Statistical Asymptotics
Part III: Higher-order Theory

Andrew Wood

School of Mathematical Sciences
University of Nottingham

APTS, April 11-15, 2016

Structure of Chapter

Coverage of this chapter: some techniques which go beyond the first-order theory.

Topics: Edgeworth expansions, saddlepoint approximations, Laplace approximations, Bartlett correction, Bayesian asymptotics, the p^* formula, modified profile likelihood.

Motivation: to improve on first-order asymptotic results by deriving approximations whose asymptotic accuracy is higher by one or two orders.

Asymptotic expansion

An asymptotic expansion for a function $g_n(x)$ at some fixed x is expressed as

$$g_n(x) = \gamma_0(x)b_{0,n} + \gamma_1(x)b_{1,n} + \ldots + \gamma_k(x)b_{k,n} + o(b_{k,n}),$$

as $n \to \infty$, where $\{b_{r,n}\}_{r=0}^{k}$ is a sequence such as $\{1, n^{-1/2}, n^{-1}, \ldots n^{-k/2}\}$ or $\{1, n^{-1}, n^{-2}, \ldots, n^{-k}\}$.

For it to be a proper asymptotic expansion, the sequence must have the property that $b_{r+1,n} = o(b_{r,n})$ as $n \to \infty$, for each $r = 0, 1, \ldots, k-1$.

Often the function of interest $g_n(x)$ will be the exact density or distribution function of a statistic based on a sample of size n, and $\gamma_0(x)$ will be some simple first-order approximation, such as the normal density or distribution function.

One important feature of asymptotic expansions is that they are not in general convergent series for $g_n(x)$ for any fixed x: taking successively more terms, letting $k \to \infty$ for fixed n, will not necessarily improve the approximation to $g_n(x)$.

Stochastic asymptotic expansion

For a sequence of random variables $\{Y_n\}$, a stochastic asymptotic expansion is expressed as

$$Y_n = X_0 b_{0,n} + X_1 b_{1,n} + \ldots + X_k b_{k,n} + o_p(b_{k,n}),$$

where $\{b_{k,n}\}$ is a given set of sequences and $\{X_0, X_1, \ldots\}$ have distributions which only depend weakly on n.

Stochastic asymptotic expansions are not as well defined as asymptotic expansions, as there is usually considerable arbitrariness in the choice of the coefficient random variables $\{X_0, X_1, \ldots\}$.

A simple application of stochastic asymptotic expansion is the proof of asymptotic normality of the maximum likelihood estimator.

Asymptotic expansion of log-likelihood

An important example we have already seen is the asymptotic expansion of the score statistic $U(\theta) = \partial l(\theta)/\partial\theta$:

$$\begin{aligned} 0 = l^{(1)}(\hat{\theta}) &= n^{-1/2} l^{(1)}(\theta_0) + n^{-1/2}(\hat{\theta} - \hat{\theta}_0) l^{(2)}(\theta_0) \\ &\quad + n^{-1/2}\frac{1}{2!}(\hat{\theta} - \theta_0)^2 l^{(3)}(\theta_0) \\ &\quad + n^{-1/2}\frac{1}{3!}(\hat{\theta} - \theta_0)^3 l^{(4)}(\theta_0) + \ldots \\ &= Z_1 + \delta Z_2 + n^{-1/2}\frac{1}{2!}\delta^2 Z_3 + n^{-1}\frac{1}{3!}\delta^3 Z_4 + \ldots \\ &= A_0 + n^{-1/2} A_1 + n^{-1} A_2 + \ldots \end{aligned}$$

where $l^{(i)} = \partial^i l(\theta)/\partial\theta^i$, $\delta = n^{1/2}(\hat{\theta} - \theta_0)$, $Z_1 = n^{-1/2} l^{(1)}(\theta_0)$, $Z_i = n^{-1} l^{(i)}(\theta_0)$, $i \geq 2$, $A_0 = Z_1 + \delta Z_2$, $A_1 = \delta^2 Z_3/2$ and $A_2 = \delta^3 Z_4/6$.

In a regular IID framework, δ, the Z_i and the A_i are all $O_p(1)$.

Tools of asymptotic analysis

- Edgeworth expansions.
- Saddlepoint approximations.
- Laplace's method.

Edgeworth expansion

Let $Y_1, Y_2, \ldots, Y_n$ be IID univariate with cumulant generating function $K_Y(t)$ and cumulants κ_r.

Let $S_n = \sum_1^n Y_i$, $S_n^* = n^{-1/2}(S_n - n\mu)/\sigma$

where $\mu \equiv \kappa_1 = E(Y_1), \sigma^2 \equiv \kappa_2 = \mathrm{Var}(Y_1)$.

Define the rth standardised cumulant by $\rho_r = \kappa_r/\kappa_2^{r/2}$.

The Edgeworth expansions for the density of S_n^* is:

$$f_{S_n^*}(x) = \phi(x)\left\{1 + n^{-1/2}\frac{\rho_3}{6}H_3(x) + n^{-1}\left[\frac{\rho_4 H_4(x)}{24} + \frac{\rho_3^2 H_6(x)}{72}\right]\right\} + O(n^{-3/2}).$$

The orders of the terms in the expansion decrease in powers of $n^{-1/2}$.

Here $\phi(x)$ is the standard normal density and $H_r(x)$ is the rth degree Hermite polynomial defined by

$$\begin{aligned} H_r(x) &= (-1)^r \frac{d^r \phi(x)}{dx^r} \Big/ \phi(x) \\ &= (-1)^r \phi^{(r)}(x)/\phi(x), \quad \text{say.} \end{aligned}$$

We have $H_3(x) = x^3 - 3x$, $H_4(x) = x^4 - 6x^2 + 3$ and $H_6(x) = x^6 - 15x^4 + 45x^2 - 15$.

Comments

The leading term in the expansion is the standard normal density, as is appropriate from the CLT.

The $n^{-1/2}$ term is an adjustment for skewness, via the standardised skewness ρ_3.

The n^{-1} term is a simultaneous adjustment for skewness and kurtosis.

If the density of Y_1 is symmetric, $\rho_3 = 0$ and the normal approximation is accurate to order n^{-1}, rather than the usual $n^{-1/2}$ for $\rho_3 \neq 0$.

The accuracy of the Edgeworth approximation, which truncates the expansion, will depend on the value of x.

