LTCC/Reid: Approximations November 12, 2012

Laplace approximation of integrals

1. $\int_a^b e^{-ng(y)}dy = e^{-ng(\tilde{y})}\sqrt{\frac{2\pi}{n}}\{g''(\tilde{y})\}^{-1/2}\{1 + \frac{5\tilde{\rho}_3^2 - 3\tilde{\rho}_4}{24n} + O(n^{-2})\},$

   where $g'(\tilde{y}) = 0$, $g''(\tilde{y}) > 0$, $\tilde{\rho}_3 = g'''(\tilde{y})/\{g''(\tilde{y})\}^{3/2}$, $\tilde{\rho}_4 = g^{(4)}(\tilde{y})/\{g''(\tilde{y})\}^2$, and we assume here and in the following that the function g has a unique non-zero minimum in the interval (a, b).

2. $$\int h(y)e^{-ng(y)} = h(\tilde{y})e^{-ng(\tilde{y})}\sqrt{\frac{2\pi}{n}}\{g''(\tilde{y})\}^{-1/2}\{1 + \frac{5\tilde{\rho}_3^2 - 3\tilde{\rho}_4}{24n} + \frac{h''(\tilde{y})}{2g''(\tilde{y})h(\tilde{y})n} - \frac{\tilde{\rho}_3 h'(\tilde{y})/h(\tilde{y})}{2\{g''(\tilde{y})\}^{1/2}n} + O(n^{-2})\}\ ,$$

   for which we need to assume that $h(\tilde{y}) \neq 0$.

3. $$\begin{aligned} \int_{R^d} h(y)e^{-ng(y)}dy &= e^{-ng(\tilde{y})}h(\tilde{y})\left(\sqrt{\frac{2\pi}{n}}\right)^d |g''(\tilde{y})|^{-1/2}\{1 + O(n^{-1})\} \\ &= e^{-ng(y^*)}h(y^*)\left(\sqrt{\frac{2\pi}{n}}\right)^d |g''(y^*)|^{-1/2}\{1 + O(n^{-1})\} \end{aligned}$$

4. $$\begin{aligned} \pi(\theta \mid y) &= \frac{\exp\{\ell(\theta; y)\}\pi(\theta)}{\int \exp\{\ell(\theta; y)\}\pi(\theta)d\theta} \\ &\doteq \frac{\exp\{\ell(\theta; y)\}\pi(\theta)}{\exp\{\ell(\hat{\theta}; y)\}\pi(\hat{\theta})|j(\hat{\theta})|^{-1/2}\sqrt{(2\pi)^d}} \\ &= \frac{1}{\sqrt{(2\pi)^d}}e^{\ell(\theta)-\ell(\hat{\theta})}|j(\hat{\theta})|^{1/2}\frac{\pi(\theta)}{\pi(\hat{\theta})} \qquad (1) \end{aligned}$$

5. $$\begin{aligned} \pi_m(\psi \mid y) &= \frac{\int \exp\{\ell(\psi, \lambda; y)\}\pi(\psi, \lambda)d\lambda}{\int \exp\{\ell(\psi, \lambda; y)\}\pi(\psi, \lambda)d\psi d\lambda} \\ &\doteq \frac{\exp\{\ell(\psi, \hat{\lambda}_\psi)\}\pi(\psi, \hat{\lambda}_\psi)|j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|^{-1/2}\sqrt{(2\pi)^{d-q}}}{\exp\{\ell(\hat{\psi}, \hat{\lambda})\}\pi(\hat{\psi}, \hat{\lambda})|j(\hat{\psi}, \hat{\lambda})|^{-1/2}\sqrt{(2\pi)^d}} \\ &= \frac{1}{\sqrt{(2\pi)^q}}e^{\ell_P(\psi)-\ell_P(\hat{\psi})}j_P^{1/2}(\hat{\psi})\frac{\pi(\psi, \hat{\lambda}_\psi)}{\pi(\hat{\psi}, \hat{\lambda})}\frac{|j_{\lambda\lambda}(\hat{\psi}, \hat{\lambda})|^{1/2}}{|j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|^{1/2}} \qquad (2) \end{aligned}$$

