HIGHLY ACCURATE MODEL APPROXIMATIONS

1.1 Laplace integration

Some parameters in a model may be of prime interest and are called interest parameters. While other parameters can be of little interest and yet disrupt the assessment of interest parameters; these are often called nuisance parameters. The method of Laplace (1774) integration examined in this section has had profound influence on the reduction of adverse effects from nuisance parameters.

With increasing amounts of data we are of course well aware of the approach to Normality of many common variables: the Central Limit Theorem effect! Laplace integration acts as if a function of interest is approximately Normal, and then uses the integral of a best fitting Normal shape as the approximate integral of the given function: for this the best fitting is taken to be the one with the same maximum value and the same curvature at the maximum.

Consider a nonnegative function \( f_n(y) \) on the real line and suppose it has smoothness, a unique maximum value, and depends on data size \( n \) so that \( \log f_n(y) \) is \( O(n) \) and thus grows like the data size \( n \). Now let \( f_n(y) \) have maximum value \( \hat{f} \) at the point \( \hat{y} \) and curvature \( j_{yy} = -(d^2/dy^2) \log f_n(y) \) at that maximum; and consider the centered and scaled variable:

\[
z = (y - \hat{y}) j_{yy}^{1/2}, \tag{1.1}
\]

Then in the neighbourhood of \( \hat{y} \) the function \( f_n(y) \) has the following log-second-derivative form

\[
f_n(y) = f_n(\hat{y}) \exp\{- (y - \hat{y})^2 j_{yy}/2 \} = \hat{f} \exp\{- z^2/2 \}, \tag{1.2}
\]

to accuracy \( O(n^{-1/2}) \). The integral of this with respect to \( y \) gives
HIGHLY ACCURATE MODEL APPROXIMATIONS

the Laplace integration formula:

\[ \int f_n(y)dy = \hat{f} (2\pi)^{1/2} \hat{j}_{yy}^{-1/2} \exp\{k/n\} + O(n^{-2}), \quad (1.3) \]

where the first factor comes from the coefficient of \( \exp\{-z^2/2\} \) in (1.2), the next factor is the integral of \( \exp\{-z^2/2\} \) with respect to \( z \), the third factor \( \hat{j}_{yy}^{-1/2} \) comes from the Jacobian \( dy/dz \), and the final factor comes from non-Normality as illustrated in the following example. And then correspondingly the normed version of the function \( f_n(y) \) has the probability density function form:

\[ f^*_n(y) = (2\pi)^{-1/2} \hat{j}_{yy}^{1/2} \exp\{-k/n\} \hat{f}^{-1} f_n(y) + O(n^{-2}), \quad (1.4) \]

Some details are clarified in the following example.

Example with cubic and quartic adjustments. Consider the function \( f_n(z) = \phi(z) \exp\{a_3 z^3/6n^{1/2} + a_4 z^4/24n\} \) and its integral with respect to \( z \), where \( \phi(z) = (2\pi)^{-1/2} \exp\{-z^2/2\} \) is the standard Normal density function which of course has integral equal to unity. Consider the integral of \( f_n(z) \):

\[ \int f_n(z)dz = \int \phi(z) \exp\{a_3 z^3/6n^{1/2} + a_4 z^4/24n\}dz \]

\[ = \int \phi(z)\{1 + a_3 z^3/6n^{1/2} + a_4 z^4/24n + a_3^2 z^6/72n\}dz \]

\[ = 1 + a_3 3/24n + a_4^2 15/72n + \]

\[ = \exp\left\{ 3a_4 + 5a_3^2 \right\/ 24n \}; \]

this uses Taylor expansions for the functions log and exp to move terms up to and down from the exponent, uses the first few moments \( 1, 0, 1, 0, 3, 0, 15, 0, \ldots \) of the standard Normal starting here with the \( \hat{0} - \hat{th} \), and omits terms of order \( O(n^{-3/2}) \) and higher.

The function \( f_n(z) \) was chosen for simplicity and is typically not integrable as it stands due to the cubic and quartic terms in the exponent. Details for a verification would include bounding conditions on the tails of the function to ensure integrability.

