Stability and uniqueness for likelihood inference

Thomas J. DiCiccio, Todd A. Kuffner, G. Alastair Young and Russell Zaretzki

Cornell University, Washington University in St. Louis, Imperial College London and University of Tennessee, Knoxville

Supplementary Material

S1 Proof of Lemma 2

The unconditional variance of $T(\psi)$ is

$$\begin{aligned}
\mathrm{var}\{T(\psi)\} &= E[\{T(\psi)\}^2] - [E\{T(\psi)\}]^2 = E[\{T(\psi)\}^2] + O(n^{-1}) \\
&= \eta E\{T_1^2 + 2T_1T_2 + O_p(n^{-2})\} + O(n^{-1}) \\
&= \eta E\{\lambda^{1r}\lambda^{1s} l_r l_s - 2\lambda^{1r}\xi^{stu} l_r l_{st} l_u + 2\lambda^{1r}\xi^{st} l_r l_s l_t + O_p(n^{-2})\} + O(n^{-1}) \\
&= -\eta\{\lambda^{1r}\lambda^{1s}\lambda_{rs} + O(n^{-2})\} + O(n^{-1}) \\
&= 1 + O(n^{-1}).
\end{aligned}$$

Correspondingly, the conditional variance of $T(\psi)$ is

$$\begin{aligned}
\mathring{\mathrm{var}}\{T(\psi)\} &= \mathring{E}[\{T(\psi)\}^2] - [\mathring{E}\{T(\psi)\}]^2 = \mathring{E}[\{T(\psi)\}^2] + O_p(n^{-1}) \\
&= \eta \mathring{E}\{T_1^2 + 2T_1T_2 + O_p(n^{-2})\} + O_p(n^{-1}) \\
&= \eta \mathring{E}\{\lambda^{1r}\lambda^{1s}\mathring{l}_r\mathring{l}_s - 2\lambda^{1r}\xi^{stu}\mathring{l}_r(\mathring{l}_{st} + \mathring{\Delta}_{st})\mathring{l}_u + 2\lambda^{1r}\xi^{st}\mathring{l}_r\mathring{l}_s\mathring{l}_t + O_p(n^{-2})\} + O_p(n^{-1}) \\
&= -\eta\{\lambda^{1r}\lambda^{1s}\mathring{\lambda}_{rs} - 2\lambda^{1r}\xi^{stu}\mathring{\lambda}_{ru}\mathring{\Delta}_{st} + O_p(n^{-2})\} + O_p(n^{-1}) \\
&= -\eta\{\lambda^{1r}\lambda^{1s}(\lambda_{rs} + \mathring{\Delta}_{rs}) - 2\lambda^{1r}\xi^{stu}\lambda_{ru}\mathring{\Delta}_{st}\} + O_p(n^{-1}) \\
&= 1 - \eta(\lambda^{1r}\lambda^{1s}\mathring{\Delta}_{rs} - 2\xi^{st1}\mathring{\Delta}_{st}) + O_p(n^{-1}) \\
&= 1 - \eta\{(\lambda^{1r}\lambda^{1s} - 2\xi^{rs1})\mathring{\Delta}_{rs}\} + O_p(n^{-1}).
\end{aligned}$$

It follows that $\mathring{\mathrm{var}}\{T(\psi)\} = \mathrm{var}\{T(\psi)\} + O_p(n^{-1})$ provided $\xi^{rs1} = \frac{1}{2}\lambda^{1r}\lambda^{1s}$. $\square$