Saddlepoint approximation for sums

1. Assume $Y_1, \ldots, Y_n$ are iid from a density $f_Y(y), y \in \mathbb{R}$, with moment generating function $M_Y(t) = E\{\exp(tY)\} = \int \exp(ty) f_Y(y;\theta)$ and cumulant generating function $K_Y(t) = \log M_Y(t)$ (each for a single observation). We assume enough smoothness in the model that $K_Y(t)$ has a series expansion about 0 of the form

$$K_Y(t) = \kappa_1 t + \frac{1}{2}\kappa_2 t^2 + \frac{1}{6}\kappa_3 t^3 + \frac{1}{24}\kappa_4 t^4 + \ldots;$$

note that $\kappa_1 = \mu = E(Y_1)$, $\kappa_2 = \sigma^2 = \text{Var}(Y_1)$, $\kappa_3 = E(Y_1 - \mu)^3$, and $\kappa_4 = E(Y_1 - \mu)^4 - 3\sigma^4$. The standardized 3rd and 4th cumulants are

$$\rho_3 = \frac{\kappa_3}{\sigma^{3/2}}, \quad \rho_4 = \frac{\kappa_4}{\sigma^2}.$$

2. Let $S_n = \sum_{i=1}^n Y_i$. The saddlepoint approximation for the density of S_n is

$$\hat{f}_{S_n}(s) = \frac{1}{\sqrt{2\pi}} \frac{1}{\{nK_Y''(\hat{\phi})\}^{1/2}} \exp\{nK_Y(\hat{\phi}) - \hat{\phi}s\} \tag{3}$$

where $\hat{\phi}$ satisfies the equation $nK_Y'(\hat{\phi}) = s$.

(a) Under smoothness conditions on the density $f(\cdot)$, it can be shown that

$$f_{S_n}(s) = \hat{f}_{S_n}(s)\{1 + O(n^{-1})\},$$

and that the $O(n^{-1})$ term in (3) is $(3\hat{\rho}_4 - 5\hat{\rho}_3^2)/(24n)$ where $\hat{\rho}_j = \rho_j(\hat{\phi})$.

(b) If (3) is renormalized to integrate to 1, it approximates the density of S_n with relative error $O(n^{-3/2})$. It is more usual to assume this renormalization, and write either

$$\hat{f}_{S_n}(s) = \frac{c}{\{nK_Y''(\hat{\phi})\}^{1/2}} \exp\{nK_Y(\hat{\phi}) - \hat{\phi}s\} \qquad \text{or}$$

$$\hat{f}_{S_n}(s) = \frac{c}{\sqrt{2\pi}} \frac{1}{\{nK_Y''(\hat{\phi})\}^{1/2}} \exp\{nK_Y(\hat{\phi}) - \hat{\phi}s\}$$

where c here is used as a 'generic' constant. The second form is a useful reminder that the leading term in the renormalization constant is $1/\sqrt{(2\pi)}$.

(c) A simple change of variables gives a saddlepoint approximation to the density of $\bar{Y}_n = S_n/n$.

(d) If Y_i are d-dimensional vectors, then $M_Y(t) = E\exp(t^T Y)$ and

$$\hat{f}_{S_n}(s) = \frac{c}{\sqrt{2\pi}^d} \frac{1}{\{n|K_Y''(\hat{\phi})|\}^{1/2}} \exp\{nK_Y(\hat{\phi}) - \hat{\phi}^T s\} \tag{4}$$

where $K_Y'(\phi)$ is a $d \times 1$ vector and $K_Y''(\phi)$ is a $d \times d$ matrix.

3. If $f_Y(y) = f_Y(y;\theta) = \exp\{\theta^T y - k(\theta) - d(y)\}$, then $K(\phi) = k(\theta+\phi) - k(\theta)$, $\hat\phi = \hat\theta - \theta$, and (3) becomes

$$\hat f_{\hat\Theta}(\hat\theta;\theta) = \frac{c}{\sqrt{2\pi}^d} |j(\hat\theta)|^{1/2} \exp\{\ell(\theta) - \ell(\hat\theta)\} \tag{5}$$

using the change of variable $nk'(\hat\theta) = s$: compare (1). In (5) $\ell(\theta) = \ell(\theta; y_1, \ldots, y_n) = \ell(\theta;\hat\theta) = \ell(\theta; s) = \theta^T s - nk(\theta) = \theta^T nk'(\hat\theta) - nk(\theta)$.[1] Equivalently

$$\hat f_S(s;\theta) = \frac{c}{\sqrt{(2\pi)^d}} |j(\hat\theta)|^{-1/2} \exp\{\ell(\theta) - \ell(\hat\theta)\}. \tag{6}$$

