

## 3.1. Introduction

- Methods proposed for Bayesian inference when the necessary calculations cannot be performed analytically
- The main techniques presented are Normal and Laplace Approximations based on Asymptotics, Quadrature Approximations, Monte Carlo Integration and Resampling Techniques

## 3.2. Asymptotic Approximations

- Rely on results obtained when the sample size is large
- Consider a parameter  $\theta = (\theta_1, \dots, \theta_d)^T$  with posterior dist.

$$\pi(\theta)$$

## 3.2.1. Normal Approximation

- Based on the Taylor series expansion of the log  $\pi(\theta)$  density around  $\mu$  (absolutely unique) mode  $m$

$$\begin{aligned} \log \pi(\theta) &= \log \pi(m) + \left[ \frac{\partial \log \pi(m)}{\partial \theta} \right]^T (\theta - m) - \frac{1}{2} (\theta - m)^T \left[ \frac{\partial^2 \log \pi(m)}{\partial \theta^2} \right] (\theta - m) \\ &\quad + R(\theta) \\ &\approx \log \pi(m) - \frac{1}{2} (\theta - m)^T \left[ -\frac{\partial^2 \log \pi(m)}{\partial \theta^2} \right] (\theta - m) \end{aligned}$$

$R(\theta)$ : Higher order components

- Typically the posterior is known up to a constant  $k$ , that is, we have  $\pi^*(\theta) = \pi(\theta) / p(\theta) = k / p(\theta)$ ,  $k = \int \pi^*(\theta) d\theta$
- The expansion for this approximation is given by:

$$\begin{aligned} \pi^*(\theta) &\approx \pi^*(m) \exp \left\{ -\frac{1}{2} (\theta - m)^T \left[ -\frac{\partial^2 \log \pi^*(m)}{\partial \theta \partial \theta^T} \right] (\theta - m) \right\} \\ &= \pi^*(m) \exp \left\{ -\frac{1}{2} (\theta - m)^T V^{-1} (\theta - m) \right\} \\ \Rightarrow k &= \pi^*(m) (2\pi)^{d/2} |V|^{1/2}, \quad V = \left[ -\frac{\partial^2 \log \pi^*(m)}{\partial \theta \partial \theta^T} \right]^{-1} \end{aligned}$$

- This approximation is similar to the asymptotic result of the MLE, if the prior is uniform
- IN THIS CASE,  $m = \hat{\theta}$ , AND  $V = I(\theta)^{-1}$
- It's important to note that the normal approximation ignores skewness and secondary modes and it will work well if the posterior is similar in shape to the normal distribution

## 3.2.2. Mode calculation

- To obtain the mode to use in the normal approximation, we must solve for the below:

$$\frac{\partial \log \pi(\theta)}{\partial \theta} = 0$$

- In situations where the above equation cannot be solved analytically, we must use numerical methods as, for example, the Newton-Raphson algorithm

## 3.2.3. Standard Laplace Approximation

- Include higher order terms in the Taylor expansion:

$$\pi(\theta) \approx \pi^*(m) \exp \left\{ -\frac{1}{2} (\theta - m)^T V^{-1} (\theta - m) + \frac{1}{3!} R(\theta) \right\}$$

$$R(\theta) = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \pi^{ijk} (\theta_i - m_i) (\theta_j - m_j) (\theta_k - m_k)$$

$$\pi^{ijk} = \frac{\partial^3 \log \pi^*(m)}{\partial \theta_i \partial \theta_j \partial \theta_k}$$

## 3.3. Approximations by Gaussian Quadrature

- Suppose a one-dimensional problem of evaluating  $I = \int_a^b g(\theta) d\theta$

- Quadrature rules approximate  $I$  by  $\hat{I} = \sum_{i=1}^n w_i g(\theta_i)$ , for some weights  $w_i$  and grid points  $\theta_i$ ,  $i=1, \dots, n$

