivity in log-linear models by using two-stage tests procedure as a method of testing non-additivity. Another study can be performed to see the size and power properties of the penalized power-divergence test statistics, especially for small samples.

ACKNOWLEDGEMENTS

Authors are grateful to Ayenandranath Basu for his precious support.

Table 1. Exact sizes of $S(O)^{(1)}_{\phi(\lambda),\phi(\lambda)}$ under null hypothesis $H_{null}: \boldsymbol{p_0} \in H_2$

	λ							
n	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
15	0.1090	0.0560	0.0650	0.1520	0.1710	0.2050	0.2750	0.4050
25	0.0690	0.0520	0.0580	0.1110	0.1320	0.1670	0.2320	0.3390
35	0.0490	0.0370	0.0400	0.0710	0.0860	0.1050	0.1430	0.1790

Table 2. Exact sizes of $S(P^1)^{(1)}_{\phi(\lambda),\phi(\lambda)}$ under null hypothesis $H_{null}: \boldsymbol{p_0} \in H_2$

	λ							
n	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
15	0.1090	0.1090	0.1070	0.0910	0.0880	0.0870	0.0840	0.0810
25	0.0690	0.0660	0.0670	0.0710	0.0690	0.0650	0.0670	0.0650
35	0.0490	0.0510	0.0500	0.0530	0.0530	0.0530	0.0530	0.0520

Table 3. Exact sizes of $S(P^{0.5})^{(1)}_{\phi(\lambda),\phi(\lambda)}$ under null hypothesis $H_{null}: \boldsymbol{p_0} \in H_2$

	λ							
n	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
15	0.0500	0.0560	0.0530	0.0410	0.0400	0.0370	0.0340	0.0340
25	0.0570	0.0520	0.0530	0.0510	0.0500	0.0490	0.0490	0.0470
35	0.0370	0.0370	0.0360	0.0410	0.0420	0.0420	0.0430	0.0430

Table 4. (Exact power-Asymptotic power) of $S(O)^{(1)}_{\phi_{(\lambda)},\phi_{(\lambda)}}$ for testing $H_{null}: \boldsymbol{p_0} \in H_2$ against $H_{alt}: \boldsymbol{p}^{*}_{1,n}$ where $\boldsymbol{p_1} \in H_1$

	λ							
n	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
15	0.1489	0.0359	0.0609	0.2349	0.2569	0.3009	0.3349	0.4209
25	-0.0071	-0.0601	-0.0441	0.0509	0.0779	0.1119	0.1519	0.1989
35	-0.0103	-0.0573	-0.0403	-0.0013	-0.0023	-0.0013	-0.0013	-0.0023

Table 5. (Exact power-Asymptotic power) of $S(P^1)^{(1)}_{\phi_{(\lambda)},\phi_{(\lambda)}}$ for testing $H_{null}: \boldsymbol{p_0} \in H_2$ against $H_{alt}: \boldsymbol{p}^{*}_{1,n}$ where $\boldsymbol{p_1} \in H_1$

	λ							
n	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
15	0.1489	0.1559	0.1519	0.0949	0.0919	0.0909	0.0839	0.0799
25	-0.0071	-0.0131	-0.0101	-0.0041	-0.0021	-0.0001	-0.0001	-0.0001
35	-0.0103	-0.0293	-0.0233	-0.0013	-0.0023	-0.0013	-0.0013	-0.0013

Table 6. (Exact power-Asymptotic power) of $S(P^{0.5})^{(1)}_{\phi_{(\lambda)},\phi_{(\lambda)}}$ for testing $H_{null}: \boldsymbol{p_0} \in H_2$ against $H_{alt}: \boldsymbol{p}^{*}_{1,n}$ where $\boldsymbol{p_1} \in H_1$,

