Biometrika (2012), **99**, 1, *pp*. 1–5
© 2012 Biometrika Trust
Printed in Great Britain
Advance Access publication on 31 July 2012

Supplementary material for On asymptotic validity of naive inference with an approximate likelihood

BY H. E. OGDEN
Mathematical Sciences, University of Southampton, Southampton SO17 1BJ, U.K.
h.e.ogden@soton.ac.uk

1. FINDING $\delta_i(\theta)$ IN EXAMPLE 3.1

Recall Y_i is the number of successes out of $m = m_n$ trials on item i. We study how $\delta_i(\theta) = \|(d/d\theta)\epsilon_i(\theta)\|$ varies with m.

Write $f(b; y_i) = -y_i \log\left\{\text{logit}^{-1}(b)\right\} + (m - y_i)\log\left\{1 - \text{logit}^{-1}(b)\right\}$ and $g(b; \theta, y_i) = f(b; y_i) - \log\phi(b; 0, \theta)$, so that

$$L_i(\theta) = \int_{-\infty}^{\infty} \exp\{-g(b; \theta, y_i)\} db.$$

In the following, we drop the data y_i from the notation for convenience. Write $\hat{b}(\theta)$ for the maximizer of $g(\cdot, \theta)$, and

$$\hat{g}_r(\theta) = \frac{\partial^r g}{\partial b^k}(\hat{b}(\theta); \theta).$$

By equation (4) of Shun & McCullagh (1995), the error in the Laplace approximation to the log-likelihood $\ell_i(\theta)$ is

$$\epsilon_i(\theta) = \sum_{l=1}^{\infty} \frac{1}{2l!} \sum_{P \in \mathcal{P}_{2l}} n_2(P)(-1)^v \hat{g}_{|p_1|}(\theta) \ldots \hat{g}_{|p_v|}(\theta)\{\hat{g}_2(\theta)\}^{-l}, \tag{1}$$

where $P = p_1|\ldots|p_v$ is a partition of $2l$ indices into v blocks of size 3 or more, and $n_2(P)$ is the number of partitions Q of $2l$ indices into l blocks of size 2, such that Q is complementary to P.

Write $h_P(\theta) = \hat{g}_{|p_1|}(\theta) \ldots \hat{g}_{|p_v|}(\theta)\{\hat{g}_2(\theta)\}^{-l}$. Then $h_P(\theta) = O_p(m^{v-l})$, since $\hat{g}_r(\theta) = O_p(m)$ for each r.

Differentiating (1) gives

$$\frac{d}{d\theta}\epsilon_i(\theta) = \sum_{l=1}^{\infty} \frac{1}{2l!} \sum_{P \in \mathcal{P}_{2l}} n_2(P)(-1)^v \frac{d}{d\theta} h_P(\theta), \tag{2}$$

and

$$\frac{d}{d\theta} h_P(\theta) = \sum_{i=1}^{v} \left[\hat{g}'_{|p_i|}(\theta) \prod_{j \neq i} \hat{g}_{|p_j|}(\theta)\{\hat{g}_2(\theta)\}^{-l} - \prod_{j=1}^{v} l\hat{g}_{|p_j|}(\theta)\hat{g}'_2(\theta)\{\hat{g}_2(\theta)\}^{-(l+1)}\right], \tag{3}$$

where $\hat{g}'_r(\theta) = (d/d\theta)\hat{g}_r(\theta)$.

For each r, we have

$$\hat{g}_r'(\theta) = \frac{d}{d\theta}\left\{g^{(r)}(\hat{b}(\theta);\theta)\right\}$$

$$= \frac{\partial g^{(r)}}{\partial\theta}\{\hat{b}(\theta);\theta\} + \frac{d\hat{b}(\theta)}{d\theta}\hat{g}_{r+1}(\theta). \tag{4}$$

We now study the size of each of the terms in (4). We have

$$\frac{\partial g^{(r)}}{\partial\theta}(b;\theta) = \frac{\partial}{\partial\theta}\left\{-\frac{\partial^r}{\partial b^r}\log\phi(b;0,\theta)\right\} = O_p(1), \tag{5}$$

and

$$\hat{g}_{r+1}(\theta) = O_p(m). \tag{6}$$

For each θ, $\hat{b}(\theta)$ satisfies $g_1\{\hat{b}(\theta);\theta\} = 0$. Differentiating this with respect to θ,

$$\frac{d\hat{b}(\theta)}{d\theta}g_2\{\hat{b}(\theta);\theta\} + \frac{\partial g_1}{\partial\theta}\{\hat{b}(\theta);\theta\} = 0.$$

But

$$\frac{\partial g_1}{\partial\theta}(b;\theta) = -2b\theta^{-3},$$

so

$$\frac{d\hat{b}(\theta)}{d\theta} = 2\hat{b}(\theta)\theta^{-3}\{\hat{g}_2(\theta)\}^{-1} = O_p(m^{-1}). \tag{7}$$

Substituting (5), (6) and (7) into (4) gives that $\hat{g}_r'(\theta) = O_p(1)$ for each r. From (3) we then have

$$\frac{d}{d\theta}h_P(\theta) = O_p(m^{v-l-1}).$$

The highest order terms in (2) come from partitions with $(l,v) = (2,1)$ or $(3,2)$, and so $\delta_i(\theta) = O_p(m^{-2})$.

