Laplace's Method

Laplace's method is an elementary technique for approximating an integral of the form

$$I = \int f(t) \exp(nh(t))dt \tag{1}$$

where $f(t)$ and $h(t)$ are smooth real-valued functions and $h(t)$ has a single maximum in the interior of the domain of integration. In many statistical applications the loglikelihood function, or the log posterior density, plays the role of $nh(t)$ in (1), with n becoming the sample size. Although a log-likelihood $\ell(\theta)$ (or log posterior $\tilde{\ell}(\theta)$) based on distinct data values will not have the form of a fixed function multiplied by n, $\ell(\theta)/n$ (or $\tilde{\ell}(\theta)$) will often behave like a fixed function of θ for large samples (it converges to its expectation). In this situation, under regularity conditions, Laplace's method produces accurate approximations as long as the loglikelihood function (log posterior density) is not too far from being quadratic or, equivalently, when the likelihood function (posterior density) has roughly the form of a Normal density function.

Letting $\hat{t}$ denote the value that maximizes $h(t)$, and assuming $f(\hat{t}) \neq 0$, the formula produced by Laplace's method is

$$\hat{I} = (2\pi)^{m/2} | - nh''(\hat{t})|^{-1/2} f(\hat{t}) \exp(nh(\hat{t})) \tag{2}$$

where m is the dimension of t and $h''(\hat{t})$ is the Hessian matrix of $h(t)$ at $\hat{t}$. Under suitable regularity conditions,

$$I = \hat{I} \cdot \{1 + O(n^{-1})\}, \tag{3}$$

as $n \to \infty$.

The argument used in obtaining (2) is pretty straightforward. In outline, taking, for simplicity, the one-dimensional case ($m = 1$) with the domain of integration being the whole real line, a quadratic Taylor series expansion of $h(t)$ about $\hat{t}$ produces the factor

$$\exp(nh(t)) \doteq \exp(nh(\hat{t})) \exp(\frac{1}{2} nh''(\hat{t})(t - \hat{t})^2)$$

in the integrand, which is recognized to be proportional to a Normal density having standard deviation $1/\sqrt{-nh''(\hat{t})}$. Importantly, as $n \to \infty$, the integrand becomes increasingly concentrated near $\hat{t}$. Now, using $f(t) \doteq f(\hat{t})$, (1) becomes approximated by

$$\int f(t) \exp(nh(t))dt \doteq f(\hat{t}) \exp(nh(\hat{t})) \int \exp(\frac{1}{2} nh''(\hat{t})(t - \hat{t})^2)dt \tag{4}$$

and the integral on the right-hand side of (4) is $\sqrt{2\pi/(-nh''(\hat{t}))}$. This gives the one-dimensional version of (2). Examination of the remainder terms in the Taylor series expansions shows that the order of accuracy of the approximation is as given in (2). The key is that terms of order $O(n^{-1/2})$ that appear under the integral sign drop out when integrated against the Normal density factor $\exp(\frac{1}{2}nh''(\hat{t})(t-\hat{t})^2)$.

Laplace's method has been known at least since Laplace. Because of its simplicity, approximation (2) and its variants have been used by many statistical researchers, often without any apparent awareness of the eponym. For additional details see Kass, Tierney, and Kadane (1990), The validity of posterior expansions based on Laplace's method, which is on my website.

Laplace's Method for State-Space Models

We may write the filtering and prediction equations as

$$post_t(x_t|y_{1:t}) \propto f(y_t|x_t)pred_t(x_t|y_{1:t-1}).$$

and

$$pred_t(x_t|y_{1:t-1}) = \int f(x_t|x_{t-1})post_{t-1}(x_{t-1}|y_{1:t-1})dx_{t-1}.$$

The problem is to compute

$$E(X_t|Y_{1:t}) = \int x_t post_t(x_t|y_{1:t})dx_t$$

and in order to do so there must be an expression for the predictive density $pred_t(x_t|y_{1:t-1})$. Laplace's method may be used. All that needs to be done is to approximate each distribution by a Gaussian.

