MT4531/5731: (Advanced) Bayesian Inference

Rejection Sampling

Nicolò Margaritella

School of Mathematics and Statistics, University of St Andrews

–

Outline

Outline

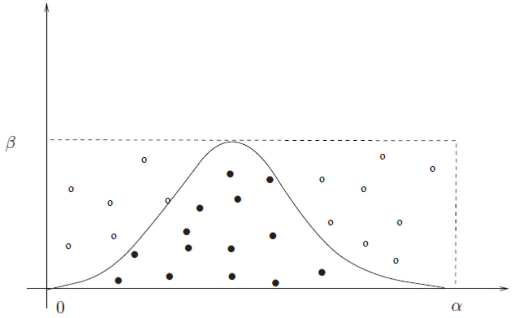
Sampling from Univariate distributions

- Another widely applicable method (more so than Inversion) for sampling from a distribution is Rejection sampling.
- One similarity with the Metropolis-Hastings algorithm, is that in both algorithms we reject proposed values.
- However, these are two **very different** algorithms!
- The Rejection sampling algorithm generates **independent** samples. It does **not** generate a Markovian chain.

Outline

Rectangle envelope

- Suppose that we wish to generate observations $\theta^1, \theta^2, \ldots, \theta^n$ from the posterior distribution $\pi(\theta|\boldsymbol{x})$. (Any distribution in fact.)
- One approach is to enclose the density within a rectangular box and generate points uniformly at random within the box:



Rectangle envelope - The algorithm

- Any points above the density function are rejected and any underneath are accepted. We take the x-coordinate of the accepted point to be the random sample from $\pi(\theta|\boldsymbol{x})$.
- Thus, for $0 \leq \theta \leq \alpha$ and $0 \leq \pi(\theta|\boldsymbol{x}) \leq \beta$.

STEP 1. GENERATE θ_* AS A REALISATION FROM $\Theta_* \sim U[0, \alpha]$.
STEP 2. GENERATE y AS A REALISATION FROM $Y \sim U[0, \beta]$.
STEP 3. ACCEPT θ_* IF $y \leq \pi(\theta_*|\boldsymbol{x})$, ELSE IF $y > \pi(\theta_*|\boldsymbol{x})$ GO BACK TO STEP 1 AND REPEAT.

Rectangle envelope - Notes

- Note that simulating $\theta_* \sim U[0, \alpha]$ is equivalent to simulating v from $V \sim U[0, 1]$ and setting $\theta_* = v\alpha$.
- We can then replace steps 2 and 3 above by simulating u from $U \sim U[0, 1]$, and accepting the simulated value of θ_* if $u \leq \frac{\pi(\theta_* | \boldsymbol{x})}{\beta}$.

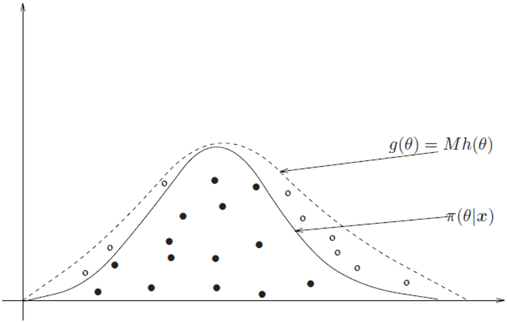
Outline

Limitations of the rectangle envelope

- There are a number of problems associated with using a rectangle envelope.
- A rectangle cannot be used when $\pi(\theta|\boldsymbol{x})$ has an infinite range.
- Also, the probability of rejection can become quite large if the posterior probability is concentrated in a small subspace of the parameter space.

The general method

- *Envelope* the posterior density, $\pi(\theta|\boldsymbol{x})$, *not* by a rectangle, but another (more general) curve, $g(\theta)$.
- $g(\theta)$ is some multiple of a second p.d.f. $h(\theta)$ i.e., $g(\theta) = Mh(\theta)$, from which it easy to sample.



The algorithm

- If we let $g(\theta) = Mh(\theta)$, where $M \geq 1$, then we can use the following algorithm.