↓  
integrand

- A simple rule would be to take  $n$  equally spaced points and equal weights given by  $w_i = \frac{b-a}{n}$

- Other simple rules like the trapezium rule, where the two endpoints  $\theta_0$  and  $\theta_n$  receive half weight, and Simpson's rule, where weights alternate between  $\frac{4c}{3}$  and  $\frac{2c}{3}$  apart from the endpoints that receive  $c$

- Gaussian rules were developed when the integrand is well approximated by a form  $h(\theta)p(\theta)$ , where  $h(\theta)$  is a polynomial function of  $\theta$  and  $p$  is a density

(a)  $p = U[-1, 1] \Rightarrow$  GAUSS-JACOBI RULE

(b)  $p = G(x, 1) \Rightarrow$  GAUSS-LAGUERRE

(c)  $p = G(0, 1) \Rightarrow$  GAUSS-HERMITE

## 3.4. Monte Carlo integration

- Consider the problem of solving  $I = \int t(\theta) \pi(\theta) d\theta$ , where  $t(\theta)$  is any function of the parameter  $\theta$  and  $\pi(\theta)$  is the posterior

- If we had samples  $\theta_1, \dots, \theta_m$  from  $\pi$ , the simple Monte Carlo estimator is given by:

$$\hat{I}_1 = \frac{1}{m} \sum_{j=1}^m t(\theta_j)$$

- Often sampling from  $\pi(\theta)$  is complicated, and so MC must be extended by the use of draws from auxiliary distributions

- Let  $q(\theta)$  be a density for  $\theta$  with the same support of  $\pi(\theta)$ . Then:

$$I = \int \frac{t(\theta) \pi(\theta)}{q(\theta)} q(\theta) d\theta = \mathbb{E}_{q(\theta)} \left[ \frac{t(\theta) \pi(\theta)}{q(\theta)} \right]$$

- If samples  $\theta_1, \dots, \theta_m$  from  $q$  are available we have:

$$\hat{I}_2 = \frac{1}{m} \sum_{j=1}^m \frac{t(\theta_j) \pi(\theta_j)}{q(\theta_j)}$$

is another estimator of  $I$ .

- This estimators enjoy good frequentist properties:

(a) unbiased estimators

(b)  $V_q(\hat{I}_2) = \sigma^2 / m$

(c) CLT states that:  $\sqrt{m} (\hat{I}_2 - I) \xrightarrow{d} N(0, 1)$ ,  $m \rightarrow \infty$

(d) They are strongly consistent estimators

- The generating of density  $q$  is usually called importance density as samples from  $q$  is called importance samples

- There are no restrictions on  $q$ , and the simplest choice is the uniform dist. when the support of  $\theta$  is compact

- The optimal choice of  $q$  in terms of minimizing  $\sigma^2$  is to take  $q(\theta) \propto t(\theta) \pi(\theta)$

(a)  $q(\theta) = t(\theta) \pi(\theta) \Rightarrow$  GAUSS-HERMITE

(b)  $q(\theta) = G(x, 1) \Rightarrow$  GAUSS-LAGUERRE

(c)  $q(\theta) = G(0, 1) \Rightarrow$  GAUSS-JACOBI

## 3.5. Methods based on stochastic simulation

- Unlike previously presented methods, here the methods are based on direct samples from  $\pi$ , instead of samples from  $q$  or any other density

- It's important to note that, no matter how large the sample is, this method provides only approximations to  $\pi$ , and should be only used when  $\pi$  cannot be derived analytically

- Assume that the posterior density is only known up to a constant, therefore  $\pi^*(\theta) = l(\theta) p(\theta)$  is available but  $\pi = k \pi^*$  is not.