Edgeworth approximations tend to be poor, and may even be negative, in the tails of the distribution, as $|x|$ increases.

Distribution function

Integrating the Edgeworth expansion using the properties of the Hermite polynomials, gives an expansion for the distribution function of S_n^*:

$$F_{S_n^*}(x) = \Phi(x) - \phi(x)\left\{n^{-1/2}\frac{\rho_3}{6}H_2(x) + \frac{\rho_4}{24n}H_3(x) + \frac{\rho_3^2}{72n}H_5(x)\right\} + O(n^{-3/2}).$$

Also, if T_n is a sufficiently smooth function of S_n^*, then a formal Edgeworth expansion can be obtained for the density of T_n.

Cornish-Fisher expansion

We might wish to determine an x, x_α say, such that $F_{S_n^*}(x_\alpha) = \alpha$, to the order considered in the Edgeworth approximation to the distribution function of S_n^*.

The solution is known as the Cornish-Fisher expansion and the formula is

$$\begin{aligned} x_\alpha &= z_\alpha + \frac{1}{6\sqrt{n}}(z_\alpha^2 - 1)\rho_3 + \frac{1}{24n}(z_\alpha^3 - 3z_\alpha)\rho_4 \\ &\quad - \frac{1}{36n}(2z_\alpha^3 - 5z_\alpha)\rho_3^2 + O(n^{-3/2}), \end{aligned}$$

where $\Phi(z_\alpha) = \alpha$.

Derivation*

The density of a random variable can be obtained by inversion of its characteristic function.

In particular, the density for $\bar{X}$, the sample mean of a set of IID random variables $X_1, \ldots, X_n$, can be obtained as

$$f_{\bar{X}}(\bar{x}) = \frac{n}{2\pi i} \int_{\tau - i\infty}^{\tau + i\infty} \exp\big[n\{K(\phi) - \phi\bar{x}\}\big] d\phi,$$

where K is the cumulant generating function of X, and τ is any point in the open interval around 0 in which the moment generating function M exists.

Edgeworth expansions are obtained by expanding the cumulant generating function in a Taylor series around 0, exponentiating and inverting term by term.

Saddlepoint approximations

Saddlepoint density approximations are now considered.

First of all, we study the saddlepoint density approximation for the density of a single random variable X. Then we will see that it is easy to generalise the approximation to a sum of IID random variables.

Later we will consider further topics including multivariate saddlepoint density approximations, the Lugananni-Rice tail probability approximation and the p^* formula.

Exponential tilting

Let X denote a random variable with cumulant generating function (CGF) $K_0(t)$ and pdf $f_0(x)$.

It is assumed that $K_0(t)$ is finite for all $t \in (-a, b)$ for some $a, b > 0$.

The exponentially tilted pdf $f_t(x)$ is defined by

$$f_t(x) = \exp\{xt - K_0(t)\} f_0(x). \tag{1}$$

Check that $f_t(x)$ is a pdf: clearly $f_t(x) \geq 0$, and

$$\int f_t(x)dx = 1 \quad \text{because} \quad E_0[\exp(tX)] = \exp\{K_0(t)\}.$$

Exponential tilting (continued)

Let $K_t(\phi)$ denote the CGF of the distribution with pdf $f_t(x)$.

Then it is easy to see that

$$K_t(\phi) = K_0(t+\phi) - K_0(t).$$

Note that the mean and variance of this distribution are given by

$$\left.\frac{\partial K_t}{\partial \phi}(\phi)\right|_{\phi=0} = K_0'(t).$$

and

$$\left.\frac{\partial^2 K_t}{\partial \phi^2}(\phi)\right|_{\phi=0} = K_0''(t).$$

Main idea

The main idea used in the real-variable derivation of the saddlepoint approximation can be expressed in two steps.

Step 1.
Given the x at which we wish to calculate $f_0(x)$, choose $\hat{t}$ so that the mean of the distribution with pdf $f_{\hat{t}}$ is x.

Step 2.
Then approximate $f_{\hat{t}}(x)$ [the tilted density $f_{\hat{t}}$ evaluated at its mean] by a normal density evaluated at its mean [where the mean and variance of the approximating normal are the same as the mean and variance of $f_{\hat{t}}$].

The SP approximation

For Step 1, choose $\hat{t}$ to solve the saddlepoint equation $K_0'(\hat{t}) = x$.

Note: the solution is unique because cumulant generating functions are convex.

In Step 2, we approximate $f_{\hat{t}}(x)$ by $\frac{1}{\{2\pi K_0''(\hat{t})\}^{1/2}}$.

Rearranging (1), we obtain the first-order saddlepoint density approximation

$$\hat{f}_0(x) = \frac{1}{\{2\pi K_0''(\hat{t})\}^{1/2}} \exp\{K_0(\hat{t}) - x\hat{t}\}. \qquad (2)$$

SP approximation for IID sums

Let us now return to the IID sum $S_n = \sum_{i=1}^n Y_i$ where each Y_i has CGF $nK_Y(\phi)$ and $f_{S_n}(s)$ is the pdf of S_n at s.

Substituting into (2), it is seen that the first-order saddlepoint density approximation to $f_{S_n}(s)$ is given by

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

where $\hat{\phi}$ is chosen to solve the saddlepoint equation $nK_Y'(\hat{\phi}) = s$.

Accuracy of the SP approximation

In the case where the Edgeworth approximation to the tilted distribution is valid, the error in Step 2 is $O(n^{-1})$.

Consequently, in this case, we may write

$$f_{S_n}(s) = \hat{f}_{S_n}(s)\{1 + O(n^{-1})\}, \tag{4}$$

where $\hat{f}_{S_n}(s)$ is defined in (3).

Three important points to note about (4):

1. The $O(n^{-1})$ error is a relative error.

2. The $O(n^{-1})$ relative error is uniform for s such that $\hat{\phi} = \hat{\phi}(s)$ lies in a fixed, open interval containing 0.

3. As a consequence, the accuracy statement is valid uniformly in a large deviation region for S_n, i.e. and interval of the form $s \in (\mu n - cn, \mu n + cn)$ for some $c > 0$.

The $O(n^{-1})$ term is actually

$$(3\hat{\rho}_4 - 5\hat{\rho}_3^2)/(24n),$$

where

$$\hat{\rho}_j \equiv \hat{\rho}_j(\hat{\phi}) = \frac{K_Y^{(j)}(\hat{\phi})}{\{K_Y''(\hat{\phi})\}^{j/2}} \tag{5}$$

is the jth standardised derivative of the cumulant generating function for Y_1 evaluated at $\hat{\phi}$.

