

An expository supplement to the paper “A family of the information criteria using the phi-divergence for categorical data”

Haruhiko Ogasawara

This article supplements Ogasawara (2018) and gives some expository results for the saturated model under parametrization with $\hat{\boldsymbol{\theta}} = \mathbf{p}_{(K-1)}$, $\boldsymbol{\pi}_{0(K-1)} = (\pi_{01}, \dots, \pi_{0(K-1)})'$ and $\boldsymbol{\theta}_0 = \boldsymbol{\pi}_{0(K-1)}$. Additional numerical results are also shown in Tables S1 to S5.

$$\text{S.1} \quad \frac{\partial \boldsymbol{\pi}_0}{\partial \boldsymbol{\theta}_0'} \frac{\partial \boldsymbol{\theta}_0}{\partial \boldsymbol{\pi}_0'}$$

Define $\mathbf{e}_{(k)}$ as the $K \times 1$ vector, whose k -th element is 1 with the remaining elements being zero. Then, using $\hat{\pi}_K = p_K = 1 - \mathbf{1}_{(K-1)}' \hat{\boldsymbol{\theta}}$, we have $\frac{\partial \pi_{0k}}{\partial \boldsymbol{\theta}_0'} = \mathbf{e}_{(k)}' \ (k = 1, \dots, K-1)$ and $\frac{\partial \pi_{0K}}{\partial \boldsymbol{\theta}_0'} = -\mathbf{1}_{(K-1)}'$, which give

$$\begin{aligned} \frac{\partial \boldsymbol{\pi}_0}{\partial \boldsymbol{\theta}_0'} &= \begin{pmatrix} \mathbf{I}_{(K-1)} \\ -\mathbf{1}_{(K-1)}' \end{pmatrix} = (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})', \\ \frac{\partial \boldsymbol{\theta}_0}{\partial \boldsymbol{\pi}_0'} &= \left\{ \frac{\partial \boldsymbol{\pi}_0'}{\partial \boldsymbol{\theta}_0'} \text{diag}^{-1}(\boldsymbol{\pi}_0) \frac{\partial \boldsymbol{\pi}_0}{\partial \boldsymbol{\theta}_0'} \right\}^{-1} \frac{\partial \boldsymbol{\pi}_0'}{\partial \boldsymbol{\theta}_0'} \text{diag}^{-1}(\boldsymbol{\pi}_0) \end{aligned} \quad (\text{S.1})$$

$$\begin{aligned}
&= \left\{ (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)}) \text{diag}^{-1}(\boldsymbol{\pi}_0) \begin{pmatrix} \mathbf{I}_{(K-1)} \\ -\mathbf{1}_{(K-1)} \end{pmatrix} \right\}^{-1} (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)}) \text{diag}^{-1}(\boldsymbol{\pi}_0) \\
&= \{ \text{diag}^{-1}(\boldsymbol{\pi}_{0(K-1)}) + \pi_{0K}^{-1} \mathbf{1}_{(K-1)} \mathbf{1}_{(K-1)}' \}^{-1} \{ \text{diag}^{-1}(\boldsymbol{\pi}_{0(K-1)}) - \pi_{0K}^{-1} \mathbf{1}_{(K-1)} \} \\
&= [\text{diag}(\boldsymbol{\pi}_{0(K-1)}) - \text{diag}(\boldsymbol{\pi}_{0(K-1)}) \mathbf{1}_{(K-1)} \{ \pi_{0K} + \mathbf{1}_{(K-1)}' \text{diag}(\boldsymbol{\pi}_{0(K-1)}) \mathbf{1}_{(K-1)} \}^{-1} \\
&\quad \times \mathbf{1}_{(K-1)}' \text{diag}(\boldsymbol{\pi}_{0(K-1)})] \{ \text{diag}^{-1}(\boldsymbol{\pi}_{0(K-1)}), -\pi_{0K}^{-1} \mathbf{1}_{(K-1)} \} \\
&= \{ \text{diag}(\boldsymbol{\pi}_{0(K-1)}) - \boldsymbol{\pi}_{0(K-1)} \boldsymbol{\pi}_{0(K-1)}' \} \{ \text{diag}^{-1}(\boldsymbol{\pi}_{0(K-1)}), -\pi_{0K}^{-1} \mathbf{1}_{(K-1)} \} \\
&= [\mathbf{I}_{(K-1)} - \boldsymbol{\pi}_{0(K-1)} \mathbf{1}_{(K-1)}', \{ -\pi_{0K}^{-1} + (1 - \pi_{0K}) \pi_{0K}^{-1} \} \boldsymbol{\pi}_{0(K-1)}] \\
&= (\mathbf{I}_{(K-1)} - \boldsymbol{\pi}_{0(K-1)} \mathbf{1}_{(K-1)}', -\boldsymbol{\pi}_{0(K-1)}).
\end{aligned}$$

Consequently,

$$\begin{aligned}
\frac{\partial \boldsymbol{\pi}_0}{\partial \boldsymbol{\theta}_0'} \frac{\partial \boldsymbol{\theta}_0}{\partial \boldsymbol{\pi}_0'} &= \begin{pmatrix} \mathbf{I}_{(K-1)} \\ -\mathbf{1}_{(K-1)} \end{pmatrix} (\mathbf{I}_{(K-1)} - \boldsymbol{\pi}_{0(K-1)} \mathbf{1}_{(K-1)}', -\boldsymbol{\pi}_{0(K-1)}) \\
&= \begin{pmatrix} \mathbf{I}_{(K-1)} - \boldsymbol{\pi}_{0(K-1)} \mathbf{1}_{(K-1)}' & -\boldsymbol{\pi}_{0(K-1)} \\ -\pi_{0K} \mathbf{1}_{(K-1)}' & 1 - \pi_{0K} \end{pmatrix} \\
&= \mathbf{I}_{(K)} - \boldsymbol{\pi}_{0(K)} \mathbf{1}_{(K)}',
\end{aligned} \tag{S.2}$$

which is symmetric with respect to the K categories though $\hat{\boldsymbol{\theta}} = \mathbf{p}_{(K-1)}$ depends on the parametrization. The matrix of (S.2) is written as

$$\begin{aligned}
\frac{\partial \boldsymbol{\pi}_0}{\partial \boldsymbol{\theta}_0'} \frac{\partial \boldsymbol{\theta}_0}{\partial \boldsymbol{\pi}_0'} &= \begin{pmatrix} \mathbf{I}_{(K-1)} \\ -\mathbf{1}_{(K-1)} \end{pmatrix} \left\{ (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)}) \text{diag}^{-1}(\boldsymbol{\pi}_0) \begin{pmatrix} \mathbf{I}_{(K-1)} \\ -\mathbf{1}_{(K-1)} \end{pmatrix} \right\}^{-1} \\
&\quad \times (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)}) \text{diag}^{-1}(\boldsymbol{\pi}_0),
\end{aligned} \tag{S.3}$$

which is an idempotent asymmetric projection matrix onto the space spanned by the columns of $(\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})'$, whose rank is $K-1$. Note that (S.2) is written as

$$\left(\frac{\partial \boldsymbol{\pi}_0}{\partial \boldsymbol{\theta}_0'} \frac{\partial \boldsymbol{\theta}_0}{\partial \boldsymbol{\pi}_0'} \right)_{ij} = \delta_{ij} - \pi_{0i} \quad (i, j = 1, \dots, K), \quad (\text{S.4})$$

where $(\cdot)_{ij}$ is the (i, j) th element of a matrix.