The "Laplace-Gauss Filter" algorithm, based on first-order Laplace approximation, is as follows.

1. At time $t = 0$, initialize the predictive distribution of the state, $\hat{f}(x_0) = f(x_0)$.
2. Observe y_t.
3. (Filtering) Obtain the approximate posterior mean $\tilde{x}_t$ and variance $\tilde{v}_t$ by Laplace's method:

$$\tilde{x}_t = \hat{x}_t \equiv \text{argmax}_{x_t} l(x_t)$$

and

$$\tilde{v}_t = [-l''(\hat{x}_t)]^{-1},$$

where

$$l(x_t) = \log f(y_t|x_t)\hat{f}(x_t|y_{1:t-1}).$$

Set $\hat{p}(x_t|y_{1:t})$ to be a Gaussian distribution with the same mean and variance.

4. (Prediction) Calculate the predictive distribution,

$$\hat{f}(x_{t+1}|y_{1:t}) = \int f(x_{t+1}|x_t)\hat{f}(x_t|y_{1:t})dx_t. \tag{5}$$

Assuming that $f(x_{t+1}|x_t)$ is a Gaussian pdf, this integral may be performed analytically. Otherwise, Laplace's method must be used again.

5. Increment t and go to step 2.

Example Brown *et al.* (1998, A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patters of rat hipocampal place cells, *J. Neurosci.*) used the following state-space model. The ith neuron was assumed to follow Poisson process intensity given by

$$\lambda^i(t|x_t) = \lambda^i_x(t|x_t, \alpha^i, \mu^i, \Sigma^i)\lambda^i_\theta(t|\phi(t), \phi^i, \beta^i)$$

where $\phi(t)$ is a measured theta rhythm, the dependence on the theta rhythm is

$$\lambda^i_\theta(t|\phi(t), \phi^i, \beta^i) = \exp(\beta^i \cos(\phi(t) - \phi^i)).$$

and the place field is defined by

$$\lambda^i_x(t|x_t, \alpha^i, \mu^i, \Sigma^i) = \exp\left(\alpha^i - \frac{1}{2}(x_t - \mu^i)^T(\Sigma^i)^{-1}(x_t - \mu^i)\right).$$

The state model was taken to be the random walk

$$X_t = X_{t-1} + W_t$$

where W_t is a bivariate Normal with mean zero and variance matrix Σ_W. The neurons were assumed to be independent, given x_t.

In this case we have

$$f(y_t|x_t) = \prod_i f(y^i_t|x_t)$$

where

$$f(y_t^i|x_t) = e^{-\mu_t^i} \frac{(\mu_t^i)^{y_t^i}}{y_t^i!}$$

and

$$\mu_t^i = \lambda^i(t|x_t)\Delta t.$$

The random walk model gives

$$f(x_t|x_{t-1}) = \frac{1}{2\pi|\Sigma_w|^{1/2}} \exp\left(-\frac{1}{2}(x_t - x_{t-1})^T \Sigma_W^{-1}(x_t - x_{t-1})\right).$$

The matrix Σ_W was assumed to be diagonal. Maximum likelihood based on initial data was used to estimate the parameters $\Sigma_{W,11}, \Sigma_{W,22}$ and $\beta^i, \alpha^i, \mu_1^i, \mu_2^i, \Sigma_{11}^i, \Sigma_{12}^i, \Sigma_{22}^i$, for all i. The additional phase parameters ϕ^i were determined by an averaging of spike phases. Plugging all of the estimates into the formulas above for $f(y_t|x_t)$ and $f(x_t|x_{t-1})$ gives recursively computable forms once $post_t$ and $pred_t$ are approximated by Gaussians. All we need to do is (1) find second derivatives and (2) maximize.