

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),

$$\begin{aligned}\kappa_1(\hat{\theta}_{\text{ML}} - \theta_0) &= -\frac{n^{-1}}{2} i_0^{-3} i_0^{(\text{D1})} i_0 + O(n^{-2}) = -\frac{n^{-1}}{2} i_0^{-2} i_0^{(\text{D1})} + O(n^{-2}) \\ &= -\frac{n^{-1}}{2} \{\text{var}(s)\}^{-1/2} \text{sk}(s) + O(n^{-2}),\end{aligned}$$

$$\begin{aligned}\kappa_2(\hat{\theta}_{\text{ML}}) &= n^{-1} i_0^{-1} + n^{-2} \left[\frac{1}{2} i_0^{-6} (i_0^{(\text{D1})})^2 i_0^2 - i_0^{-4} i_0^{(\text{D1})} \text{E}\{(s - \eta_0)^3\} \right. \\ &\quad \left. + 2 \left\{ \frac{1}{2} i_0^{-5} (i_0^{(\text{D1})})^2 - \frac{1}{6} i_0^{-4} i_0^{(\text{D2})} \right\} i_0^{-1} \text{E}\{(s - \eta_0)^4\} \right] + O(n^{-3}) \\ &= n^{-1} i_0^{-1} + n^{-2} \left\{ \frac{5}{2} i_0^{-4} (i_0^{(\text{D1})})^2 - i_0^{-3} i_0^{(\text{D2})} \right\} + O(n^{-3}), \\ &= n^{-1} \{\text{var}(s)\}^{-1} + n^{-2} \{\text{var}(s)\}^{-1} \left[\frac{5}{2} \{\text{sk}(s)\}^2 - \text{kt}(s) \right] + O(n^{-3})\end{aligned}$$

$$\begin{aligned}
\kappa_3(\hat{\theta}_{ML}) &= n^{-2} \left[i_0^{-3} E\{(s - \eta_0)^3\} - \frac{3}{2} (i_0^{-1})^2 i_0^{-3} i_0^{(D1)} E\{(s - \eta_0)^4\} - 3\alpha_{ML1}\alpha_{ML2} \right] \\
&\quad + O(n^{-3}) \\
&= n^{-2} \left(1 - \frac{9}{2} + \frac{3}{2} \right) i_0^{-3} i_0^{(D1)} + O(n^{-3}) \\
&= -n^{-2} 2 \{ \text{var}(s) \}^{-3/2} \text{sk}(s) + O(n^{-3}),
\end{aligned}$$

$$\begin{aligned}
\kappa_4(\hat{\theta}_{ML}) &= n^{-3} \left[(i_0^{-1})^4 \kappa_4(s) + 4(i_0^{-1})^3 \left\{ -\frac{1}{2} i_0^{-3} i_0^{(D1)} E\{(s - \eta_0)^5\} \right\} \right. \\
&\quad + 6(i_0^{-1})^2 \left\{ -\frac{1}{2} i_0^{-3} i_0^{(D1)} \right\}^2 E\{(s - \eta_0)^6\} \\
&\quad + 4(i_0^{-1})^3 \left\{ \frac{1}{2} i_0^{-5} (i_0^{(D1)})^2 - \frac{1}{6} i_0^{-4} i_0^{(D2)} \right\} E\{(s - \eta_0)^6\} \\
&\quad \left. - 4\alpha_{ML1}\alpha_{ML3} - 6\alpha_{ML2}\alpha_{ML\Delta 2} - 6\alpha_{ML2}\alpha_{ML1}^2 \right] + O(n^{-4}) \\
&= n^{-3} \left[(i_0^{-1})^4 i_0^{(D2)} - 20i_0^{-6} i_0^{(D1)} (i_0 i_0^{(D1)}) + \frac{45}{2} (i_0^{-1})^8 (i_0^{(D1)})^2 i_0^3 \right. \\
&\quad + (30i_0^{-8} (i_0^{(D1)})^2 - 10i_0^{-7} i_0^{(D2)}) i_0^3 - 4 \left(-\frac{1}{2} i_0^{-2} i_0^{(D1)} \right) (-2i_0^{-3} i_0^{(D1)}) \\
&\quad \left. - 6i_0^{-1} \left\{ \frac{5}{2} i_0^{-4} (i_0^{(D1)})^2 - i_0^{-3} i_0^{(D2)} \right\} - 6i_0^{-1} \left(-\frac{1}{2} i_0^{-2} i_0^{(D1)} \right)^2 \right] + O(n^{-4})
\end{aligned}$$