S2 Proof of Lemma 3

The unconditional skewness of $T(\psi)$ is

$$\begin{aligned}
\text{skew}\{T(\psi)\} &= E([T(\psi) - E\{T(\psi)\}]^3) = E[\{T(\psi)\}^3] - 3E[\{T(\psi)\}^2]E\{T(\psi)\} + O(n^{-1}) \\
&= \eta^{3/2}[E\{(T_1+T_2)^3\} - 3E\{(T_1+T_2)^2\}E(T_1+T_2)] + O(n^{-1}) \\
&= \eta^{3/2}[E\{T_1^3 + 3T_1^2T_2 + O_p(n^{-5/2})\} - 3E\{T_1^2 + O_p(n^{-3/2})\}E(T_2)] + O(n^{-1}) \\
&= \eta^{3/2}[E\{-\lambda^{1r}\lambda^{1s}\lambda^{1t}l_rl_sl_t + 3\lambda^{1r}\lambda^{1s}(\xi^{tuv}l_{tu}l_v - \xi^{tu}l_tl_u)l_rl_s + O_p(n^{-5/2})\} \\
&\quad - 3E\{\lambda^{1r}\lambda^{1s}l_rl_s + O_p(n^{-3/2})\}\{\xi^{rst}\lambda_{rs,t} + \xi^{rs}\lambda_{rs} + O_p(n^{-3/2})\}] + O(n^{-1}) \\
&= \eta^{3/2}\{E(-\lambda^{1r}\lambda^{1s}\lambda^{1t}l_rl_sl_t + 3\lambda^{1r}\lambda^{1s}\xi^{tuv}l_rl_sl_{tu}l_v - 3\lambda^{1r}\lambda^{1s}\xi^{tu}l_rl_sl_tl_u \\
&\quad - 3\lambda^{1r}\lambda^{1s}\xi^{tuv}l_rl_s\lambda_{tu,v} - 3\lambda^{1r}\lambda^{1s}\xi^{tu}l_rl_s\lambda_{tu})\} + O(n^{-1}).
\end{aligned}$$

To continue the calculation, we make use of the following identities:

$$\begin{aligned}
-E(l_rl_sl_t) &= \lambda_{rs,t} + \lambda_{rt,s} + \lambda_{st,r} + \lambda_{rst}, \\
E(l_rl_sl_{tu}l_v) &= -\lambda_{rs}\lambda_{tu,v} - \lambda_{rv}\lambda_{tu,s} - \lambda_{sv}\lambda_{tu,r} + O(n^{3/2}), \\
E(l_rl_sl_tl_u) &= \lambda_{rs}\lambda_{tu} + \lambda_{rt}\lambda_{su} + \lambda_{ru}\lambda_{st} + O(n^{3/2}).
\end{aligned}$$

By using these identities, we obtain

$$\begin{aligned}
\text{skew}\{T(\psi)\} &= \eta^{3/2}(3\lambda^{1r}\lambda^{1s}\lambda^{1t}\lambda_{rs,t} + \lambda^{1r}\lambda^{1s}\lambda^{1t}\lambda_{rst} \\
&\quad - 3\lambda^{11}\xi^{tuv}\lambda_{tu,v} - 3\lambda^{1s}\xi^{tu1}\lambda_{tu,s} - 3\lambda^{1r}\xi^{tu1}\lambda_{tu,r} \\
&\quad - 3\lambda^{11}\xi^{tu}\lambda_{tu} - 3\xi^{11} - 3\xi^{11} \\
&\quad + 3\lambda^{11}\xi^{tuv}\lambda_{tu,v} + 3\lambda^{11}\xi^{tu}\lambda_{tu}) + O(n^{-1}) \\
&= \eta^{3/2}(\lambda^{1r}\lambda^{1s}\lambda^{1t}\lambda_{rst} + 3\lambda^{1r}\lambda^{1s}\lambda^{1t}\lambda_{rs,t} - 6\xi^{rs1}\lambda^{1t}\lambda_{rs,t} - 6\xi^{11}) + O(n^{-1}).
\end{aligned}$$

