# A Note on Penalized Power-Divergence Test Statistics 

Aylin Alin


#### Abstract

In this paper, penalized power-divergence test statistics have been defined and their exact size properties to test a nested sequence of log-linear models have been compared with ordinary power-divergence test statistics for various penalization, $\lambda$ and main effect values. Since the ordinary and penalized powerdivergence test statistics have the same asymptotic distribution, comparisons have been only made for small and moderate samples. Three-way contingency tables distributed according to a multinomial distribution have been considered. Simulation results reveal that penalized power-divergence test statistics perform much better than their ordinary counterparts.


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

## I. Introduction

LET $\mathrm{S}_{1}, \ldots, \mathrm{~S}_{\mathrm{n}}$ be an independent and identically distributed _sample of size $\mathrm{n} \geq 1$ with the probability distribution $\boldsymbol{P}\left(\boldsymbol{\theta}_{0}\right)$. This distribution is assumed to be unknown but belonging to a known family,

$$
\boldsymbol{P}=\left\{\boldsymbol{p}(\boldsymbol{\theta})=\left(p_{1}(\boldsymbol{\theta}), \ldots, p_{M}(\boldsymbol{\theta})\right)^{T}: \boldsymbol{\theta} \in \boldsymbol{\Theta}\right\}
$$

of distributions with components taking values on $\chi=\{1,2$, $\ldots, M\}$ with parameter space $\Theta \subseteq R^{t}(t<M-1)$. Hence, the true value, $\boldsymbol{\theta}_{0}$, of parameters vector $\boldsymbol{\theta}=\left(\theta_{1}, \ldots, \theta_{t}\right)^{T} \in \Theta$ is assumed to be fixed but unknown. The statistic interested in is $\left(N_{1}, \ldots, N_{M}\right)^{\mathrm{T}} . N_{s}$, for $s=1, \ldots, M$, denotes the cell count in the $j$-th cell of the contingency table. For the rest of the paper, it will be assumed that $\left(N_{1}, \ldots, N_{M}\right)^{\mathrm{T}}$ has the multinomial distribution with probabilities belonging to a general class of log-linear models, that is ,

$$
\begin{equation*}
P\left(N_{1}=n_{1}, \ldots, N_{M}=n_{M}\right)=\frac{n!}{n_{1}!\ldots n_{M}!} p_{1}(\boldsymbol{\theta})^{n_{1}} \ldots p_{M}(\boldsymbol{\theta})^{n_{M}} \tag{1}
\end{equation*}
$$

for $n_{s} \geq 0$ such that $\sum_{s=1}^{M} n_{s}=n$ and

[^0]\[

$$
\begin{equation*}
p_{s}(\boldsymbol{\theta})=\frac{\exp \left(\boldsymbol{w}_{S}^{T} \boldsymbol{\theta}^{*}\right)}{\sum_{s=1}^{M} \exp \left(\boldsymbol{w}_{\mathrm{S}}^{\boldsymbol{T}} \boldsymbol{\theta}^{*}\right)} \text { for } s=1, \ldots, M \tag{2}
\end{equation*}
$$

\]

with $\boldsymbol{\theta}^{*}=\left(u, \theta_{1}, \ldots, \theta_{t}\right)^{T}$ being $(t+1) \times 1$ column vector with

$$
\begin{equation*}
u=-\log \left(\sum_{s=1}^{M} \exp \left(\boldsymbol{w}_{S}^{\boldsymbol{T}} \boldsymbol{\theta}\right)\right) \tag{3}
\end{equation*}
$$

which is overall mean effect parameter calculated as the normalizing constant to insure $\sum_{s=1}^{M} p_{S}(\boldsymbol{\theta})=1.1 \mathrm{x} t$ vector $\boldsymbol{w}_{s}^{T}=\left(w_{1(s)}, \ldots, w_{t(s)}\right)$ in equation (2) forms the $M \mathrm{x} t \log$ linear model matrix of explanatory variables $\boldsymbol{W}=\left(\boldsymbol{w}_{1}, \ldots\right.$, $\left.\boldsymbol{w}_{M}\right)^{T}$ which is assumed to have full column rank $t<M-1$. Elements of this matrix are determined according to linear constraints on the parameter vector $\boldsymbol{\theta}$. Columns of $\boldsymbol{W}$ are linearly independent of the $M \times 1$ column vector $(1, \ldots, 1)^{T}$.
[1] defined the family of power-divergence measures, $I_{\lambda}(\hat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\theta}))$ as below.

