Chapter 6. Bayesian Computation Methods

C. Devon Lin

Queen's University, Nov 6, 2015

References

Read

- Course notes
- Chapters 1-3, 6, 7 of *Monte Carlo Statistical Methods* by Robert and Casella.
- Chapters 10 - 13 of *Bayesian Data Analysis* by Gelman.

Outline

- Background and concepts
- Monte Carlo method for computing integrals
- Rejection sampling
- Importance sampling
- Metropolis-Hastings algorithm
- Gibbs sampling

(1) Background and concepts

(a) **Prior**: In Bayesian statistics, the parameter θ is considered to be a quantity whose variation can be described by a probability distribution. This distribution is called *prior distribution* $\pi(\theta)$.

(b) **Posterior**: A sample is taken from a population indexed by θ and the prior distribution is updated with this sample information. The updated prior is called the *posterior distribution* $\pi(\theta|x)$.

(c) **Bayes' Rule**

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{m(x)} = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)d(\theta)}$$

Example 1. *Let $X_1, \ldots, X_n$ be n i.i.d random variables with mean zero and unknown variance σ^2. The likelihood function is then give by*

$$L(X|\sigma^2) \propto (\sigma^2)^{-n/2} \mathit{exp}\{-\sum_{i=1}^{n} X_i^2/(2\sigma^2)\}.$$

Suppose the prior distribution for σ^2 is noninformative, that is, $\pi(\sigma^2) \propto 1/\sigma^2$ Then the posterior density of σ^2 is

$$\pi(\sigma^2|X_1, \ldots, X_n) = \frac{\pi(\sigma^2) f(X_1, \ldots, X_n|\sigma^2)}{\int \pi(\sigma^2) f(X_1, \ldots, X_n|\sigma^2) d(\sigma^2)} \propto (\sigma^2)^{-n/2-1} \mathit{exp}\{-\sum_{i=1}^{n} X_i^2/(2\sigma^2)\}.$$

It can be shown σ^2 follows scaled inverse chi-squared distribution.

Conjugate priors

- Conjugate prior: belonging to a specific distributional family $\pi(\theta)$, with the likelihood $f(x|\theta)$, it leads to a posterior distribution $p(\theta|x)$ belong to the same distribution family as the prior.

Example 2. *Suppose X is the number of pregnant women arriving at a particular hospital to deliver their babies during a given month. The discrete count nature of the data plus its natural interpretation as an arrival rate suggest adopting a Poisson likelihood,*

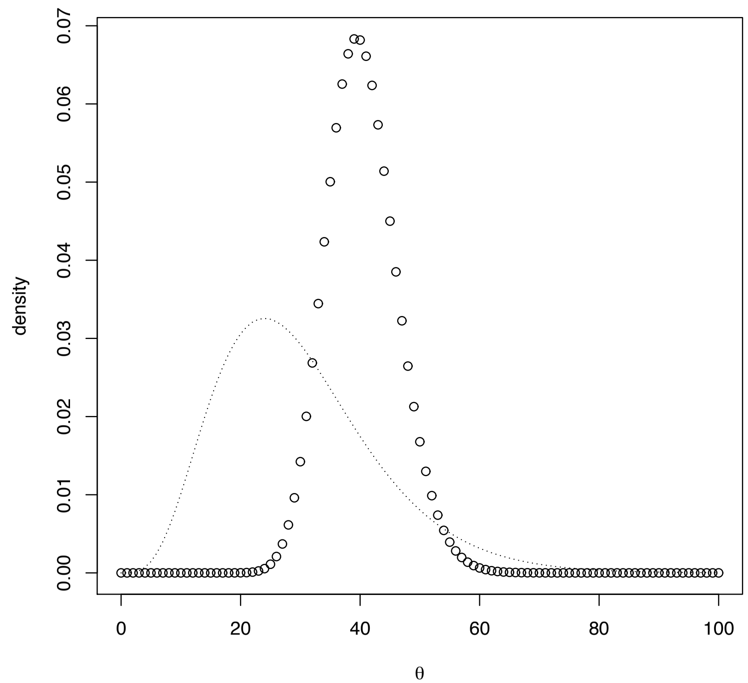
$$f(x|\theta) = \frac{e^{-\theta}\theta^x}{x!}, x = 0, 1, \ldots, \theta > 0.$$