S.2 $\partial^2 \boldsymbol{\theta}_0 / (\partial \boldsymbol{\pi}_0')^{<2>}$

Since $\boldsymbol{\pi}$ is linear with respect to $\boldsymbol{\theta}$ in the saturated model, (A.2) yields

$$\begin{aligned} & \frac{\partial^2 \boldsymbol{\theta}_0}{\partial \pi_{0i} \partial \pi_{0j}} \\ &= -\frac{1}{\phi_1''} \mathbf{I}_0^{-1} \left[\sum_{a=1}^K \left\{ -\frac{1}{\pi_{0a}^2} (\phi_1^{(3)} + 3\phi_1'') \frac{\partial \pi_{0a}}{\partial \boldsymbol{\theta}_0} \left(\frac{\partial \pi_{0a}}{\partial \boldsymbol{\theta}_0'} \right)^{<2>} \frac{\partial \boldsymbol{\theta}_0}{\partial \pi_{0i}} \otimes \frac{\partial \boldsymbol{\theta}_0}{\partial \pi_{0j}} \right\} \right. \\ & \quad \left. + \sum_{(i,j)}^2 \frac{1}{\pi_{0i}^2} (\phi_1^{(3)} + 2\phi_1'') \frac{\partial \pi_{0i}}{\partial \boldsymbol{\theta}_0} \frac{\partial \pi_{0i}}{\partial \boldsymbol{\theta}_0'} \frac{\partial \boldsymbol{\theta}_0}{\partial \pi_{0j}} - \frac{\delta_{ij}}{\pi_{0i}^2} (\phi_1^{(3)} + \phi_1'') \frac{\partial \pi_{0i}}{\partial \boldsymbol{\theta}_0} \right] \\ &= -\frac{1}{\phi_1''} \mathbf{I}_0^{-1} \left[\sum_{a=1}^K \left\{ -\frac{1}{\pi_{0a}^2} (\phi_1^{(3)} + 3\phi_1'') (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\cdot a} (\delta_{ai} - \pi_{0a}) (\delta_{aj} - \pi_{0a}) \right\} \right. \\ & \quad \left. + \sum_{(i,j)}^2 \frac{1}{\pi_{0i}^2} (\phi_1^{(3)} + 2\phi_1'') (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\cdot i} (\delta_{ij} - \pi_{0i}) \right. \\ & \quad \left. - \frac{\delta_{ij}}{\pi_{0i}^2} (\phi_1^{(3)} + \phi_1'') (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\cdot i} \right] \\ &= -\mathbf{I}_0^{-1} \left\{ \frac{\delta_{ij}}{\pi_{0i}^2} (-3 + 4 - 1) (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\cdot i} + \sum_{(i,j)}^2 \frac{1}{\pi_{0i}} (3 - 2) (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\cdot i} \right\} \\ &= -\{ \text{diag}(\boldsymbol{\pi}_{0(K-1)}) - \boldsymbol{\pi}_{0(K-1)} \boldsymbol{\pi}_{0(K-1)}' \} \sum_{(i,j)}^2 \frac{1}{\pi_{0i}} (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\cdot i} \\ &= -\sum_{(i,j)}^2 \{ \text{diag}(\boldsymbol{\pi}_{0(K-1)}) - \boldsymbol{\pi}_{0(K-1)} \boldsymbol{\pi}_{0(K-1)}', -\boldsymbol{\pi}_{0K} \boldsymbol{\pi}_{0(K-1)}' \}_{\cdot i} \frac{1}{\pi_{0i}} \quad (i, j = 1, \dots, K). \end{aligned} \quad (\text{S.5})$$

In the above expression, the terms with $\phi_1^{(3)}$ cancel as is expected since the

saturated model does not depend on ϕ_1 .

When $1 \leq i \leq K-1$ and $1 \leq j \leq K-1$,

$$\frac{\partial^2 \boldsymbol{\theta}_0}{\partial \pi_{0i} \partial \pi_{0j}} = -(\mathbf{I}_{(K-1)})_{\cdot i} - (\mathbf{I}_{(K-1)})_{\cdot j} + 2\boldsymbol{\pi}_{0(K-1)}; \quad (\text{S.6})$$

$$\text{when } 1 \leq i \leq K-1 \text{ and } j = K, \quad \frac{\partial^2 \boldsymbol{\theta}_0}{\partial \pi_{0i} \partial \pi_{0j}} = -(\mathbf{I}_{(K-1)})_{\cdot i} + 2\boldsymbol{\pi}_{0(K-1)};$$

$$\text{when } i = K \text{ and } 1 \leq j \leq K-1, \quad \frac{\partial^2 \boldsymbol{\theta}_0}{\partial \pi_{0i} \partial \pi_{0j}} = -(\mathbf{I}_{(K-1)})_{\cdot j} + 2\boldsymbol{\pi}_{0(K-1)};$$

$$\text{and when } i = K \text{ and } j = K, \quad \frac{\partial^2 \boldsymbol{\theta}_0}{\partial \pi_{0i} \partial \pi_{0j}} = 2\boldsymbol{\pi}_{0(K-1)},$$

which gives

$$\frac{\partial^2 \boldsymbol{\theta}_0}{\partial \pi_{0i} \partial \pi_{0j}} = -(\mathbf{I}_{(K-1)}, \mathbf{0})_{\cdot i} - (\mathbf{I}_{(K-1)}, \mathbf{0})_{\cdot j} + 2\boldsymbol{\pi}_{0(K-1)} \quad (i, j = 1, \dots, K). \quad (\text{S.7})$$

Consequently, we obtain

$$\begin{aligned} \frac{\partial \boldsymbol{\pi}_0}{\partial \boldsymbol{\theta}_0'} \frac{\partial^2 \boldsymbol{\theta}_0}{\partial \pi_{0i} \partial \pi_{0j}} &= \begin{pmatrix} \mathbf{I}_{(K-1)} \\ -\mathbf{1}_{(K-1)} \end{pmatrix} \{ -(\mathbf{I}_{(K-1)}, \mathbf{0})_{\cdot i} - (\mathbf{I}_{(K-1)}, \mathbf{0})_{\cdot j} + 2\boldsymbol{\pi}_{0(K-1)} \} \\ &= \begin{pmatrix} -(\mathbf{I}_{(K-1)}, \mathbf{0})_{\cdot i} - (\mathbf{I}_{(K-1)}, \mathbf{0})_{\cdot j} + 2\boldsymbol{\pi}_{0(K-1)} \\ (\mathbf{1}_{(K-1)}', \mathbf{0})_i + (\mathbf{1}_{(K-1)}', \mathbf{0})_j - 2(1 - \pi_{0K}) \end{pmatrix} \\ &= \begin{pmatrix} -(\mathbf{I}_{(K-1)}, \mathbf{0})_{\cdot i} - (\mathbf{I}_{(K-1)}, \mathbf{0})_{\cdot j} + 2\boldsymbol{\pi}_{0(K-1)} \\ -2 + (\mathbf{1}_{(K-1)}', \mathbf{0})_i + (\mathbf{1}_{(K-1)}', \mathbf{0})_j + 2\pi_{0K} \end{pmatrix} \quad (\text{S.8}) \\ &= \{ -(\mathbf{I}_{(K)})_{\cdot i} - (\mathbf{I}_{(K)})_{\cdot j} + 2\boldsymbol{\pi}_0 \} \quad (i, j = 1, \dots, K), \end{aligned}$$

where $(\cdot)_i$ is the i -th element of a vector. The last result of (S.8) is symmetric with respect to the K categories as is expected.