$$
=\left\{\begin{array}{l}
\frac{1}{\lambda(\lambda+1)} \sum_{\mathrm{s}=1}^{M} \hat{p}_{s}\left[\left(\frac{\hat{p}_{s}}{p_{s}(\boldsymbol{\theta})}\right)^{\lambda}-1\right] \text { for } \lambda \neq 0, \lambda \neq-1  \tag{4}\\
\sum_{\mathrm{s}=1}^{M} \hat{p}_{s} \log \left(\frac{\hat{p}_{s}}{p_{s}(\boldsymbol{\theta})}\right)-\left(\frac{\hat{p}_{s}}{p_{s}(\boldsymbol{\theta})}-1\right) \text { for } \lambda=0 \\
\sum_{\mathrm{s}=1}^{M} p_{s}(\boldsymbol{\theta}) \log \left(\frac{\hat{p}_{s}}{p_{s}(\boldsymbol{\theta})}\right)-\left(\frac{\hat{p}_{s}}{p_{s}(\boldsymbol{\theta})}-1\right) \text { for } \lambda=-1
\end{array}\right.
$$

where $\hat{\boldsymbol{p}}=\left(\hat{p}_{1}, \ldots, \hat{p}_{M}\right)^{T}$, with $\hat{p}_{S}=\frac{N s}{n}$. The minimum powerdivergence estimator, as the value minimizing $I_{\lambda}(\hat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\theta}))$ with respect to $\boldsymbol{\theta}$, is defined by

$$
\begin{equation*}
\hat{\boldsymbol{\theta}}_{\lambda}^{I} \equiv \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\arg \min } I_{\lambda}(\hat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\theta})) \tag{5}
\end{equation*}
$$

To deal with the large weight that power-divergence measures put on empty cells, [2] have proposed an empty cell penalty for $I_{\lambda}(\hat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\theta}))$ in multinomial models and defined the
family of penalized power-divergence measures as given by Eq.(6);

$$
\begin{align*}
P_{\lambda}^{w}(\hat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\theta})) & =\sum_{\substack{s=1 \\
\hat{p}_{s} \neq 0}}^{M}\left[\frac{1}{\lambda(\lambda+1)} \hat{p}_{s}\left[\left(\frac{\hat{p}_{s}}{p_{s}(\boldsymbol{\theta})}\right)^{\lambda}-1\right]+\frac{p_{s}(\boldsymbol{\theta})-\hat{p}_{s}}{\lambda+1}\right] \\
& +w \sum_{\substack{s=1 \\
\hat{p}_{s}=0}}^{M} p_{s}(\boldsymbol{\theta}) \tag{6}
\end{align*}
$$

which is comprised of two parts; the part for the nonempty cells and the part for the empty cells. $w$ is the penalty put on empty cells. Eq. (6) is for the $\lambda$ values not equal to 0 or -1 . For $\lambda=0$ and $\lambda=-1$, only the part for the nonempty cells in Eq. (6) will be replaced by the corresponding power-divergence measures given in Eq. (4). The penalized minimum powerdivergence estimator can be defined by

$$
\begin{equation*}
\hat{\boldsymbol{\theta}}_{\lambda}^{P^{w}} \equiv \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\arg \min } P_{\lambda}^{w}(\hat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\theta})) \tag{7}
\end{equation*}
$$

Note that reweighting empty cells will not alter the asymptotic properties of the corresponding estimator [3]. It should be noted that 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 $\lambda$. As mentioned by [2], the disproportionately large weight that power-divergence measures put on empty cells causes the unfortunate trade-off between robustness and small sample efficiency properties of the minimum powerdivergence estimators. As seen from the simulation results, this also affects the small sample properties of the ordinary power-divergence test statistics which are based on powerdivergence measures and developed by [4] for testing a nested sequence of log-linear models. In this study, the new family of penalized power-divergence test statistics has been defined to deal with this problem. The aim is to show that penalization improves the exact sizes of ordinary power-divergence test statistics. The rest of the paper is laid out as follows: After giving the brief description of the ordinary and penalized power-divergence test statistics in Section two, the simulation results will be presented in Section three.

## II. Ordinary and Penalized Power-Divergence Test Statistics

To test nested sequence of log-linear models

$$
\begin{gather*}
H_{\text {null }}: H_{L+1}: \boldsymbol{p}=\boldsymbol{p}(\boldsymbol{\theta}) ; \boldsymbol{\theta} \in \Theta_{L+1} \\
\begin{array}{c}
\text { against }
\end{array}  \tag{8}\\
H_{\text {alt }}: H_{L}: \boldsymbol{p}=\boldsymbol{p}(\boldsymbol{\theta}) ; \boldsymbol{\theta} \in \Theta_{L}
\end{gather*}
$$

