Economic Review (Otaru University of Commerce),
Vol.65, No.2 & 3, 3-16, December, 2014.

Supplement to the paper "Asymptotic cumulants of the estimator of the canonical parameter in the exponential family"

Haruhiko Ogasawara

This article is to supplement Ogasawara (2013) with proofs, examples and a correction.

A.1 Proof of Theorem 1

From (4.1),

$$\kappa_1(\hat{\theta}_{\mathrm{ML}} - \theta_0) = -\frac{n^{-1}}{2} i_0^{-3} i_0^{(\mathrm{D1})} i_0 + O(n^{-2}) = -\frac{n^{-1}}{2} i_0^{-2} i_0^{(\mathrm{D1})} + O(n^{-2})$$

$$= -\frac{n^{-1}}{2} \{\mathrm{var}(s)\}^{-1/2} \mathrm{sk}(s) + O(n^{-2}),$$

$$\kappa_2(\hat{\theta}_{\mathrm{ML}}) = n^{-1} i_0^{-1} + n^{-2} \left[\frac{1}{2} i_0^{-6} (i_0^{(\mathrm{D1})})^2 i_0^2 - i_0^{-4} i_0^{(\mathrm{D1})} \mathrm{E}\{(s-\eta_0)^3\} \right.$$

$$\left. + 2\left\{ \frac{1}{2} i_0^{-5} (i_0^{(\mathrm{D1})})^2 - \frac{1}{6} i_0^{-4} i_0^{(\mathrm{D2})} \right\} i_0^{-1} \mathrm{E}\{(s-\eta_0)^4\} \right] + O(n^{-3})$$

$$= n^{-1} i_0^{-1} + n^{-2} \left\{ \frac{5}{2} i_0^{-4} (i_0^{(\mathrm{D1})})^2 - i_0^{-3} i_0^{(\mathrm{D2})} \right\} + O(n^{-3}),$$

$$= n^{-1} \{\mathrm{var}(s)\}^{-1} + n^{-2} \{\mathrm{var}(s)\}^{-1} \left[\frac{5}{2} \{\mathrm{sk}(s)\}^2 - \mathrm{kt}(s) \right] + O(n^{-3})$$

$$\kappa_3(\hat{\theta}_{\mathrm{ML}}) = n^{-2}\left[i_0^{-3}\mathrm{E}\{(s-\eta_0)^3\} - \frac{3}{2}(i_0^{-1})^2 i_0^{-3} i_0^{(\mathrm{D1})}\mathrm{E}\{(s-\eta_0)^4\} - 3\alpha_{\mathrm{ML1}}\alpha_{\mathrm{ML2}} \right]$$

$$+O(n^{-3})$$

$$= n^{-2}\left(1 - \frac{9}{2} + \frac{3}{2}\right) i_0^{-3} i_0^{(\mathrm{D1})} + O(n^{-3})$$

$$= -n^{-2} 2\{\mathrm{var}(s)\}^{-3/2}\mathrm{sk}(s) + O(n^{-3}),$$

$$\kappa_4(\hat{\theta}_{\mathrm{ML}}) = n^{-3}\left[(i_0^{-1})^4 \kappa_4(s) + 4(i_0^{-1})^3 \left\{ -\frac{1}{2} i_0^{-3} i_0^{(\mathrm{D1})}\mathrm{E}\{(s-\eta_0)^5\} \right\} \right.$$

$$+ 6(i_0^{-1})^2 \left\{ -\frac{1}{2} i_0^{-3} i_0^{(\mathrm{D1})} \right\}^2 \mathrm{E}\{(s-\eta_0)^6\}$$

$$+ 4(i_0^{-1})^3 \left\{ \frac{1}{2} i_0^{-5} (i_0^{(\mathrm{D1})})^2 - \frac{1}{6} i_0^{-4} i_0^{(\mathrm{D2})} \right\} \mathrm{E}\{(s-\eta_0)^6\}$$

$$\left. - 4\alpha_{\mathrm{ML1}}\alpha_{\mathrm{ML3}} - 6\alpha_{\mathrm{ML2}}\alpha_{\mathrm{ML\Delta 2}} - 6\alpha_{\mathrm{ML2}}\alpha_{\mathrm{ML1}}^2 \right] + O(n^{-4})$$

$$= n^{-3}\left[(i_0^{-1})^4 i_0^{(\mathrm{D2})} - 20 i_0^{-6} i_0^{(\mathrm{D1})} (i_0 i_0^{(\mathrm{D1})}) + \frac{45}{2} (i_0^{-1})^8 (i_0^{(\mathrm{D1})})^2 i_0^3 \right.$$

$$+ (30 i_0^{-8} (i_0^{(\mathrm{D1})})^2 - 10 i_0^{-7} i_0^{(\mathrm{D2})}) i_0^3 - 4\left(-\frac{1}{2} i_0^{-2} i_0^{(\mathrm{D1})} \right)(-2 i_0^{-3} i_0^{(\mathrm{D1})})$$

$$\left. - 6 i_0^{-1} \left\{ \frac{5}{2} i_0^{-4} (i_0^{(\mathrm{D1})})^2 - i_0^{-3} i_0^{(\mathrm{D2})} \right\} - 6 i_0^{-1} \left(-\frac{1}{2} i_0^{-2} i_0^{(\mathrm{D1})} \right)^2 \right] + O(n^{-4})$$

$$= n^{-3}\{12 i_0^{-5} (i_0^{(\mathrm{D1})})^2 - 3 i_0^{-4} i_0^{(\mathrm{D2})}\} + O(n^{-4})$$

$$= n^{-3}\{\mathrm{var}(s)\}^{-2}[12\{\mathrm{sk}(s)\}^2 - 3\mathrm{kt}(s)] + O(n^{-4}).$$

The above results are also given from (3.1) to (3.4).