S.3 $\partial^3 \boldsymbol{\theta}_0 / (\partial \boldsymbol{\pi}_0')^{<3>}$

As in Subsection S.2, from (A.2) we have

$$\begin{aligned}
& \frac{\partial^3 \boldsymbol{\theta}_0}{\partial \pi_{0i} \partial \pi_{0j} \partial \pi_{0k}} \\
&= -\frac{1}{\phi_1''} \mathbf{I}_0^{-1} \left[\sum_{(A)}^K \left[\frac{1}{\pi_{0a}^3} (\phi_1^{(4)} + 8\phi_1^{(3)} + 12\phi_1'') \frac{\partial \pi_{0a}}{\partial \boldsymbol{\theta}_0} \left(\frac{\partial \pi_{0a}}{\partial \boldsymbol{\theta}_0'} \right)^{<3>} \right. \right. \\
& \quad \times \frac{\partial \boldsymbol{\theta}_0}{\partial \pi_{0i}} \otimes \frac{\partial \boldsymbol{\theta}_0}{\partial \pi_{0j}} \otimes \frac{\partial \boldsymbol{\theta}_0}{\partial \pi_{0k}} \\
& \quad + \sum_{(i,j,k)}^3 \left\{ -\frac{1}{\pi_{0a}^2} (\phi_1^{(3)} + 3\phi_1'') \frac{\partial \pi_{0a}}{\partial \boldsymbol{\theta}_0} \left(\frac{\partial \pi_{0a}}{\partial \boldsymbol{\theta}_0'} \right)^{<2>} \right\} \frac{\partial \boldsymbol{\theta}_0}{\partial \pi_{0i}} \otimes \frac{\partial^2 \boldsymbol{\theta}_0}{\partial \pi_{0j} \partial \pi_{0k}} \Bigg]_{(B)} \\
& \quad + \sum_{(i,j,k)}^3 \left\{ -\frac{1}{\pi_{0i}^3} (\phi_1^{(4)} + 6\phi_1^{(3)} + 6\phi_1'') \frac{\partial \pi_{0i}}{\partial \boldsymbol{\theta}_0} \left(\frac{\partial \pi_{0i}}{\partial \boldsymbol{\theta}_0'} \right)^{<2>} \frac{\partial \boldsymbol{\theta}_0}{\partial \pi_{0j}} \otimes \frac{\partial \boldsymbol{\theta}_0}{\partial \pi_{0k}} \right. \\
& \quad + \frac{1}{\pi_{0i}^2} (\phi_1^{(3)} + 2\phi_1'') \frac{\partial \pi_{0i}}{\partial \boldsymbol{\theta}_0} \frac{\partial \pi_{0i}}{\partial \boldsymbol{\theta}_0'} \frac{\partial^2 \boldsymbol{\theta}_0}{\partial \pi_{0j} \partial \pi_{0k}} \\
& \quad + \frac{\delta_{ij}}{\pi_{0i}^3} (\phi_1^{(4)} + 4\phi_1^{(3)} + 2\phi_1'') \frac{\partial \pi_{0i}}{\partial \boldsymbol{\theta}_0} \frac{\partial \pi_{0i}}{\partial \boldsymbol{\theta}_0'} \frac{\partial \boldsymbol{\theta}_0}{\partial \pi_{0k}} \\
& \quad \left. - \delta_{ijk} \frac{1}{\pi_{0i}^3} (\phi_1^{(4)} + 2\phi_1^{(3)}) \frac{\partial \pi_{0i}}{\partial \boldsymbol{\theta}_0} \right]_{(A)} \quad (i, j, k = 1, \dots, K).
\end{aligned} \tag{S.9}$$

The six terms in $\left[\cdot \right]_{(A)}^{(A)}$ of (S.9) become

$$\begin{aligned}
& (i) \sum_{a=1}^K \frac{1}{\pi_{0a}^3} (\phi_1^{(4)} + 8\phi_1^{(3)} + 12\phi_1'') \frac{\partial \pi_{0a}}{\partial \boldsymbol{\theta}_0} \left(\frac{\partial \pi_{0a}}{\partial \boldsymbol{\theta}_0'} \right)^{<3>} \frac{\partial \boldsymbol{\theta}_0}{\partial \pi_{0i}} \otimes \frac{\partial \boldsymbol{\theta}_0}{\partial \pi_{0j}} \otimes \frac{\partial \boldsymbol{\theta}_0}{\partial \pi_{0k}} \\
&= \sum_{a=1}^K \frac{1}{\pi_{0a}^3} (\phi_1^{(4)} + 8\phi_1^{(3)} + 12\phi_1'') (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\cdot a} \\
& \quad \times (\delta_{ai} - \pi_{0a})(\delta_{aj} - \pi_{0a})(\delta_{ak} - \pi_{0a})
\end{aligned} \tag{S.10}$$

$$\begin{aligned}
&= (\phi_1^{(4)} + 8\phi_1^{(3)} + 12\phi_1'') \left[\frac{\delta_{ijk}}{\pi_{0i}^3} (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\bullet i} \right. \\
&\quad + \sum_{(i,j,k)}^3 \left\{ -\frac{\delta_{ij}}{\pi_{0i}^2} (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\bullet i} + \frac{1}{\pi_{0i}} (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\bullet i} \right\} \\
&\quad \left. - \sum_{a=1}^K (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\bullet a} \right], \\
(ii) &\sum_{a=1}^K \sum_{(i,j,k)}^3 \left\{ -\frac{1}{\pi_{0a}^2} (\phi_1^{(3)} + 3\phi_1'') \frac{\partial \pi_{0a}}{\partial \theta_0} \left(\frac{\partial \pi_{0a}}{\partial \theta_0'} \right)^{<2>} \right\} \frac{\partial \theta_0}{\partial \pi_{0i}} \otimes \frac{\partial^2 \theta_0}{\partial \pi_{0j} \partial \pi_{0k}} \\
&= - \sum_{(i,j,k)}^3 \sum_{a=1}^K \frac{1}{\pi_{0a}^2} (\phi_1^{(3)} + 3\phi_1'') (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\bullet a} (\delta_{ai} - \pi_{0a}) \\
&\quad \times \{ -(\mathbf{I}_{(K-1)})_{aj} - (\mathbf{I}_{(K-1)})_{ak} + 2\pi_{0a} \} \\
&= (\phi_1^{(3)} + 3\phi_1'') \sum_{(i,j,k)}^3 \left\{ \frac{1}{\pi_{0i}^2} (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\bullet i} (\delta_{ij} + \delta_{ik}) - \frac{2}{\pi_{0i}} (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\bullet i} \right. \\
&\quad \left. - \frac{1}{\pi_{0j}} (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\bullet j} - \frac{1}{\pi_{0k}} (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\bullet k} + 2 \sum_{a=1}^K (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\bullet a} \right\}, \\
(iii) &- \sum_{(i,j,k)}^3 \frac{1}{\pi_{0i}^3} (\phi_1^{(4)} + 6\phi_1^{(3)} + 6\phi_1'') \frac{\partial \pi_{0i}}{\partial \theta_0} \left(\frac{\partial \pi_{0i}}{\partial \theta_0'} \right)^{<2>} \frac{\partial \theta_0}{\partial \pi_{0j}} \otimes \frac{\partial \theta_0}{\partial \pi_{0k}} \\
&= -(\phi_1^{(4)} + 6\phi_1^{(3)} + 6\phi_1'') \sum_{(i,j,k)}^3 \frac{1}{\pi_{0i}^3} (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\bullet i} (\delta_{ij} - \pi_{0i})(\delta_{ik} - \pi_{0i}) \\
&= (\phi_1^{(4)} + 6\phi_1^{(3)} + 6\phi_1'') \sum_{(i,j,k)}^3 \left\{ -\frac{\delta_{ijk}}{\pi_{0i}^3} (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\bullet i} \right. \\
&\quad \left. + (\delta_{ij} + \delta_{ik}) \frac{1}{\pi_{0i}^2} (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\bullet i} - \frac{1}{\pi_{0i}} (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\bullet i} \right\},
\end{aligned}$$