where $L=1, \ldots, m-1, m \leq \mathrm{t}<M-1$ and $\Theta_{m} \subset \Theta_{m-1} \subset \ldots \subset \Theta_{1} \equiv$ $\Theta \subseteq R^{t}(t<M-1)$ with $d_{m}<d_{m-1}<\ldots<d_{1}=t$ where $d_{L}=\operatorname{dim}\left(\Theta_{L}\right)$ for $L=1,2, \ldots, m,[4]$ have suggested the following family of test statistics:

$$
\begin{equation*}
S(O)_{\lambda_{1}, \lambda_{2}}^{(L)}=2 n\left\{I_{\lambda_{1}}\left(\hat{\boldsymbol{p}}, \boldsymbol{p}\left(\hat{\boldsymbol{\theta}}_{\lambda_{2}}^{I(L+1)}\right)\right)-I_{\lambda_{1}}\left(\hat{\boldsymbol{p}}, \boldsymbol{p}\left(\hat{\boldsymbol{\theta}}_{\lambda_{2}}^{I(L)}\right)\right)\right\} \tag{9}
\end{equation*}
$$

where $\hat{\boldsymbol{\theta}}_{\lambda_{2}}^{I(L+1)}$ and $\hat{\boldsymbol{\theta}}_{\lambda_{2}}^{I(L)}$ are the minimum powerdivergence estimators as defined by Eq.(5) under the models of $H_{L+1}$ and $H_{L}$, respectively. When $S(O)_{\lambda_{1}, \lambda_{2}}^{(L)}>\mathrm{c}, \mathrm{H}_{\text {null }}$ is rejected, where c is specified so that the size of the test is $\alpha$;

$$
\begin{equation*}
P\left(S(O)_{\lambda_{1}, \lambda_{2}}^{(L)}>c \mid H_{L+1}\right)=\alpha ; \alpha \in(0,1) \tag{10}
\end{equation*}
$$

[4] have shown that under multinomial sampling with probabilities belonging to a log-linear model and $H_{\text {null }}=H_{L+1}$, the test statistics $S(O)_{\lambda_{1}, \lambda_{2}}^{(L)}$ converges in distribution to a chisquared distribution with degrees of freedom $\mathrm{d}_{L}-\mathrm{d}_{L+1} ; L=1$, ..., m-1. Hence,

$$
\begin{equation*}
c=\chi_{d_{L}-d_{L+1}}^{2}(1-\alpha), \tag{11}
\end{equation*}
$$

where $P\left(\chi_{f}^{2} \leq \chi_{f}^{2}(p)\right)=p . S(O)_{\lambda_{1}, \lambda_{2}}^{(L)}$ gives the well known likelihood ratio test statistic when $\lambda_{1}=\lambda_{2}=0$. It should be pointed out that the nonnegativity of $S(O)_{\lambda_{1}, \lambda_{2}}^{(L)}$ does not hold when $\lambda_{1} \neq \lambda_{2}$ [4]. The large weight put on empty cells by the family of power-divergence measures effects the exact size properties of $S(O)_{\lambda_{1}, \lambda_{2}}^{(L)}$ for testing the nested sequence of loglinear models given with Eq.(8). To deal with this problem, the following family of penalized power-divergence test statistic has been proposed.

$$
\begin{equation*}
S\left(P^{w}\right)_{\lambda_{1}, \lambda_{2}}^{(L)}=2 n\left\{P_{\lambda_{1}}^{w}\left(\hat{\boldsymbol{p}}, \boldsymbol{p}\left(\hat{\boldsymbol{\theta}}_{\lambda_{2}}^{P^{w}}(L+1)\right)\right)-P_{\lambda_{1}}^{w}\left(\hat{\boldsymbol{p}}, \boldsymbol{p}\left(\hat{\boldsymbol{\theta}}_{\lambda_{2}}^{P^{w}(L)}\right)\right)\right\} \tag{12}
\end{equation*}
$$

where $\hat{\boldsymbol{\theta}}_{\lambda_{2}}^{P^{w}(L+1)}$ and $\hat{\boldsymbol{\theta}}_{\lambda_{2}}^{P^{w}(L)}$ are the penalized minimum power-divergence estimators as defined by (7) under the models of $H_{L+1}$ and $H_{L}$, respectively. As mentioned by [2], 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, only the small and moderate sample exact size properties of $S(O)_{\lambda_{1}, \lambda_{2}}^{(L)}$ and $S\left(P^{w}\right)_{\lambda_{1}, \lambda_{2}}^{(L)}$ have been studied via simulation study. The next section gives the results.

