4.4 Approximating posterior densities

4.4.1 A Bernstein-von Mises theorem

Theorem 4.4.1 Let the prior $f(w)$ be continuous and non-zero at $w_0 = \text{a.s.} - \lim \hat{w}_n$ and such that (0) $\int f(w)L_n(w)dw < \infty$.
Assume that
(1) for all $\delta > 0$ $\overline{\lim_{n\to\infty}} \sup_{\{s:|s-w_0|\geq\delta\}} (l_n(s) - l_n(w_0) < 0$ a.s.
(2) $l_1(w)$ is continuously differentiable with $0 < \mathbf{W}(-l_1''(w)) < \infty$.
(3) for all $\varepsilon > 0$ there is a $\delta > 0$ such that

$$\overline{\lim_{n\to\infty}} \frac{1}{\mathbf{E}(-l_n''(w))} \sup_{\{s:|s-w|<\delta\}} |l_n''(s) - l_n''(w)| < \varepsilon \quad \text{a. s.}$$

Let $\hat{\sigma}_n^2 = -l_n''(\hat{w})$. Then

$$\frac{L_n(\hat{w} + \hat{\sigma}_n t)}{L_n(\hat{w})} \to e^{-t^2/2} \text{ uniformly for } t \text{ bounded, and}$$

$$\frac{f(\hat{w} + \hat{\sigma}_n t)L_n(\hat{w} + \hat{\sigma}_n t)}{\hat{\sigma}_n \int f(s)L_n(s)ds} \overset{\text{a.s.}}{\to} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}.$$

Corollary: Under the conditions of the theorem, let w_n^* be the Bayes estimator for squared error loss. Then
(i) $\hat{\sigma}_n^{-1}(w_n^* - w_0) \overset{d}{\to} \mathrm{N}(0, 1)$
(ii) $(w_n^* - \hat{w}_n) \overset{\mathbf{P}}{\to} 0$.

4.4.2 The Laplace approximation

The Laplace approximation is

$$\int \exp(nh(\theta))d\theta \approx \left(-\frac{1}{h''(\hat{\theta})}\right)^{\frac{1}{2}} \left(\frac{2\pi}{n}\right)^{\frac{1}{2}} \exp(nh(\hat{\theta}))$$

where $\hat{\theta}$ is the unique maximum of $h(\theta)$. The accuracy is $\mathrm{O}(1/n)$. When using the Laplace approximation (in both numerator and denominator) for calculating a posterior mean, the accuracy is $\mathrm{O}(1/n^2)$.

4.5 Markov Chain Monte Carlo

The Metropolis-Hastings algorithm simulates from a distribution π by selecting a candidate chain governed by a transition matrix $Q(x, y)$. When the simulated chain is in state x a candidate is drawn according to the distribution $Q(x, \cdot)$. The candidate state is accepted (and becomes the next state of the chain) with probability

$$\alpha(x, y) = \min\left(\frac{\pi_y Q(x, y)}{\pi_x Q(x, y)}, 1\right)$$

and is rejected (so the next state is x with the complementary probability. The resulting Markov chain has stationary distribution π, so integrals of the form $\int g(x)\pi_x dx$ can be estimated by $\sum g(X_i)/n$, where the X_i are successive states of the chain, usually ignoring the early values of the simulation (the *burnin* period).

Independence chain. Here $Q(x, y) = f(y)$. Let $w(x) = \pi_{x/} f(x)$. Then $\alpha(x, y) = \min(w(y)/w(x), 1)$. The w are the weights used in importance sampling.

Random walk chain. Here $Y = x + Z$ where $Z \sim f$, so $Q(x, y) = f(y - x)$.

Gibbs sampler. For a k-dimensional vector, one step consists of succesively for each i draw $x_i^{(n+1)}$ from $\pi(x_i|x_1^{(n+1)}, \ldots, x_{i-1}^{(n+1)}, x_{i+1}^{(n)}, \ldots, x_k^{(n)})$. All proposals are accepted in this simulation.

4.6 Bayes factors

Consider data D coming from hypotheses H_1 or H_2. The Bayes factor, or the odds for H_1 based on the data, is defined as

$$BF_{12} = \frac{\mathbf{P}(D|H_1)}{\mathbf{P}(D|H_2)} = \frac{I_1}{I_2}$$

where $I_k = \int \mathbf{P}(D|\theta_k, H_k)\pi(\theta_k|H_k)d\theta_k$.

Let $S = \log \mathbf{P}(d|\hat{\theta}_1, H_1) - \log \mathbf{P}(D|\hat{\theta}_2, H_2) - \frac{1}{2}(d_1 - d_2)\log n$ where $\hat{\theta}_k$ is the mle of θ_k under H_k, and $d_k = dim(\theta_{k)}$. As $n \to \infty$ $S - \log BF_{12} = o(\log BF_{12})$, so $S \approx \log BF_{12}$. S is called the **Schwartz criterion**, and is often used in the form $-2\log S$, when it is called BIC or the **Bayes Information Criterion**.