$$\begin{aligned}
&= n^{-3} \{ 12i_0^{-5} (i_0^{(D1)})^2 - 3i_0^{-4} i_0^{(D2)} \} + O(n^{-4}) \\
&= n^{-3} \{ \text{var}(s) \}^{-2} [12 \{ \text{sk}(s) \}^2 - 3\text{kt}(s)] + O(n^{-4}).
\end{aligned}$$

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

$$\begin{aligned}
 \hat{i}^{1/2} &= i_0^{1/2} + \frac{1}{2} i_0^{-1/2} i_0^{(D1)} (\hat{\theta}_{ML} - \theta_0) + \left\{ \frac{1}{4} i_0^{-1/2} i_0^{(D2)} - \frac{1}{8} i_0^{-3/2} (i_0^{(D1)})^2 \right\} (\hat{\theta}_{ML} - \theta_0)^2 \\
 &\quad + O_p(n^{-2}) \\
 &= i_0^{1/2} + \frac{1}{2} i_0^{-1/2} i_0^{(D1)} \{i_0^{-1} (\hat{\eta} - \eta_0)\} + \frac{1}{2} i_0^{-1/2} i_0^{(D1)} \left\{ -\frac{1}{2} i_0^{-3} i_0^{(D1)} (\hat{\eta} - \eta_0)^2 \right\} \\
 &\quad + \left\{ \frac{1}{4} i_0^{-1/2} i_0^{(D2)} - \frac{1}{8} i_0^{-3/2} (i_0^{(D1)})^2 \right\} \{i_0^{-1} (\hat{\eta} - \eta_0)\}^2 + O_p(n^{-3/2}) \\
 &= i_0^{1/2} + \frac{1}{2} i_0^{-3/2} i_0^{(D1)} (\hat{\eta} - \eta_0) + \left\{ \frac{1}{4} i_0^{-5/2} i_0^{(D2)} - \frac{3}{8} i_0^{-7/2} (i_0^{(D1)})^2 \right\} (\hat{\eta} - \eta_0)^2 \\
 &\quad + O_p(n^{-3/2}) \\
 &\equiv i_0^{1/2} + i_0^{(t1)} (\hat{\eta} - \eta_0) + i_0^{(t2)} (\hat{\eta} - \eta_0)^2 + O_p(n^{-3/2}) \\
 &= i_0^{1/2} + i_0^{(t1)} I_0^{(1)} + i_0^{(t2)} I_0^{(2)} + O_p(n^{-3/2}).
 \end{aligned}$$

Then, using (4.1) we have

$$\begin{aligned}
 t &= n^{1/2} \left\{ \sum_{j=1}^3 \lambda^{(j)} I_0^{(j)} + O_p(n^{-2}) \right\} (i_0^{1/2} + i_0^{(t1)} I_0^{(1)} + i_0^{(t2)} I_0^{(2)} + O_p(n^{-3/2})) \\
 &= n^{1/2} \{ \lambda^{(1)} I_0^{(1)} i_0^{1/2} + (\lambda^{(2)} I_0^{(2)} i_0^{1/2} + \lambda^{(1)} I_0^{(1)} i_0^{(t1)} I_0^{(1)}) \\
 &\quad + (\lambda^{(3)} I_0^{(3)} i_0^{1/2} + \lambda^{(2)} I_0^{(2)} i_0^{(t1)} I_0^{(1)} + \lambda^{(1)} I_0^{(1)} i_0^{(t2)} I_0^{(2)}) \} + O_p(n^{-3/2}) \\
 &\equiv n^{1/2} (\lambda^{(t1)} I_0^{(1)} + \lambda^{(t2)} I_0^{(2)} + \lambda^{(t3)} I_0^{(3)}) + O_p(n^{-3/2}),
 \end{aligned}$$