## III. Simulation Study

In the simulation study, the case of $2 \times 2 \times 2$ contingency tables has been considered, so $M=8$. To distinguish the categorical variables, separate indices have been used for each variable. Let $X, Y$ and $Z$ be three categorical response variables having $I, J$ and $K$ levels, respectively. $p_{i j k}(\theta)=\mathrm{P}(X$ $=i, Y=j, Z=k), i=1,2 ; j=1,2 ; k=1,2$ is a probability
distribution of the responses $(X, Y, Z)$ of a subject randomly chosen from a population. The hypotheses considered are;
$H_{1}: p_{i j k}(\theta)=\exp \left[\mathbf{u}+\theta_{1(i)}+\theta_{2(j)}+\theta_{3(k)}+\theta_{12(i j)}+\theta_{13(i k)}+\theta_{3(j k)}\right]$
$H_{2}: p_{i j k}(\theta)=\exp \left[u+\theta_{1(i)}+\theta_{2(j)}+\theta_{3(k)}+\theta_{12(i j)}+\theta_{13(i k)}\right]$
$H_{3}: p_{i j k}(\theta)=\exp \left[\mathrm{u}+\theta_{1(i)}+\theta_{2(j)}+\theta_{3(k)}+\theta_{12(i j)}\right]$
$H_{4}: p_{i j k}(\theta)=\exp \left[u+\theta_{1(i)}+\theta_{2(j)}+\theta_{3(k)}\right]$
for $i, j, k=1,2$ with the following linear constraints on the parameters;

$$
\begin{gather*}
\sum_{i=1}^{2} \theta_{1(i)}=\sum_{j=1}^{2} \theta_{2(j)}=\sum_{k=1}^{2} \theta_{3(k)}=0, \sum_{i=1}^{2} \theta_{12(i j)}=\sum_{j=1}^{2} \theta_{12(i j)}=0 \\
\sum_{i=1}^{2} \theta_{13(i k)}=\sum_{k=1}^{2} \theta_{13(i k)}=0, \sum_{j=1}^{2} \theta_{23(j k)}=\sum_{k=1}^{2} \theta_{23(j k)} \tag{13}
\end{gather*}
$$

Since the conclusions are the same for testing $\mathrm{H}_{3}$ versus $\mathrm{H}_{2}$ and $H_{2}$ versus $H_{1}$, only the results for testing $H_{4}$ versus $H_{3}$ are given because of the space considerations. $\exp \left[\theta_{1(1)}\right]=$ $\exp \left[\theta_{2(1)}\right]=\exp \left[\theta_{3(1)}\right]=5 / 6$ and $\exp \left[\theta_{1(1)}\right]=\exp \left[\theta_{2(1)}\right]=$ $\exp \left[\theta_{3(1)}\right]=3 / 4$ are the moderate and big main effects considered for this study, respectively. The big interaction effect $\exp \left[\theta_{12(11)}\right]=3 / 4$ has been chosen. Overall mean effect parameter $u$ is calculated as given by (3).

Let $\boldsymbol{p}_{0} \in H_{4} \in H_{\text {null }}$ and $\boldsymbol{p}_{1} \in H_{3} \in H_{\text {alt }}$. By this simulation study, the following exact probabilities will be obtained:

$$
\begin{align*}
\alpha_{n}^{S(O)^{(3)}} \equiv P\left(S(O)_{\lambda_{1}, \lambda_{2}}^{(3)}>c \mid \boldsymbol{p}_{\mathbf{0}}\right) \\
\alpha_{n}^{S\left(P^{w}\right)^{(3)}} \equiv P\left(S\left(P^{w}\right)_{\lambda_{1}, \lambda_{2}}^{(3)}>c \mid \boldsymbol{p}_{\mathbf{0}}\right) \tag{14}
\end{align*}
$$