A change of variable gives an expansion for the density of $\bar{Y}_n = S_n/n$:

$$\hat{f}_{\bar{Y}_n}(y) = \left\{ \frac{n}{(2\pi)K''_Y(\hat{\phi})} \right\}^{1/2} \exp\big\{ n[K_Y(\hat{\phi}) - \hat{\phi}y] \big\},$$

where now $K'_Y(\hat{\phi}) = y$.

Again, we have

$$f_{\bar{Y}_n}(y) = \hat{f}_{\bar{Y}_n}(y)\{1 + O(n^{-1})\},$$

which is value uniformly in a large deviation region for $\bar{Y}_n$, i.e. and interval of the form $y \in (\mu - c, \mu + c)$ for some $c > 0$.

Second-order SP approximation

A second-order SP approximation for $f_{S_n}(s)$ is given by

$$\hat{f}^{[2]}_{S_n}(s) = \hat{f}_{S_n}(s)\left\{1 + \frac{1}{n}\left(\frac{3\hat{\rho}_4 - 5\hat{\rho}_3^2}{24}\right)\right\}, \qquad (6)$$

where $\hat{f}_{S_n}(s)$ is the first-order SP approximation given in (3), and $\hat{\rho}_3$ and $\hat{\rho}_4$ are defined in (5).
In this case,

$$f_{S_n}(s) = \hat{f}^{[2]}_{S_n}(s)\{1 + O(n^{-2})\}.$$

The approximation (6) can be derived by using an Edgeworth expansion, with the n^{-1} term included, for the approximation of the relevant tilted pdf evaluated at s (see Step 2 above).

The second-order SP approximation will not be considered further in these lectures, but it is typically more accurate in practice than the first-order approximation.

Multivariate saddlepoint approximation

Suppose that $X = (X^1, \ldots, X^m)^\top$ is now a random vector with pdf $f_0(x)$ and CGF $K_0(t)$ where now $t = (t^1, \ldots, t^m)^T$.

The first-order multivariate saddlepoint approximation for $f_0(x)$ is:

$$\hat{f}_0(x) = \frac{1}{(2\pi)^{m/2}|\nabla_t \nabla_t^\top K_0(\hat{t})|^{1/2}} \exp\{K_0(\hat{t}) - x^\top \hat{t}\},$$

where now $\hat{t}$ solves the vector saddlepoint equation

$$\nabla_t K_0(\hat{t}) = x.$$

and $|\nabla_t \nabla_t K_0(\hat{t})|$ is the determinant of the Hessian of $K_0(t)$ evaluated at $t = \hat{t}$.

Multivariate saddlepoint approximation (continued)

The derivation via exponential tilting is similar to that in the univariate case. The main difference is that in Step 2 we now approximate $f_{\hat{t}}(x)$ by

$$\frac{1}{(2\pi)^{m/2}|\Sigma|^{1/2}},$$

which is the density of the multivariate normal distribution $N_p(\mu, \Sigma)$ evaluated at its mean μ, where

$$\Sigma = \nabla_t \nabla_t^\top K_0(\hat{t})$$

is the Hessian of $K_0(t)$ evaluated at $t = \hat{t}$.

SP approximation for IID sums: multivariate case

Suppose that $S_n = \sum_{i=1}^n Y_i$ where now the Y_i are IID m-dimensional random vectors with common CGF $K_Y(\phi)$.

Then, from above,

$$\hat{f}_{S_n}(s) = \frac{1}{(2\pi)^{m/2} n^{m/2} |\nabla_\phi \nabla_\phi^\top K_Y(\hat{\phi})|^{1/2}} \exp\{nK_Y(\hat{\phi}) - s^\top \hat{\phi}\},$$

where $\hat{\phi}$ solves the multivariate SP equation $n\nabla_\phi K_Y(\hat{\phi}) = s$; and

$$\hat{f}_{\bar{Y}_n}(y) = \frac{n^{m/2}}{(2\pi)^{m/2} |\nabla_\phi \nabla_\phi^\top K_Y(\hat{\phi})|^{1/2}} \exp\{n[K_Y(\hat{\phi}) - y^\top \hat{\phi}]\},$$

where now $\bar{Y}_n = n^{-1} S_n$ and again $\hat{\phi}$ solves $\nabla_\phi K_Y(\hat{\phi}) = y = n^{-1}s$.

In the multivariate case, as in the univariate case,

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

and

$$f_{\bar{Y}_n}(y) = \hat{f}_{\bar{Y}_n}(y)\{1 + O(n^{-1})\}.$$

As in the univariate case, these relative error statements are valid uniformly in large deviation regions.

Comparison with Edgeworth expansion

The comments below apply to the univariate case; similar comments apply to the multivariate case.

To use the saddlepoint expansion to approximate $f_{\bar{Y}_n}(y)$ it is necessary to know the whole cumulant generating function, not just the first four cumulants.

Also necessary to solve the equation $K'_Y(\hat{\phi}) = y$ for each value of y.

The leading term in saddlepoint expansion is not the normal (or any other) density; in fact it will not usually integrate to 1, although it can be renormalised to do so.

The saddlepoint expansion is an asymptotic expansion in powers of n^{-1}, rather than $n^{-1/2}$ as in the Edgeworth expansion. The main correction for skewness has been absorbed by the leading term.

More on Accuracy

Saddlepoint approximations are generally very accurate.

Even when $n = 1$, SP approximations often do very well in practice.

With distributions that differ from the normal density through asymmetry, such as the gamma distribution, the saddlepoint approximation is extremely accurate throughout the range of s.

In many important cases, the relative error of the SP approximation remains bounded throughout the domain of the distribution.

Renormalisation

We may consider using a renormalised version of the approximation to $f_{\bar{Y}_n}(y)$:

$$\tilde{f}_{\bar{Y}_n}(y) = c\{n/K''_Y(\hat{\phi})\}^{1/2} \exp\big[n\{K_Y(\hat{\phi}) - \hat{\phi}y\}\big]$$

where c is determined, usually numerically, so that the right-hand side integrates to 1.

If the $O(n^{-1})$ correction term is constant in y, the renormalised approximation will be exact. For scalar random variables this happens only in the case of the normal, gamma and inverse Gaussian distributions.