where

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

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

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

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

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

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

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

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

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

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

$$\alpha_{\text{JM}\Delta 2} \leq (>) \alpha_{\text{ML}\Delta 2} \text{ when } k/ss \leq (>) 2. \quad (\text{A1})$$

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

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

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

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

order $O(n^{-2})$ is smaller than that of $\hat{\theta}_{ML}$.

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

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

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

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

while

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

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

Example 1. Bernoulli distribution. The single population canonical parameter (logit) is $\theta_0 = \log\{\pi_0 / (1 - \pi_0)\}$ ($\pi_0 \neq 0, 1$), where

$$s = s(X) = X, \Pr(X = 1) = \pi_0,$$

$$\Pr(X = 0) = 1 - \pi_0, i_0 = \pi_0(1 - \pi_0), \kappa_3(s) = (1 - 2\pi_0)\pi_0(1 - \pi_0) \text{ and}$$

$$\kappa_4(s) = (1 - 6\pi_0 + 6\pi_0^2)\pi_0(1 - \pi_0). \text{ Then,}$$

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

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

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

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.,

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

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

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

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

which give

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

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

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 $f(X = x | \lambda, \gamma) = \frac{x^{\gamma-1} e^{-x/\lambda}}{\Gamma(\gamma) \lambda^\gamma}$,

where γ is assumed to be known. The single population canonical parameter (negative rate) is $\theta_0 = -1/\lambda$, which gives

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

$$s = s(X) = X, \quad E(X) = \partial a(\theta_0) / \partial \theta_0 = -\gamma / \theta_0,$$

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

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

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

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

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

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

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

when $\hat{\mathbf{q}}^* = -k \hat{\mathbf{I}} \hat{\alpha}_{ML1}$. Then, from (5.3) and

$\hat{\theta}_{ML} - \theta_0 = \sum_{j=1}^3 \Lambda^{(j)} \mathbf{I}_0^{(j)} + O_p(n^{-1})$, we have

$$\hat{\theta}_C = \hat{\theta}_W + O_p(n^{-2}). \tag{A11}$$

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

$$\alpha_{C1} = \alpha_{W1} \text{ and } \mathbf{A}_{C\Delta 2} = \mathbf{A}_{W\Delta 2} \text{ with } \alpha_{Cj} = \alpha_{MLj} (j = 2, 3, 4), \tag{A12}$$

where $\alpha_{Cj} (j = 2, 3, 4)$ are defined similarly.

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

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

That is, the estimator asymptotically equal to the Jeffreys Bayes modal estimator ($\hat{\mathbf{q}}^* = -\hat{\mathbf{I}} \hat{\alpha}_{ML1}$) up to order $O_p(n^{-3/2})$ is also obtained by

bias-correction of $\hat{\theta}_{ML}$. A scalar version of (A9) is

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

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

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

which is minimized when k is

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

and

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

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

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

which yields

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

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

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

and

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

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 $k_{\min} = 5 - (k/ss)$ due to a typographical error should be $k_{\min} = 5 - 2(k/ss)$ 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),

$$\begin{aligned}
 k_{\min} &= 5 - 2 \frac{kt(s)}{\{sk(s)\}^2} = 5 - 2 \left\{ 1 - \frac{2\pi_0(1-\pi_0)}{(1-2\pi_0)^2} \right\} \\
 &= 3 + \frac{4\pi_0(1-\pi_0)}{(1-2\pi_0)^2} > 3 \quad (\pi_0 \neq 0, 0.5, 1),
 \end{aligned}
 \tag{A22}$$