Five different penalization values ( $w$ ) are chosen: $0,0.25$, $0.5,0.75$ and 1 . As mentioned above, the nonnegativity of $S(O)_{\lambda_{1}, \lambda_{2}}^{(l)}$ does not hold when $\lambda_{1} \neq \lambda_{2}$. Thus, only the combinations with $\lambda_{1}=\lambda_{2}=\lambda$ have been considered. For convenience, only one subscript $\lambda$ will be used with test statistic instead of two $\lambda \mathrm{s}$ for the rest of the paper. For $S(O)_{\lambda}^{(3)}, \lambda$ values have been chosen as $-0.9,-0.8,-0.7,-0.6$, $-0.5,0,2 / 3,1$. One of the advantages of $S\left(P^{w}\right)_{\lambda}^{(L)}$ over its ordinary counterparts is its being able to be defined for $\lambda \leq-1$. Hence $\lambda=-1,-1.5$ and -2 are also considered for $S\left(P^{w}\right)_{\lambda}^{(3)}$. Sample sizes considered for this simulation study are $n=25$, 35, 45, 55. Exact probability estimations given with Eq.(14) are obtained using 10000 simulations from the multinomial distributions with ( $n, \boldsymbol{p}_{\mathbf{0}}$ ). All calculations have been done using Mathematica 5.2. The aims are: 1) To show the effects of main effects on $\alpha_{n}^{S(O)^{(3)}}$ and $\alpha_{n}^{S\left(P^{w}\right)^{(3)}}$, 2) To determine the
test statistic with closest $\alpha_{n}^{S(i)^{(3)}}$ for $i=O, P^{w}$, to the nominal size of 0.05 for each sample size. Results are given at the end of the paper. Tables I-VI and Tables VII-XII give the exact sizes of the test statistics for moderate and big main effects, respectively. As it is seen from Tables I and VII, $\alpha_{n}^{S(O)^{(3)}}$ for $\lambda<0$ are not even close to 0.05 especially for $n=25$ and $n=$ 35. However, penalization seems to improve the exact sizes of these test statistics. Moreover, exact sizes of $S(O)_{\lambda}^{(3)}$ for $\lambda<0$ get bigger as main effects get bigger whereas it turns the otherwise for their penalized counter parts. There is no obvious pattern for $\alpha_{n}^{S(O)^{(3)}}$ for $\lambda \geq 0$ in terms of main effect changes while $\alpha_{n}^{S\left(P^{w}\right)^{(3)}}$ for $\lambda \geq 0$ and $w=1$ are bigger for big main effects than for moderate main effects. However, as $w$ decreases $\alpha_{n}^{S\left(P^{w}\right)^{(3)}}$ for $\lambda \geq 0$ start to get smaller for big main effects. In general, departure of $\alpha_{n}^{S\left(P^{w}\right)^{(3)}}$ from 0.05 is bigger for big main effects for negative values of $\lambda$ with bigger departure for $\lambda \leq-1$ and $w \leq 0.5$, especially for small samples. Figs. 1a and 1b illustrate the $\left(\alpha_{n}^{S(i)^{(3)}}-0.05\right)$ for $i=O, P^{w}$ for $n$ $=35$ with moderate and big main effects, respectively. Lines represent the $\lambda$ values, and penalization values are replaced on $x$-axes. $w=2$ represents the no penalization case, i.e. ordinary power-divergence case, for each $\lambda$. It is easily seen that as $w$ decreases $\alpha_{n}^{S\left(P^{w}\right)^{(3)}}$ for all $\lambda$ get smaller for big main effects resulting with bigger departure from 0.05 than moderate main effects. As mentioned this difference is bigger for $\lambda \leq-1$ and $w \leq 0.5$. The best penalization value is given in circle for each $\lambda$. When best penalization value is checked for $S(O)_{\lambda}^{(3)}$, for both main effects and for each sample size in terms of closeness to the 0.05 , it is seen that $w=0.75$ and $w=1$ seem better for $S(O)_{\lambda}^{(3)}$ for $\lambda>0$. On the other hand, $w$ needed decreases as $n$ increases for $\lambda \leq 0$ for moderate main effects. Moreover, there seems no difference between moderate and big main effects in terms of penalization value needed for $n=$ 25 and $n=55$. However for $n=35$ and $n=45, w$ needed increases for $\lambda<0$ as main effects get bigger. The best test statistics with smallest $\left|\alpha_{n}^{S(i)^{(3)}}-0.05\right|$ for $i=O, P^{w}$ are given in bold for each sample sizes on Tables I-XII. Table XIII summarizes these test statistics. It seems that as sample size gets larger, penalization value needed decreases and $\lambda$ increases and after $n=45 S\left(P^{0}\right)_{0}^{(1)}$ performs best in terms of exact size for moderate main effects. While there seems no obvious pattern for big main effects as it is for moderate main effects, $w$ tends to increase for small samples and then it decreases as $n$ increases, and penalized power-divergence test statistic with $\lambda=0$ performs better.

TABLE XIII

| The Test Statistics with Smallest $\left\|\alpha_{n}^{S(i)^{(3)}}{ }^{4} 0.05\right\|$ |  |  |  |
| :---: | :---: | :---: | :---: |
| WITH MODERATE MAIN EFFECTS |  |  |  |
| $n=25$ | $n=35$ | $n=45$ | $n=55$ |
| $S\left(P^{1}\right)_{-2}^{(3)}$ | $S\left(P^{0.5}\right)_{-0.9}^{(3)}$ | $S\left(P^{0}\right)_{0}^{(3)}$ | $S\left(P^{0}\right)_{0}^{(3)}$ |
|  | WITH BIG MAIN EFFECTS |  |  |
| $n=25$ | $n=35$ | $n=45$ | $n=55$ |
| $S\left(P^{0.75}\right)_{0}^{(3)}$ | $S\left(P^{1}\right)_{-1.5}^{(3)}$ | $S\left(P^{0.75}\right)_{-1.5}^{(3)}$ | $S\left(P^{0.5}\right)_{0}^{(3)}$ |



