

Monte Carlo algorithms project: DNA methylation of cytosines

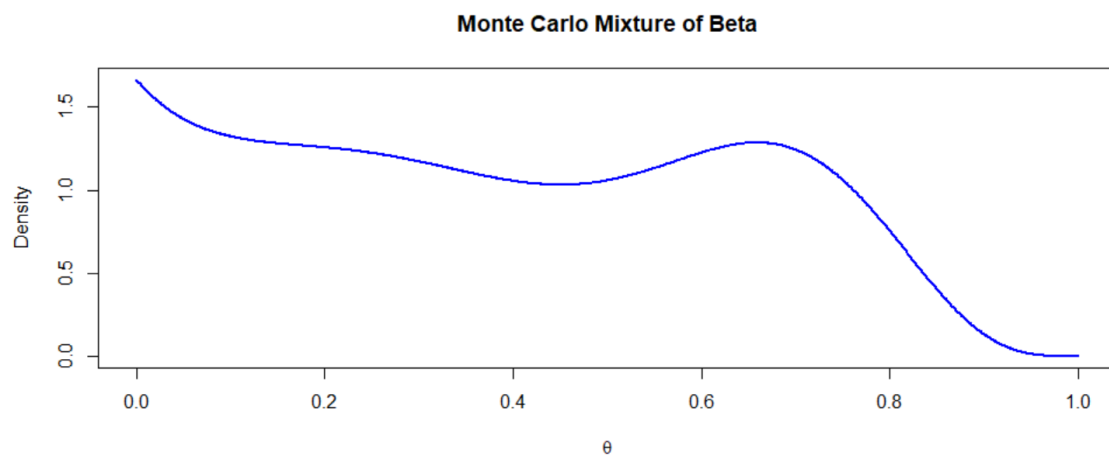
Waikei Lau (Z5349878)

Question 1

Using dbeta and plot functions in R, produce a sequence from 0 to 1 of 1000 increments and plot the density from the mixture distribution – a weighted sum of three distinct beta distributions.

$$f(x) = \frac{1}{3} \text{beta}(1,5) + \frac{1}{3} \text{beta}(3,5) + \frac{1}{3} \text{beta}(10,5)$$

$$f(x) = \frac{1}{3} * (\text{dbeta}(\text{seq}, 1,5) + \text{dbeta}(\text{seq}, 3,5) + \text{dbeta}(\text{seq}, 10,5))$$



Question 2

Calculate K as maximum of stored f(x) density = 1.67.

Calculate x_{\max} as the corresponding x value = 0.

Simulate random values for target & proposal distributions using runif(1,0,1).

Define the acceptance probability and accept if proposal is less than acceptance probability:

$$\text{Acceptance prob} = \frac{\pi(\theta|x)}{Kg(\theta)} = \frac{\text{dbeta}(x, 1,5) + \text{dbeta}(x, 3,5) + \text{dbeta}(x, 10,5)}{\text{dbeta}(x_{\max}, 1,5) + \text{dbeta}(x_{\max}, 3,5) + \text{dbeta}(x_{\max}, 10,5)}$$

Question 3

Compute acceptance rate as total number of accepted values divided by 10,000 = 0.06003

Compute theoretical acceptance rate as $1/K = 0.6$

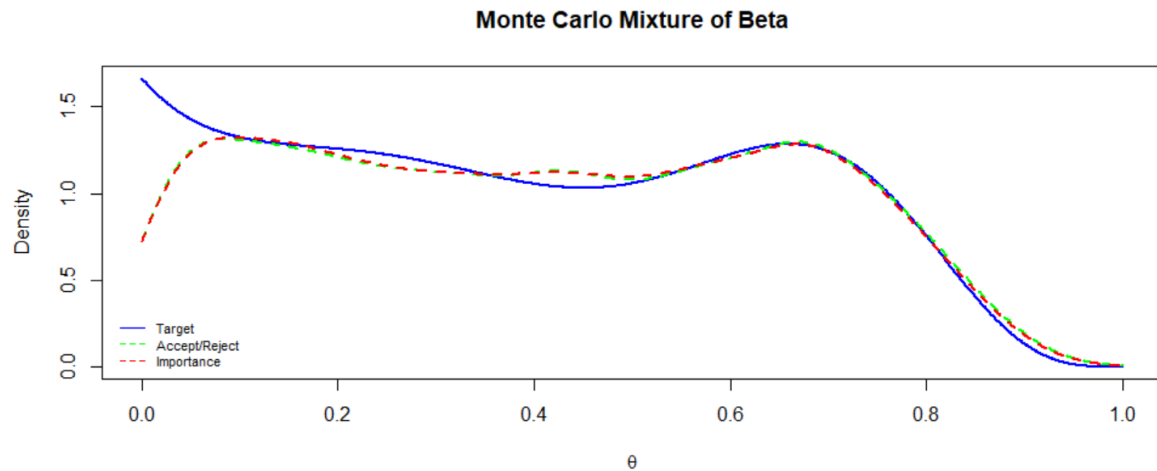
Question 4

Compute importance weights as:

$$\text{Importance weight } i = \frac{w^i}{\text{sum}(w)} \text{ where } w^i = \frac{\frac{1}{3} * (\text{dbeta}(x, 1,5) + \text{dbeta}(x, 3,5) + \text{dbeta}(x, 10,5))}{\text{dunif}(x)}$$

Question 5

Use density function plotting Accept/reject sampling in green and importance sampling in red.



Question 6

Both algorithms are good approximations of the target density however both suffer from edge effects due to the bounded support of the posterior. From visual inspection, it could be said that accept/reject sampling is marginally better than importance sampling, however the overall difference is visually negligible. Due to the similarity in results, importance sampling may be preferred in this scenario as it exhibits some advantages over accept/reject sampling.

Importance sampling does *not*:

- Exclude simulations.
- Need to compute K constant.
- Need to find a distribution $g(\theta)$

Hence importance sampling is more efficient – as it uses all samples – and is less computationally expensive as it does not require the computation of K constant or the distribution $g(\theta)$.