where k_{\min} depends on π_0 , but 3 is a lower bound of practical use. That is,

when $k = 3$, from (A17), (A19), $\alpha_{ML1} = \frac{1}{2} \left(-\frac{1}{\pi_0} + \frac{1}{1-\pi_0} \right)$ (see (A6)) and

$$\begin{aligned}
 n \operatorname{acov}(\hat{\theta}_{ML}, \hat{\alpha}_{ML1}) &= \frac{\operatorname{var}(s)}{\pi_0(1-\pi_0)} \frac{1}{2} \left(\frac{1}{\pi_0^2} + \frac{1}{(1-\pi_0)^2} \right) \\
 &= \frac{1}{4} \left(\frac{\{sk(s)\}^2}{\operatorname{var}(s)} + \frac{1}{\{\operatorname{var}(s)\}^2} \right), \text{ we have} \\
 \operatorname{MSE}_{O(n^{-2})}(\hat{\theta}_{C(3)}) &= n^{-1}i_0^{-1} + n^{-2}\{\alpha_{ML\Delta 2} + 4\alpha_{ML1}^2 - 6n \operatorname{acov}(\hat{\theta}_{ML}, \hat{\alpha}_{ML1})\} \\
 &= n^{-1}i_0^{-1}
 \end{aligned}
 \tag{A23}$$

$$\begin{aligned}
 &+ n^{-2}i_0^{-1} \left[\frac{5}{2}\{sk(s)\}^2 - kt(s) + 4 \left\{ -\frac{sk(s)}{2} \right\}^2 - \frac{3}{2} \left\{ \{sk(s)\}^2 + \frac{1}{\operatorname{var}(s)} \right\} \right] \\
 &= n^{-1}i_0^{-1} + n^{-2}i_0^{-1} [2\{sk(s)\}^2 - kt(s) - (3/2)\{\operatorname{var}(s)\}^{-1}] \\
 &= n^{-1}i_0^{-1} + n^{-2}i_0^{-1} [\{sk(s)\}^2 + 2 - (3/2)\{\operatorname{var}(s)\}^{-1}] \\
 &= n^{-1}i_0^{-1} + n^{-2}i_0^{-1} [(1/2)\{sk(s)\}^2 - \{\operatorname{var}(s)\}^{-1}] \\
 &= n^{-1}\{\pi_0(1-\pi_0)\}^{-1} + n^{-2}\{\pi_0(1-\pi_0)\}^{-2} (1/2)\{(1-2\pi_0)^2 - 2\}
 \end{aligned}$$

(note that $kt(s) = \{sk(s)\}^2 - 2$ in this case), which is smaller than that of

$$\begin{aligned}
 &\hat{\theta}_{ML} \text{ by } -3\alpha_{ML1}^2 + 6n \operatorname{acov}(\hat{\theta}_{ML}, \hat{\alpha}_{ML1}) \\
 &= \frac{3\{sk(s)\}^2}{4\operatorname{var}(s)} + \frac{3}{2\{\operatorname{var}(s)\}^2} = \frac{3(1-2\pi_0)^2 + 6}{4\{\pi_0(1-\pi_0)\}^2}. \text{ Note that (A23) is smaller than}
 \end{aligned}$$

the usual asymptotic variance $n^{-1}\{\pi_0(1-\pi_0)\}^{-1}$ or equivalently $\operatorname{MSE}_{O(n^{-1})}(\cdot)$.

Example 2 (continued). Poisson distribution.

From (A7) and (A18),

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

and as in (A19) or (A21).