Fig. 1(a) $\left(\alpha_{n}^{S(i)^{(3)}}-0.05\right)$ of test statistics vs penalization values for $n=35$ with moderate main effects

## References

[1] Cressie, N., and Read, T. R. C., "Multinomial Goodness-of-Fit Tests," Journal of Royal Statistical Society B, vol. 46, pp. 440-464, 1984.
[2] Basu, A., and Basu, S., "Penalized Minimum Disparity Methods for Multinomial Models," Statistica Sinica, vol. 8, pp. 841-860, 1998.
[3] Pardo, L., and Pardo, M. D. C., "Minimum Power-Divergence Estimator in Three-Way Contingency Tables," Journal of Statistical Computation Simulation, vol. 73, pp. 819-831, 2003.
[4] Cressie, N., and Pardo, L.," $M$ Minimum $\phi$-divergence Estimator and Hierarchical Testing in Loglinear models," Statistica Sinica, vol.10, pp. 867-884, 2000.

In conclusion, even though the likelihood ratio test statistic $\left(S(O)_{0}^{(3)}\right)$ is the mostly used test statistic for testing nested log-linear models, it does not perform well in the case of empty cells for the $2 \times 2 \times 2$ contingency tables. But, penalization improves the exact sizes of this test statistic and its penalized counter parts perform better among other test statistics as sample size increases for both main effects. Penalization has also great effect on the exact sizes of ordinary power-divergence test statistics for $\lambda<0$ such that penalized ones for $\lambda<-0.8$ perform better for small sample sizes for both main effects.


Fig. 1 (b) $\left(\alpha_{n}^{S(i)^{(3)}}-0.05\right)$ of test statistics vs penalization values for $n=35$ with big main effects

## Exact Sizes of Test Statistics for Moderate Main Effects

TABLE I
$\alpha^{S(O)^{(3)}}$

|  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | TABLE I <br> $\alpha^{S(O)^{(3)}}$ <br> $n$ |  |  |
|  |  |  |  |  | $\lambda$ |

TABLE II
$\alpha^{S\left(P^{1}\right)^{(3)}}$

| $\alpha_{n}^{S\left(P^{1}\right)^{(3)}}$ |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\lambda$ |  |  |  |  |  |  |  |  |  |  |  |
| $n$ | 1 | 2/3 | 0 | -0.5 | -0.6 | -0.7 | -0.8 | -0.9 | -1.0 | -1.5 | -2.0 |
| 25 | 0.0576 | 0.0592 | 0.0604 | 0.0570 | 0.0563 | 0.0554 | 0.0554 | 0.0548 | 0.0543 | 0.0523 | 0.0493 |
| 35 | 0.0492 | 0.0508 | 0.0539 | 0.0561 | 0.0570 | 0.0564 | 0.0567 | 0.0566 | 0.0571 | 0.0578 | 0.0573 |
| 45 | 0.0481 | 0.0509 | 0.0561 | 0.0596 | 0.0605 | 0.0618 | 0.0619 | 0.0625 | 0.0624 | 0.0677 | 0.0677 |
| 55 | 0.0447 | 0.0471 | 0.0527 | 0.0573 | 0.0582 | 0.0588 | 0.0591 | 0.0598 | 0.0601 | 0.0640 | 0.0655 |

TABLE III
$\alpha_{n}^{S\left(P^{0.75}\right)^{(3)}}$

|  | $\lambda$ |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $n$ | 1 | $2 / 3$ | 0 | -0.5 | -0.6 | -0.7 | -0.8 | -0.9 | -1.0 | -1.5 | -2.0 |
| 25 | 0.0477 | 0.0483 | 0.0485 | 0.0469 | 0.0466 | 0.0458 | 0.0453 | 0.0444 | 0.0438 | 0.0423 | 0.0365 |
| 35 | 0.0455 | 0.0468 | 0.0510 | 0.0526 | 0.0533 | 0.0529 | 0.0533 | 0.0539 | 0.0542 | 0.0543 | 0.0518 |
| 45 | 0.0464 | 0.0492 | 0.0547 | 0.0583 | 0.0591 | 0.0604 | 0.0606 | 0.0611 | 0.0608 | 0.0646 | 0.0641 |
| 55 | 0.0438 | 0.0463 | 0.0523 | 0.0569 | 0.0577 | 0.0579 | 0.0582 | 0.0589 | 0.0593 | 0.0628 | 0.0642 |