In general, the n^{-1} correction term $\{3\hat{\rho}_4(\hat{\phi}) - 5\hat{\rho}_3^2(\hat{\phi})\}/24$ varies only slowly with y and the relative error in the renormalised approximation is $O(n^{-3/2})$.

Distribution function approximation

It is usually not possible to integrate the saddlepoint approximation theoretically to obtain an approximation to the distribution function of S_n.

However, one can do this by numerical integration of the saddlepoint density approximation.

Lugannani-Rice

A useful alternative approach is given by the Lugannani-Rice approximation:

$$P[S_n \leq s] \equiv F_{S_n}(s) = \Phi(r_s) + \phi(r_s)\left(\frac{1}{r_s} - \frac{1}{v_s}\right) + O(n^{-1}),$$

where

$$\begin{aligned} r_s &= \mathrm{sgn}(\hat{\phi})\sqrt{2n\{\hat{\phi}K'_Y(\hat{\phi}) - K_Y(\hat{\phi})\}} \\ v_s &= \hat{\phi}\sqrt{nK''_Y(\hat{\phi})}, \end{aligned}$$

and $\hat{\phi} \equiv \hat{\phi}(s)$ is the saddlepoint, satisfying $nK'_Y(\hat{\phi}) = s$.

An alternative approximation

The expansion can be expressed in the asymptotically equivalent form

$$F_{S_n}(s) = \Phi(r_s^*)\{1 + O(n^{-1})\},$$

with

$$r_s^* = r_s - \frac{1}{r_s} \log \frac{r_s}{v_s}.$$

Exponential family case

Suppose $f(y)$ is itself in the exponential family,

$$f(y;\theta) = \exp\{y\theta - c(\theta) - h(y)\}.$$

Then since $K_Y(t) = c(\theta + t) - c(\theta)$, it follows that
$\hat{\lambda} \equiv \hat{\lambda}(s) = \hat{\theta} - \theta$, where $\hat{\theta}$ is the MLE based on $s = y_1 + \cdots + y_n$.

The SP approximation for $f_{S_n}(s;\theta)$ is

$$\hat{f}_{S_n}(s;\theta) = \frac{1}{\{2\pi n c''(\hat{\theta})\}^{1/2}} \exp\big[n\{c(\hat{\theta}) - c(\theta)\} - (\hat{\theta} - \theta)s\big]$$

which can be expressed as

$$\hat{f}_{S_n}(s;\theta) = c \exp\{l(\theta) - l(\hat{\theta})\}|j(\hat{\theta})|^{-1/2}$$

where $l(\theta)$ is the log-likelihood function based on $(y_1, \ldots, y_n)$, or s, and $j(\hat{\theta})$ is the observed information.

Since $\hat{\theta} = \hat{\theta}(s)$ is a one-to-one function of s, with Jacobian $|j(\hat{\theta})|$, we can obtain an approximation to the density of $\hat{\theta}$

$$\hat{f}_{\hat{\theta}}(\hat{\theta}; \theta) = c \exp\{l(\theta) - l(\hat{\theta})\}|j(\hat{\theta})|^{1/2}.$$

This is a particular case of Barndorff-Nielsen's p^* formula, considered later.

Laplace approximation of integrals

The aim here is to obtain an approximation for the integral

$$g_n = \int_a^b e^{-ng(y)} dy.$$

The main contribution, for large n, will come from values of y near the minimum of $g(y)$, which may occur at a or b, or in the interior of the interval (a, b).

Assume that $g(y)$ has a unique global minimum over (a, b) at $\tilde{y} \in (a, b)$ and that $g'(\tilde{y}) = 0$, $g''(\tilde{y}) > 0$.

We can write

$$\begin{aligned} g_n &= \int_a^b e^{-n\{g(\tilde{y})+\frac{1}{2}(\tilde{y}-y)^2 g''(\tilde{y})+\cdots\}} dy \\ &\approx e^{-ng(\tilde{y})} \int_a^b e^{-\frac{n}{2}(\tilde{y}-y)^2 g''(\tilde{y})} dy \\ &\approx e^{-ng(\tilde{y})} \sqrt{\frac{2\pi}{ng''(\tilde{y})}} \int_{-\infty}^{\infty} \phi\left(y - \tilde{y}; \frac{1}{ng''(\tilde{y})}\right) dy \end{aligned}$$

where $\phi(y - \mu; \sigma^2)$ is the density of $N(\mu, \sigma^2)$.

Since ϕ integrates to one,

$$g_n \approx e^{-ng(\tilde{y})} \sqrt{\frac{2\pi}{ng''(\tilde{y})}}.$$

A more detailed analysis gives

$$g_n = e^{-ng(\tilde{y})} \sqrt{\frac{2\pi}{ng''(\tilde{y})}} \left\{ 1 + \frac{5\tilde{\rho}_3^2 - 3\tilde{\rho}_4}{24n} + O(n^{-2}) \right\},$$

where

$$\begin{aligned} \tilde{\rho}_3 &= g^{(3)}(\tilde{y}) / \{g''(\tilde{y})\}^{3/2}, \\ \tilde{\rho}_4 &= g^{(4)}(\tilde{y}) / \{g''(\tilde{y})\}^{2}. \end{aligned}$$

A similar analysis gives

$$\int_a^b h(y)e^{-ng(y)}dy = h(\tilde{y})e^{-ng(\tilde{y})}\sqrt{\frac{2\pi}{ng''(\tilde{y})}}\{1+O(n^{-1})\}.$$

Provided $h(y) > 0$, a further refinement of the method is possible:

$$\begin{aligned}
&\int_a^b e^{-n\{g(y)-\frac{1}{n}\log h(y)\}}\,dy \\
&= \int_a^b e^{-nq_n(y)}\,dy, \quad \text{say}, \\
&= e^{-ng(y^*)}h(y^*)\sqrt{\frac{2\pi}{nq_n''(y^*)}} \\
&\quad \times \{1 + n^{-1}(5\rho_3^{*2} - 3\rho_4^*)/24 + O(n^{-2})\},
\end{aligned}$$

where

$$q_n'(y^*) = 0, \rho_j^* = q_n^{(j)}(y^*)/\{q_n''(y^*)\}^{j/2}.$$

Multivariate Laplace approximation

Suppose now that $y \in D \subseteq \mathbb{R}^m$ where D is a connected open region.