Similar reasoning shows that the conditional skewness of $T(\psi)$ is

$$\begin{aligned}
\mathring{\text{skew}}\{T(\psi)\} &= \eta^{3/2}[\mathring{E}\{-\lambda^{1r}\lambda^{1s}\lambda^{1t}l_rl_sl_t + 3\lambda^{1r}\lambda^{1s}\xi^{tuv}l_rl_sl_{tu}l_v - 3\lambda^{1r}\lambda^{1s}\xi^{tu}l_rl_sl_tl_u \\
&\quad - 3\lambda^{1r}\lambda^{1s}\xi^{tuv}l_rl_s\lambda_{tu,v} - 3\lambda^{1r}\lambda^{1s}\xi^{tu}l_rl_s\lambda_{tu} + O_p(n^{-5/2})\}] + O_p(n^{-1}) \\
&= \eta^{3/2}[\mathring{E}\{-\lambda^{1r}\lambda^{1s}\lambda^{1t}\mathring{l}_r\mathring{l}_s\mathring{l}_t + 3\lambda^{1r}\lambda^{1s}\xi^{tuv}\mathring{l}_r\mathring{l}_s(\mathring{l}_{tu} + \mathring{\Delta}_{tu})\mathring{l}_v - 3\lambda^{1r}\lambda^{1s}\xi^{tu}\mathring{l}_r\mathring{l}_s\mathring{l}_t\mathring{l}_u \\
&\quad - 3\lambda^{1r}\lambda^{1s}\xi^{tuv}\mathring{l}_r\mathring{l}_s\lambda_{tu,v} - 3\lambda^{1r}\lambda^{1s}\xi^{tu}\mathring{l}_r\mathring{l}_s\lambda_{tu}\}] + O_p(n^{-1}) \\
&= \eta^{3/2}\{\mathring{E}(-\lambda^{1r}\lambda^{1s}\lambda^{1t}\mathring{l}_r\mathring{l}_s\mathring{l}_t + 3\lambda^{1r}\lambda^{1s}\xi^{tuv}\mathring{l}_r\mathring{l}_s\mathring{l}_{tu}\mathring{l}_v - 3\lambda^{1r}\lambda^{1s}\xi^{tu}\mathring{l}_r\mathring{l}_s\mathring{l}_t\mathring{l}_u \\
&\quad - 3\lambda^{1r}\lambda^{1s}\xi^{tuv}\mathring{l}_r\mathring{l}_s\lambda_{tu,v} - 3\lambda^{1r}\lambda^{1s}\xi^{tu}\mathring{l}_r\mathring{l}_s\lambda_{tu})\} + O_p(n^{-1}).
\end{aligned}$$

Now we use the following identities:

$$\begin{aligned}
-\mathring{E}(\mathring{l}_r\mathring{l}_s\mathring{l}_t) &= \mathring{\lambda}_{rs,t} + \mathring{\lambda}_{rt,s} + \mathring{\lambda}_{st,r} + \mathring{\lambda}_{rst} \\
&= \lambda_{rs,t} + \lambda_{rt,s} + \lambda_{st,r} + \lambda_{rst} + O_p(n^{1/2}) \\
&= \lambda_{r,s,t} + O_p(n^{1/2}) \\
&= -E(l_r l_s l_t) + O_p(n^{1/2}),
\end{aligned}$$

$$\begin{aligned}
\mathring{E}(\mathring{l}_r\mathring{l}_s\mathring{l}_{tu}\mathring{l}_v) &= -\mathring{\lambda}_{rs}\mathring{\lambda}_{tu,v} - \mathring{\lambda}_{rv}\mathring{\lambda}_{tu,s} - \mathring{\lambda}_{sv}\mathring{\lambda}_{tu,r} + O_p(n^{3/2}) \\
&= -\lambda_{rs}\lambda_{tu,v} - \lambda_{rv}\lambda_{tu,s} - \lambda_{sv}\lambda_{tu,r} + O_p(n^{3/2}) \\
&= E(l_r l_s l_{tu} l_v) + O_p(n^{3/2}),
\end{aligned}$$

$$\begin{aligned}
\mathring{E}(\mathring{l}_r\mathring{l}_s\mathring{l}_t\mathring{l}_u) &= \mathring{\lambda}_{rs}\mathring{\lambda}_{tu} + \mathring{\lambda}_{rt}\mathring{\lambda}_{su} + \mathring{\lambda}_{ru}\mathring{\lambda}_{st} + O_p(n^{3/2}) \\
&= \lambda_{rs}\lambda_{tu} + \lambda_{rt}\lambda_{su} + \lambda_{ru}\lambda_{st} + O_p(n^{3/2}) \\
&= E(l_r l_s l_t l_u) + O_p(n^{3/2}).
\end{aligned}$$