TABLE IV
$\alpha^{S\left(P^{0.5}\right)^{(3)}}$

| $\lambda$ |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $n$ | 1 | 2/3 | 0 | -0.5 | -0.6 | -0.7 | -0.8 | -0.9 | -1.0 | -1.5 | -2.0 |
| 25 | 0.0372 | 0.0381 | 0.0396 | 0.0375 | 0.0374 | 0.0368 | 0.0368 | 0.0367 | 0.0365 | 0.0334 | 0.0276 |
| 35 | 0.0401 | 0.0417 | 0.0459 | 0.0484 | 0.0492 | 0.0493 | 0.0498 | 0.0499 | 0.0505 | 0.0502 | 0.0472 |
| 45 | 0.0439 | 0.0472 | 0.0530 | 0.0566 | 0.0574 | 0.0584 | 0.0585 | 0.0591 | 0.0588 | 0.0631 | 0.0622 |
| 55 | 0.0434 | 0.0459 | 0.0516 | 0.0562 | 0.0571 | 0.0575 | 0.0577 | 0.0585 | 0.0589 | 0.0625 | 0.0634 |

TABLE V
$\alpha^{S\left(P^{0.25}\right)^{(3)}}$

|  |  |  |  | $n$ |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $n$ | 1 | $2 / 3$ | 0 | -0.5 | -0.6 | -0.7 | -0.8 | -0.9 | -1.0 | -1.5 |  |  |
| 25 | 0.0265 | 0.0278 | 0.0321 | 0.0313 | 0.0313 | 0.0308 | 0.0306 | 0.0304 | 0.0300 | 0.0279 |  |  |
| 35 | 0.0366 | 0.0381 | 0.0427 | 0.0455 | 0.0464 | 0.0464 | 0.0469 | 0.0476 | 0.0482 | 0.0477 |  |  |
| 45 | 0.0420 | 0.0451 | 0.0511 | 0.0547 | 0.0555 | 0.0569 | 0.0571 | 0.0576 | 0.0573 | 0.0614 |  |  |
| 55 | 0.0424 | 0.0449 | 0.0508 | 0.0554 | 0.0563 | 0.0567 | 0.0570 | 0.0577 | 0.0581 | 0.0615 |  |  |

TABLE VI
$\alpha_{n}{ }_{n}\left(P^{0}\right)^{(3)}$

| $\lambda$ |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $n$ | 1 | 2/3 | 0 | -0.5 | -0.6 | -0.7 | -0.8 | -0.9 | -1.0 | -1.5 | -2.0 |
| 25 | 0.0207 | 0.0234 | 0.0285 | 0.0287 | 0.0288 | 0.0284 | 0.0284 | 0.0282 | 0.0279 | 0.0264 | 0.0214 |
| 35 | 0.0336 | 0.0362 | 0.0413 | 0.0446 | 0.0458 | 0.046 | 0.0467 | 0.0471 | 0.0477 | 0.0476 | 0.0441 |
| 45 | 0.0410 | 0.0443 | 0.0506 | 0.0545 | 0.0555 | 0.0568 | 0.057 | 0.0576 | 0.0575 | 0.0618 | 0.0616 |
| 55 | 0.0421 | 0.0446 | 0.0507 | 0.0555 | 0.0564 | 0.0568 | 0.0571 | 0.0578 | 0.0583 | 0.0617 | 0.0628 |

Exact Sizes of Test Statistics for Big Main Effects
TABLE VII

|  | TABLE VII$\alpha_{n}^{S(O)^{(3)}}$ |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\lambda$ |  |  |  |  |  |  |  |
| $n$ | 1 | 2/3 | 0 | -0.5 | -0.6 | -0.7 | -0.8 | -0.9 |
| 25 | 0.0364 | 0.0440 | 0.0665 | 0.1023 | 0.1233 | 0.1589 | 0.2152 | 0.3251 |
| 35 | 0.0393 | 0.0426 | 0.0543 | 0.0797 | 0.0952 | 0.1227 | 0.1628 | 0.2437 |
| 45 | 0.0419 | 0.0452 | 0.0550 | 0.0728 | 0.0854 | 0.1004 | 0.1268 | 0.1665 |
| 55 | 0.0448 | 0.0472 | 0.0547 | 0.0688 | 0.0756 | 0.0852 | 0.1022 | 0.1262 |

TABLE VIII
$\alpha^{S\left(P^{1}\right)^{(3)}}$

| $n$ |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $n$ | 1 | $2 / 3$ | 0 | -0.5 | -0.6 | -0.7 | -0.8 | -0.9 | -1.0 |
| 25 | 0.0662 | 0.0659 | 0.0665 | 0.0605 | 0.0598 | 0.0591 | 0.0574 | 0.0563 | 0.0551 |
| 35 | 0.0534 | 0.0540 | 0.0543 | 0.0523 | 0.0521 | 0.0523 | 0.0523 | 0.0516 | 0.0512 |
| 45 | 0.0518 | 0.0530 | 0.0550 | 0.0544 | 0.0545 | 0.0545 | 0.0549 | 0.0545 | 0.0546 |
| 55 | 0.0505 | 0.0516 | 0.0547 | 0.0580 | 0.0590 | 0.0601 | 0.0608 | 0.0613 | 0.0568 |