A.2 Proof of Theorem 2

Expand $\hat{i}^{1/2}$ about $i_0^{1/2}$ as

$$\hat{i}^{1/2} = i_0^{1/2} + \frac{1}{2} i_0^{-1/2} i_0^{(\mathrm{D1})} (\hat{\theta}_{\mathrm{ML}} - \theta_0) + \left\{ \frac{1}{4} i_0^{-1/2} i_0^{(\mathrm{D2})} - \frac{1}{8} i_0^{-3/2} (i_0^{(\mathrm{D1})})^2 \right\} (\hat{\theta}_{\mathrm{ML}} - \theta_0)^2$$

$$+ O_p(n^{-2})$$

$$= i_0^{1/2} + \frac{1}{2} i_0^{-1/2} i_0^{(\mathrm{D1})} \{ i_0^{-1} (\hat{\eta} - \eta_0) \} + \frac{1}{2} i_0^{-1/2} i_0^{(\mathrm{D1})} \left\{ -\frac{1}{2} i_0^{-3} i_0^{(\mathrm{D1})} (\hat{\eta} - \eta_0)^2 \right\}$$

$$+ \left\{ \frac{1}{4} i_0^{-1/2} i_0^{(\mathrm{D2})} - \frac{1}{8} i_0^{-3/2} (i_0^{(\mathrm{D1})})^2 \right\} \{ i_0^{-1} (\hat{\eta} - \eta_0) \}^2 + O_p(n^{-3/2})$$

$$= i_0^{1/2} + \frac{1}{2} i_0^{-3/2} i_0^{(\mathrm{D1})} (\hat{\eta} - \eta_0) + \left\{ \frac{1}{4} i_0^{-5/2} i_0^{(\mathrm{D2})} - \frac{3}{8} i_0^{-7/2} (i_0^{(\mathrm{D1})})^2 \right\} (\hat{\eta} - \eta_0)^2$$

$$+ O_p(n^{-3/2})$$

$$\equiv i_0^{1/2} + i_0^{(t1)} (\hat{\eta} - \eta_0) + i_0^{(t2)} (\hat{\eta} - \eta_0)^2 + O_p(n^{-3/2})$$

$$= i_0^{1/2} + i_0^{(t1)} l_0^{(1)} + i_0^{(t2)} l_0^{(2)} + O_p(n^{-3/2}).$$

Then, using (4.1) we have

$$t = n^{1/2} \left\{ \sum_{j=1}^{3} \lambda^{(j)} l_0^{(j)} + O_p(n^{-2}) \right\} (i_0^{1/2} + i_0^{(t1)} l_0^{(1)} + i_0^{(t2)} l_0^{(2)} + O_p(n^{-3/2}))$$

$$= n^{1/2} \{ \lambda^{(1)} l_0^{(1)} i_0^{1/2} + (\lambda^{(2)} l_0^{(2)} i_0^{1/2} + \lambda^{(1)} l_0^{(1)} i_0^{(t1)} l_0^{(1)})$$

$$+ (\lambda^{(3)} l_0^{(3)} i_0^{1/2} + \lambda^{(2)} l_0^{(2)} i_0^{(t1)} l_0^{(1)} + \lambda^{(1)} l_0^{(1)} i_0^{(t2)} l_0^{(2)}) \} + O_p(n^{-3/2})$$

$$\equiv n^{1/2} (\lambda^{(t1)} l_0^{(1)} + \lambda^{(t2)} l_0^{(2)} + \lambda^{(t3)} l_0^{(3)}) + O_p(n^{-3/2}),$$

where

$$\lambda^{(t1)} l_0^{(1)} = i_0^{-1/2}(\hat{\eta} - \eta_0),$$

$$\begin{aligned}\lambda^{(t2)} l_0^{(2)} &= \lambda^{(2)} i_0^{1/2} l_0^{(2)} + \lambda^{(1)} i_0^{(t1)} (l_0^{(1)})^2 \\ &= \left(-\frac{1}{2} i_0^{-5/2} i_0^{(\mathrm{D1})} + \frac{1}{2} i_0^{-5/2} i_0^{(\mathrm{D1})} \right) (\hat{\eta} - \eta_0)^2 \\ &= 0,\end{aligned}$$

$$\begin{aligned}\lambda^{(t3)} l_0^{(3)} &= \left[\frac{1}{2} i_0^{-9/2} (i_0^{(\mathrm{D1})})^2 - \frac{1}{6} i_0^{-7/2} i_0^{(\mathrm{D2})} + \left(-\frac{1}{2} i_0^{-3} i_0^{(\mathrm{D1})} \right) \left(\frac{1}{2} i_0^{-3/2} i_0^{(\mathrm{D1})} \right) \right. \\ &\quad \left. + \frac{1}{4} i_0^{-7/2} i_0^{(\mathrm{D2})} - \frac{3}{8} i_0^{-9/2} (i_0^{(\mathrm{D1})})^2 \right] (\hat{\eta} - \eta_0)^3 \\ &= \left\{ -\frac{1}{8} i_0^{-9/2} (i_0^{(\mathrm{D1})})^2 + \frac{1}{12} i_0^{-7/2} i_0^{(\mathrm{D2})} \right\} (\hat{\eta} - \eta_0)^3.\end{aligned}$$