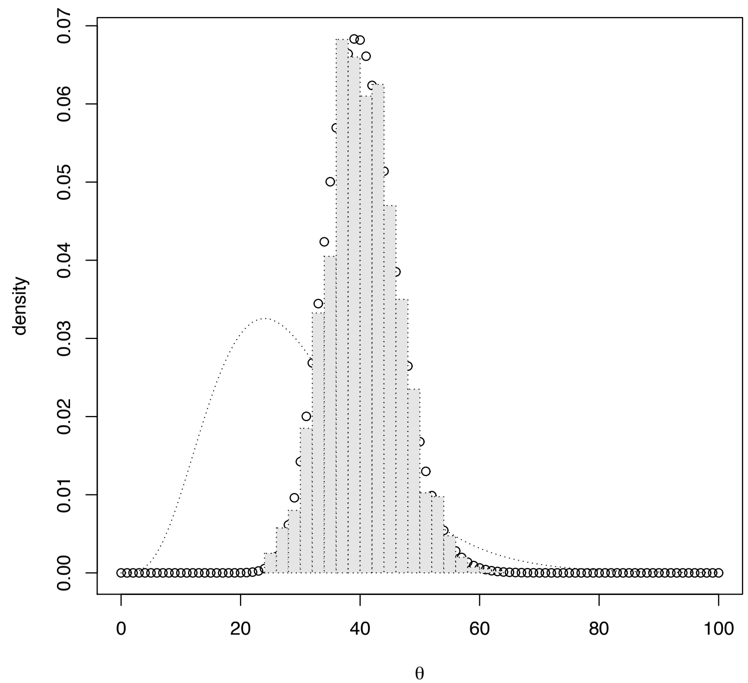
Suppose the prior is gamma distribution,

$$\pi(\theta) = \frac{\theta^{\alpha-1}e^{-\theta/\beta}}{\Gamma(\alpha)\beta^\alpha}, \theta > 0, \alpha > 0, \beta > 0.$$

The posterior distribution would be

$$p(\theta|x) \propto \theta^{x+\alpha-1}e^{-\theta(1+1/\beta)}.$$


density
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0
20
40
60
80
100
θ


density
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0
20
40
60
80
100
θ

Conjugate Priors

- Beta
- Gamma
- Dirichlet
- Gaussian
- Inverse Gamma
- Wishart
- Inverse-Wishart

Non-informative priors

- Non-informative priors: a prior that contains no information about the parameter θ, that is, the prior is "flat" relative to the likelihood function.
  - If $0 \leq \theta \leq 1$, $Uniform(0, 1)$ is a non-informative prior for θ.
  - If $-\infty < \theta < \infty$, $N(\theta_0, \sigma_0^2)$ and $\sigma_0^2 \to \infty$ forms a non-informative prior.

Improper priors

- Improper priors: $\int \pi(\theta)d(\theta) = \infty$
- Improper priors can lead to proper or improper posterior.
- Example: $y_1, \ldots, y_n \overset{iid}{\sim} N(\theta, 1)$ and $\pi(\theta) \propto 1$. Drive the posterior distribution of θ.

Jeffreys' priors

- Jeffery's Rule: a rule for the choice of a non-informative prior
- Jeffreys' priors: the prior is given by

$$\pi(\theta) \propto |I(\theta)|^{1/2},$$

where $I(\theta)$ is the expected Fisher information.

Properties of Jeffreys' priors

- Invariant to transformation
- Non-informative
- Can be improper for many models

Examples of Jeffreys' priors

Example 3. *Suppose* $y_1, \ldots, y_n \overset{iid}{\sim}$ *Binomial*$(1, \theta)$. *Derive Jeffreys' prior.*

Examples of Jeffreys' priors

Example 4. *Let* $y_1, \ldots, y_n \stackrel{iid}{\sim} N(\theta, \sigma^2)$. *Drive Jeffreys' prior for* θ *and* σ^2.

Informative priors

- An informative prior is a type of prior that is not dominated by the likelihood function and has an impact on the posterior.
- For example, $y_1, \ldots, y_n \overset{iid}{\sim} N(\theta, 5)$, $\theta \sim N(0, 1)$.