2. FINDING $\delta_m^{(k)}(\beta)$ IN EXAMPLE 3.2

Kaufman (1949) provides an exact expression for the normalizing constant for an Ising model on an $n \times m$ lattice, with $\alpha = 0$ and periodic boundary, as

$$Z_{n\times m}(0,\beta) = \{2\sinh(2\beta)\}^{nm/2}\,\bar{A}_{n,m}(\beta)/2,$$

where

$$\bar{A}_{n,m}(\beta) = A_{n,m}^{(1)}(\beta) + A_{n,m}^{(2)}(\beta) + A_{n,m}^{(3)}(\beta) + A_{n,m}^{(4)}(\beta),$$

and

$$A_{n,m}^{(1)}(\beta) = \prod_{q=0}^{n} 2\cosh\{m\,a_{2q+1,n}(\beta)/2\}, \qquad A_{n,m}^{(2)}(\beta) = \prod_{q=0}^{n} 2\sinh\{m\,a_{2q+1,n}(\beta)/2\},$$

$$A_{n,m}^{(3)}(\beta) = \prod_{q=0}^{n} 2\cosh\{m\,a_{2q,n}(\beta)/2\}, \qquad A_{n,m}^{(4)}(\beta) = \prod_{q=0}^{n} 2\sinh\{m\,a_{2q,n}(\beta)/2\}$$

where

$$a_{l,n}(\beta) = \cosh^{-1}\left\{\cosh(2\beta)^2/\sinh(2\beta) - \cos(\pi l/n)\right\}$$

for $l \geq 1$, and $a_{0,n}(\beta) = a_0(\beta) = 2\beta + \log\{\tanh(\beta)\}$.

Using the approximation $Z^{(k)}_{m\times m}(\beta)$ to $Z_{m\times m}(\beta)$, the error in the log-likelihood is

$$\epsilon^{(k)}_m(\beta) = (m-k+1)\log \bar{A}_{k,m}(\beta) - (m-k)\log \bar{A}_{k-1,m}(\beta) - \log \bar{A}_{m,m}(\beta).$$

Differentiating this with respect to β,

$$\frac{d}{d\beta}\epsilon^{(k)}_m(\beta) = (m-k+1)\frac{d}{d\beta}\left\{\log \bar{A}_{k,m}(\beta)\right\} - (m-k)\frac{d}{d\beta}\left\{\log \bar{A}_{k-1,m}(\beta)\right\} - \frac{d}{d\beta}\left\{\log \bar{A}_{m,m}(\beta)\right\}, \tag{8}$$

and

$$\frac{d}{d\beta}\left\{\log \bar{A}_{n,m}(\beta)\right\} = \sum_{i=1}^{4}\frac{d}{d\beta}\left\{\log A^{(i)}_{n,m}(\beta)\right\} r^{(i)}_{n,m}(\beta),$$

where $r^{(i)}_{n,m}(\beta) = A^{(i)}_{n,m}(\beta)/\bar{A}_{n,m}(\beta)$. We have

$$\begin{aligned}\frac{d}{d\beta}\left\{\log A^{(1)}_{n,m}(\beta)\right\} &= m/2\sum_{q=0}^{n} a'_{2q+1,n}(\beta)\tanh\{m a_{2q+1,n}(\beta)/2\} \\ &= m/2\sum_{q=0}^{n} a'_{2q+1,n}(\beta) + O[m\exp\{-a_0(\beta)m\}]\end{aligned}$$

as $m \to \infty$, since $\tanh(x) = 1 + O\{\exp(-2x)\}$ as $x \to \infty$, and $a_{2q+1,n}(\beta) \geq a_0(\beta) > 0$.

Similar expressions may be obtained for the derivatives of the other $\{\log A^{(i)}_{n,m}(\beta)\}$ terms, and combining these gives

$$\frac{d}{d\beta}\left\{\log \bar{A}_{n,m}(\beta)\right\} = mS^{(o)}_n(\beta)r^{(o)}_{n,m}(\beta) + mS^{(e)}_n(\beta)r^{(e)}_{n,m}(\beta) + O[m\exp\{-a_0(\beta)m\}]$$

where $S^{(o)}_n = \sum_{q=0}^{n} a'_{2q+1,n}(\beta)$, $S^{(e)}_n = \sum_{q=0}^{n} a'_{2q,n}(\beta)$, $r^{(o)}_{n,m}(\beta) = r^{(1)}_{n,m}(\beta) + r^{(2)}_{n,m}(\beta)$ and $r^{(e)}_{n,m}(\beta) = r^{(3)}_{n,m}(\beta) + r^{(4)}_{n,m}(\beta)$. Define