By using these identities in the preceding expression for $\mathring{\text{skew}}\{T(\psi)\}$, it is apparent that $\mathring{\text{skew}}\{T(\psi)\} = \text{skew}\{T(\psi)\} + O_p(n^{-1})$, and hence, the conditional third cumulant agrees with the unconditional one to error of order $O_p(n^{-1})$, as required. □

S3 Proof of Theorem 5

To establish the stability of $\bar{W}(\psi)$ to error of order $O(n^{-3/2})$, we need only show that $\mathring{E}\{\bar{W}(\psi)\} = E\{\bar{W}(\psi)\} + O_p(n^{-3/2})$. For full generality, the previous notation, which is applicable when ψ is a scalar, must be extended. In the expressions that follow, it is assumed that subscripts and superscripts $a, b, \ldots$ have the range $1, \ldots, q$, while $r, s, \ldots$ range over $1, \ldots, d$. Let (η_{ab}) be the $q \times q$ matrix inverse of $(-\lambda^{ab})$, let $\tau^{rs} = \eta_{ab}\lambda^{ar}\lambda^{bs}$, and let $\nu^{rs} = \lambda^{rs} + \tau^{rs}$. In addition, let $B_a(\psi) = \partial B(\psi)/\partial\psi^a, B_{ab}(\psi) = \partial^2 B(\psi)/\partial\psi^a\partial\psi^b, \beta_a = E\{B_a(\psi)\}, \beta_{ab} = E\{B_{ab}(\psi)\}, b_a = B_a(\psi) - \beta_a, b_{ab} = B_{ab}(\psi) - \beta_{ab}$, and so forth. The constants β_a, β_{ab} etc. are assumed to be of order $O(1)$ and the variables b_a, b_{ab} etc. are assumed to be of order $O_p(n^{-1/2})$. Finally, it is assumed that the joint cumulants of $nb_a, nb_{ab}, l_r, l_{rs}$, and so forth are of order $O(n)$.

DiCiccio & Stern (1994b) showed that

$$\begin{aligned}
\bar{W}(\psi) = W(\psi) &- 2\lambda^{ar}\beta_a l_r - 2\lambda^{ar}b_a l_r + 2\lambda^{ar}\lambda^{st}\beta_a l_{rs}l_t - \lambda^{ar}\lambda^{su}\lambda^{tv}\beta_a\lambda_{rst}l_u l_v \\
&+ \lambda^{ar}\lambda^{bs}\beta_{ab}l_r l_s - \lambda^{ab}\beta_a\beta_b + O_p(n^{-3/2}),
\end{aligned}$$

and it follows that

$$\begin{aligned}
E\{\bar{W}(\psi)\} &= E\{W(\psi)\} - 2\lambda^{ar}E(b_a l_r) + \lambda^{ar}\lambda^{st}\beta_a(2\lambda_{rs,t} + \lambda_{rst}) - \lambda^{ab}(\beta_{ab} + \beta_a\beta_b) + O(n^{-3/2}) \\
&= E\{W(\psi)\} + \lambda^{ar}\lambda^{st}\beta_a(2\lambda_{rs,t} + \lambda_{rst}) - 2\lambda^{ar}\beta_{a/r} + \lambda^{ab}(\beta_{ab} - \beta_a\beta_b) + O(n^{-3/2}),
\end{aligned}$$