Informative priors

Some choices of information priors

- $\theta \in R$: normal distribution or t distribution
- $\theta > 0$: gamma, inverse gamma, lognnormal
- $\theta \in (0, 1)$: Beta distribution

Hierarchical priors

Example 5. *An insect lays a large number of eggs, each surviving with probability p. On average, how many eggs will survive? Let X be the number of survivors and Y be the number of eggs laid.*

Hierarchical priors

Example 6. *Consider a generalization Example 5, where instead of one mother insect there are a large number of mothers, and one mother is chosen at random.*

Tsutakawa et al. (1985) describe the problem of simultaneously estimating the rates of death from stomach cancer for maLes at risk in the age of 45 - 64 for the largest cities in Missouri. Let n_j be the number at risk and y_j be the number of cancer deaths for a given city. A model assumes that y_j is distributed from a beta-binomial model with mean η and precision K, that is,

$$f(y_j|\eta, K) = \binom{n_j}{y_j} \frac{B(K\eta + y_j, K(1-\eta) + n_j - y_j)}{B(K\eta, K(1-\eta))}.$$

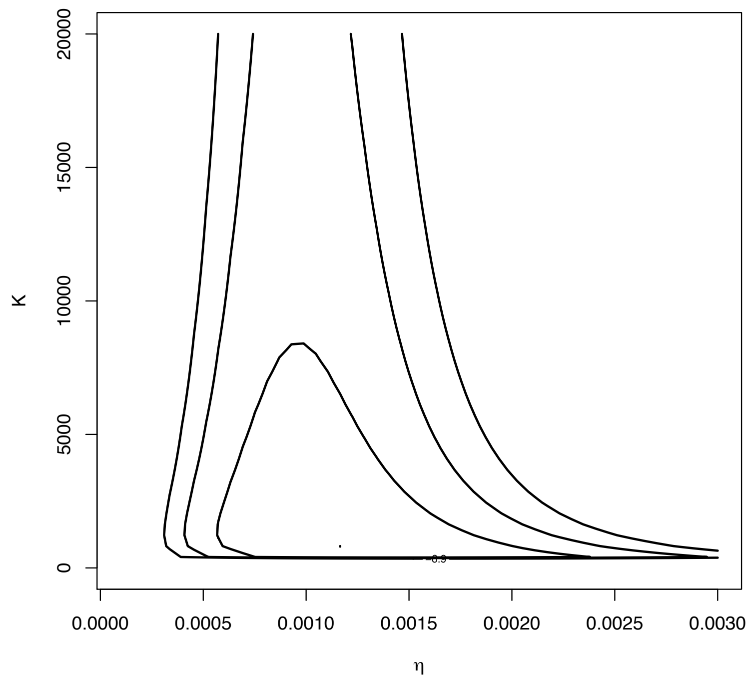
The prior is chosen to be,

$$\pi(\eta, K) \propto \frac{1}{\eta(1-\eta)} \frac{1}{(1+K)^2}.$$

Then the posterior density of (η, K) is given, up to a proportionality constant, by

$$\pi(\eta, K)|n_j, y_j \propto \frac{1}{\eta(1-\eta)} \frac{1}{(1+K)^2} \prod_{j=1}^{20} \frac{B(K\eta + y_j, K(1-\eta) + n_j - y_j)}{B(K\eta, K(1-\eta))}.$$