Note that $\lambda^{(t2)} l_0^{(2)} = 0$. Then, from the above results,

$$\kappa_1(t) = O(n^{-3/2}),$$

$$\begin{aligned}\kappa_2(t) &= 1 + n^{-1} 2 \lambda^{(t1)} \lambda^{(t3)} \mathrm{E}(n^2 l_0^{(1)} l_0^{(3)}) + O(n^{-2}) \\ &= 1 + n^{-1} \left\{ -\frac{1}{4} i_0^{-5} (i_0^{(\mathrm{D1})})^2 + \frac{1}{6} i_0^{-4} i_0^{(\mathrm{D2})} \right\} 3 i_0^2 + O(n^{-2}) \\ &= 1 + n^{-1} \left\{ -\frac{3}{4} i_0^{-3} (i_0^{(\mathrm{D1})})^2 + \frac{1}{2} i_0^{-2} i_0^{(\mathrm{D2})} \right\} + O(n^{-2}) \\ &= 1 + n^{-1} \left[-\frac{3}{4} \{\mathrm{sk}(s)\}^2 + \frac{1}{2} \mathrm{ks}(s) \right] + O(n^{-2}),\end{aligned}$$

$$\kappa_3(t) = n^{-1/2} i_0^{-3/2} \mathrm{E}\{n^2(\hat{\eta}-\eta_0)^3\} + O(n^{-3/2})$$
$$= n^{-1/2}\mathrm{sk}(s) + O(n^{-3/2}),$$
$$\kappa_4(t) = \mathrm{E}(t^4) - 3 - n^{-1} 6\alpha^{(t)}_{\mathrm{ML}\Delta 2} + O(n^{-2})$$
$$= n^{-1}[\ \mathrm{E}\{n^3(\lambda^{(t1)} l_0^{(1)})^4\} - 3n + 4\mathrm{E}\{n^3(\lambda^{(t1)} l_0^{(1)})^3 \lambda^{(t3)} l_0^{(3)}\} - 6\alpha^{(t)}_{\mathrm{ML}\Delta 2}\ ]$$
$$+ O(n^{-2})$$
$$= n^{-1}\left[\ i_0^{-2}\mathrm{E}\{n^2(\hat{\eta}-\eta_0)^4\} - 3n \right.$$
$$\left. + 4\left\{-\frac{1}{8} i_0^{-9/2}(i_0^{(\mathrm{D1})})^2 + \frac{1}{12} i_0^{-7/2} i_0^{(\mathrm{D2})}\right\} i_0^{-3/2}\mathrm{E}\{n^3(\hat{\eta}-\eta_0)^6\} - 6\alpha^{(t)}_{\mathrm{ML}\Delta 2}\right]$$
$$+ O(n^{-2})$$
$$= n^{-1}\left[\mathrm{kt}(s) + \left\{-\frac{15}{2}\{\mathrm{sk}(s)\}^2 + 5\mathrm{kt}(s)\right\} - 6\left\{-\frac{3}{4}\{\mathrm{sk}(s)\}^2 + \frac{1}{2}\mathrm{kt}(s)\right\}\right]$$
$$+ O(n^{-2})$$
$$= n^{-1}[-3\{\mathrm{sk}(s)\}^2 + 3\mathrm{kt}(s)] + O(n^{-2}).$$

A.3 Derivation of the results in Table 1 and examples

A relationship between $\alpha_{\mathrm{JM}\Delta 2}$ and $\alpha_{\mathrm{ML}\Delta 2}$ is given by

Theorem 3. *A Bayes modal estimator* $\hat{\theta}_{\mathrm{JM}}$ *using the Jeffreys prior for a canonical parameter in the exponential family gives*

$$\alpha_{\mathrm{JM}\Delta 2} \le (>) \alpha_{\mathrm{ML}\Delta 2} \quad \textit{when} \quad \mathrm{k/ss} \le (>) 2\,. \tag{A1}$$

Proof. From (5.10), the inequalities (A1) hold when

$$\begin{aligned} -2n\,\mathrm{acov}(\hat{\theta}_{\mathrm{ML}}, \hat{\alpha}_{\mathrm{ML1}}) &= -2 i_0^{-4}(i_0^{(\mathrm{D1})})^2 + i_0^{-3} i_0^{(\mathrm{D2})} \\ &= i_0^{-1}[-2\{\mathrm{sk}(s)\}^2 + \mathrm{kt}(s)] \le (>) 0, \end{aligned} \tag{A2}$$

which gives (A1). Q.E.D.

Theorem 3 indicates that when $\mathrm{kt}(s)$ is smaller than $2\{\mathrm{sk}(s)\}^2$, not only the bias of order $O(n^{-1})$ for $\hat{\theta}_{\mathrm{JM}}$ vanishes, but also the variance up to order $O(n^{-2})$ is smaller than that of $\hat{\theta}_{\mathrm{ML}}$.

Theorem 4. *The mean square error of* $\hat{\theta}_{\mathrm{JM}}$ *up to order* $O(n^{-2})$ *is smaller than or equal to (greater than) that of* $\hat{\theta}_{\mathrm{ML}}$ *when*

$$\mathrm{k/ss} \le (>) \frac{9}{4}. \tag{A3}$$

Proof. Since $\alpha_{\mathrm{JM1}} = 0$, the mean square error of $\hat{\theta}_{\mathrm{JM}}$ up to order $O(n^{-2})$ denoted by $\mathrm{MSE}_{O(n^{-2})}(\cdot)$ is equal to the higher-order asymptotic variance up to order $O(n^{-2})$ i.e.,

$$\begin{aligned}\mathrm{MSE}_{O(n^{-2})}(\hat{\theta}_{\mathrm{JM}}) &= n^{-1}i_0^{-1} + n^{-2}\alpha_{\mathrm{JM}\Delta 2} \\ &= n^{-1}\{\mathrm{var}(s)\}^{-1} + \frac{n^{-2}}{2}\{\mathrm{var}(s)\}^{-1}\{\mathrm{sk}(s)\}^2,\end{aligned} \tag{A4}$$

while

$$\begin{aligned}\mathrm{MSE}_{O(n^{-2})}(\hat{\theta}_{\mathrm{ML}}) &= n^{-1}i_0^{-1} + n^{-2}(\alpha_{\mathrm{ML1}}^2 + \alpha_{\mathrm{ML}\Delta 2}) \\ &= n^{-1}\{\mathrm{var}(s)\}^{-1} + n^{-2}\{\mathrm{var}(s)\}^{-1}\left[\frac{1}{4}\{\mathrm{sk}(s)\}^2 + \frac{5}{2}\{\mathrm{sk}(s)\}^2 - \mathrm{kt}(s)\right] \\ &= n^{-1}\{\mathrm{var}(s)\}^{-1} + n^{-2}\{\mathrm{var}(s)\}^{-1}\left[\frac{11}{4}\{\mathrm{sk}(s)\}^2 - \mathrm{kt}(s)\right]\end{aligned} \tag{A5}$$

(see (4.2)), which gives (A3). Q.E.D.