1.2 The \( p^* \)-approximation for a statistical model

Regular statistical models are usually presented in terms of density functions and sometimes the densities arrive in a form not convenient for statistical analysis. The \( p^* \)-formula developed by
1.3 The saddlepoint approximation for an exponential model

Barndorff-Nielsen (1980), Barndorff-Nielsen (1986) offers an approximate version, with third order accuracy and also the correct likelihood function at all points. We suppose the given model is \( f(y; \theta) \) and that the dimensions of the variable and the parameter are equal, here one, with a smooth relationship between variable \( y \) and maximum likelihood value \( \hat{\theta} \); for more general cases where \( y \) has dimension larger than the parameter, conditioning is needed and examined separately in Chapter . The \( p^* \)-approximation would seemingly make available a density version of the model for computation using only widely accessible likelihood. However to be useful the points where the computations are to be made must be specified and often this is not possible unless appropriate conditioning is specified. Fortunately in many contexts full conditioning may not be needed, just the form of particular local conditioning .

For the model \( f(y; \theta) \) consider the expression

\[
f_n(\hat{\theta}; \theta) = (2\pi)^{-1/2} \exp\{\ell(\theta; \theta) - \ell(\hat{\theta}; \theta)\} \frac{j_{\hat{\theta}^2}}{\hat{\theta}^2} \exp\{k/n\} \tag{1.5}
\]

for the density of \( \hat{\theta} \). The first exponential factor assign the correct likelihood at each point, the initial factor gets the norming constant as if the first exponential were \(-z^2/2\), and the \( j_{\hat{\theta}^2} \) factor then scales from \( z \) to \( \hat{\theta} \). Expansions (Cakmak et al., 1998) of \( \exp\{\ell(\theta; \theta) - \ell(\hat{\theta}; \theta)\} \) then show that the accuracy is retained if \( j_{\hat{\theta}^2} \) is replaced by \( j_{\hat{\theta}^2}^2 \) giving the \( p^* \)-formula

\[
f_n(\hat{\theta}; \theta) = (2\pi)^{-1/2} \exp\{\ell(\theta; \theta) - \ell(\hat{\theta}; \theta)\} j_{\hat{\theta}^2}^2 \exp\{k/n\} \tag{1.6}
\]

for the density of the maximum likelihood variable \( \hat{\theta} \). The \( p^* \)-formula is often coupled with conditioning on an ancillary but this conditioning is a separate issue and is examined later.

As an example we consider in the next section an exponential model where the \( p^* \)-formula becomes the earlier saddlepoint approximation of Daniels (1954) and Barndorff-Nielsen and Cox (1979).

**1.3 The saddlepoint approximation for an exponential model**

Consider the general exponential model at (2.2) in Chapter 2. The parameter affects the distribution entirely through the vari-
HIGHLY ACCURATE MODEL APPROXIMATIONS

An approximate Laplace inversion (Daniels, 1954) of the cumulant generating function \( \kappa(\varphi) \) gives the saddlepoint approximation which has a long standing background in applied mathematics. Then in statistical notation using log-likelihood \( \ell(\varphi; y) = a + \log \{ f(y; \varphi) \} = \varphi s - \kappa(\varphi) \), the saddlepoint approximation for the statistical model is

\[
f^*(s; \varphi) = (2\pi)^{-1/2} \exp\{\ell(\varphi; s) - \ell(\hat{\varphi}; s)\} \hat{j}_{\varphi\varphi}^{-1/2} \exp\{k/n\} \tag{1.7}
\]

and is similar to the \( p^* \) formula but describes the canonical variable \( s \) rather than the maximum likelihood variable \( \hat{\varphi} \). These are however closely connected. For consider the score equation \( \ell_{s}(\hat{\varphi}; s) = s - \kappa_{s}(\hat{\varphi}) = 0 \). If we differentiate with respect to \( \hat{\varphi} \) we obtain \( ds/d\hat{\varphi} = \kappa_{s\varphi} = \hat{j}_{\varphi\varphi} \) which in going from (1.6) to (1.7) has the effect of changing the sign attached to the root information; also we see that \( \hat{j}_{\varphi\varphi}^{1/2} d\varphi \) is parameterization invariant.