TABLE IX
$\alpha_{n}^{S\left(P^{0.75}\right)^{(3)}}$

|  |  |  |  |  | $\lambda$ |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $n$ | 1 | $2 / 3$ | 0 | -0.5 | -0.6 | -0.7 | -0.8 | -0.9 | -1.0 |
| 25 | 0.0516 | 0.0515 | $\mathbf{0 . 0 5 0 0}$ | 0.0478 | 0.0469 | 0.0469 | 0.0450 | 0.0440 | 0.0429 |
| 35 | 0.0479 | 0.0480 | 0.0474 | 0.0459 | 0.0456 | 0.0459 | 0.0457 | 0.0450 | 0.0447 |
| 45 | 0.0463 | 0.0480 | 0.0493 | 0.0489 | 0.0494 | 0.0495 | 0.0497 | 0.0494 | 0.0498 |
| 55 | 0.0477 | 0.0488 | 0.0520 | 0.0546 | 0.0555 | 0.0567 | 0.0575 | 0.0580 | 0.048 |

TABLE X
$\alpha^{S\left(P^{0.5}\right)^{(3)}}$

| $n$ |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 1 | $2 / 3$ | 0 | -0.5 | -0.6 | -0.7 | -0.8 | -0.9 | -1.0 |
| $n$ | 1 | 0.0372 | 0.0388 | 0.0350 | 0.0337 | 0.0338 | 0.0332 | 0.0322 | 0.0315 |
| 25 | 0.0364 | 0.0260 | 0.0191 |  |  |  |  |  |  |
| 35 | 0.0393 | 0.0402 | 0.0417 | 0.0413 | 0.0411 | 0.0409 | 0.0413 | 0.0407 | 0.0403 |
| 45 | 0.0419 | 0.0433 | 0.0457 | 0.0453 | 0.0458 | 0.0459 | 0.0463 | 0.0462 | 0.0463 |
| 55 | 0.0448 | 0.0460 | $\mathbf{0 . 0 5 0 0}$ | 0.0527 | 0.0537 | 0.0548 | 0.0554 | 0.0559 | 0.0561 |

TABLE XI
$\alpha_{n}^{\alpha_{n}^{S\left(P^{0.25}\right)^{(3)}}}$

|  |  |  |  |  | $\lambda$ |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $n$ | 1 | $2 / 3$ | 0 | -0.5 | -0.6 | -0.7 | -0.8 | -0.9 | -1.0 |
| 25 | 0.0225 | 0.0241 | 0.0270 | 0.0253 | 0.0251 | 0.0252 | 0.0246 | 0.0240 | 0.0235 |
| 35 | 0.0322 | 0.0328 | 0.0348 | 0.0349 | 0.0348 | 0.0348 | 0.0351 | 0.0347 | 0.0341 |
| 45 | 0.0370 | 0.0393 | 0.0428 | 0.0431 | 0.0435 | 0.0435 | 0.0442 | 0.0442 | 0.0445 |
| 55 | 0.0427 | 0.0441 | 0.0480 | 0.0515 | 0.0525 | 0.0535 | 0.0541 | 0.0543 | 0.0 .0139 |

TABLE XII
$\alpha_{n}^{S\left(P^{0}\right)^{(3)}}$

|  |  |  |  |  |  | $\lambda$ |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $n$ | 1 | $2 / 3$ | 0 | -0.5 | -0.6 | -0.7 | -0.8 | -0.9 | -1.0 | -1.5 |
| 25 | 0.0169 | 0.0191 | 0.0241 | 0.0239 | 0.0240 | 0.0250 | 0.0244 | 0.0238 | 0.0232 | 0.0187 |
| 35 | 0.0301 | 0.0312 | 0.0345 | 0.0356 | 0.0352 | 0.0356 | 0.0358 | 0.0357 | 0.0355 | 0.0343 |
| 45 | 0.0347 | 0.0372 | 0.0412 | 0.0415 | 0.0419 | 0.0423 | 0.0430 | 0.0434 | 0.0437 | 0.0443 |
| 55 | 0.0417 | 0.0433 | 0.0473 | 0.0512 | 0.0522 | 0.0536 | 0.0543 | 0.0548 | 0.0551 | 0.0560 |


[^0]:    Author is with the Department of Statistics, Dokuz Eylul University, 35160, Buca, Izmir, Turkey (phone: +90-232-4128600; fax: +90-2324534265; e-mail: aylin.alin@deu.edu.tr).