Example 1. Bernoulli distribution. The single population canonical parameter (logit) is $\theta_0 = \log\{\pi_0/(1-\pi_0)\}\ (\pi_0 \ne 0,\ 1)$, where
$s = s(X) = X,\ \Pr(X=1) = \pi_0,$
$\Pr(X=0) = 1-\pi_0,\ i_0 = \pi_0(1-\pi_0),\ \kappa_3(s) = (1-2\pi_0)\pi_0(1-\pi_0)$ and
$\kappa_4(s) = (1-6\pi_0+6\pi_0^2)\pi_0(1-\pi_0)$. Then,

$$\alpha_{\mathrm{ML1}} = -\frac{i_0^{-2}}{2}\kappa_3(s) = -\frac{1-2\pi_0}{2\pi_0(1-\pi_0)},\quad \mathrm{sk}(s) = \frac{1-2\pi_0}{\{\pi_0(1-\pi_0)\}^{1/2}}\ \text{ and}$$

$$\mathrm{kt}(s) = \frac{1-6\pi_0+6\pi_0^2}{\pi_0(1-\pi_0)},\ \text{which give}$$

$$\text{k/ss of } s = \frac{1-6\pi_0+6\pi_0^2}{(1-2\pi_0)^2} = 1 - \frac{2\pi_0(1-\pi_0)}{(1-2\pi_0)^2} < 1 \ \ (\pi_0 \neq 0,\ 0.5,\ 1) \qquad \text{(A6)}$$

belonging to Area A in Table 1 with advantages for $\hat{\theta}_{\text{JM}}$ over $\hat{\theta}_{\text{ML}}$.

Example 2. Poisson distribution. The single population canonical parameter is given by $\theta_0 = \log\lambda$, where λ is the source parameter i.e.,

$\Pr(X = x) = \exp(-e^{\theta_0}) e^{\theta_0 x} / x! = \exp(\theta_0 x - e^{\theta_0}) / x! \ \ (e^{\theta_0} = \lambda)$

$(x = 0, 1, \ldots)$,

$s = s(X) = X,\ \mathrm{E}(X) = \mathrm{var}(X) = \kappa_3(X) = \kappa_4(X) = e^{\theta_0} = \lambda,$

$i_0 = (\partial\lambda / \partial\theta_0)^2 \lambda^{-1} = \lambda = e^{\theta_0},$

which give

$$\alpha_{\text{ML1}} = -\frac{e^{-\theta_0}}{2},\ \mathrm{sk}(s) = \frac{e^{\theta_0}}{(e^{\theta_0})^{3/2}} = e^{-\theta_0/2},\ \mathrm{kt}(s) = \frac{e^{\theta_0}}{(e^{\theta_0})^2} = e^{-\theta_0} \qquad \text{(A7)}$$

and $\text{k/ss of } s = e^{-\theta_0} / e^{-\theta_0} = 1$,

where the k/ss ratio does not depend on θ_0 and belongs to advantageous Area A in Table 1 for $\hat{\theta}_{\text{JM}}$.

Example 3. Gamma distribution with known shape parameter. The density function using the source parameter is $f(X = x \mid \lambda, \gamma) = \dfrac{x^{\gamma-1} e^{-x/\lambda}}{\Gamma(\gamma)\lambda^{\gamma}}$, where γ is assumed to be known. The single population canonical parameter (negative rate) is $\theta_0 = -1/\lambda$, which gives

$f(X = x \mid \theta_0, \gamma) = x^{\gamma-1}(-\theta_0)^{\gamma} e^{x\theta_0} / \Gamma(\gamma)$ and $a(\theta_0) = -\gamma\log(-\theta_0)$. Then,

$s = s(X) = X,\ \mathrm{E}(X) = \partial a(\theta_0) / \partial\theta_0 = -\gamma / \theta_0,$

$i_0 = \mathrm{var}(X) = \gamma / \theta_0^2,\ \kappa_3(X) = -2\gamma / \theta_0^3$ and $\kappa_4(X) = 6\gamma / \theta_0^4$, which give

$$\alpha_{\text{ML1}} = \gamma^{-1}\theta_0,\ \mathrm{sk}(s) = -\frac{2\gamma / \theta_0^3}{(\gamma / \theta_0^2)^{3/2}} = -2\gamma^{-1/2},\ \mathrm{kt}(s) = \frac{6\gamma / \theta_0^4}{(\gamma / \theta_0^2)^2} = 6\gamma^{-1} \ \text{ and}$$

$$\text{k/ss of } s = \frac{6\gamma^{-1}}{4\gamma^{-1}} = \frac{3}{2}, \qquad \text{(A8)}$$

where the k/ss ratio again does not depend on θ_0 and belongs to Area A in Table 1.