$$\begin{aligned}
& \text{(iv)} \sum_{(i,j,k)}^3 \frac{1}{\pi_{0i}^2} (\phi_1^{(3)} + 2\phi_1'') \frac{\partial \pi_{0i}}{\partial \boldsymbol{\theta}_0} \frac{\partial \pi_{0i}}{\partial \boldsymbol{\theta}_0'} \frac{\partial^2 \boldsymbol{\theta}_0}{\partial \pi_{0j} \partial \pi_{0k}} \\
&= (\phi_1^{(3)} + 2\phi_1'') \sum_{(i,j,k)}^3 \frac{1}{\pi_{0i}^2} (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\bullet i} \{ -(\mathbf{I}_{(K-1)})_{ij} - (\mathbf{I}_{(K-1)})_{ik} + 2\pi_{0i} \} \\
&= (\phi_1^{(3)} + 2\phi_1'') \sum_{(i,j,k)}^3 \left\{ -\frac{1}{\pi_{0i}^2} (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\bullet i} (\delta_{ij} + \delta_{ik}) + \frac{2}{\pi_{0i}} (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\bullet i} \right\}, \\
& \text{(v)} \sum_{(i,j,k)}^3 \frac{\delta_{ij}}{\pi_{0i}^3} (\phi_1^{(4)} + 4\phi_1^{(3)} + 2\phi_1'') \frac{\partial \pi_{0i}}{\partial \boldsymbol{\theta}_0} \frac{\partial \pi_{0i}}{\partial \boldsymbol{\theta}_0'} \frac{\partial \boldsymbol{\theta}_0}{\partial \pi_{0k}} \\
&= (\phi_1^{(4)} + 4\phi_1^{(3)} + 2\phi_1'') \sum_{(i,j,k)}^3 \frac{\delta_{ij}}{\pi_{0i}^3} (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\bullet i} (\delta_{ik} - \pi_{0i}) \\
&= (\phi_1^{(4)} + 4\phi_1^{(3)} + 2\phi_1'') \sum_{(i,j,k)}^3 \left(\frac{\delta_{ijk}}{\pi_{0i}^3} - \frac{\delta_{ij}}{\pi_{0i}^2} \right) (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\bullet i} \\
&= (\phi_1^{(4)} + 4\phi_1^{(3)} + 2\phi_1'') \left\{ \frac{3\delta_{ijk}}{\pi_{0i}^3} (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\bullet i} - \sum_{(i,j,k)}^3 \frac{\delta_{ij}}{\pi_{0i}^2} (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\bullet i} \right\} \\
& \text{and (vi)} -\delta_{ijk} \frac{1}{\pi_{0i}^3} (\phi_1^{(4)} + 2\phi_1^{(3)}) \frac{\partial \pi_{0i}}{\partial \boldsymbol{\theta}_0} = -(\phi_1^{(4)} + 2\phi_1^{(3)}) \frac{\delta_{ijk}}{\pi_{0i}^3} (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\bullet i}.
\end{aligned}$$

From (i) to (vi) of (S.10), we find that the results for $\phi_1^{(3)}$ and $\phi_1^{(4)}$ cancel as is expected. Then, we have

$$\begin{aligned}
& \frac{\partial^3 \boldsymbol{\theta}_0}{\partial \pi_{0i} \partial \pi_{0j} \partial \pi_{0k}} \\
&= -\mathbf{I}_0^{-1} \left[12 \left\{ \frac{\delta_{ijk}}{\pi_{0i}^3} (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\bullet i} + \sum_{(i,j,k)}^3 \left(-\frac{\delta_{ij}}{\pi_{0i}^2} + \frac{1}{\pi_{0i}} \right) (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\bullet i} \right\} \right. \\
&+ 3 \sum_{(i,j,k)}^3 \left(\frac{\delta_{ij} + \delta_{ik}}{\pi_{0i}^2} - \frac{4}{\pi_{0i}} \right) (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\bullet i} - 18 \frac{\delta_{ijk}}{\pi_{0i}^3} (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\bullet i} \\
&+ 6 \sum_{(i,j,k)}^3 \left(\frac{\delta_{ij} + \delta_{ik}}{\pi_{0i}^2} - \frac{1}{\pi_{0i}} \right) (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\bullet i} \\
&+ 2 \sum_{(i,j,k)}^3 \left(-\frac{\delta_{ij} + \delta_{ik}}{\pi_{0i}^2} + \frac{2}{\pi_{0i}} \right) (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\bullet i} \left. \right\} \quad \text{(S.11)}
\end{aligned}$$

$$\begin{aligned}
& +6 \frac{\delta_{ijk}}{\pi_{0i}^3} (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\bullet i} - 2 \sum_{(i,j,k)}^3 \frac{\delta_{ij}}{\pi_{0i}^2} (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\bullet i} \Big] \\
& = -\mathbf{I}_0^{-1} \sum_{(i,j,k)}^3 \left\{ -\frac{2}{\pi_{0i}} (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\bullet i} \right\} \\
& = 2 \{ \text{diag}(\boldsymbol{\pi}_{0(K-1)}) - \boldsymbol{\pi}_{0(K-1)} \boldsymbol{\pi}_{0(K-1)}' \} \sum_{(i,j,k)}^3 (\mathbf{I}_{(K-1)}, -\mathbf{1}_{(K-1)})_{\bullet i} \frac{1}{\pi_{0i}} \\
& = \sum_{(i,j,k)}^3 2 \{ \text{diag}(\boldsymbol{\pi}_{0(K-1)}) - \boldsymbol{\pi}_{0(K-1)} \boldsymbol{\pi}_{0(K-1)}', -\pi_{0K} \boldsymbol{\pi}_{0(K-1)} \}_{\bullet i} \frac{1}{\pi_{0i}} \\
& = \sum_{(i,j,k)}^3 (2\mathbf{I}_{(K-1)}, \mathbf{0})_{\bullet i} - 6\pi_{0(K-1)} \quad (i, j, k = 1, \dots, K)
\end{aligned}$$

Consequently, we obtain

$$\begin{aligned}
& \frac{\partial \boldsymbol{\pi}_0}{\partial \boldsymbol{\theta}_0'} \frac{\partial^3 \boldsymbol{\theta}_0}{\partial \pi_{0i} \partial \pi_{0j} \partial \pi_{0k}} \\
& = 2 \left(\begin{array}{c} \mathbf{I}_{(K-1)} \\ -\mathbf{1}_{(K-1)} \end{array} \right)_{(i,j,k)} \sum_{(i,j,k)}^2 \{ \text{diag}(\boldsymbol{\pi}_{0(K-1)}) - \boldsymbol{\pi}_{0(K-1)} \boldsymbol{\pi}_{0(K-1)}', -\pi_{0K} \boldsymbol{\pi}_{0(K-1)} \}_{\bullet i} \frac{1}{\pi_{0i}} \\
& = 2 \sum_{(i,j,k)}^3 \left[\begin{array}{cc} \text{diag}(\boldsymbol{\pi}_{0(K-1)}) - \boldsymbol{\pi}_{0(K-1)} \boldsymbol{\pi}_{0(K-1)}' & -\pi_{0K} \boldsymbol{\pi}_{0(K-1)} \\ -\pi_{0K} \boldsymbol{\pi}_{0(K-1)}' & \pi_{0K} - \pi_{0K}^2 \end{array} \right]_{\bullet i} \frac{1}{\pi_{0i}} \\
& = 2 \sum_{(i,j,k)}^3 \{ \text{diag}(\boldsymbol{\pi}_0) - \boldsymbol{\pi}_0 \boldsymbol{\pi}_0' \}_{\bullet i} \frac{1}{\pi_{0i}} \tag{S.12} \\
& = 2 \sum_{(i,j,k)}^3 (\mathbf{I}_{(K)})_{\bullet i} - 6\pi_0 \\
& = 2(\mathbf{I}_{(K)})_{\bullet i} + 2(\mathbf{I}_{(K)})_{\bullet j} + 2(\mathbf{I}_{(K)})_{\bullet k} - 6\pi_0 \quad (i, j, k = 1, \dots, K),
\end{aligned}$$

which is symmetric with respect to the K categories.