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

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

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

From (A8) and (A18),

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

and from (4.2) and (A17)

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

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

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 $2^{c-1}(c-1)!\nu$ for the c -th cumulant of the chi-squared variable with ν degrees of freedom,

$$\begin{aligned}
 E(s_d) &= \frac{1}{2}(\mu^2 - \sigma^2), \quad i_0^* \equiv \text{var}(s_d) = \frac{\sigma^4}{2}, \quad \kappa_3(s_d) = \left(-\frac{\sigma^2}{2}\right)^3 8 = -\sigma^6, \\
 \kappa_4(s_d) &= \left(-\frac{\sigma^2}{2}\right)^4 48 = 3\sigma^8, \quad \text{sk}(s_d) = -\frac{\sigma^6}{(\sigma^4/2)^{3/2}} = -\sqrt{8}, \\
 \text{kt}(s_d) &= \frac{3\sigma^8}{(\sigma^4/2)^2} = 12, \quad \alpha_{\text{ML}1}^* = -\frac{1}{2}\{\text{var}(s_d)\}^{-1/2}\text{sk}(s_d) = \frac{2}{\sigma^2} = 2\theta_0^*,
 \end{aligned} \tag{A28}$$

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

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

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

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),

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

which is smaller than that of $\hat{\theta}_{\text{ML}}^*$ by $n^{-2}4(\alpha_{\text{ML}1}^*)^2 = n^{-2}16\sigma^{-4}$.

From (A28) and (A29),

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

On the other hand, the bias-corrected estimator is

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

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 $\text{MSE}_{O(n^{-2})}(c^* \hat{\sigma}^2)$ is given by

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

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 $\hat{\theta}_{\text{ML}}^* = 1 / \hat{\sigma}^2$ 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

$$\begin{aligned} \alpha_{\text{ML1}}^{(t)} &= 0, \quad \alpha_{\text{ML2}}^{(t)} = 1, \\ \alpha_{\text{ML}\Delta 2}^{(t)} &= -\frac{3}{4} \{\text{sk}(s_d)\}^2 + \frac{\text{kt}(s_d)}{2} = -\frac{3}{4} \times 8 + \frac{12}{2} = 0, \\ \alpha_{\text{ML3}}^{(t)} &= \text{sk}(s_d) = -\sqrt{8}, \\ \alpha_{\text{ML4}}^{(t)} &= -3 \{\text{sk}(s_d)\}^2 + 3 \text{kt}(s_d) = -3 \times 8 + 3 \times 12 = 12 = \text{kt}(s_d), \end{aligned} \quad (\text{A35})$$

where it is of interest to find that $\alpha_{\text{ML}\Delta 2}^{(t)} = 0$ and $\alpha_{\text{ML4}}^{(t)} = \text{kt}(s_d)$ in this example.

Though the results of (A28) were directly obtained, they can also be given

$$\begin{aligned} \text{by } a(\theta^*) &= \frac{\mu^2}{2\sigma^2} + \frac{1}{2} \log \sigma^2 = \frac{\theta^*}{2} \mu^2 - \frac{1}{2} \log \theta^* \quad \text{as} \\ \frac{\partial a(\theta_0^*)}{\partial \theta_0^*} &= \frac{\mu^2}{2} - \frac{1}{2\theta_0^*} = \frac{1}{2} (\mu^2 - \sigma^2) = E(s_d), \\ i_0^* &= \frac{\partial^2 a(\theta_0^*)}{\partial \theta_0^{*2}} = \frac{1}{2\theta_0^{*2}} = \frac{\sigma^4}{2} = \text{var}(s_d), \end{aligned}$$

$$\begin{aligned}
 i_0^{*(D1)} &= \frac{\partial^3 a(\theta_0^*)}{\partial \theta_0^{*3}} = -\frac{1}{\theta_0^{*3}} = -\sigma^6 = \kappa_3(s_d), \\
 i_0^{*(D2)} &= \frac{\partial^4 a(\theta_0^*)}{\partial \theta_0^{*4}} = \frac{3}{\theta_0^{*4}} = 3\sigma^8 = \kappa_4(s_d).
 \end{aligned}
 \tag{A36}$$

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.