A.4 Bias adjustment reducing the mean square error

Define

$$\hat{\boldsymbol{\theta}}_{\mathrm{C}(k)} = \hat{\boldsymbol{\theta}}_{\mathrm{ML}} - n^{-1}k\,\hat{\boldsymbol{\alpha}}_{\mathrm{ML1}}, \tag{A9}$$

where k is a constant. When $k = 1$, (A9) gives the familiar bias-corrected estimator. Note that (A9) is a special case of

$$\hat{\boldsymbol{\theta}}_{\mathrm{C}} \equiv \hat{\boldsymbol{\theta}}_{\mathrm{ML}} + n^{-1}\hat{\mathbf{I}}^{-1}\hat{\mathbf{q}}^{*} \tag{A10}$$

when $\hat{\mathbf{q}}^{*} = -k\hat{\mathbf{I}}\hat{\boldsymbol{\alpha}}_{\mathrm{ML1}}$. Then, from (5.3) and $\hat{\boldsymbol{\theta}}_{\mathrm{ML}} - \boldsymbol{\theta}_0 = \sum_{j=1}^{3}\boldsymbol{\Lambda}^{(j)}\mathbf{l}_0^{(j)} + O_p(n^{-1})$, we have

$$\hat{\boldsymbol{\theta}}_{\mathrm{C}} = \hat{\boldsymbol{\theta}}_{\mathrm{W}} + O_p(n^{-2}). \tag{A11}$$

From (5.4) and (A11), $\boldsymbol{\alpha}_{\mathrm{C1}}$ and $\mathbf{A}_{\mathrm{C}\Delta 2}$ in $\kappa_1(\hat{\boldsymbol{\theta}}_{\mathrm{C}})$ and $\kappa_2(\hat{\boldsymbol{\theta}}_{\mathrm{C}})$, defined similarly to $\boldsymbol{\alpha}_{\mathrm{W1}}$ and $\mathbf{A}_{\mathrm{W}\Delta 2}$ in $\kappa_1(\hat{\boldsymbol{\theta}}_{\mathrm{W}})$ and $\kappa_2(\hat{\boldsymbol{\theta}}_{\mathrm{W}})$, respectively, are found to be

$$\boldsymbol{\alpha}_{\mathrm{C1}} = \boldsymbol{\alpha}_{\mathrm{W1}} \ \text{ and } \ \mathbf{A}_{\mathrm{C}\Delta 2} = \mathbf{A}_{\mathrm{W}\Delta 2} \ \text{ with } \ \boldsymbol{\alpha}_{\mathrm{C}j} = \boldsymbol{\alpha}_{\mathrm{ML}j}\,(j = 2,\,3,\,4), \tag{A12}$$

where $\boldsymbol{\alpha}_{\mathrm{C}j}\,(j = 2,\,3,\,4)$ are defined similarly.

In the case of the canonical parameter, (A11) gives

$$\hat{\boldsymbol{\theta}}_{\mathrm{C}(1)} = \hat{\boldsymbol{\theta}}_{\mathrm{ML}} - n^{-1}\hat{\boldsymbol{\alpha}}_{\mathrm{ML1}} = \hat{\boldsymbol{\theta}}_{\mathrm{JM}} + O_p(n^{-2}). \tag{A13}$$

That is, the estimator asymptotically equal to the Jeffreys Bayes modal estimator ($\hat{\mathbf{q}}^{*} = -\hat{\mathbf{I}}\hat{\boldsymbol{\alpha}}_{\mathrm{ML1}}$) up to order $O_p(n^{-3/2})$ is also obtained by bias-correction of $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$. A scalar version of (A9) is

$$\hat{\theta}_{\mathrm{C}(k)} = \hat{\theta}_{\mathrm{ML}} - n^{-1}k\hat{\alpha}_{\mathrm{ML1}}, \tag{A14}$$

whose $\mathrm{MSE}_{O(n^{-2})}$ is

$$\begin{aligned}&\mathrm{MSE}_{O(n^{-2})}(\hat{\theta}_{\mathrm{C}(k)})\\ &= n^{-1}i_0^{-1} + n^{-2}\{\alpha_{\mathrm{ML}\Delta 2} + (1-k)^2\alpha_{\mathrm{ML1}}^2 - 2k\,n\,\mathrm{acov}(\hat{\theta}_{\mathrm{ML}},\hat{\alpha}_{\mathrm{ML1}})\},\end{aligned} \tag{A15}$$

which is minimized when k is

$$k_{\min} \equiv 1 + \alpha_{\mathrm{ML1}}^{-2}n\,\mathrm{acov}(\hat{\theta}_{\mathrm{ML}},\hat{\alpha}_{\mathrm{ML1}}), \tag{A16}$$

and

$$\mathrm{MSE}_{O(n^{-2})}(\hat{\theta}_{\mathrm{C}(k_{\min})}) = n^{-1}i_0^{-1} + n^{-2}\{\alpha_{\mathrm{ML}\Delta 2} + \alpha_{\mathrm{ML}1}^2(1-k_{\min}^2)\}$$
$$= n^{-1}i_0^{-1} + n^{-2}[\alpha_{\mathrm{ML}\Delta 2} - 2n\,\mathrm{acov}(\hat{\theta}_{\mathrm{ML}}, \hat{\alpha}_{\mathrm{ML}1}) - \alpha_{\mathrm{ML}1}^{-2}\{n\,\mathrm{acov}(\hat{\theta}_{\mathrm{ML}}, \hat{\alpha}_{\mathrm{ML}1})\}^2], \quad \text{(A17)}$$

where (A16) and (A17) were given by Ogasawara (2014). From the above results,

Theorem 5. *For a scalar canonical parameter, a constant for bias adjustment minimizing the* MSE *up to order* $O(n^{-2})$ *in* (A15) *is given by*

$$k_{\min} = 5 - 2\frac{\mathrm{kt}(s)}{\{\mathrm{sk}(s)\}^2}, \quad \text{(A18)}$$

which yields

$$\mathrm{MSE}_{O(n^{-2})}(\hat{\theta}_{\mathrm{C}(k_{\min})}) = n^{-1}\{\mathrm{var}(s)\}^{-1}$$
$$+ n^{-2}\{\mathrm{var}(s)\}^{-1}\{\mathrm{sk}(s)\}^2\left[-\frac{\{\mathrm{kt}(s)\}^2}{\{\mathrm{sk}(s)\}^4} + 4\frac{\mathrm{kt}(s)}{\{\mathrm{sk}(s)\}^2} - \frac{7}{2}\right]. \quad \text{(A19)}$$

