

Stat314/461Term 4:

Rejection sampling: A simple simulation method for non-standard (posterior) densities

Patrick Graham

September, 2021

Rejection sampling: Problem 1 - non-conjugate Binomial model

$$\begin{aligned} Y_i &\overset{\text{indep}}{\sim} \text{Binomial}(10, \theta) \\ \text{logit}(\theta) &\sim \text{Normal}(\mu, \sigma^2); \left[\text{logit}(\theta) = \log \left(\frac{\theta}{(1-\theta)} \right) \right] \\ \Rightarrow p(\theta) &= \frac{1}{\sqrt{2\pi}\sigma\theta(1-\theta)} \exp \left\{ -\frac{1}{2\sigma^2} \left[\log \left(\frac{\theta}{(1-\theta)} \right) - \mu \right]^2 \right\} \end{aligned} \quad (1)$$

$$\mathbf{Y} = (6, 7, 5, 5, 4, 8)$$

$$\begin{aligned} p(\theta|\mathbf{Y}) &= \frac{\prod_{i=1}^6 \text{binomial}(Y_i|10, \theta) \text{logitN}(\theta|\mu, \sigma^2)}{\int \prod_{i=1}^6 \text{binomial}(Y_i|10, \theta) \text{logitN}(\theta|\mu, \sigma^2) d\theta} \\ &\propto \prod_{i=1}^6 \theta^{Y_i} (1-\theta)^{10-Y_i} \times \text{logitN}(\theta|\mu, \sigma^2) \end{aligned}$$

How can we simulate $p(\theta|\mathbf{Y})$?

Rejection sampling (or acceptance sampling)

- Simulate from an approximation; then accept with a probability that depends on the ratio of the true density to the approximating density.
- This idea re-appears in more advanced Bayesian computation methods e.g. Metropolis-Hastings.

Rejection sampling algorithm

(stated in a general way, assuming some parameter of interest θ , and data \mathbf{Y})

- Suppose

$$p(\theta|\mathbf{Y}) = \frac{p(\mathbf{Y}|\theta)p(\theta)}{\int p(\mathbf{Y}|\theta)p(\theta) d\theta} = \frac{q(\theta)}{\int p(\mathbf{Y}|\theta)p(\theta) d\theta}$$

- Let $g(\theta)$ be an approximation to $p(\theta|\mathbf{Y})$, that can be easily simulated from.
- Set $r(\theta) = q(\theta|\mathbf{Y})/g(\theta)$
- Find M such that $r(\theta) \leq M$, for all θ .
- ① Sample θ^* from $g(\theta)$
- ② Accept θ^* with probability $r(\theta^*)/M$.
- ③ Repeat until the desired number of draws is obtained.

To implement step 2:

- draw $u \sim \text{Uniform}(0, 1)$
- Accept if $u \leq r(\theta)/M$

Some characteristics of rejection sampling (i)

- a good approximation $g(\theta)$ is roughly proportional to $q(\theta|\mathbf{Y})$ which is, of course, proportional to $p(\theta|\text{data})$.
- if proportionality does hold

$$\begin{aligned} g(\theta) &= Kq(\theta|\mathbf{Y}) \\ r(\theta) &= \frac{q(\theta|\mathbf{Y})}{Kq(\theta|\mathbf{Y})} \\ &= 1/K \end{aligned} \tag{2}$$

- so if we set $M = 1/K$ all draws from $g(\theta)$ will be accepted.

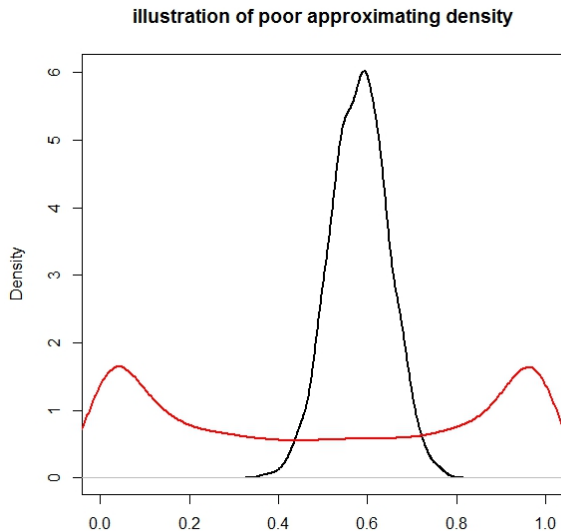
Some characteristics of rejection sampling (ii): Consequences of non-proportionality