where $\beta_{a/r} = \partial\beta_a/\partial\theta^r$. For calculating $E\{\bar{W}(\psi)\}$, we assume that $B(\psi)$ is a function of Y and ψ only, so, in particular, it does not depend on ϕ. Thus, differentiation of the identity $\beta_a = E\{B_a(\psi)\}$ yields $\beta_{a/b} = E(b_a l_b)+\beta_{ab}$ and $\beta_{a/i} = E(b_a l_i)$ for $i = q+1, \ldots, d$. It follows that $\lambda^{ar}E(b_a l_r) = \lambda^{ar}\beta_{a/r} - \lambda^{ab}\beta_{ab}$.

To calculate $\mathring{E}\{W(\psi)\}$, some care is required about the conditional properties of $B_a(\psi)$, $B_{ab}(\psi)$, and so forth. The quantities $\mathring{\beta}_a = \mathring{E}\{B_a(\psi)\}$, $\mathring{\beta}_{ab} = \mathring{E}\{B_{ab}(\psi)\}$, etc. are assumed to be of order $O_p(1)$, while $\mathring{b}_a = B_a(\psi) - \mathring{\beta}_a$, $\mathring{b}_{ab} = B_{ab}(\psi) - \mathring{\beta}_{ab}$, etc. are assumed to be of order $O_p(n^{-1/2})$. Finally, it is assumed that the joint conditional cumulants of $n\mathring{b}_a$, $n\mathring{b}_{ab}$, $\mathring{l}_r$, $\mathring{l}_{rs}$, and so forth are of order $O_p(n)$.

Under the preceding assumptions, it is possible to determine the orders of the differences $\mathring{\beta}_a - \beta_a$ and $\mathring{\beta}_{ab} - \beta_{ab}$. Since $E(\mathring{\beta}_a) = E[\mathring{E}\{B_a(\psi)\}] = E\{B_a(\psi)\} = \beta_a$ and $\mathrm{var}(\mathring{\beta}_a) = \mathrm{var}[\mathring{E}\{B_a(\psi)\}] = \mathrm{var}\{B_a(\psi)\} - E[\mathring{\mathrm{var}}\{B_a(\psi)\}] = O(n^{-1}) - E\{\mathring{\mathrm{var}}(\mathring{b}_a)\} = O(n^{-1}) - E\{O_p(n^{-1})\} = O(n^{-1})$, it follows that $\mathring{\beta}_a = \beta_a + O_p(n^{-1/2})$. A similar argument shows that $\mathring{\beta}_{ab} = \beta_{ab} + O_p(n^{-1/2})$. We assume that differentiation of the identity $\mathring{\beta}_a = \beta_a + O_p(n^{-1/2})$ yields $\mathring{\beta}_{a/r} = \beta_{a/r} + O_p(n^{-1/2})$.

Now, define $\mathring{\delta}_a = \mathring{\beta}_a - \beta_a$, so that $\mathring{\delta}_a$ is a function of θ and A of order $O_p(n^{-1/2})$. Furthermore, $b_a = B_a(\psi) - \beta_a = B_a(\psi) - \mathring{\beta}_a + \mathring{\delta}_a = \mathring{b}_a + \mathring{\delta}_a$. To calculate $\mathring{E}\{\bar{W}(\psi)\}$, we observe that