## 3.5.1. Bayes theorem via the rejection method

- A value is drawn from  $\pi$  using the rejection method by drawing samples from  $q$  and accepting it with probability  $\frac{\pi^*(\theta)}{A q(\theta)}$   $\Rightarrow$  proposal

- The constant  $A < \infty$  satisfies the bounding condition  $\frac{\pi^*(\theta)}{q(\theta)} \leq A$ , for all  $\theta$

- Efficiency of the method is improves as  $A \rightarrow \infty$

- Take for instance the following conditions:

$$q(\theta) = p(\theta) \quad (\text{the prior}) \quad \pi^*(\theta) = l(\theta) q(\theta)$$

$$\Rightarrow \frac{\pi^*(\theta)}{q(\theta)} = l(\theta) \Rightarrow \text{the smallest } A \text{ enclosing the envelope is}$$

$$L_{\max} = \max_{\theta} l(\theta) = L(\hat{\theta}), \text{ where } \hat{\theta} \text{ is MLE}$$

$$\Rightarrow \text{Acceptance ratio} = \frac{\pi^*(\theta)}{A q(\theta)} = \frac{l(\theta) p(\theta)}{A l(\theta) p(\theta)} = \frac{l(\theta)}{L(\hat{\theta})} < 1$$

- In this context, a summary of this method would be:

(1)  $\theta$  is drawn from  $p(\theta)$

(2)  $\theta$  is accepted with probab.  $w_i = \frac{l(\theta)}{L(\hat{\theta})}$

• Important remarks:

(1) The number of samples after step 2 will be less than or equal the number of samples in the first step. This could be a problem for dynamic models, and when prior and likelihood provide conflicting information

(2) The requirement to maximize  $l(\theta)$  could be a problem if the likelihood function is too complex

## 3.5.2. Bayes Theorem via weighted sampling (SMC)

- Does not require likelihood maximization

- A sample is generated from  $\pi$  by drawing a sample  $\theta_1, \dots, \theta_m$  from  $q$ , and resampling from the discrete distribution in  $\{\theta_1, \dots, \theta_m\}$  with probabilities  $w_i$  given by

$$w_i = \frac{\pi(\theta_i)}{\sum_{j=1}^m \pi(\theta_j)}, \quad i=1, \dots, m$$

- Let's take for instance the following conditions:

$$q(\theta) = p(\theta) \quad \text{as} \quad \frac{\pi(\theta)}{q(\theta)} = \frac{l(\theta) p(\theta)}{p(\theta)} = k l(\theta)$$

$$\Rightarrow w_i = \frac{l(\theta_i)}{\sum_{j=1}^m l(\theta_j)}, \quad i=1, \dots, m$$

- Therefore SMC methods proceeds as follows:

(1) Sample  $\theta_1, \dots, \theta_m$  is drawn from  $q(\theta) = p(\theta)$

(2)  $\{\theta_1, \dots, \theta_m\}$  are resampled with prob.  $w_i = \frac{l(\theta_i)}{\sum_{j=1}^m l(\theta_j)}$

## Example 3.6

- $n$  animals are categorized as.  $y = (y_1, y_2, y_3)$  are the counts per category with cell prob. given by  $\frac{1}{4}(z+\theta, 1-\theta, \theta)$

$$\Rightarrow p(y|\theta) \propto (2+\theta)^{y_1} (1-\theta)^{y_2} \theta^{y_3} l(\theta)$$

- Let  $\theta \sim U(0, 1)$ , and  $y = (20, 38, 84)$  we have the following posterior:

$$\pi(\theta) \propto (2+\theta)^{20} (1-\theta)^{38} \theta^{84}$$

- Let the proposal be  $q(\theta) = f_N(\theta; 0.68, 0.05^2)$ , then the SMC algorithm would be:

(1) Sample  $\theta_1, \dots, \theta_m$  is drawn from  $f_N(\theta; 0.68, 0.05^2)$

(2)  $\{\theta_1, \dots, \theta_m\}$  are resampled with prob.  $w_i = \frac{f_N(\theta_i)}{\sum_{j=1}^m f_N(\theta_j)}$