S.4 b_Δ

The final results of Subsections (S.1) to (S.3) are repeated as follows

$$\begin{aligned}
\frac{\partial \boldsymbol{\pi}_0}{\partial \boldsymbol{\theta}_0} \frac{\partial \boldsymbol{\theta}_0}{\partial \boldsymbol{\pi}_0} &= \mathbf{I}_{(K)} - \boldsymbol{\pi}_0 \mathbf{1}_{(K)}', \\
\frac{\partial \boldsymbol{\pi}_0}{\partial \boldsymbol{\theta}_0} \frac{\partial^2 \boldsymbol{\theta}_0}{\partial \pi_{0i} \partial \pi_{0j}} &= \{-(\mathbf{I}_{(K-1)})_{\cdot i} - (\mathbf{I}_{(K-1)})_{\cdot j} + 2\boldsymbol{\pi}_0\}, \\
\frac{\partial \boldsymbol{\pi}_0}{\partial \boldsymbol{\theta}_0} \frac{\partial^3 \boldsymbol{\theta}_0}{\partial \pi_{0i} \partial \pi_{0j} \partial \pi_{0k}} &= 2(\mathbf{I}_{(K)})_{\cdot i} + 2(\mathbf{I}_{(K)})_{\cdot j} + 2(\mathbf{I}_{(K)})_{\cdot k} - 6\boldsymbol{\pi}_0 \\
&\quad (i, j, k = 1, \dots, K).
\end{aligned} \tag{S.13}$$

Then, b_Δ in Theorem 1 becomes

$$\begin{aligned}
b_\Delta &= \frac{2}{\phi_2''} \sum_{k=1}^K \left[\underset{(A)}{-\frac{\phi_2''}{\pi_{0k}}} \left[\underset{(B)}{\frac{1}{2} \sum_{a,b=1}^K (-\delta_{ka} - \delta_{kb} + 2\pi_{0k}) \kappa_3(a, b, k)} \right. \right. \\
&\quad \left. \left. + \frac{1}{6} \sum_{a,b,c=1}^K \{2(\mathbf{I}_{(K)})_{ka} + 2(\mathbf{I}_{(K)})_{kb} + 2(\mathbf{I}_{(K)})_{kc} - 6\pi_{0k}\} m_4(a, b, c, k) \right] \right] \\
&\quad - \frac{1}{2\pi_{0k}^2} (\phi_2^{(3)} + \phi_2'') \left\{ \sum_{a=1}^K (\delta_{ka} - \pi_{0k}) \kappa_3(a, k, k) \right. \\
&\quad \left. + \frac{1}{2} \sum_{a,b=1}^K (-\delta_{ka} - \delta_{kb} + 2\pi_{0k}) m_4(a, b, k, k) \right\} \\
&\quad + \frac{1}{2\pi_{0k}^2} (\phi_2^{(3)} + 2\phi_2'') \sum_{a,b=1}^K \{(\delta_{ka} - \pi_{0k})(\delta_{kb} - \pi_{0k}) \kappa_3(a, b, k) \\
&\quad + \sum_{c=1}^K (\delta_{ka} - \pi_{0k})(-\delta_{kb} - \delta_{kc} + 2\pi_{0k}) m_4(a, b, c, k)\} \\
&\quad - \frac{1}{6\pi_{0k}^3} (\phi_2^{(4)} + 2\phi_2^{(3)}) \sum_{a,b=1}^K (\delta_{ka} - \pi_{0k}) m_4(a, k, k, k) \\
&\quad + \frac{1}{4\pi_{0k}^3} (\phi_2^{(4)} + 4\phi_2^{(3)} + 2\phi_2'') \sum_{a,b=1}^K (\delta_{ka} - \pi_{0k})(\delta_{kb} - \pi_{0k}) m_4(a, b, k, k)
\end{aligned} \tag{S.14}$$

$$-\frac{1}{6\pi_{0k}^3}(\phi_2^{(4)} + 6\phi_2^{(3)} + 6\phi_2'') \times \sum_{a,b,c=1}^K (\delta_{ka} - \pi_{0k})(\delta_{kb} - \pi_{0k})(\delta_{kc} - \pi_{0k}) m_4(a,b,c,k) \Bigg]_{(A)}.$$

In (S.14), noting that the terms with the factors $\sum_{a=1}^K \kappa_3(a,b,k)$, $\sum_{a,b=1}^K m_4(a,b,k,k)$ and similar ones vanish, we find that the remaining terms give b_Δ in (3.1) of Theorem 2 using (A.9).

Reference

Ogasawara, H. (2018). A family of the information criteria using the phi-divergence for categorical data. *Computational Statistics and Data Analysis*, 124, 87-103.

Table S1. Simulated and asymptotic biases of the power divergences (the number of replications = 10,000)

The MLEs ($\lambda = 0$) are used for all power divergences.										
The genetics of plants (Fisher, 1970; 4 categories)										
Model 1	S.B.	A.B.	H.A.B.	S. b_{Δ}	b_{Δ}	S.B.	A.B.	H.A.B.	S. b_{Δ}	b_{Δ}
	$n = 50$					$n = 200$				
$\lambda = 0$ (G^2)	-2.06	-2	-2.13	-2.9	-6.6	-2.05	-2	-2.03	-9.7	-6.6
$\lambda = 2/3$ (C-R)	-2.19	-2	-2.15	-9.3	-7.6	-2.08	-2	-2.04	-15.4	-7.6
$\lambda = 1$ (X^2)	-2.31	-2	-2.18	-15.5	-9.2	-2.10	-2	-2.05	-20.7	-9.2
$\lambda = 2$	-3.03	-2	-2.37	-51.3	-18.4	-2.24	-2	-2.09	-47.3	-18.4
E_g^2	-2.14	-2	-2.12	-7.0	-5.9	-2.07	-2	-2.03	-13.4	-5.9
	$n = 800$					(b_{Δ}^*)				
$\lambda = 0$ (G^2)	-1.91	-2	-2.01	74.4	-6.6	(-7.0)				
$\lambda = 2/3$ (C-R)	-1.92	-2	-2.01	67.4	-7.6	(-10.3)				
$\lambda = 1$ (X^2)	-1.92	-2	-2.01	61.8	-9.2	(-13.0)				
$\lambda = 2$	-1.96	-2	-2.02	35.8	-18.4	(-25.0)				
E_g^2	-1.91	-2	-2.01	68.4	-5.9	(-10.0)				
Model 2	$n = 50$					$n = 200$				
$\lambda = 0$ (G^2)	-4.17	-4	-4.24	-8.4	-12.0	-4.07	-4	-4.06	-13.2	-12.0
$\lambda = 2/3$ (C-R)	-4.49	-4	-4.35	-24.7	-17.3	-4.14	-4	-4.09	-27.3	-17.3
$\lambda = 1$ (X^2)	-4.81	-4	-4.44	-40.3	-22.0	-4.20	-4	-4.11	-39.2	-22.0
$\lambda = 2$	-6.84	-4	-4.89	-142.2	-44.3	-4.49	-4	-4.22	-98.4	-44.3
E_g^2	-4.42	-4	-4.32	-20.9	-15.8	-4.12	-4	-4.08	-24.8	-15.8
	$n = 800$					(b_{Δ}^*)				
$\lambda = 0$ (G^2)	-3.95	-4	-4.02	42.4	-12.0	(-14.0)				
$\lambda = 2/3$ (C-R)	-3.97	-4	-4.02	26.5	-17.3	(-20.7)				
$\lambda = 1$ (X^2)	-3.98	-4	-4.03	14.3	-22.0	(-26.0)				
$\lambda = 2$	-4.05	-4	-4.06	-40.0	-44.3	(-50.0)				
E_g^2	-3.97	-4	-4.02	27.0	-15.8	(-20.0)				