$$\begin{aligned}\bar{W}(\psi) &= W(\psi) - 2\lambda^{ar}\beta_a l_r - 2\lambda^{ar}b_a l_r + 2\lambda^{ar}\lambda^{st}\beta_a l_{rs}l_t - \lambda^{ar}\lambda^{su}\lambda^{tv}\beta_a\lambda_{rst}l_u l_v \\ &\quad + \lambda^{ar}\lambda^{bs}\beta_{ab}l_r l_s - \lambda^{ab}\beta_a\beta_b + O_p(n^{-3/2}) \\ &= W(\psi) - 2\lambda^{ar}\beta_a\mathring{l}_r - 2\lambda^{ar}(\mathring{b}_a + \mathring{\delta}_a)\mathring{l}_r + 2\lambda^{ar}\lambda^{st}\beta_a(\mathring{l}_{rs} + \mathring{\Delta}_{rs})\mathring{l}_t - \lambda^{ar}\lambda^{su}\lambda^{tv}\beta_a\lambda_{rst}\mathring{l}_u\mathring{l}_v \\ &\quad + \lambda^{ar}\lambda^{bs}\beta_{ab}\mathring{l}_r\mathring{l}_s - \lambda^{ab}\beta_a\beta_b + O_p(n^{-3/2}),\end{aligned}$$

and thus

$$\begin{aligned}\mathring{E}\{\bar{W}(\psi)\} = \mathring{E}\{W(\psi)\} &- 2\lambda^{ar}\mathring{b}_a\mathring{l}_r + 2\lambda^{ar}\lambda^{st}\beta_a\mathring{\lambda}_{rs,t} + \lambda^{ar}\lambda^{su}\lambda^{tv}\beta_a\lambda_{rst}\mathring{\lambda}_{uv} \\ &- \lambda^{ar}\lambda^{bs}\beta_{ab}\mathring{\lambda}_{rs} - \lambda^{ab}\beta_a\beta_b + O_p(n^{-3/2}).\end{aligned}$$

Barndorff-Nielsen & Cox (1984) showed that $\mathring{E}\{W(\psi)\} = E\{W(\psi)\} + O_p(n^{-3/2})$; recall that $\mathring{\lambda}_{rs} = \lambda_{rs} + O_p(n^{1/2})$ and $\mathring{\lambda}_{rs,t} = \lambda_{rs,t} + O_p(n^{1/2})$. Then, $\lambda^{ru}\lambda^{st}\mathring{\lambda}_{ut} = \lambda^{rs} + O_p(n^{-3/2})$, and

$$\mathring{E}\{\bar{W}(\psi)\} = E\{W(\psi)\} + \lambda^{ar}\lambda^{st}\beta_a(2\lambda_{rs,t} + \lambda_{rst}) - 2\lambda^{ar}\mathring{E}(\mathring{b}_a\mathring{l}_r) - \lambda^{ab}(\beta_{ab} + \beta_a\beta_b) + O_p(n^{-3/2}).$$

Now, using the result that $\lambda^{ar}\mathring{E}(\mathring{b}_a\mathring{l}_r) = \lambda^{ar}\mathring{\beta}_{a/r} - \lambda^{ab}\mathring{\beta}_{ab} = \lambda^{ar}\beta_{a/r} - \lambda^{ab}\beta_{ab} + O_p(n^{-3/2})$, which holds since $\mathring{\beta}_{a/r} = \beta_{a/r} + O_p(n^{-1/2})$ and $\mathring{\beta}_{ab} = \beta_{ab} + O_p(n^{-1/2})$, we have

$$\begin{aligned}\mathring{E}\{\bar{W}(\psi)\} &= E\{W(\psi)\} + \lambda^{ar}\lambda^{st}\beta_a(2\lambda_{rs,t} + \lambda_{rst}) - 2\lambda^{ar}\beta_{a/r} + \lambda^{ab}(\beta_{ab} - \beta_a\beta_b) + O_p(n^{-3/2}) \\ &= E\{\bar{W}(\psi)\} + O_p(n^{-3/2}),\end{aligned}$$

as required. □

References

Barndorff-Nielsen, O. E. & Cox, D. R. (1984). Bartlett adjustments to the likelihood ratio statistic and the distribution of the maximum likelihood estimator. *J.R. Statist. Soc.* B **46**, 483–95.

DiCiccio, T. J. & Stern, S. E. (1994b). Frequentist and Bayesian Bartlett correction of test statistics based on adjusted profile likelihoods. *J.R. Statist. Soc.* B **56**, 397–408.