Assume that $g(y)$ is a (smooth) real-valued function with a unique global minimum over D at $y = \tilde{y} \in D$.

Then the multivariate Laplace approximation is given by

$$\int_{y \in D} h(y) g^{-ng(y)} dy = h(\tilde{y}) e^{-ng(\tilde{y})} \frac{(2\pi)^{m/2}}{|\nabla_y \nabla_y^\top g(\tilde{y})|^{1/2}} \{1 + O(n^{-1})\}, \tag{7}$$

where $|.|$ is the determinant of the Hessian of $g(y)$ evaluated at $y = \tilde{y}$.

Comments

Assume that $h(y) > 0$ for all $y \in D$.

1. In (7), the convention is usually adopted that $h(y)$ does not play a role in the minimisation; it is $g(y)$ in the exponent that is minimised. However, the value of the Laplace approximation does depend on how we define g and h. Clearly g and h are not uniquely defined, so some care is needed.

2. Even when $g(y)$ has a unique global minimum over D at $y = \tilde{y}$, it is possible that g has other local minima whose contribution to the integral on the LHS of (7) is non-neglible unless n is very large. In such cases, the Laplace approximation may not do so well for moderate values of n.

Comments (continued)

3. In many examples, the Laplace approximation does well even when $n = 1$.

4. Inclusion of the n^{-1} term in the approximation reduces the theoretical relative error to $O(n^{-2})$, and often improves the numerical accuracy too.

Bayesian asymptotics

The key result is that the posterior distribution is asymptotically normal. Write

$$\pi_n(\theta \mid y) = f(y;\theta)\pi(\theta) \Big/ \int f(y;\theta)\pi(\theta)d\theta$$

for the posterior density. Denote by $\hat{\theta}$ the MLE.

Proof

For θ in a neighbourhood of $\hat{\theta}$ we have, by Taylor expansion,

$$\log\left\{\frac{f(y;\theta)}{f(y;\hat{\theta})}\right\} \approx -\tfrac{1}{2}(\theta-\hat{\theta})^T j(\hat{\theta})(\theta-\hat{\theta}).$$

Provided the likelihood dominates the prior, we can approximate $\pi(\theta)$ in a neighbourhood of $\hat{\theta}$ by $\pi(\hat{\theta})$.

Then we have

$$f(y;\theta)\pi(\theta) \approx f(y;\hat{\theta})\pi(\hat{\theta})\exp\{-\tfrac{1}{2}(\theta-\hat{\theta})^T j(\hat{\theta})(\theta-\hat{\theta})\}.$$

Then, to first order,

$$\pi_n(\theta \mid y) \sim N\big(\hat{\theta}, j(\hat{\theta})^{-1}\big).$$

Another approximation

When the likelihood does not dominate the prior, expand about the posterior mode $\hat{\theta}_\pi$, which maximises $f(y;\theta)\pi(\theta)$.

Then

$$\pi_n(\theta \mid y) \sim N\big(\hat{\theta}_\pi, j_\pi(\hat{\theta}_\pi)^{-1}\big),$$

where j_π is minus the matrix of second derivatives of $f(y;\theta)\pi(\theta)$.

A more accurate approximation

We have

$$\begin{aligned}\pi_n(\theta \mid y) &= f(y;\theta)\pi(\theta) \Big/ \int f(y;\theta)\pi(\theta)d\theta \\ &\approx \frac{c\exp\{l(\theta;y)\}\pi(\theta)}{\exp\{l(\hat{\theta};y)\}|j(\hat{\theta})|^{-1/2}\pi(\hat{\theta})},\end{aligned}$$

by Laplace approximation of the denominator.

We can rewrite as

$$\pi_n(\theta \mid y) \approx c|j(\hat{\theta})|^{1/2}\exp\{l(\theta) - l(\hat{\theta})\} \times \{\pi(\theta)/\pi(\hat{\theta})\}.$$

Posterior expectations

To approximate to the posterior expectation of a function $g(\theta)$ of interest,

$$E\{g(\theta) \mid y\} = \frac{\int g(\theta) e^{n\bar{l}_n(\theta)} \pi(\theta) d\theta}{\int e^{n\bar{l}_n(\theta)} \pi(\theta) d\theta},$$

where $\bar{l}_n = n^{-1} \sum_{i=1}^{n} \log f(y_i; \theta)$ is the average log-likelihood function.

Rewrite the integrals as

$$E\{g(\theta) \mid y\} = \frac{\int e^{n\{\bar{l}_n(\theta) + q(\theta)/n\}} d\theta}{\int e^{n\{\bar{l}_n(\theta) + p(\theta)/n\}} d\theta}$$

and use the modified version of the Laplace approximation.

Applying this to the numerator and denominator gives

$$\begin{aligned} E\{g(\theta) \mid y\} &\approx \frac{e^{n\bar{l}_n(\theta^*)+q(\theta^*)}}{e^{n\bar{l}_n(\tilde{\theta})+p(\tilde{\theta})}} \\ &\times \frac{\{-n\bar{l}_n''(\tilde{\theta}) - p''(\tilde{\theta})\}^{1/2}}{\{-n\bar{l}_n''(\theta^*) - q''(\theta^*)\}^{1/2}} \frac{\{1+O(n^{-1})\}}{\{1+O(n^{-1})\}} \end{aligned}$$

where θ^* maximises $n\bar{l}_n(\theta) + \log g(\theta) + \log \pi(\theta)$ and $\tilde{\theta}$ maximises $n\bar{l}_n(\theta) + \log \pi(\theta)$.

Detailed analysis shows that the relative error is, in fact, $O(n^{-2})$. If the integrals are approximated in their unmodified form the result is not as accurate.

Logistic regression example: approximate Bayesian analysis

Let us return to the logistic regression model for binary data.

Here,

$$L(\beta|y) = \prod_{i=1}^{n} p_i^{y_i}(1-p_i)^{1-y_i}, \qquad y_i = 0, 1,$$

where

$$p_i = \frac{e^{\beta^\top x_i}}{1 + e^{\beta^\top x_i}},$$

$\beta = (\beta^1, \ldots, \beta^d)^\top$ is the parameter vector and x_i is a d-dimensional covariate vector, $i = 1, \ldots, n$.

Suppose we wish to perform a Bayesian analysis with a non-informative prior for β.