Table S1. (continued)

The MLEs ($\lambda = 0$) are used for all power divergences.											
The genetics of plants (Fisher, 1970; 4 categories)											
Model 3	S.B.	A.B.	H.A.B.	$S.b_{\Delta}$	b_{Δ}	S.B.	A.B.	H.A.B.	$S.b_{\Delta}$	b_{Δ}	
	$n = 50$					$n = 200$					
$\lambda = 0 (G^2)$	-6.39	-6	-6.42	-19.7	-21.0	-6.12	-6	-6.11	-24.0	-21.0	
$\lambda = 2/3$ (C-R)	-6.99	-6	-6.62	-49.5	-31.0	-6.24	-6	-6.16	-47.2	-31.0	
$\lambda = 1 (X^2)$	-7.54	-6	-6.78	-77.2	-39.0	-6.33	-6	-6.20	-66.2	-39.0	
$\lambda = 2$	-11.26	-6	-7.50	-263.0	-75.0	-6.79	-6	-6.38	-158.0	-75.0	
E_g^2	-6.88	-6	-6.60	-44.1	-30.0	-6.22	-6	-6.15	-44.8	-30.0	
	$n = 800$					$(b_{\Delta}^* = b_{\Delta})$					
$\lambda = 0 (G^2)$	-5.96	-6	-6.03	29.7	-21.0	(-21.0)					
$\lambda = 2/3$ (C-R)	-6.00	-6	-6.04	3.4	-31.0	(-31.0)					
$\lambda = 1 (X^2)$	-6.02	-6	-6.05	-16.0	-39.0	(-39.0)					
$\lambda = 2$	-6.13	-6	-6.09	-100.7	-75.0	(-75.0)					
E_g^2	-6.00	-6	-6.04	2.8	-30.0	(-30.0)					

Note. n = the number of observations, S.B. = simulated bias, A.B. = asymptotic bias = $-2q$, H.A.B. = $b + n^{-1}b_{\Delta} = -2q + n^{-1}b_{\Delta}$, $S.b_{\Delta}$ = simulated $b_{\Delta} = n(S.B. + 2q)$, G^2 = the log-likelihood ratio statistic, C-R = the Cressie-Read statistic, X^2 = Pearson's statistic, E_g^2 = Eguchi's divergence. The number for model identification is the number of independent parameters.

Table S2. Simulated and asymptotic biases of the power divergences (the number of replications = 10,000)

The parameter estimators by $\lambda = 1 (X^2)$ are used for all power divergences.										
The genetics of plants (Fisher, 1970; 4 categories)										
Model 1	S.B.	A.B.	H.A.B.	S. b_Δ	b_Δ	S.B.	A.B.	H.A.B.	S. b_Δ	b_Δ
	$n = 50$					$n = 200$				
$\lambda = 0 (G^2)$	-2.31	-2	-2.08	-15.3	-4.1	-2.05	-2	-2.02	-9.9	-4.1
$\lambda = 2/3 (C-R)$	-2.48	-2	-2.14	-24.0	-6.8	-2.09	-2	-2.03	-17.2	-6.8
$\lambda = 1 (X^2)$	-2.64	-2	-2.19	-31.9	-9.3	-2.12	-2	-2.05	-23.2	-9.3
$\lambda = 2$	-3.56	-2	-2.42	-78.1	-21.2	-2.26	-2	-2.11	-51.7	-21.2
E_g^2	-2.44	-2	-2.12	-21.9	-6.0	-2.08	-2	-2.03	-16.2	-6.0
	$n = 800$					(b_Δ^*)				
$\lambda = 0 (G^2)$	-2.00	-2	-2.01	2.6	-4.1	(-7.0)				
$\lambda = 2/3 (C-R)$	-2.01	-2	-2.01	-4.8	-6.8	(-10.3)				
$\lambda = 1 (X^2)$	-2.01	-2	-2.01	-10.8	-9.3	(-13.0)				
$\lambda = 2$	-2.05	-2	-2.03	-38.1	-21.2	(-25.0)				
E_g^2	-2.00	-2	-2.01	-3.9	-6.0	(-10.0)				
Model 2	$n = 50$					$n = 200$				
$\lambda = 0 (G^2)$	-3.77	-4	-4.19	11.5	-9.7	-4.02	-4	-4.05	-3.9	-9.7
$\lambda = 2/3 (C-R)$	-3.85	-4	-4.34	7.3	-17.2	-4.08	-4	-4.09	-15.5	-17.2
$\lambda = 1 (X^2)$	-3.97	-4	-4.46	1.7	-23.0	-4.13	-4	-4.12	-25.7	-23.0
$\lambda = 2$	-4.64	-4	-4.97	-31.9	-48.7	-4.38	-4	-4.24	-75.3	-48.7
E_g^2	-3.75	-4	-4.34	12.4	-16.8	-4.06	-4	-4.08	-12.9	-16.8
	$n = 800$					(b_Δ^*)				
$\lambda = 0 (G^2)$	-4.03	-4	-4.01	-27.0	-9.7	(-14.0)				
$\lambda = 2/3 (C-R)$	-4.05	-4	-4.02	-42.8	-17.2	(-20.7)				
$\lambda = 1 (X^2)$	-4.07	-4	-4.03	-55.1	-23.0	(-26.0)				
$\lambda = 2$	-4.14	-4	-4.06	-110.3	-48.7	(-50.0)				
E_g^2	-4.05	-4	-4.02	-42.0	-16.8	(-20.0)				

Table S2. (continued)

The parameter estimators by $\lambda = 1$ (χ^2) are used for all power divergences.											
The genetics of plants (Fisher, 1970; 4 categories)											
Model 3	S.B.	A.B.	H.A.B.	S. b_Δ	b_Δ	S.B.	A.B.	H.A.B.	S. b_Δ	b_Δ	
	$n = 50$					$n = 200$					
$\lambda = 0$ (G^2)	-6.08	-6	-6.42	-4.0	-21.0	-6.04	-6	-6.11	-7.7	-21.0	
$\lambda = 2/3$ (C-R)	-6.43	-6	-6.62	-21.3	-31.0	-6.13	-6	-6.16	-25.8	-31.0	
$\lambda = 1$ (χ^2)	-6.78	-6	-6.78	-39.2	-39.0	-6.21	-6	-6.20	-41.5	-39.0	
$\lambda = 2$	-9.18	-6	-7.50	-159.1	-75.0	-6.59	-6	-6.38	-118.0	-75.0	
E_g^2	-6.28	-6	-6.60	-14.1	-30.0	-6.11	-6	-6.15	-22.1	-30.0	
	$n = 800$					$(b_\Delta^* = b_\Delta)$					
$\lambda = 0$ (G^2)	-6.08	-6	-6.03	-66.6	-21.0	(-21.0)					
$\lambda = 2/3$ (C-R)	-6.11	-6	-6.04	-88.1	-31.0	(-31.0)					
$\lambda = 1$ (χ^2)	-6.13	-6	-6.05	-105.4	-39.0	(-39.0)					
$\lambda = 2$	-6.23	-6	-6.09	-183.7	-75.0	(-75.0)					
E_g^2	-6.11	-6	-6.04	-86.1	-30.0	(-30.0)					

Note. n = the number of observations, S.B. = simulated bias, A.B. = asymptotic bias = $-2q$, H.A.B. = $b + n^{-1}b_\Delta = -2q + n^{-1}b_\Delta$, S. b_Δ = simulated $b_\Delta = n(\text{S.B.} + 2q)$, G^2 = the log-likelihood ratio statistic, C-R = the Cressie-Read statistic, χ^2 = Pearson's statistic, E_g^2 = Eguchi's divergence. The number for model identification is the number of independent parameters.