Proof. Using (4.2) and (5.10), (A16) and (A17) become

$$k_{\min} = 1 + \left[-\frac{1}{2}\{\mathrm{var}(s)\}^{-1/2}\mathrm{sk}(s)\right]^{-2}\{\mathrm{var}(s)\}^{-1}\left[\{\mathrm{sk}(s)\}^2 - \frac{1}{2}\mathrm{kt}(s)\right] \quad \text{(A20)}$$

and

$$\mathrm{MSE}_{O(n^{-2})}(\hat{\theta}_{\mathrm{C}(k_{\min})}) = n^{-1}i_0^{-1} + n^{-2}i_0^{-1}\left[\frac{5}{2}\{\mathrm{sk}(s)\}^2 - \{\mathrm{kt}(s)\}\right.$$
$$\left. + \left\{-\frac{1}{2}\mathrm{sk}(s)\right\}^2\left(1 - \left\{5 - 2\frac{\mathrm{kt}(s)}{\{\mathrm{sk}(s)\}^2}\right\}^2\right)\right], \quad \text{(A21)}$$

which give (A18) and (A19), respectively. Q.E.D.

It is found that $k_{\min}$ in (A18) depends only on the k/ss ratio.

Correction 1. In Result 2 of Ogasawara (2013), the incorrect equation $k_{\min} = 5 - (\mathrm{k/ss})$ due to a typographical error should be $k_{\min} = 5 - 2(\mathrm{k/ss})$ as in Theorem 5.

Example 1 (continued). Bernoulli distribution. The result in this continued example without using the k/ss ratio is known (Ogasawara, 2014), but is repeated here using the ratio. From (A6) and (A18),

$$k_{\min} = 5 - 2\frac{\mathrm{kt}(s)}{\{\mathrm{sk}(s)\}^2} = 5 - 2\left\{1 - \frac{2\pi_0(1-\pi_0)}{(1-2\pi_0)^2}\right\}$$

$$= 3 + \frac{4\pi_0(1-\pi_0)}{(1-2\pi_0)^2} > 3\ (\pi_0 \neq 0,\ 0.5,\ 1), \qquad \text{(A22)}$$

where $k_{\min}$ depends on π_0, but 3 is a lower bound of practical use. That is, when $k = 3$, from (A17), (A19), $\alpha_{\mathrm{ML1}} = \frac{1}{2}\left(-\frac{1}{\pi_0} + \frac{1}{1-\pi_0}\right)$ (see (A6)) and

$$n\,\mathrm{acov}(\hat{\theta}_{\mathrm{ML}}, \hat{\alpha}_{\mathrm{ML1}}) = \frac{\mathrm{var}(s)}{\pi_0(1-\pi_0)}\frac{1}{2}\left(\frac{1}{\pi_0^2} + \frac{1}{(1-\pi_0)^2}\right)$$

$$= \frac{1}{4}\left(\frac{\{\mathrm{sk}(s)\}^2}{\mathrm{var}(s)} + \frac{1}{\{\mathrm{var}(s)\}^2}\right)$$, we have

$$\mathrm{MSE}_{O(n^{-2})}(\hat{\theta}_{\mathrm{C(3)}}) = n^{-1}i_0^{-1} + n^{-2}\{\alpha_{\mathrm{ML\Delta 2}} + 4\alpha_{\mathrm{ML1}}^2 - 6n\,\mathrm{acov}(\hat{\theta}_{\mathrm{ML}}, \hat{\alpha}_{\mathrm{ML1}})\}$$

$$= n^{-1}i_0^{-1}$$

$$+n^{-2}i_0^{-1}\left[\frac{5}{2}\{\mathrm{sk}(s)\}^2 - \mathrm{kt}(s) + 4\left\{-\frac{\mathrm{sk}(s)}{2}\right\}^2 - \frac{3}{2}\left\{\{\mathrm{sk}(s)\}^2 + \frac{1}{\mathrm{var}(s)}\right\}\right] \qquad \text{(A23)}$$

$$= n^{-1}i_0^{-1} + n^{-2}i_0^{-1}[\ 2\{\mathrm{sk}(s)\}^2 - \mathrm{kt}(s) - (3/2)\{\mathrm{var}(s)\}^{-1}]$$

$$= n^{-1}i_0^{-1} + n^{-2}i_0^{-1}[\ \{\mathrm{sk}(s)\}^2 + 2 - (3/2)\{\mathrm{var}(s)\}^{-1}]$$

$$= n^{-1}i_0^{-1} + n^{-2}i_0^{-1}[(1/2)\{\mathrm{sk}(s)\}^2 - \{\mathrm{var}(s)\}^{-1}]$$

$$= n^{-1}\{\pi_0(1-\pi_0)\}^{-1} + n^{-2}\{\pi_0(1-\pi_0)\}^{-2}(1/2)\{(1-2\pi_0)^2 - 2\}$$

(note that $\mathrm{kt}(s) = \{\mathrm{sk}(s)\}^2 - 2$ in this case), which is smaller than that of $\hat{\theta}_{\mathrm{ML}}$ by $-3\alpha_{\mathrm{ML1}}^2 + 6n\,\mathrm{acov}(\hat{\theta}_{\mathrm{ML}}, \hat{\alpha}_{\mathrm{ML1}})$

$= \frac{3\{\mathrm{sk}(s)\}^2}{4\,\mathrm{var}(s)} + \frac{3}{2\{\mathrm{var}(s)\}^2} = \frac{3(1-2\pi_0)^2 + 6}{4\{\pi_0(1-\pi_0)\}^2}$. Note that (A23) is smaller than the usual asymptotic variance $n^{-1}\{\pi_0(1-\pi_0)\}^{-1}$ or equivalently $\mathrm{MSE}_{O(n^{-1})}(\cdot)$.