$$f(x;\beta) = d_\beta\left\{-1 + c_\beta - \cos(x)\right\}^{-1/2}\{1 + c_\beta - \cos(x)\}^{-1/2}$$

where $d_\beta = 4\cosh(2\beta) - 2\cosh(2\beta)\coth(2\beta)^2$ and $c_\beta = \cosh(2\beta)^2/\sinh(2\beta)$. Then $a'_{j,n}(\beta) = f(j\pi/n;\beta)$, and $n^{-1}S^{(o)}_n(\beta)$ and $n^{-1}S^{(e)}_n(\beta)$ are both trapezium rule approximations to $I(\beta) = \frac{1}{2\pi}\int_0^{2\pi} f(x;\beta)dx$. Write $R^{(o)}_n(\beta) = n^{-1}S^{(o)}_n(\beta) - I(\beta)$ and $R^{(e)}_n(\beta) = n^{-1}S^{(e)}_n(\beta) - I(\beta)$ for the error in each of these approximations to the integral.

LEMMA 1. *For each* $\beta < \beta_c$, $R_n(\beta) = \max\{|R^{(e)}_n(\beta)|, |R^{(o)}_n(\beta)|\} = O\{\exp(-b_\beta n)\}$, *where* $b_\beta = 2\cosh^{-1}\{-1 + \cosh(2\beta)^2/\sinh(2\beta)\}$.

Proof. We apply the results of Trefethen & Weideman (2014) to show exponentially fast convergence of these trapezium rule approximations to $I(\beta)$. These results depend on properties of the integrand $f(z,\beta)$, considered as a function of complex-valued z. There are a branch points of $f(z,\beta)$ at a distance $a_\beta = \cosh^{-1}\{-1 + \cosh(2\beta)^2/\sinh(2\beta)\}$ from the real axis, and the function is analytic for $-a_\beta < \operatorname{Im} z < a_\beta$, so by Theorem 3.2 of Trefethen & Weideman (2014), $|R_n^{(o)}(\beta)| = O\{\exp(-2a_\beta n)\} = O\{\exp(-b_\beta n)\}$.

The same argument holds with $R_n^{(e)}(\beta)$ in place of $R_n^{(o)}(\beta)$, so $R_n(\beta) = O\{\exp(-b_\beta n)\}$, as required. □

We now prove the main result.

LEMMA 2. *If* $k \to \infty$ *as* $m \to \infty$, $\delta_m^{(k)}(\beta) = O\{m^2 k \exp(-b_\beta k)\} + o(1)$.

Proof. We have

$$\frac{d}{d\beta}\left\{\log \bar{A}_{n,m}(\beta)\right\} = mnI(\beta) + mnt_{n,m}(\beta) + O[m\exp\{-a_0(\beta)m\}]$$

where $t_{n,m}(\beta) = R_n^{(o)}(\beta)r_{n,m}^{(o)}(\beta) + R_n^{(e)}(\beta)r_{n,m}^{(e)}(\beta)$.

Substituting this into (8), the contributions from the $mnI(\beta)$ terms cancel, and the combined remainder terms are always $o(1)$, since $m^2\exp\{-a_0(\beta)m\} = o(1)$. We are left with

$$\frac{d}{d\beta}\epsilon_m^{(k)}(\beta) = (m-k+1)mt_{k,m}(\beta) - (m-k)mt_{k-1,m}(\beta) - mt_{m,m}(\beta) + o(1).$$

Then

$$\begin{aligned}\delta_m^{(k)}(\beta) &= \left|\frac{d}{d\beta}\epsilon_m^{(k)}(\beta)\right| \\ &\le (m-k+1)mk|t_{k,m}(\beta)| + (m-k)m(k-1)|t_{k-1,m}(\beta)| + m^2|t_{m,m}(\beta)| + o(1) \\ &\le (m-k+1)mkR_k(\beta) + (m-k)m(k-1)R_{k-1}(\beta) + m^2R_m(\beta) + o(1)\end{aligned}$$

since $|t_{n,m}(\beta)| \le |R_n^{(o)}(\beta)|r_{n,m}^{(o)}(\beta) + |R_n^{(e)}(\beta)|r_{n,m}^{(e)}(\beta) \le R_n(\beta)$

$$= O\{m^2k\exp(-b_\beta k)\} + o(1)$$

by Lemma 1, as required. □

REFERENCES

KAUFMAN, B. (1949). Crystal statistics. II. Partition function evaluated by spinor analysis. *Physical Review* **76**, 1232.

SHUN, Z. & MCCULLAGH, P. (1995). Laplace approximation of high dimensional integrals. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 749–760.

TREFETHEN, L. N. & WEIDEMAN, J. (2014). The exponentially convergent trapezoidal rule. *SIAM Review* **56**, 385–458.

[*Received April* 2012. *Revised September* 2012]