1 Proof of rejection sampling

We will show that rejection sampling results in draws from the target distribution for the case of a continuous random variable, θ. The proof in the discrete case follows very similar lines.

In this proof we will use the symbol data to denote the observed data. So if we observed $\mathbf{Y}^{obs} = (Y_1, \ldots, Y_n)'$ data $= (Y_1, \ldots, Y_n)'$.

For Bayesian inference the target density is the posterior

$$\begin{aligned} p(\theta|\text{data}) &= \frac{p(\text{data}|\theta)p(\theta)}{\int p(\text{data}|\theta)p(\theta)\,\mathrm{d}\theta} \\ &= \frac{q(\theta|\text{data})}{\int q(\theta|\text{data})\,\mathrm{d}\theta} \end{aligned} \tag{1}$$

where $q(\theta|\text{data}) = p(\text{data}|\theta)p(\theta)$ is the unnormalised posterior density which is the likelihood multiplied by the prior. For convenience we let $I = \int q(\theta|\text{data})\,\mathrm{d}\theta$ which is a constant not depending on θ. Let $g(\theta)$ denote an appoximation to the posterior density from which we can easily sample. If we want to refer to the value of this density at a particular value $\theta = a, say$, we will write $g_\theta(a)$. (The notation $g(\theta)$ is really just shorthand for the density function $g_\theta(a), \forall a$). Similarly, if we want to refer to the posterior density evaluated at a particular value, $\theta = a$, say we will use the notation $p_{\theta|\text{data}}(a)$, and similarly for the unnormalised posterior density $q_{\theta|\text{data}}(a)$.

U will denote a Uniform$(0, 1)$ random variable.

$r(\theta) = q(\theta|\text{data})/g(\theta)$ is the importance ratio

M is an upper bound for the importance ratio i.e $r(\theta) \leq M, \forall\theta$

The rejection sampling algorithm generates draws of θ from $g(\theta)$, then draws U and accepts θ if $U \leq \frac{q(\theta|\text{data})}{Mg(\theta)}$ i.e if $U \leq r(\theta)/M$.

We note:

1. if $U \sim$ Uniform$(0, 1)$, and $0 \leq a \leq 1$, $\Pr(U \leq a) = \int_0^a 1\,\mathrm{du} = a$.

2. For continuous random variables X and Y

$$\begin{aligned}\Pr(X \leq Y) &= \int_y \Pr(X \leq Y|Y = y)p_Y(y)\mathrm{d}y \\ &= \int_y \Pr(X \leq y)p_Y(y)\mathrm{d}y \qquad (2)\end{aligned}$$

3. For some function $h(Y)$

$$\Pr(X \leq h(Y), Y < c) = \int_{-\infty}^{c} Pr(X \leq h(Y), Y < c|Y = a)p_Y(a)\,\mathrm{d}a \qquad (3)$$

$$= \int_{-\infty}^{c} Pr(X \leq h(Y), Y = a)p_Y(a)\,\mathrm{d}a \qquad (4)$$

We want to show

$$p\left(\theta|U \leq \frac{q(\theta|\text{data})}{Mg(\theta)}\right) = p(\theta|\text{data}).$$

We need to keep in mind that θ is originally generated from $g(\theta)$. It is easiest to work with the distribution function, hence we consider

$$\Pr\left(\theta < c|U \leq \frac{q(\theta|\text{data})}{Mg(\theta)}\right)$$

and want to show

$$\Pr\left(\theta < c|U \leq \frac{q(\theta|\text{data})}{Mg(\theta)}\right) = \Pr_{\theta|\text{data}}(\theta < c|\text{data}), \forall c$$

where $\Pr_{\theta|\text{data}}(\theta < c|\text{data}) = \int_{-\infty}^{c} p(\theta|\text{data})\,\mathrm{d}\theta$. Using Bayes theorem, we can write

$$\Pr\left(\theta < c|U \leq \frac{q(\theta|\text{data})}{Mg(\theta)}\right) = \frac{\Pr(U \leq \frac{q(\theta|\text{data})}{Mg(\theta)}|\theta < c)\Pr_g(\theta < c)}{\Pr(U \leq \frac{q(\theta|\text{data})}{Mg(\theta)})} \qquad (5)$$

where $\Pr_g(\theta < c) = \int_{-\infty}^{c} g(\theta)\,\mathrm{d}\theta$.

Now noting that U and θ (and hence $\frac{q(\theta|\text{data})}{Mg(\theta)}$) are both random variables we can use (2) to write the denominator of (5) as

$$\begin{aligned}
\Pr\left(U \leq \frac{q(\theta|\text{data})}{Mg(\theta)}\right) &= \int_a \Pr(U \leq \frac{q(\theta|data)}{Mg(\theta)}|\theta = a) g_\theta(a)\,\mathrm{d}a \\
&\int_a \frac{q_{\theta|\text{data}}(a)}{Mg_\theta(a)} g_\theta(a)\,\mathrm{d}a \\
&\int_a \frac{q_{\theta|\text{data}}(a)}{M}\mathrm{d}a \\
&= \frac{I}{M} \qquad (6)
\end{aligned}$$

Turning, now to the numerator of (5) we note that

$$\begin{aligned}
\Pr(U \leq \frac{q(\theta|\text{data})}{Mg(\theta)}|\theta < c) &= \frac{\Pr(U \leq \frac{q(\theta|\text{data})}{Mg(\theta)}, \theta < c)}{\Pr_g(\theta < c)} \qquad (7) \\
&= \frac{\int_{-\infty}^{c} \Pr(U \leq \frac{q(\theta|\text{data})}{Mg(\theta)}, \theta = a)\,\mathrm{d}a}{\Pr_g(\theta < c)} \qquad (8) \\
&= \frac{\int_{-\infty}^{c} \Pr(U \leq \frac{q(\theta|\text{data})}{Mg(\theta)}|\theta = a) g_\theta(a)\,\mathrm{d}a}{\Pr_g(\theta < c)} \qquad (9) \\
&= \frac{\int_{-\infty}^{c} \frac{q_{\theta|\text{data}}(a)}{Mg_\theta(a)} g_\theta(a)\,\mathrm{d}a}{\Pr_g(\theta < c)} \qquad (10) \\
&= \frac{I \Pr_{\theta|\text{data}}(\theta < c|\text{data})}{M \Pr_g(\theta < c)} \qquad (11)
\end{aligned}$$

Substituting (11) and (6) in (5) gives

$$\begin{aligned}
\Pr\left(\theta < c | U \leq \frac{q(\theta|\text{data})}{Mg(\theta)}\right) &= \frac{\frac{I \Pr_{\theta|data}(\theta < c|\text{data})}{M \Pr_g(\theta < c)} \Pr_g(\theta < c)}{I/M} \\
&= \Pr_{\theta|\text{data}}(\theta < c|\text{data})
\end{aligned}$$

as required.