Take $\pi(\beta) \equiv 1$, so that $\pi(\beta)$ is an improper prior for β.

Logistic regression example (continued)

The posterior $\pi(\beta|y)$ for β is given by

$$\pi(\beta|y) = c^{-1}L(\beta),$$

where c is the normalising constant. Applying Laplace's approximation we find

$$c = \int_{\beta\in\mathbb{R}^d} L(\beta)d\beta \approx L(\hat{\beta})\frac{(2\pi)^{d/2}}{|j(\hat{\beta})|^{1/2}},$$

where $\hat{\beta}$ is the MLE of β and $|j(\hat{\beta})|$ is the determinant of the observed information for β evaluated at $\beta = \hat{\beta}$, with

$$j(\hat{\beta}) = \sum_{i=1}^{n} \hat{p}_i(1-\hat{p}_i)x_i x_i^\top.$$

Logistic regression example (continued)

Now suppose we would like to approximate the marginal posterior pdf of β^d, the final component of $\beta = (\beta^1, \ldots, \beta^d)^\top$.

Define $\beta_\gamma = (\beta^1, \ldots, \beta^{d-1}, \gamma)^\top$ and write

$$L_P(\gamma) = \sup_{(\beta^1,\ldots,\beta^{d-1})^\top \in \mathbb{R}^{d-1}} L(\beta_\gamma) = L(\hat{\beta}_\gamma),$$

so that $\hat{\beta}_\gamma$ is the MLE of β under the hypothesis $H_\gamma : \beta^d = \gamma$.

Applying Laplace's approximation again,

$$\int_{(\beta^1,\ldots,\beta^{d-1})^\top \in \mathbb{R}^{d-1}} L(\beta_\gamma) d\beta^1 \cdots d\beta^{d-1} \approx L(\hat{\beta}_\gamma) \frac{(2\pi)^{(d-1)/2}}{|j_{d-1}(\hat{\beta}_\gamma)|^{1/2}},$$

where j_{d-1} is the $(d-1) \times (d-1)$ submatrix of j, with column d and row d removed.

Logistic regression example (continued)

So we may approximate the posterior density of $\beta^d = \gamma$ by

$$\pi(\gamma|y) \approx \frac{L(\hat{\beta}_\gamma)(2\pi)^{(d-1)/2}}{|j_{d-1}(\hat{\beta}_\gamma)|^{1/2}} \Big/ \frac{L(\hat{\beta})(2\pi)^{d/2}}{|j(\hat{\beta})|^{1/2}}$$
$$= \frac{L(\hat{\beta}_\gamma)}{(2\pi)^{1/2} L(\hat{\beta})} \left\{ \frac{|j(\hat{\beta})|}{|j_{d-1}(\hat{\beta}_\gamma)|} \right\}^{1/2}.$$

Note that calculation of $\hat{\beta}$ and $\hat{\beta}_\gamma$ is routine in a standard generalised linear model analysis using e.g. glm in R.

Under mild conditions on the sequence of covariate vectors $x_1, x_2, \ldots$, the relative error in the above Laplace approximation for $\pi(\gamma|y)$ is $O(n^{-1})$.

Bartlett correction

The first-order approximation to the distribution of the likelihood ratio statistic $w(\psi)$ is

$$\Pr_\theta\{w(\psi) \leq \omega^\circ\} = P\{\chi^2_q \leq \omega^\circ\}\{1 + O(n^{-1})\},$$

where $q = d_\psi, \theta = (\psi, \lambda)$, say.

In the case of IID sampling, it can be shown that

$$E_\theta[w(\psi)] = q\{1 + b(\theta)/n + O(n^{-2})\},$$

and so $E_\theta[w'(\psi)] = q\{1 + O(n^{-2})\}$, where $w' = w/\{1 + b(\theta)/n\}$.

The adjustment procedure of replacing w by w', is known as Bartlett correction.

Discussion

Bartlett correction yields remarkably good results under continuous models.

Division by $\{1 + b(\theta)/n\}$ adjusts not only the mean but simultaneously all the cumulants—and hence the whole distribution—of w towards those of χ^2_q. It can be shown that

$$P_\theta\{w'(\psi) \leq \omega^\circ\} = P\{\chi^2_q \leq \omega^\circ\}\{1 + O(n^{-2})\}.$$

In practice, $b(\theta)$ will be replaced by $b(\psi, \hat{\lambda}_\psi)$. The above result still holds, even to $O(n^{-2})$.

The small $O(n^{-2})$ error resulting from Bartlett correction depends on the special character of the likelihood ratio statistic, and the same device applied to, for instance, the score test does not have a similar effect.

Also, under discrete models this type of adjustment does not generally lead to an improved χ^2 approximation.

The p^* formula

The p^* formula, due to Barndorff-Nielsen, is a general approximation for the conditional pdf of the maximum likelihood estimator $\hat{\theta}$.

Consider a log-likelihood $l(\theta)$ with data vector Y.

The formula is:

$$p^*(\hat{\theta}|a,\theta) = c|\hat{j}|^{1/2} e^{l(\theta)-l(\hat{\theta})}, \tag{8}$$

where $\hat{\theta} = (\hat{\theta}^1, \ldots, \hat{\theta}^d)^\top$ is the MLE, c is a normalising constant, $\hat{j} = j(\hat{\theta})$ is the observed information matrix evaluated at the MLE, $|.|$ denotes determinant, and $a = a(Y)$ is a random vector such that

- $(\hat{\theta}, a)$ is a minimal sufficient reduction of the data Y; and
- a is an ancillary statistic for θ, i.e. the distribution of $a(Y)$ should be independent of θ.

Note that p^* gives the pdf of the distribution of $\hat{\theta}$ conditional on a.

Transformation models

A class of models in which a minimal sufficient reduction of the data $(\hat{\theta}, a)$ exists with a ancillary for θ is the class of transformation models.

The simplest cases of transformation models are the univariate location models and location-scale models.

A location model is a model of the form

$$f_\mu(x) = g(x - \mu),$$

where g is a known pdf and μ is an unknown location parameter.

Transformation models (continued)

A location-scale model has the form

$$f_{\mu,\sigma}(x) = \frac{1}{\sigma} g\left(\frac{x-\mu}{\sigma}\right),$$

where g is a known pdf, μ is an unknown location parameter, and σ is an unknown scale parameter, with $\theta = (\mu, \sigma)^\top$.