  STEP 1. SIMULATE θ_* AS A REALISATION FROM $\Theta_* \sim h(\theta)$.
  STEP 2. GENERATE y FROM $Y \sim U[0, g(\theta_*)]$.
  STEP 3. ACCEPT θ_* AS A REALISATION FROM $\pi(\theta|\boldsymbol{x})$ IF AND ONLY IF $y \leq \pi(\theta_*|\boldsymbol{x})$.

- We will prove in this lecture that this algorithm samples from $\pi(\theta|\boldsymbol{x})$. See separate short video with this proof, and the lecture notes.

Notes

- We can again replace steps 2 and 3 above by simulating u from $U \sim U[0,1]$, and accepting the simulated value of θ_* if $g(\theta_*)u \leq \pi(\theta_*|\boldsymbol{x}) \Rightarrow u \leq \frac{\pi(\theta_*|\boldsymbol{x})}{g(\theta_*)}$.
- For a specific θ_*,
$P(\theta_* \text{ is accepted }) = \pi(\theta_*|\boldsymbol{x})/g(\theta_*) = \pi(\theta_*|\boldsymbol{x})/[Mh(\theta_*)]$.
- The probability of acceptance (not for a specific θ_*),

$$\mathbb{P}(\text{accept}) = P\left(U \leq \frac{\pi(\Theta_*|\boldsymbol{x})}{Mh(\Theta_*)}\right)$$

$$= \int_{-\infty}^{\infty} P\left(U \leq \frac{\pi(\Theta_*|\boldsymbol{x})}{Mh(\Theta_*)}|\Theta_* = \theta_*\right) h(\theta_*)d\theta_*$$

$$= \int_{-\infty}^{\infty} \frac{\pi(\theta_*|\boldsymbol{x})}{Mh(\theta_*)} h(\theta_*)d\theta_* = \frac{1}{M} \times 1 = \frac{1}{M}$$

The importance of the choice of M

- Given that the acceptance probability in general is $1/M$, we would like M to be as close to 1 as possible, subject to $M \geq 1$.
- But, we also need that $g(\theta) = Mh(\theta) \geq \pi(\theta|\boldsymbol{x})$, for all θ, since g must "envelope" π.
- Therefore,

$$M \geq \frac{\pi(\theta|\boldsymbol{x})}{h(\theta)} \ \forall\theta.$$

  As we want M to be as small as possible, the optimal M is $M^* = \sup_\theta \left(\frac{\pi(\theta|\boldsymbol{x})}{h(\theta)}\right)$, where the Supremum denotes the smallest upper bound.
- In practice, the optimal M can be evaluated as the maximum of $\frac{\pi(\theta|\boldsymbol{x})}{h(\theta)}$.

Example (1)

- Suppose that we wish to sample from a distribution $f(\theta)$, which is the $Beta(3,2)$ distribution.
- The support for the distribution is $[0,1]$ and so we can consider a rectangle envelope, where $g(\theta)$ is a straight line. (i.e. $h \equiv U[0,1]$ with $h(\theta) = 1$).
- The optimal value of M, is the maximum of,

$$\frac{f(\theta)}{h(\theta)} = \frac{\Gamma(5)}{\Gamma(3)\Gamma(2)}\theta^2(1-\theta).$$

- The maximum is obtained when $\theta = \frac{2}{3}$ giving the value of $M = \frac{16}{9}$. (Check!).
- Now, $g(\theta) = Mh(\theta) = 16/9$.

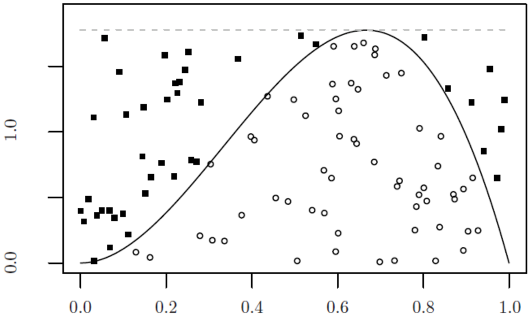
Example (2)

- The rejection sampling algorithm is:

STEP 1. SIMULATE θ_* FROM $\sim U[0, 1]$.
STEP 2. GENERATE y FROM $Y \sim U[0, 16/9]$.
STEP 3. ACCEPT θ_* AS A REALISATION FROM $Beta(3, 2)$
IF AND ONLY IF $y \leq \frac{\Gamma(5)}{\Gamma(3)\Gamma(2)} \theta_*^2 (1 - \theta_*)$.