Example 2 (continued). Poisson distribution.
From (A7) and (A18),

$$k_{\min} = 5 - 2 = 3 \qquad \text{(A24)}$$

and as in (A19) or (A21).

$$\begin{aligned}\mathrm{MSE}_{O(n^{-2})}(\hat{\theta}_{\mathrm{C}(k_{\min})}) &= n^{-1}i_0^{-1} + n^{-2}i_0^{-1}\left[\frac{1}{2}\{\mathrm{sk}(s)\}^2 - \{\mathrm{kt}(s)\}\right] \\ &= n^{-1}\lambda^{-1} - n^{-2}\frac{\lambda^{-2}}{2}\end{aligned} \qquad \text{(A25)}$$

(see (A7)), which is smaller than that of $\hat{\theta}_{\mathrm{ML}}$ by $n^{-2}9\alpha_{\mathrm{ML1}}^2 = n^{-2}i_0^{-1}(9/4)\{\mathrm{sk}(s)\}^2 = n^{-2}(9/4)\lambda^{-2}$ (see (A17)). Note that (A25) is smaller than the usual asymptotic variance $n^{-1}\lambda^{-1}$ or equivalently $\mathrm{MSE}_{O(n^{-1})}(\cdot)$.

Example 3 (continued). Gamma distribution with known shape parameter.

From (A8) and (A18),

$$k_{\min} = 5 - 2 \times \frac{3}{2} = 2 \qquad \text{(A26)}$$

and from (4.2) and (A17)

$$\begin{aligned}&\mathrm{MSE}_{O(n^{-2})}(\hat{\theta}_{\mathrm{C}(k_{\min})}) \\ &= n^{-1}i_0^{-1} + n^{-2}i_0^{-1}\left[\frac{5}{2}\{\mathrm{sk}(s)\}^2 - \{\mathrm{kt}(s)\} + \frac{1}{4}\{\mathrm{sk}(s)\}^2(1-2^2)\right] \\ &= n^{-1}\gamma^{-1}\lambda^{-2} + n^{-2}\gamma^{-1}\lambda^{-2}\left\{\left(\frac{5}{2} - \frac{3}{4}\right)4 - 6\right\}\gamma^{-1} \\ &= n^{-1}\gamma^{-1}\lambda^{-2} + n^{-2}\gamma^{-2}\lambda^{-2},\end{aligned} \qquad \text{(A27)}$$

which is smaller than that of $\hat{\theta}_{\mathrm{ML}}$ by $n^{-2}4\alpha_{\mathrm{ML1}}^2 = n^{-2}4\gamma^{-2}\lambda^{-2}$.

A.5 The asymptotic cumulants and associated results for Example 4

Noting that $-(X-\mu)^2/2$ is a negatively scaled chi-square distributed variable with 1 degree of freedom and using $2^{c-1}(c-1)!\upsilon$ for the *c*-th cumulant of the chi-squared variable with υ degrees of freedom,

$$E(s_d) = \frac{1}{2}(\mu^2 - \sigma^2), \;\; i_0^* \equiv \mathrm{var}(s_d) = \frac{\sigma^4}{2}, \;\; \kappa_3(s_d) = \left(-\frac{\sigma^2}{2}\right)^3 8 = -\sigma^6,$$

$$\kappa_4(s_d) = \left(-\frac{\sigma^2}{2}\right)^4 48 = 3\sigma^8, \; \mathrm{sk}(s_d) = -\frac{\sigma^6}{(\sigma^4/2)^{3/2}} = -\sqrt{8}, \tag{A28}$$

$$\mathrm{kt}(s_d) = \frac{3\sigma^8}{(\sigma^4/2)^2} = 12, \; \alpha_{\mathrm{ML1}}^* = -\frac{1}{2}\{\mathrm{var}(s_d)\}^{-1/2}\mathrm{sk}(s_d) = \frac{2}{\sigma^2} = 2\theta_0^*,$$

where $\alpha_{\mathrm{ML}j}^*$ is defined for $\hat{\theta}_{\mathrm{ML}}^*$ (the MLE of θ_0^*) similarly to $\alpha_{\mathrm{ML}j}$ for $\hat{\theta}_{\mathrm{ML}}$, giving

$$\frac{\mathrm{kt}(s_d)}{\{\mathrm{sk}(s_d)\}^2} = \frac{3}{2}, \tag{A29}$$

which belongs to advantageous Area A in Table 1 for $\hat{\theta}_{\mathrm{JM}}^*$ over $\hat{\theta}_{\mathrm{ML}}^*$.

From (A29),

$$k_{\min} = 5 - 2 \times \frac{3}{2} = 2, \tag{A30}$$

which does not depend on θ_0^*, and gives as in (A27),

$$\begin{aligned} \mathrm{MSE}_{O(n^{-2})}(\hat{\theta}_{\mathrm{C}(k_{\min})}^*) &= n^{-1}(i_0^*)^{-1} + n^{-2}\{\alpha_{\mathrm{ML}\Delta 2}^* + (\alpha_{\mathrm{ML1}}^*)^2(1 - k_{\min}^2)\} \\ &= n^{-1}2\sigma^{-4} + n^{-2}2\sigma^{-4}\left\{\left(\frac{5}{2} - \frac{3}{4}\right)(-\sqrt{8})^2 - 12\right\} \\ &= n^{-1}2\sigma^{-4} + n^{-2}4\sigma^{-4}, \end{aligned} \tag{A31}$$

which is smaller than that of $\hat{\theta}_{\mathrm{ML}}^*$ by $n^{-2}4(\alpha_{\mathrm{ML1}}^*)^2 = n^{-2}16\sigma^{-4}$.