In transformation models

- an ancillary statistic a can be constructed such that $(\hat{\theta}, a)$ is minimal sufficient; and
- the p^* formula is typically exact, a result that goes back to R.A. Fisher.

(m, m) exponential family models

A second class of models for which p^* is immediately applicable is the (m, m) exponential family class.

In this case there is no need for an ancillary a because $\hat{\theta}$ is of the same dimension of the natural statistic and in fact is a smooth $1:1$ function of it.

Let $Y_1, \ldots, Y_n$ be an independent sample from a full (m, m) exponential density

$$\exp\{y^\top \phi - K(\phi) + D(y)\}.$$

We have already seen that the saddlepoint approximation to the density of natural parameter vector $\hat{\phi}$ is

$$p^*(\hat{\phi}|\phi) = c|j(\hat{\phi})|^{1/2} e^{l(\phi)-l(\hat{\phi})}, \tag{9}$$

which is of the same form as p^*.

(m, m) exponential family models (continued)

What happens if we consider an alternative parametrisation $\phi = \phi(\theta)$, where the relationship between ϕ and θ is smooth and $1:1$?

Changing variables from $\hat{\phi}$ to $\hat{\theta}$ in (9), we obtain

$$p^*(\hat{\theta}|\theta) = p^*(\hat{\phi}|\phi)\left|\frac{\partial \phi^\top}{\partial \theta}(\hat{\theta})\right| \tag{10}$$

which is not obviously of the same form as (8).

(m, m) exponential family models (continued)

However, we shall show on the next slide that

$$j(\hat{\theta}) = \frac{\partial \phi^\top}{\partial \theta}(\hat{\theta}) j(\hat{\phi}) \frac{\partial \phi}{\partial \theta^\top}(\hat{\theta}), \tag{11}$$

where $j(\hat{\theta})$ is the observed information for θ at $\theta = \hat{\theta}$ and $j(\hat{\phi})$ is the observed information for ϕ at $\phi = \hat{\phi}$.

Consequently,

$$|j(\hat{\theta})|^{1/2} = |j(\hat{\phi})|^{1/2} \left| \frac{\partial \phi}{\partial \theta^\top}(\hat{\theta}) \right|,$$

and so (10) is indeed of the form (8).

Jacobian result for p^*

Using the summation convention, and the chain and product rules for partial differentiation,

$$\frac{\partial l}{\partial \theta^r} = l_i \phi^i_r$$

and

$$\frac{\partial^2 l}{\partial \theta^r \theta^s} = l_i \phi^i_{rs} + l_{ij} \phi^i_r \phi^j_s, \tag{12}$$

where $l_i = \partial l / \partial \phi^i$, $l_{ij} = \partial^2 l / \partial \phi^i \partial \phi^j$, $\phi^i_r = \partial \phi^i / \partial \theta^r$ and $\phi^i_{rs} = \partial^2 l / \partial \theta^r \partial \theta^s$.

The key point is that at the MLE $\hat{\theta}$, $l_i\{\phi(\hat{\theta})\} = 0$, so that the RHS of (12) is equal to (11) when written in matrix form.

Discussion

1. The p^* formula is very useful for transformation models, for which it is generally exact, with $c = c(\theta, a)$ is independent of θ.

2. In $(m.m)$ exponential models, p^* is equivalent to a saddlepoint approximation, perhaps after a smooth $1:1$ transformation of the parameter vector.

3. Outside these two model classes, p^* is generally more difficult to implement, because of the difficulty of constructing ancillary statistics.

4. In general models, ancillary statistics may not exist, or may exist but not be unique. However, there are methods for constructing approximate ancillaries, and also methods for constructing approximations to $p*$ which are discussed later.

Discussion (continued)

5. In general, $c = c(\theta, a) = (2\pi)^{-d/2}\bar{c}$, where $\bar{c} = 1 + O(n^{-1})$.

6. Outside the realm of exactness cases, the formula is quite generally accurate to relative error of order $O(n^{-3/2})$:

$$f(\hat{\theta}; \theta \mid a) = p^*(\hat{\theta}; \theta \mid a)\big(1 + O(n^{-3/2})\big),$$

provided a is exactly ancillary, or approximately ancillary to a suitably high order.

7. For a detailed discussion of the theoretical accuracy of p^* in curved exponential families with different choices of ancillaries, see Barndorff-Nielsen and Wood (Bernoulli, 1998).

Distribution function approximation

Suppose we wish to evaluate $\Pr(\hat{\theta} \leq t; \theta \mid a)$, where θ is assumed to be a scalar.

Exact integration of the p^* formula to obtain an approximation to the distribution function of the MLE is generally not possible.

However, accuracte theoretical approximations to the integral of p^* can be derived; cf. the Lugananni-Rice formula.

Notation

Write

$$r_t \equiv r_t(\theta) = \mathrm{sgn}(t-\theta)\sqrt{2(l(t;t,a) - l(\theta;t,a))},$$

and let

$$v_t \equiv v_t(\theta) = j(t;t,a)^{-1/2}\{l_{;\hat{\theta}}(t;t,a) - l_{;\hat{\theta}}(\theta;t,a)\},$$

in terms of the sample space derivative $l_{;\hat{\theta}}$ defined by

$$l_{;\hat{\theta}}(\theta;\hat{\theta},a) = \frac{\partial}{\partial\hat{\theta}} l(\theta;\hat{\theta},a),$$

and with j the observed information.

The formula

Then

$$P(\hat{\theta} \leq t; \theta \mid a) = \Phi\{r_t^*(\theta)\}\{1 + O(n^{-3/2})\},$$

where $r_t^*(\theta) = r_t(\theta) + r_t(\theta)^{-1} \log\{v_t(\theta)/r_t(\theta)\}$.

The random variable $r^*(\theta)$ corresponding to $r_t^*(\theta)$ [replace fixed t by random $\hat{\theta}$] is an adjusted form of the signed root likelihood ratio statistic, $N(0,1)$ to (relative) error $O(n^{-3/2})$, conditional on ancillary a.

Conditional inference in exponential families

Suppose that $Y_1, \ldots, Y_n$ are independent, identically distributed from the exponential family density

$$f(y; \psi, \lambda) = \exp\{\psi\tau_1(y) + \lambda\tau_2(y) - d(\psi, \lambda) - Q(y)\},$$

where we will suppose for simplicity that the parameter of interest ψ and the nuisance parameter λ are both scalar.