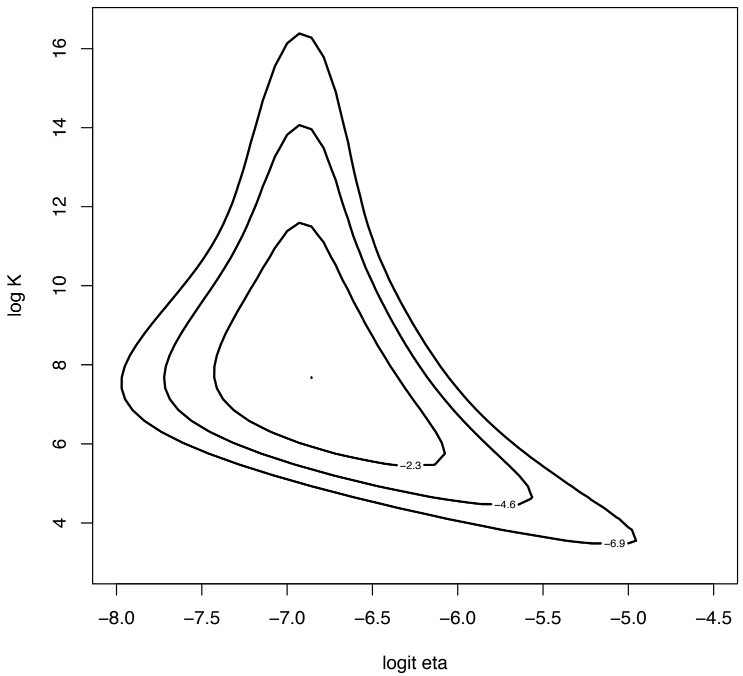
where $0 < \eta < 1$ and $K > 0$.



0
5000
10000
15000
20000
K
0.0000
0.0005
0.0010
0.0015
0.0020
0.0025
0.0030
η

Due to the strong skewness in the density, we transform each parameter to the real line be

$$\theta_1 = logit(\eta) = log(\frac{\eta}{1-\eta}), \theta_2 = log(K).$$



log K
16
14
12
10
8
6
4
−2.3
−4.6
−6.9
−8.0
−7.5
−7.0
−6.5
−6.0
−5.5
−5.0
−4.5
logit eta

(d) Computing integrals

Suppose we wish to compute a complex integral,

$$\int_a^b w(x)d(x).$$

If we can decompose $w(x)$ into the product of a function $h(x)$ and a probability density function $f(x)$ defined over the interval (a, b), we then have

$$\int_a^b w(x)d(x) = \int_a^b h(x)f(x)d(x) = \mathsf{E}_{f(x)}[h(x)].$$

Monte Carlo integration draws a large number of $x_1, \ldots, x_n$ of random variables from $f(x)$, then

$$\int_a^b w(x)d(x) = \mathsf{E}_{f(x)}[h(x)] \simeq \frac{1}{n}\sum_{i=1}^{n} h(x_i)$$

```
> # compute normal cdf by Monte Carlo integration
> t<-0
> n<-10000
> mean(rnorm(n)<t)
[1] 0.5021
> mean(rnorm(n)<1.96)
[1] 0.9749
```

(e) Monte Carlo method for estimating the mean of $h(\theta)$

Suppose θ has a posterior density $\pi(\theta|x)$ and we are interested in the mean of $h(\theta)$, given by

$$\mathsf{E}(h(\theta)|x) = \int h(\theta)\pi(\theta|x)d\theta.$$

To obtain a Monte Carlo estimate, we simulate an independent sample $\theta^1, \ldots, \theta^m$ from the posterior density $\pi(\theta|x)$. The Monte Carlo estimate is given by the sample mean

$$\bar{h} = \sum_{j=1}^{m} h(\theta^j)/m$$

and its associated simulation standard error is

$$se_{\bar{h}} = \sqrt{\sum_{j=1}^{m}(h(\theta^j) - \bar{h})^2/[(m-1)m]}.$$

Example: Let p be the proportion of the American college students who sleep at least eight hours. We are interested in estimating p. We now take a sample of 27 students. Among them, 11 has at least eight hours of sleep. If we regard a "success" as sleeping at least eight hours and we take a random sample with s successes and f failures, then the likelihood function is given by

$$L(p) \propto p^{S}(1-p)^{f}, 0 \leq p \leq 1.$$

The posterior density for p is $\pi(p|data) \propto \pi(p)L(p)$. Suppose that the prior distribution is chosen to be

$$\pi(p) \propto p^{a-1}(1-p)^{b-1}, 0 \leq p \leq 1.$$

The posterior density is

$$\pi(p|data) \propto p^{a+s-1}(1-p)^{b+f-1}, 0 \leq p \leq 1.$$

Now suppose $a = 3.26, b = 7.19$. If now we are interested in the posterior mean of p^2.

```
> ### estimating the mean of p^2
> p<- rbeta(1000,14.26,23.19)
> est<-mean(p^2)
> se<-sd(p^2)/sqrt(1000)
> c(est,se)
[1] 0.149490312 0.001850406
```