If $g(\theta)$ not approximately proportional to $q(\theta|\text{data})$

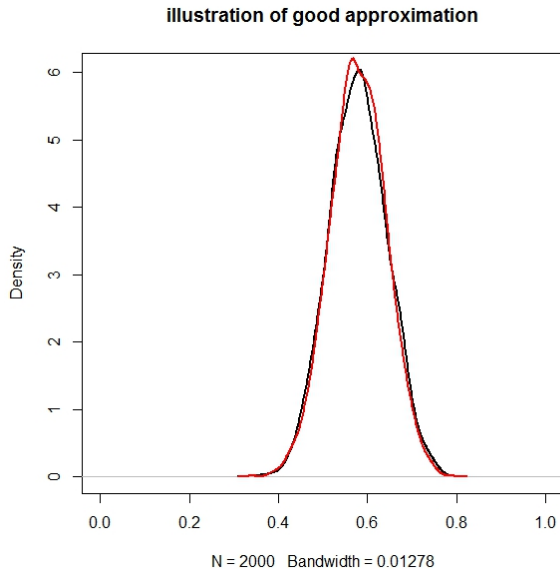
- There will be θ s.t. $g(\theta) \ll q(\theta|\text{data})$
- Hence M must be set large to ensure $r(\theta) = q(\theta|\mathbf{Y})/g(\theta) \leq M$ for all θ
- But $g(\theta) \gg q(\theta|\text{data})$ for some θ , hence $r(\theta)/M$ will be low in these high g density regions
- So most draws rejected.
- The worse the approximation the greater the expected number of draws required before acceptance.
- But self-diagnosing – many rejections implies a poor approximating density

A poor approximating density

(consider (i) red approximating black; (ii) black approximating red)



A good approximating density



Insight: what if the approximating density is the prior

This will often be inefficient because the prior is usually more diffuse than the posterior. Hence many draws are likely to be rejected (wasted).

Nevertheless it is interesting to think this through:

if $g(\theta) = p(\theta)$ then the importance sampling ratio, $r(\theta)$ is

$$\begin{aligned} r(\theta) &= \frac{q(\theta|\mathbf{Y})}{p(\theta)} \\ &= \frac{p(\mathbf{Y}|\theta)p(\theta)}{p(\theta)} \\ &= \underbrace{p(\mathbf{Y}|\theta)} \\ &= \text{Likelihood} \end{aligned} \tag{3}$$

Implications of using the prior as the approximation - cont'd

setting the approximating density $g(\theta)$ to the prior $p(\theta)$ implies the importance sampling ratio is equal to the likelihood: $r(\theta) = p(\mathbf{Y}|\theta)$. Hence M can be found by maximizing the likelihood $p(\mathbf{Y}|\theta)$ wrt θ i.e. set M to the value of the likelihood with θ set to the maximising value, $p(\mathbf{Y}|\hat{\theta})$, where $\hat{\theta}$ is the MLE.

Rejection algorithm when prior is used as the approximating density

(M is the maximum value of the likelihood)

- Sample θ^* from the prior $p(\theta)$
- Accept θ^* with probability $p(\mathbf{Y}|\theta^*)/M$.
- so values of θ close to the the value that maximises the likelihood (i.e. the MLE) have a high probability of being accepted
- values far from the MLE have low probability of being accepted.
- Not a bad insight on Bayesian inference; start with the prior; values more strongly supported by the data (high likelihood) are more likely to make it into the posterior sample;
- Like giving more weight to values more strongly supported by the data.

Rejection sampling for Binomial - logitNormal problem: Using the prior as the approximator

$$Y_i \stackrel{\text{indep}}{\sim} \text{Binomial}(10, \theta)$$

$$\text{logit}(\theta) \sim \text{Normal}(\mu, \sigma^2) \quad (4)$$

$$p(\theta | \mathbf{Y}, N) \propto \prod_i \text{Binomial}(Y_i | 10, \theta) \times \text{logitN}(\theta | \mu, \sigma^2) \quad (5)$$

$$\mathbf{Y} = (6, 7, 5, 5, 4, 8), \quad N = 10$$

If we use the prior as the approximating density:

$$\begin{aligned} r(\theta) &= \frac{\prod_i \text{Binomial}(Y_i | 10, \theta) \times \text{logitN}(\theta | \mu, \sigma^2)}{\text{logitN}(\theta | \mu, \sigma^2)} \\ &= \prod_i \text{Binomial}(Y_i | 10, \theta) \propto \theta^{\sum_i Y_i} (1 - \theta)^{\sum_i (10 - Y_i)} \end{aligned} \quad (6)$$