The natural statistics are $T = n^{-1}\sum \tau_1(y_i)$ and $S = n^{-1}\sum \tau_2(y_i)$. From the general properties of exponential families, the conditional distribution of T given $S = s$ depends only on ψ, so that inference about ψ may be derived from a conditional likelihood, given s.

The log-likelihood based on the full data $y_1, \ldots, y_n$ is

$$n\psi t + n\lambda s - nd(\psi, \lambda),$$

ignoring terms not involving ψ and λ, and a conditional log-likelihood function is the full log-likelihood minus the log-likelihood function based on the marginal distribution of S.

We consider an approximation to the marginal distribution of S, based on a saddlepoint approximation to the density of S, evaluated at its observed value s.

The cumulant generating function of $\tau_2(Y_i)$ is given by

$$K(z) = d(\psi, \lambda + z) - d(\psi, \lambda).$$

The saddlepoint equation is therefore given by

$$d_\lambda(\psi, \lambda + \hat{z}) = s.$$

With s the observed value of S, the likelihood equation for the model with ψ held fixed is

$$ns - nd_\lambda(\psi, \hat{\lambda}_\psi) = 0,$$

so that $\lambda + \hat{z} = \hat{\lambda}_\psi$, where $\hat{\lambda}_\psi$ denotes the maximum likelihood estimator of λ for fixed ψ.

Applying the saddlepoint approximation, ignoring constants, we approximate the marginal likelihood function based on S as

$$|d_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|^{-1/2} \exp\{n[d(\psi, \hat{\lambda}_\psi) - d(\psi, \lambda)] - (\hat{\lambda}_\psi - \lambda)s\};$$

the resulting approximation to the conditional log-likelihood function is given by

$$n\psi t + n\hat{\lambda}_\psi^T s - nd(\psi, \hat{\lambda}_\psi) + \frac{1}{2}\log|d_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|$$

$$\equiv l(\psi, \hat{\lambda}_\psi) + \frac{1}{2}\log|d_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|.$$

Modified profile likelihood

The profile likelihood $L_{\mathrm{p}}(\psi)$ for a parameter of interest ψ can largely be thought of as if it were a genuine likelihood.

This amounts to behaving as if the nuisance parameter over which the maximisation has been carried out were known. Inference on ψ based on treating $L_{\mathrm{p}}(\psi)$ as a proper likelihood may therefore be grossly misleading if the data contain insufficient information about χ, or if there are many nuisance parameters.

Below, we shall discuss modifications of $L_{\mathrm{p}}(\phi)$.

Definition

The modified profile likelihood $\tilde{L}_{\mathrm{p}}(\psi)$ for a parameter of interest ψ, with nuisance parameter χ, is defined by

$$\tilde{L}_{\mathrm{p}}(\psi) = M(\psi) L_{\mathrm{p}}(\psi),$$

where M is a modifying factor

$$M(\psi) = \left| \frac{\partial \hat{\chi}}{\partial \hat{\chi}_{\psi}} \right| |\hat{j}_{\psi}|^{-1/2}.$$

Here $\partial\hat{\chi}/\partial\hat{\chi}_\psi$ is the matrix of partial derivatives of $\hat{\chi}$ with respect to $\hat{\chi}_\psi$, where $\hat{\chi}$ is considered as a function of $(\hat{\psi}, \hat{\chi}_\psi, a)$ and $\hat{j}_\psi = j_{\chi\chi}(\psi, \hat{\chi}_\psi)$, the observed information on χ assuming ψ is known.

Comments

The modified profile likelihood $\tilde{L}_{\mathrm{p}}$ is, like L_{p}, parameterisation invariant.

An alternative expression for the modifying factor M is

$$M(\psi) = |l_{\chi;\hat{\chi}}(\psi, \hat{\chi}_\psi; \hat{\psi}, \hat{\chi}, a)|^{-1} \times |j_{\chi\chi}(\psi, \hat{\chi}_\psi; \hat{\psi}, \hat{\chi}, a)|^{1/2}.$$

This follows from the likelihood equation for $\hat{\chi}_\psi$:

$$l_\chi(\psi, \hat{\chi}_\psi; \hat{\psi}, \hat{\chi}, a) = 0.$$

Differentiation with respect to $\hat{\chi}$ yields

$$l_{\chi\chi}(\psi, \hat{\chi}_\psi; \hat{\psi}, \hat{\chi}, a)\frac{\partial \hat{\chi}_\psi}{\partial \hat{\chi}} + l_{\chi;\hat{\chi}}(\psi, \hat{\chi}_\psi; \hat{\psi}, \hat{\chi}, a) = 0.$$

Justification

Asymptotically, $\tilde{L}_{\mathrm{p}}$ and L_{p} are equivalent to first-order.

The reason for using $\tilde{L}_{\mathrm{p}}$ rather than L_{p} is that the former arises as a higher-order approximation to a marginal likelihood for ψ when such a marginal likelihood function is available, and to a conditional likelihood for ψ when this is available.

Details

Suppose that the density $f(\hat{\psi}, \hat{\chi}; \psi, \chi \mid a)$ factorises, either as

$$f(\hat{\psi}, \hat{\chi}; \psi, \chi \mid a) = f(\hat{\psi}; \psi \mid a) f(\hat{\chi}; \psi, \chi \mid \hat{\psi}, a)$$

or as

$$f(\hat{\psi}, \hat{\chi}; \psi, \chi \mid a) = f(\hat{\chi}; \psi, \chi \mid a) f(\hat{\psi}; \psi \mid \hat{\chi}, a).$$

In the first case modified profile likelihood can be obtained as an approximation (using the p^*-formula) to the marginal likelihood for ψ based on $\hat{\psi}$ and conditional on a, i.e. to the likelihood for ψ determined by $f(\hat{\psi}; \psi \mid a)$.

In the second case it is obtained as an approximation to the conditional likelihood for ψ given $\hat{\chi}$ and a.

Further comments

Note that if $\hat{\chi}_\psi$ does not depend on ψ,

$$\hat{\chi}_\psi = \hat{\chi},$$

then

$$\tilde{L}_{\mathrm{p}}(\psi) = |\hat{j}_\psi|^{-1/2} L_{\mathrm{p}}(\psi).$$

In the case that ψ and χ are orthogonal, which is a weaker assumption, both hold to order $O(n^{-1})$.