	λ							
n	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
15	0.0229	0.0359	0.0319	0.0079	0.0029	-0.0001	-0.0001	-0.0001
25	-0.0561	-0.0601	-0.0561	-0.0551	-0.0551	-0.0541	-0.0541	-0.0591
35	-0.0353	-0.0573	-0.0503	-0.0263	-0.0273	-0.0283	-0.0293	-0.0293

Table 7. (Exact power-Exact size) of $S(O)^{(1)}_{\phi_{(\lambda)},\phi_{(\lambda)}}$ for testing $H_{null}: \boldsymbol{p_0} \in H_2$ against $H_{alt}: \boldsymbol{p_1} \in H_1$

	λ							
n	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
15	0.0930	0.0830	0.0850	0.1250	0.1330	0.1430	0.1160	0.0710
25	0.2010	0.1650	0.1750	0.2170	0.2230	0.2220	0.1970	0.1370
35	0.3270	0.2770	0.2910	0.3300	0.3320	0.3270	0.3090	0.2900

Table 8. (Exact power-Exact size) of $S(P^1)^{(1)}_{\phi_{(\lambda)},\phi_{(\lambda)}}$ for testing $H_{null}: \boldsymbol{p_0} \in H_2$ against $H_{alt}: \boldsymbol{p_1} \in H_1$

	λ							
n	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
15	0.0930	0.0940	0.0950	0.0810	0.0780	0.0790	0.0790	0.0800
25	0.2010	0.1980	0.2000	0.2020	0.2060	0.2120	0.2100	0.2120
35	0.3270	0.2920	0.3050	0.3320	0.3340	0.3320	0.3330	0.3340

Table 9. (Exact power-Exact size) of $S(P^{0.5})^{(1)}_{\phi_{(\lambda)},\phi_{(\lambda)}}$ for testing $H_{null}: \boldsymbol{p_0} \in H_2$ against $H_{alt}: \boldsymbol{p_1} \in H_1$

	λ							
n	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
15	0.0830	0.0830	0.0830	0.0870	0.0870	0.0890	0.0920	0.0920
25	0.1640	0.1650	0.1680	0.1710	0.1720	0.1740	0.1740	0.1710
35	0.3070	0.2770	0.2880	0.3070	0.3060	0.3040	0.3040	0.3040

REFERENCES

Agresti, A. (1996). *An Introduction to Categorical Data Analysis*. New York: John Wiley.

Alin, A., Kurt, S. (2006) Testing Non-Additivity (Interaction) in Two-Way ANOVA Tables with No Replication. *Statistical Methods in Medical Research* 15, 63-85.

Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W. (1975) *Discrete Multivariate Analysis Theory and Practice*. Massachusetts: MIT Press.

Basu, A. and Basu, S. (1998). Penalized Minimum Disparity Methods for Multinomial Models. *Statistica Sinica* 8, 841-860.

Christensen, R. (1990) *Log-Linear Model*. New York: Springer-Verlag.

Christensen, R. (2000). Linear and Log-Linear Models. *Journal of the American Statistical Association* 95, 1290-1293.

Cressie, N. and Read, T.R.C. (1984) Multinomial Goodness-of-Fit Tests. *Journal of Royal Statistical Society B* 46, 440-464.

Cressie, N. and Pardo, L. (2000) Minimum ϕ-divergence Estimator and Hierarchical Testing in Loglinear models. *Statistica Sinica* 10, 867-884.

Cressie, N. and Pardo, L. (2002) Model Checking in Loglinear Models Using ϕ-divergences and MLEs. *Journal of Statistical Planning and Inference* 103, 437-453.

Cressie, N., Pardo, L., Pardo, M.C. (2003) Size and Power Considerations for Testing Loglinear Models Using ϕ-Divergence Test Statistics. *Statistica Sinica* 13, 555-570.

Csiszár, I. (1967). Information Type Measures of Difference of Probability Distributions and Indirect Observations. *Studia Scientarum Mathematicarum Hungarica* 2, 105-113.