Example (3)

- In a simulation of 100 points, 58 points are accepted, with a sample average of 0.633.
- The theoretical acceptance probability is $1/M = 0.56$ and the expectation of the distribution is 0.6.



Example (4)

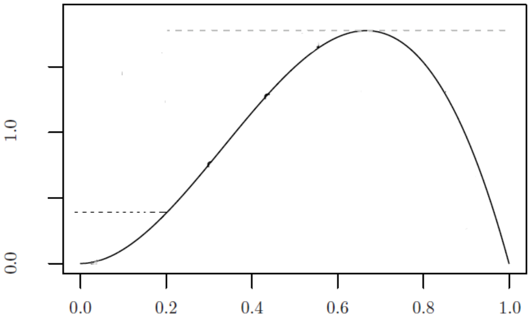
- To improve this algorithm, we can modify the 'envelope' $g(\theta)$, as a step function so that $g(\theta) = 16/9$ for $0.2 < \theta < 1$, as before, but now $g(\theta) = 0.384$ for $0 < \theta \leq 0.2$.
- Note that $f(\theta)$ is a strictly increasing function in $0 < \theta \leq 0.2$, and that $f(0.2) = 0.384$, so $g(\theta)$ still envelopes $f(\theta)$.
- Now, $\int_0^1 g(\theta) d\theta = 0.2 \times 0.384 + 0.8 \times (16/9) = 1.49902$.
- As $h(\theta)$ has to be a distribution that integrates to one,

$$h(\theta) = \frac{1}{1.49902} g(\theta),$$

and $g(\theta) = Mh(\theta) \Rightarrow M = 1.49902$.

Example (5)

- This is the new $g(\theta)$ envelope:



Example (6)

- The rejection sampling algorithm is:

  Step 1. Simulate u from $\sim U[0,1]$.
  Step 2. If $u \leq 0.2 \times 0.384/1.49902$, simulate θ_* from $\sim U[0,0.2]$.
  Step 3. If $u \leq 0.2 \times 0.384/1.49902$, generate y from $Y \sim U[0,0.384]$.
  Step 4. If $u > 0.2 \times 0.384/1.49902$, simulate θ_* from $\sim U[0.2,1]$.
  Step 5. If $u > 0.2 \times 0.384/1.49902$, generate y from $Y \sim U[0,16/9]$.
  Step 6. Accept θ_* as a realisation from $Beta(3,2)$ if and only if $y \leq \frac{\Gamma(5)}{\Gamma(3)\Gamma(2)}\theta_*^2(1-\theta_*)$.

Example (7)

- In a simulation of 100 points, 69 points are accepted, with a sample average of 0.59.
- The theoretical acceptance probability is $1/M = 0.68$ and the expectation of the distribution is 0.6.
- See the tutorial question, where you will be asked to write R code for implementing the rejection sampling algorithms for this example.

Unknown proportionality constant

- Suppose that we write the posterior distribution as,

$$\pi(\theta|\boldsymbol{x}) = \frac{f_1(\theta|\boldsymbol{x})}{f(\boldsymbol{x})},$$

where $f_1(\theta|\boldsymbol{x}) = f(\boldsymbol{x}|\theta)p(\theta)$.

- Let $g(\theta) = Mh(\theta)$ so that $g(\theta)$ envelopes $f_1(\theta|\boldsymbol{x})$. Then,

STEP 1. SIMULATE θ_* AS A REALISATION FROM $\Theta_* \sim h(\theta)$.
STEP 2. GENERATE y FROM $Y \sim U[0, g(\theta_*)]$.
STEP 3. ACCEPT θ_* AS A REALISATION FROM $\pi(\theta|\boldsymbol{x})$ IF AND ONLY IF $y \leq f_1(\theta_*|\boldsymbol{x})$.

- This works because for $u \leq \frac{f_1(\theta|\boldsymbol{X})}{g(\theta)}$, multiplying both the numerator and denominator by $1/f(\boldsymbol{x})$ does not change the algorithm.

Finally...

- In general, rejection sampling can be very wasteful and is only really feasible in one or two dimensions.
- In addition, it can be difficult to identify a suitable "g" function (as we ideally want this to be of similar shape to π for efficiency).