4. If $\theta = (\psi, \lambda)$ then we write $f_Y(y) = \exp\{\psi^T y_1 + \lambda^T y_2 - k(\psi,\lambda)\} h(y)$, where y_1 and y_2 are sub-vectors, then

$$\begin{aligned} f(s_1 \mid s_2; \psi) &= \frac{f(s_1, s_2; \psi, \lambda)}{f(s_2; \psi, \lambda)} \\ &\doteq \frac{c}{\sqrt{(2\pi)^q}} |j_p(\hat\psi)|^{-1/2} \exp\{\ell_p(\psi) - \ell_p(\hat\psi)\} \left\{ \frac{|j_{\lambda\lambda}(\hat\psi, \hat\lambda)|}{|j_{\lambda\lambda}(\psi, \hat\lambda_\psi)|} \right\}^{-1/2}, \end{aligned}$$

where $s_1 = \sum_{i=1}^n y_{1i}$, $s_2 = \sum_{i=1}^n y_{2i}$. This uses the results $|j(\hat\theta)| = |j^{\psi\psi}(\hat\theta)|^{-1} |j_{\lambda\lambda}(\hat\theta)|$, and $-\ell_p''(\hat\psi) = \{j^{\psi\psi}(\hat\theta)\}^{-1}$.

Homework problems

1. Suppose $Y_1, \ldots, Y_n$ are i.i.d. with density

$$f_{Y_i}(y;\mu) = \frac{1}{\mu} \exp(-\frac{y}{\mu}), y > 0, \mu > 0.$$

Show that the leading term in the saddlepoint approximation to the density of $\bar Y = \hat\mu$ reproduces the gamma density, with $\Gamma(n)$ replaced by Stirling's approximation to it. Deduce that the renormalized saddlepoint approximation is exact.

2. (A Neyman-Scott problem). A class of problems where maximum likelihood estimators are not consistent are those in which the number of nuisance parameters increases with the sample size. These are often called Neyman-Scott problems. For example, if $Y_{ij}, j = 1, \ldots, m_i; i = 1, \ldots, n$ follow a $N(\mu_i, \sigma^2)$ distribution, the maximum likelihood estimator of σ^2 is inconsistent as $n \to \infty$; in particular if $m_i \equiv 2$, then $\hat\sigma^2 \to \sigma^2/2$; see CH Example 9.24.

[1] This is just a cumbersome way of saying that the maximum likelihood estimator is a one-to-one function of the minimal sufficient statistic.

(a) Suppose that Y_{i1} and Y_{i2} are independent observations from exponential distributions with means $\psi\lambda_i$ and ψ/λ_i, respectively, $i = 1, \ldots, n$. Show that the maximum likelihood estimator of ψ is not consistent, but converges in probability to $(\pi/4)\psi$.

(b) A modification to the profile likelihood to account for estimation of nuisance parameters was proposed in Cox & Reid (1987):

$$\ell_m(\psi) = \ell(\psi, \hat{\lambda}_\psi) - \frac{1}{2}\log|j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|,$$

where $\lambda = (\lambda_1, \ldots, \lambda_n)$ and $\hat{\lambda}_\psi$ is the constrained maximum likelihood estimator of λ. This is to be computed using a parametrization of the nuisance parameter that is *orthogonal* to the parameter of interest ψ, with respect to expected Fisher information. (The correction term $\frac{1}{2}\log|j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|$ is not invariant to reparameterizations,) Show for the exponential case that λ is orthogonal to ψ, and that the value of ψ that solves $\ell_m'(\psi) = 0$, $\hat{\psi}_m$, say, converges to $(\pi/3)\psi$.

References

[SM] Davison, A.C. (2003). *Statistical Models.* Cambridge University Press, Cambridge. Ch 11.3

[BNC89] Barndorff-Nielsen, O.E. & Cox, D.R. (1989). *Asymptotic Techniques for use in Statistics.* Chapman & Hall.

Tierney, L.J. & Kadane, J.B. (1986). Accurate approximations for posterior moments and marginal densities. *J. Am. Statist. Assoc.* **81**, 82–86.

Neyman, J. and Scott, E.L. (1948). Consistent estimates based on partially consistent observations. *Econometrica* **16**, 1–32.

Cox, D.R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *JRSS* B **49**, 1–39.