From (A28) and (A29),

$$\hat{\theta}_{\mathrm{C}(k_{\min})}^* = \hat{\theta}_{\mathrm{ML}}^* - n^{-1}k_{\min}\hat{\alpha}_{\mathrm{ML1}}^* = (1 - n^{-1}4)\hat{\theta}_{\mathrm{ML}}^*. \tag{A32}$$

On the other hand, the bias-corrected estimator is

$$\hat{\theta}_{\mathrm{C}(1)}^* = \hat{\theta}_{\mathrm{ML}}^* - n^{-1}\hat{\alpha}_{\mathrm{ML1}}^* = (1 - n^{-1}2)\hat{\theta}_{\mathrm{ML}}^*. \tag{A33}$$

Note that the MLE $\hat{\sigma}^2 = n^{-1}\sum_{j=1}^n (x_{(j)} - \mu)^2$ with known μ is an unbiased sample mean over *n* observations. By elementary algebra (e.g., Gruber, 1998, Section 1.6; Ogasawara, 2014, Section 3.1), the exact solution of the constant minimizing $\mathrm{MSE}_{O(n^{-2})}(c^*\hat{\sigma}^2)$ is given by

$$c^*_{\min} = \frac{1}{1+n^{-1}\dfrac{\mathrm{var}\{(X-\mu)^2\}}{[\mathrm{E}\{(X-\mu)^2\}]^2}} = \frac{1}{1+n^{-1}2} = 1 - n^{-1}2 + O(n^{-2}), \quad \text{(A34)}$$

which is a fortunate case in that $c^*_{\min}$ does not depend on θ^*_0, or equivalently in that (A34) is given by the known coefficient of variation $\sqrt{2}$, which is crucial to $c^*_{\min}$. It is seen that (A34) for $\hat{\sigma}^2$ is asymptotically equal to the factor $(1-n^{-1}2)$ for $\hat{\theta}^*_{\mathrm{ML}} = 1/\hat{\sigma}^2$ in (A33) rather than that of (A32). Incidentally, when μ is unknown, the constant $1/(1+n^{-1})$ minimizing the MSE of $c^* n^{-1}\sum_{j=1}^{n}(x_{(j)} - \bar{x})^2$ under normality is well known (e.g., DeGroot & Schervish, 2002, p.431) though it is inadmissible (Stein, 1964).

Using Theorem 2, the following asymptotic cumulants are obtained

$$\begin{aligned}
&\alpha^{(t)}_{\mathrm{ML}1} = 0, \quad \alpha^{(t)}_{\mathrm{ML}2} = 1, \\
&\alpha^{(t)}_{\mathrm{ML}\Delta 2} = -\frac{3}{4}\{\mathrm{sk}(s_d)\}^2 + \frac{\mathrm{kt}(s_d)}{2} = -\frac{3}{4}\times 8 + \frac{12}{2} = 0, \\
&\alpha^{(t)}_{\mathrm{ML}3} = \mathrm{sk}(s_d) = -\sqrt{8}, \\
&\alpha^{(t)}_{\mathrm{ML}4} = -3\{\mathrm{sk}(s_d)\}^2 + 3\mathrm{kt}(s_d) = -3\times 8 + 3\times 12 = 12 = \mathrm{kt}(s_d),
\end{aligned} \quad \text{(A35)}$$

where it is of interest to find that $\alpha^{(t)}_{\mathrm{ML}\Delta 2} = 0$ and $\alpha^{(t)}_{\mathrm{ML}4} = \mathrm{kt}(s_d)$ in this example.

Though the results of (A28) were directly obtained, they can also be given by $a(\theta^*) = \dfrac{\mu^2}{2\sigma^2} + \dfrac{1}{2}\log\sigma^2 = \dfrac{\theta^*}{2}\mu^2 - \dfrac{1}{2}\log\theta^*$ as

$$\frac{\partial a(\theta^*_0)}{\partial \theta^*_0} = \frac{\mu^2}{2} - \frac{1}{2\theta^*_0} = \frac{1}{2}(\mu^2 - \sigma^2) = \mathrm{E}(s_d),$$

$$i^*_0 = \frac{\partial^2 a(\theta^*_0)}{\partial \theta^{*2}_0} = \frac{1}{2\theta^{*2}_0} = \frac{\sigma^4}{2} = \mathrm{var}(s_d),$$

$$i_0^{*(\mathrm{D1})} = \frac{\partial^3 a(\theta_0^*)}{\partial \theta_0^{*3}} = -\frac{1}{\theta_0^{*3}} = -\sigma^6 = \kappa_3(s_d),$$

$$i_0^{*(\mathrm{D2})} = \frac{\partial^4 a(\theta_0^*)}{\partial \theta_0^{*4}} = \frac{3}{\theta_0^{*4}} = 3\sigma^8 = \kappa_4(s_d). \tag{A36}$$

References

DeGroot, M. H., & Schervish, M. J. (2002). *Probability and statistics* (3rd ed.). Boston, MA: Addison-Wesley.

Gruber, M. H. J. (1998). *Improving efficiency by shrinkage: The James-Stein and ridge regression estimators*. New York: Marcel Dekker.

Ogasawara, H. (2013). Asymptotic cumulants of the estimator of the canonical parameter in the exponential family. *Journal of Statistical Planning and Inference, 143*, 2142-2150.

Ogasawara, H. (2014). Bias adjustment minimizing the asymptotic mean square error. To appear in *Communications in Statistics – Theory and Methods*.

Stein, C. (1964). Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. *Annals of the Institute of Statistical Mathematics, 16*, 155-160.