Rejection sampler using prior as the approximator cont'd

$$\sum_i Y_i = 35; \sum_i N_i = 60$$

- set $\theta_{MLE} = \sum_i Y_i / \sum_i N_i = 35/60$
- set $M = \prod_i \text{Binomial}(Y_i | 10, \theta = \theta_{MLE})$
- draw $\theta^* \sim \text{logitN}(\mu, \sigma^2)$
- Accept with probability $\prod_i \text{Binomial}(Y_i | 10, \theta = \theta^*) / M$
- Repeat until desired simulation size is reached

For acceptance step it is easiest to work on the log scale:

- draw $U \sim \text{Uniform}(0, 1)$
- accept θ^* if $\log U < \log r(\theta^*) - \log M$

Performance of the rejection sampler with prior as the approximator

-see R script for code

- $p(\theta) = \text{logitN}(\text{logit}(0.5), 5)$: acceptance rate 5.4%
- $p(\theta) = \text{logitN}(\text{logit}(0.5), 1)$: acceptance rate 24.7%

Rejection sampling for Binomial - logitNormal problem: Using a beta approximation to the posterior as the approximator

- The conjugate prior for Binomial likelihood is the beta distribution
- If we can construct a beta approximation to the logit-normal prior we can obtain a beta approximation to the posterior using the standard updating results for the Binomial - beta model.
- A beta approximation to the $\text{logitN}(\mu, \sigma^2)$ distribution can be obtained by matching the mean and variance of the two densities.

Rejection sampling using a beta approximation to the posterior, cont'd

- if $\theta \sim \text{Beta}(a, b)$

$$E(\theta) = \frac{a}{a+b} \quad (7)$$

$$V(\theta) = \frac{ab}{(a+b)^2(a+b+1)} \quad (8)$$

- obtain $E(\theta)$ and $V(\theta)$ for $\theta \sim \text{logitN}(\mu, \sigma^2)$ by simulation or numerical integration (see logitnorm package).
- Use (7) and (8) to solve for a and b in terms of $E(\theta)$ and $V(\theta)$
- $\text{Beta}(a + \sum_i Y_i, b + \sum_i (N - Y_i))$ is then an approximation to the posterior.

Rejection sampling using a beta approximation to the posterior, cont'd

$$\begin{aligned} r(\theta) &= \frac{p(\theta|\mathbf{Y}, N)}{g(\theta)} \\ &\propto \frac{\theta^{\sum_i Y_i} (1 - \theta)^{\sum_i (N - Y_i)} \times \text{logitN}(\theta|\mu, \sigma^2)}{\text{beta}(\theta|a + \sum_i Y_i, b + \sum_i (N_i - Y_i))} \\ &= r_A(\theta), \text{ say} \end{aligned}$$

- Find $\log(M) = \max \log(r_A(\theta))$
- draw θ^* from $\theta \sim \text{Beta}(a + \sum_i Y_i, b + \sum_i (N_i - Y_i))$
- compute $\log(r_A(\theta^*))$
- draw $U \sim \text{Uniform}(0, 1)$
- Accept θ^* if $\log(U) \leq \log(r_A(\theta^*)) - \log(M)$

Performance of the rejection sampler using a beta approximation to the posterior as the approximator

-see R script for code

- $p(\theta) = \text{logitN}(\text{logit}(0.5), 5)$: acceptance rate 71.5%
- $p(\theta) = \text{logitN}(\text{logit}(0.5), 1)$: acceptance rate 70.7%

Proof that rejection sampling works

see separate notes

Rejection sampler Example 2: Population estimation with known capture probability

Consider the following problem:

X “successes” observed from a sequence of Bernoulli trials with known success probability θ but unknown number of trials, N . We need to estimate N

e.g. N is the unknown size of a population, X is the observed number of “captures” from the population and θ is the capture probability -assumed known here but in practice estimated by clever methods (e.g. census coverage).

The likelihood for this problem is

$$p(X|N, \theta) = \binom{N}{X} \theta^X (1 - \theta)^{N-X}$$

i.e. a Binomial likelihood but with N as the unknown.

$$\begin{aligned} p(X|N, \theta) &= \binom{N}{X} \theta^X (1 - \theta)^{N-X} \\ &\propto \binom{N}{X} (1 - \theta)^{N-X} \end{aligned} \tag{9}$$

- ① “data” in this example is just X - a single count hence not boldfaced.
- ② the likelihood is zero for values of $N < X$ so no support from the data for values of the population size less than the observed count.

The posterior for the unknown number of trials

We want $p(N|X, \theta)$ which is shorthand for $\Pr(N = n|X, \theta)$, for $n = 0, 1, \dots$. Combining the binomial likelihood (9) with a prior for N , $p(N)$, (keep in mind N is discrete) leads to

$$\Pr(N = n|X, \theta) = \frac{\binom{N}{X} \theta^X (1 - \theta)^{N-X} \Pr(N = n)}{\sum_m \binom{m}{X} \theta^X (1 - \theta)^{m-X} \Pr(N = m)} \quad (10)$$

and we need to compute this for every possible n . How can we simulate a sample of values from this posterior pmf?