Table S3. Simulated and asymptotic biases of the power divergences (the number of replications = 10,000)

The parameter estimators by E_g^2 (Eguchi) are used for all power divergences.										
The genetics of plants (Fisher, 1970; 4 categories)										
Model 1	S.B.	A.B.	H.A.B.	S. b_A	b_A	S.B.	A.B.	H.A.B.	S. b_A	b_A
	$n = 50$					$n = 200$				
$\lambda = 0 (G^2)$	-2.24	-2	-2.16	-12.0	-7.9	-2.04	-2	-2.04	-8.9	-7.9
$\lambda = 2/3 (C-R)$	-2.44	-2	-2.16	-21.9	-8.0	-2.07	-2	-2.04	-14.9	-8.0
$\lambda = 1 (X^2)$	-2.60	-2	-2.18	-30.0	-9.1	-2.10	-2	-2.05	-20.2	-9.1
$\lambda = 2$	-3.45	-2	-2.34	-72.3	-17.0	-2.23	-2	-2.09	-46.5	-17.0
E_g^2	-2.43	-2	-2.12	-21.4	-5.8	-2.07	-2	-2.03	-13.2	-5.8
	$n = 800$					(b_A^*)				
$\lambda = 0 (G^2)$	-2.00	-2	-2.01	-2.4	-7.9	(-7.0)				
$\lambda = 2/3 (C-R)$	-2.01	-2	-2.01	-8.3	-8.0	(-10.3)				
$\lambda = 1 (X^2)$	-2.02	-2	-2.01	-13.5	-9.1	(-13.0)				
$\lambda = 2$	-2.05	-2	-2.02	-38.1	-17.0	(-25.0)				
E_g^2	-2.01	-2	-2.01	-6.8	-5.8	(-10.0)				
Model 2	$n = 50$					$n = 200$				
$\lambda = 0 (G^2)$	-3.68	-4	-4.26	15.8	-13.2	-3.98	-4	-4.07	3.5	-13.2
$\lambda = 2/3 (C-R)$	-3.78	-4	-4.35	11.2	-17.4	-4.02	-4	-4.09	-4.1	-17.4
$\lambda = 1 (X^2)$	-3.89	-4	-4.43	5.7	-21.5	-4.06	-4	-4.11	-11.9	-21.5
$\lambda = 2$	-4.52	-4	-4.84	-25.9	-42.2	-4.26	-4	-4.21	-52.8	-42.2
E_g^2	-3.69	-4	-4.31	15.3	-15.3	-4.00	-4	-4.08	0.2	-15.3
	$n = 800$					(b_A^*)				
$\lambda = 0 (G^2)$	-3.95	-4	-4.02	40.5	-13.2	(-14.0)				
$\lambda = 2/3 (C-R)$	-3.97	-4	-4.02	26.4	-17.4	(-20.7)				
$\lambda = 1 (X^2)$	-3.98	-4	-4.03	15.1	-21.5	(-26.0)				
$\lambda = 2$	-4.05	-4	-4.05	-36.3	-42.2	(-50.0)				
E_g^2	-3.97	-4	-4.02	27.7	-15.3	(-20.0)				

Table S3. (continued)

The parameter estimators by E_g^2 (Eguchi) are used for all power divergences.										
The genetics of plants (Fisher, 1970; 4 categories)										
Model 3	S.B.	A.B.	H.A.B.	S. b_Δ	b_Δ	S.B.	A.B.	H.A.B.	S. b_Δ	b_Δ
	$n = 50$					$n = 200$				
$\lambda = 0 (G^2)$	-6.03	-6	-6.42	-1.5	-21.0	-6.06	-6	-6.11	-13.0	-21.0
$\lambda = 2/3$ (C-R)	-6.43	-6	-6.62	-21.6	-31.0	-6.15	-6	-6.16	-29.8	-31.0
$\lambda = 1 (X^2)$	-6.84	-6	-6.78	-42.0	-39.0	-6.22	-6	-6.20	-44.7	-39.0
$\lambda = 2$	-9.75	-6	-7.50	-187.4	-75.0	-6.59	-6	-6.38	-118.7	-75.0
E_g^2	-6.30	-6	-6.60	-14.8	-30.0	-6.13	-6	-6.15	-25.4	-30.0
	$n = 800$					$(b_\Delta^* = b_\Delta)$				
$\lambda = 0 (G^2)$	-5.91	-6	-6.03	75.2	-21.0	(-21.0)				
$\lambda = 2/3$ (C-R)	-5.93	-6	-6.04	54.3	-31.0	(-31.0)				
$\lambda = 1 (X^2)$	-5.95	-6	-6.05	37.6	-39.0	(-39.0)				
$\lambda = 2$	-6.05	-6	-6.09	-37.7	-75.0	(-75.0)				
E_g^2	-5.93	-6	-6.04	56.0	-30.0	(-30.0)				

Note. n = the number of observations, S.B. = simulated bias, A.B. = asymptotic bias = $-2q$, H.A.B. = $b + n^{-1}b_\Delta = -2q + n^{-1}b_\Delta$, S. b_Δ = simulated $b_\Delta = n(\text{S.B.} + 2q)$, G^2 = the log-likelihood ratio statistic, C-R = the Cressie-Read statistic, X^2 = Pearson's statistic, E_g^2 = Eguchi's divergence. The number for model identification is the number of independent parameters.

Table S4. Simulated and asymptotic biases of the power divergences (the number of replications = 10,000)