Example 1 Exponential life model

Consider the exponential life model

\[
f(y; \varphi) = \varphi e^{-\varphi y}
\]

for \( y > 0 \). We have \( \ell(\varphi; y) = -\varphi y + \log \varphi, \hat{\varphi} = 1/y, \ell(\hat{\varphi}; y) = -1 - \log y \) and \( \hat{j}_{\varphi\varphi} = y^2 \) giving the saddlepoint approximation

\[
f^*(y; \varphi) = (2\pi)^{-1/2} \exp\{-\varphi y + \log \varphi + 1 + \log y\} y^{-1} \exp\{k/n\} = (2\pi)^{-1/2} \exp\{1 + k/n\} f(y; \varphi),
\]

which is exact except for the norming constant. The model is actually a location model in mild disguise and the saddlepoint method is easily shown to be exact for such models except for the norming constant; here for a case with \( n = 1 \) we might not expect high accuracy for the constant.

1.4 The vector case

The Laplace, the \( p^* \), and the saddlepoint approximations have been discussed in the context of a scalar variable but are available more generally with just obvious modifications and little extra in proof.

Consider a nonnegative function \( f_n(y) \) where \( y \) has dimension \( d \) with smoothness, a unique maximum value, and dependence on
1.4 The vector case

some data size $n$ so that $\log f_n(y)$ is $O(n)$ and thus grows like the data size $n$. The Laplace integration formula is

$$\int f_n(y)dy = \hat{f} \left( 2\pi \right)^{d/2} |\hat{j}_{yy}|^{-1/2} \exp\{k/n\} + O(n^{-2}), \quad (1.8)$$

where $\hat{j}_{yy} = -(\partial/\partial y)(\partial/\partial y) \log f_n(y)$ is the second derivative array of $\log f_n(y)$ evaluated at the maximizing value $\hat{y}$. The calculation utilizes the standard Normal in $d$ dimensions: $\phi(z) = (2\pi)^{-d/2} \exp\{-\Sigma_{i=1}^d z_i^2\}$.

Now consider a statistical model $f_n(y; \theta)$ with smoothness in $d$-dimensional variable $y$ and $d$-dimensional parameter $\theta$. The $p^*$-formula for the distribution of the maximum likelihood variable has the form

$$f^*_n(\hat{\theta}; \theta) = (2\pi)^{-d/2} \exp\{\ell(\theta; \theta) - \ell(\hat{\theta}; \theta)\} |\hat{j}_{\theta \theta}|^{1/2} \exp\{k/n\} \quad (1.9)$$

with third order accuracy, where $\hat{j}_{\theta \theta} = -(\partial/\partial \theta)(\partial/\partial \theta) \log f_n(y; \theta)$ is the Hessian or second derivative array of $-\log f_n(y; \theta)$ evaluated at the maximizing value $\hat{\theta}$. If the dimension of the original variable $y$ is larger than the dimension $d$ of the parameter then ancillary conditioning is needed to attain the same dimension $d$ and this requires special attention to retain all continuity affects of the parameter $\theta$; for some intriguing examples, see Fraser et al. (2010).

Now consider an exponential model $f_n(y; \theta)$ with smoothness in the $d$-dimensional canonical variable $s(y)$ and $d$-dimensional canonical parameter $\varphi$. The saddlepoint formula for the distribution of the maximum likelihood variable has the form

$$f^*_n(\hat{s}; \theta) = (2\pi)^{-d/2} \exp\{\ell(\theta; \theta) - \ell(\hat{s}; \theta)\} |\hat{j}_{\theta \theta}|^{-1/2} \exp\{k/n\} \quad (1.10)$$

or equivalently

$$f^*_n(s; \theta) = (2\pi)^{-d/2} \exp\{\ell(\theta; \theta) - \ell(\hat{s}; \theta)\} |\hat{j}_{\theta \theta}|^{-1/2} \exp\{k/n\} \quad (1.11)$$

with third order accuracy, where $\hat{j}_{\theta \theta} = -(\partial/\partial \theta)(\partial/\partial \theta) \log f_n(y; \theta)$ is the Hessian or second derivative array of $-\ell(\theta; s)$ evaluated at the maximizing value $\hat{\theta}$. The exceptionally attractive feature of the saddlepoint formula here is that it achieves the marginalization from the original variable $y$ to the canonical variable $s$ using only the observed information function at sample points.
1.5 Problems

1. Calculate the expression for the saddlepoint approximation for the Binomial model in Example 2; record the expressions for the various ingredients in the approximation. The canonical parameter \( \varphi = \log(p/q) \) is called the log-odds ratio.

2. Calculate the expression for the saddlepoint approximation for the gamma model \( f(y; \theta) = \Gamma^{-1}(r) \theta^{-r} y^{r-1} \); record the expressions for the various ingredients in the approximation.