Dale, J.R. (1986) Asymptotic Normality Goodness of Fit Statistics for Sparse Product Multinomials. *Journal of Royal Statistical Society Series B* 41, 48-59.

Fenech, A.P. and Westfall P.H. (1988) The Power Function of Conditional Log-Linear Models. *Journal of the American Statistical Association* 83, 198-203.

Freidlin, B. and Gastwirth, J.L. (1999) Unconditional Versions of Several Tests Commonly Used in the Analysis of Contingency Tables. *Biometrics* 55, 264-267

Haberman, S.J. (1974) *The Analysis of Frequency Data*. Chicago: University of Chicago Press.

Harris, I.R., Basu, A. (1997) *A Generalized Divergence Measure*. Technical report. Applied Statistics Unit, Indian Statistical Institute, Calcutta,700 035, India.

Lee, H., Viele, K. (2001) Loglinear Models and Goodness-of-Fit Statistics for Train Waybill Data. *Journal of Transportation and Statistics* 4, 83-87.

Lindsay, B.G. (1994) Efficiency versus Robustness: The Case for Minimum Hellinger Distance and Related Methods. *Annals of Statistic.* 22, 1081-1114.

Menéndez, M., Morales, D., Pardo, L. and Vajda, I. (2001) Minimum Divergence Estimators Based on Grouped Data. *Annals of the Institute of Statistical Mathematics*. 53, 277-288.

Morales, D., Pardo, L., Vajda, I. (1995) Asymptotic Divergence Measures of Estimates of Discrete Distributions. *Journal of Statistical Planning and Inference* 48, 347-369.

Pardo, L., Pardo, M.C. and Zografos, K. (2001) Minimum ϕ-Divergence Estimator for Homogeneity in Multinomial Populations. *The Indian Journal of Statistics* 63, 72-92.

Pardo, L. and Pardo, M. D. C. (2003) Minimum Power-Divergence Estimator in Three-Way Contingency Tables. *Journal of Statistical Computation Simulation* 73, 819-831.

Pardo, L., Pardo, M. D. C. (2005) Nonadditivity in Loglinear Models Using ϕ-Divergences and MLEs. *Journal of Statistical Planning and Inference* 127, 237-252.

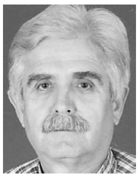
Read, T.R. and Cressie, N. (1988) *Goodness-of-Fit Statistics for Discrete Multivariate Data*. New York: Springer-Verlag,

Oler, J. (1985). Noncentrality Parameters in Chi-Squared Goodness-of-Fit Analyses with an Application to Log-Linear Procedures. *Journal of the American Statistical Association* 80, 181-189.

Tukey, J.W. (1949). One Degree of Freedom for Non-Addditivity. *Biometrics* **5**, 232-242.

Aylin ALIN, 1973 yılında Giresun'da doğdu. 1993 yılında Ondokuz Mayıs Üniversitesi İstatistik Bölümü'nde lisans, 1997 yılında Michigan State Üniversitesi İstatistik ve Olasılık Bölümü'nde yüksek lisans, 2006 yılında da Dokuz Eylül Üniversitesi İstatistik Bölümü'nde doktora eğitimini ta-

mamladı. Halen Dokuz Eylül Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü'nde öğretim üyesi olarak görev yapmaktadır.



Serdar KURT, 1949 yılında Samsun/Havza'da doğdu. 1974 yılında ODTÜ Fen Fakültesi Matematik Bölümü'nde lisans, 1979 yılında Hacettepe Üniversitesi İstatistik Bölümü'nde yüksek lisans, 1982 yılında da doktora eğitimini tamamladı. 1996 yılında Ege Üniversitesi'nde doçent, 2001 yılında da Dokuz Eylül Üniversitesi'nde profesör oldu. Halen Dokuz Eylül Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü'nde öğretim üyesi olarak çalışmaktadır.