The parameter estimators by E_g^2 (Eguchi) are used for all power divergences.										
3-category truncated Poisson variate (Bishop et al., 1975, p.503)										
Model 1	S.B.	A.B.	H.A.B.	S. b_Δ	b_Δ	S.B.	A.B.	H.A.B.	S. b_Δ	b_Δ
	$n = 50$					$n = 200$				
$\lambda = 0$ (G^2)	-2.00	-2	-2.01	0.1	-0.6	-2.01	-2	-2.00	-1.2	-0.6
$\lambda = 2/3$ (C-R)	-2.03	-2	-2.02	-1.6	-0.9	-2.01	-2	-2.00	-2.3	-0.9
$\lambda = 1$ (χ^2)	-2.07	-2	-2.03	-3.3	-1.4	-2.02	-2	-2.01	-3.7	-1.4
$\lambda = 2$	-2.25	-2	-2.09	-12.4	-4.6	-2.06	-2	-2.02	-11.3	-4.6
E_g^2	-2.02	-2	-2.01	-0.8	-0.3	-2.01	-2	-2.00	-1.3	-0.3
	$n = 800$					(b_Δ^*)				
$\lambda = 0$ (G^2)	-1.99	-2	-2.00	4.3	-0.6	(-3.1)				
$\lambda = 2/3$ (C-R)	-2.00	-2	-2.00	2.9	-0.9	(-4.3)				
$\lambda = 1$ (χ^2)	-2.00	-2	-2.00	1.4	-1.4	(-5.2)				
$\lambda = 2$	-2.01	-2	-2.01	-6.3	-4.6	(-9.4)				
E_g^2	-2.00	-2	-2.00	3.8	-0.3	(-4.2)				
Model 2	$n = 50$					$n = 200$				
$\lambda = 0$ (G^2)	-4.17	-4	-4.12	-8.3	-6.2	-4.03	-4	-4.03	-6.1	-6.2
$\lambda = 2/3$ (C-R)	-4.29	-4	-4.17	-14.7	-8.6	-4.06	-4	-4.04	-11.3	-8.6
$\lambda = 1$ (χ^2)	-4.40	-4	-4.21	-19.9	-10.4	-4.08	-4	-4.05	-15.5	-10.4
$\lambda = 2$	-4.91	-4	-4.38	-45.6	-18.9	-4.17	-4	-4.09	-34.5	-18.9
E_g^2	-4.28	-4	-4.17	-14.2	-8.3	-4.05	-4	-4.04	-10.9	-8.3
	$n = 800$					(b_Δ^*)				
$\lambda = 0$ (G^2)	-4.01	-4	-4.01	-4.2	-6.2	(-6.2)				
$\lambda = 2/3$ (C-R)	-4.01	-4	-4.01	-8.1	-8.6	(-8.6)				
$\lambda = 1$ (χ^2)	-4.01	-4	-4.01	-11.5	-10.4	(-10.4)				
$\lambda = 2$	-4.03	-4	-4.02	-27.4	-18.9	(-18.9)				
E_g^2	-4.01	-4	-4.01	-7.3	-8.3	(-8.3)				

Note. n = the number of observations, S.B. = simulated bias, A.B. = asymptotic bias = $-2q$, H.A.B. = $b + n^{-1}b_\Delta = -2q + n^{-1}b_\Delta$, S. b_Δ = simulated $b_\Delta = n(\text{S.B.} + 2q)$, G^2 = the log-likelihood ratio statistic, C-R = the Cressie-Read statistic, χ^2 = Pearson's statistic, E_g^2 = Eguchi's divergence. The number for model identification is the number of independent parameters.

Table S5. Simulated and asymptotic biases of the power divergences (the number of replications = 10,000)

The parameter estimators by E_g^2 (Eguchi) are used for all power divergences.										
4-category truncated Poisson variate										
Model 1	S.B.	A.B.	H.A.B.	S. b_A	b_A	S.B.	A.B.	H.A.B.	S. b_A	b_A
	$n = 50$					$n = 200$				
$\lambda = 0 (G^2)$	-2.06	-2	-2.00	-2.8	-0.1	-2.00	-2	-2.00	-0.7	-0.1
$\lambda = 2/3 (C-R)$	-2.11	-2	-2.00	-5.5	0.1	-2.01	-2	-2.00	-2.6	0.1
$\lambda = 1 (X^2)$	-2.17	-2	-2.01	-8.3	-0.4	-2.02	-2	-2.00	-4.9	-0.4
$\lambda = 2$	-2.45	-2	-2.09	-22.6	-4.4	-2.08	-2	-2.02	-16.8	-4.4
E_g^2	-2.09	-2	-1.97	-4.3	1.5	-2.01	-2	-1.99	-1.0	1.5
	$n = 800$					(b_A^*)				
$\lambda = 0 (G^2)$	-1.93	-2	-2.00	58.3	-0.1	(-4.2)				
$\lambda = 2/3 (C-R)$	-1.93	-2	-2.00	55.5	0.1	(-6.0)				
$\lambda = 1 (X^2)$	-1.93	-2	-2.00	53.0	-0.4	(-7.5)				
$\lambda = 2$	-1.95	-2	-2.01	40.5	-4.4	(-13.9)				
E_g^2	-1.93	-2	-2.00	56.5	1.5	(-5.8)				
Model 2	$n = 50$					$n = 200$				
$\lambda = 0 (G^2)$	-3.89	-4	-4.14	5.6	-7.0	-4.02	-4	-4.03	-3.3	-7.0
$\lambda = 2/3 (C-R)$	-3.99	-4	-4.19	0.4	-9.6	-4.05	-4	-4.05	-9.9	-9.6
$\lambda = 1 (X^2)$	-4.09	-4	-4.24	-4.6	-12.1	-4.08	-4	-4.06	-15.6	-12.1
$\lambda = 2$	-4.62	-4	-4.48	-31.0	-24.2	-4.22	-4	-4.12	-43.6	-24.2
E_g^2	-3.95	-4	-4.17	2.6	-8.5	-4.04	-4	-4.04	-8.2	-8.5
	$n = 800$					(b_A^*)				
$\lambda = 0 (G^2)$	-3.90	-4	-4.01	78.1	-7.0	(-8.5)				
$\lambda = 2/3 (C-R)$	-3.91	-4	-4.01	71.1	-9.6	(-12.0)				
$\lambda = 1 (X^2)$	-3.92	-4	-4.02	65.4	-12.1	(-14.9)				
$\lambda = 2$	-3.95	-4	-4.03	38.9	-24.2	(-27.8)				
E_g^2	-3.91	-4	-4.01	72.2	-8.5	(-11.7)				

Table S5. (continued)

The parameter estimators by E_g^2 (Eguchi) are used for all power divergences.										
4-category truncated Poisson variate										
Model 3	S.B.	A.B.	H.A.B.	S. b_Δ	b_Δ	S.B.	A.B.	H.A.B.	S. b_Δ	b_Δ
	$n = 50$					$n = 200$				
$\lambda = 0$ (G^2)	-6.01	-6	-6.25	-0.4	-12.7	-6.00	-6	-6.06	0.03	-12.7
$\lambda = 2/3$ (C-R)	-6.22	-6	-6.36	-10.8	-18.1	-6.05	-6	-6.09	-10.3	-18.1
$\lambda = 1$ (X^2)	-6.40	-6	-6.45	-20.2	-22.4	-6.09	-6	-6.11	-19.0	-22.4
$\lambda = 2$	-7.41	-6	-6.83	-70.6	-41.7	-6.30	-6	-6.21	-59.6	-41.7
E_g^2	-6.17	-6	-6.35	-8.4	-17.5	-6.04	-6	-6.09	-8.8	-17.5
	$n = 800$					$(b_\Delta^* = b_\Delta)$				
$\lambda = 0$ (G^2)	-5.92	-6	-6.02	64.1	-12.7	(-12.7)				
$\lambda = 2/3$ (C-R)	-5.93	-6	-6.02	52.1	-18.1	(-18.1)				
$\lambda = 1$ (X^2)	-5.95	-6	-6.03	43.0	-22.4	(-22.4)				
$\lambda = 2$	-6.00	-6	-6.05	2.6	-41.7	(-41.7)				
E_g^2	-5.93	-6	-6.02	52.5	-17.5	(-17.5)				

Note. n = the number of observations, S.B. = simulated bias, A.B. = asymptotic bias $= -2q$, H.A.B. $= b + n^{-1}b_\Delta = -2q + n^{-1}b_\Delta$, S. b_Δ = simulated $b_\Delta = n(\text{S.B.} + 2q)$, G^2 = the log-likelihood ratio statistic, C-R = the Cressie-Read statistic, X^2 = Pearson's statistic, E_g^2 = Eguchi's divergence. The number for model identification is the number of independent parameters.