The inverse distribution method

If our prior for N was discrete uniform or some other bounded distribution we could tabulate the entire distribution. This means we could also tabulate the cumulative distribution function which would just be

$$\Pr(N \leq q|X, \theta) = \sum_{n=nmin}^{n=q} \Pr(N = n|X, \theta) \quad (11)$$

In this case (bounded prior) we could use the inverse distribution method to simulate the posterior for N . If the cdf, F is known and $U \sim \text{Uniform}(0, 1)$ then $F^{-1}(U)$ is a draw from the target distribution (i.e. distribution for which F is the cdf).

But if there is no *a priori* bound or the bound is very large this could be computationally inefficient.

Back to the unknown N , known p (aka θ) problem

We have

$$\Pr(N = n|X, \theta) = \frac{\binom{n}{X} \theta^X (1 - \theta)^{n-X} \Pr(N = n)}{\sum_m \binom{m}{X} \theta^X (1 - \theta)^{m-X} \Pr(N = m)}, n = 0, 1, 2, \dots \quad (12)$$

Suppose $p(N)$ is defined to be discrete uniform on some suitably chosen interval: $N \sim \text{Uniform}(L_1, L_2)$ then

$$\Pr(N = n|X, \theta) = \frac{\binom{n}{X} \theta^X (1 - \theta)^{n-X}}{\sum_m \binom{m}{X} \theta^X (1 - \theta)^{m-X}}; L_1 \leq n \leq L_2 \quad (13)$$

$$\propto \binom{n}{X} \theta^X (1 - \theta)^{n-X}; L_1 \leq n \leq L_2 \quad (14)$$

$$= 0; n < L_1, n > L_2 \quad (15)$$

i.e posterior pmf is just the likelihood “normalised” to integrate to 1.

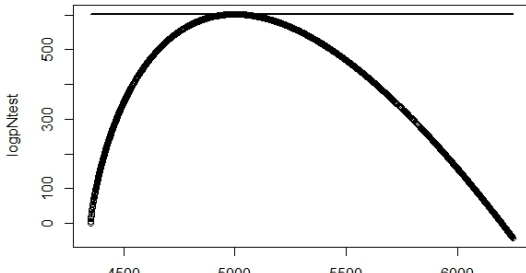
Rejection sampling for N assuming a uniform prior

For given success probability θ and observed count X the value of N that maximises the likelihood is $\hat{N} = X/\theta$.

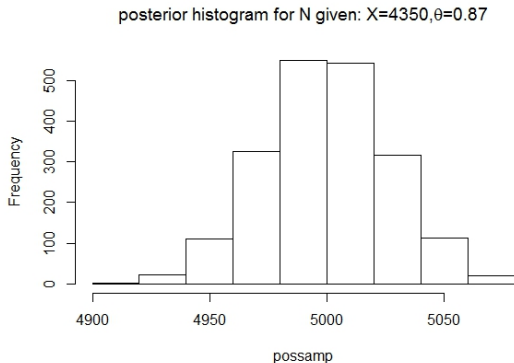
If $g(N)$ is also uniform on (L_1, L_2) , then previous argument suggests the following rejection sampling algorithm

- 1 set $M = p(X|N = \hat{N}, \theta) \propto \binom{\hat{N}}{X} (1 - \theta)^{\hat{N}-X}$
- 2 draw N^* uniformly from $U(L_1, L_2)$
- 3 accept with probability $\frac{\binom{N^*}{X} \theta^X (1 - \theta)^{N^* - X}}{M}$

Log-likelihood for N ; $n=4350, \theta=0.87$



Posterior for N given $X = 4350$, $\theta = 0.87$



posterior quantiles

```
probs=c(0.025,0.05,0.5,0.95,0.975))
quant_rejection <- quantile(possamp,probs=probs)
quant_rejection
  2.5%    5%   50%   95% 97.5%
 4945  4956 5000  5044 5053
true_quantiles
2.5%    5%   50%   95% 97.5%
4947  4956 5000  5045 5054
prior_quantiles
      2.5%      5%      50%      95%      97.5%
4400.000 4444.000 5335.000 6243.050 6288.125
```

Concluding remarks on rejection sampling

- Potentially inefficient if the approximating density is a poor approximation.
- A poor approximation is immediately revealed by low acceptance rates;
- An “adaptive” variant of rejection sampling is discussed in the literature - the idea is to improve the approximation as a simulation progresses.
- Rejection sampling struggles as the dimension of the problem increases.