David Hitchcock
STA 6934
Problem 7.28

(Part 1) \( Y = X_1 + X_2, X_1 \sim Bin(n_1, \theta_1), X_2 \sim Bin(n_2, \theta_2) \)
Assume \( X_1 \) and \( X_2 \) are independent. Then

\[
f(x_1, x_2) = C_{x_2}^{n_2} \theta_2^{x_2} (1 - \theta_2)^{n_2 - x_2} C_{x_1}^{n_1} \theta_1^{x_1} (1 - \theta_1)^{n_1 - x_1}
\]

where \( C_n^m = \frac{n!}{x!(n-x)!} \).

Letting \( y - x_1 = x_2 \), we have

\[
f(x_1, y) = C_{x_1}^{n_1} \theta_1^{x_1} (1 - \theta_1)^{n_1 - x_1} C_{y-x_1}^{n_2} \theta_2^{y-x_1} (1 - \theta_2)^{n_2 - y+x_1}
\]

\[
f(y) = \sum_{x_1} C_{x_1}^{n_1} \theta_1^{x_1} (1 - \theta_1)^{n_1 - x_1} C_{y-x_1}^{n_2} \theta_2^{y-x_1} (1 - \theta_2)^{n_2 - y+x_1}
\]

Since \( Y_1, Y_2, Y_3 \) independent,

\[
L(\theta_1, \theta_2) = \prod_{i=1}^{3} \left\{ \sum_{x_i} C_{x_i}^{n_i} C_{y_i-x_i}^{n_2} \theta_2^{y_i-x_i} (1 - \theta_2)^{n_2 - y_i+x_i} \right\}
\]

For \( i = 1, n_1 = 5, n_2 = 5, y = 7 \Rightarrow X_1 \in \{2, 3, 4, 5\} \)
For \( i = 2, n_1 = 6, n_2 = 4, y = 5 \Rightarrow X_1 \in \{1, 2, 3, 4, 5\} \)
For \( i = 3, n_1 = 4, n_2 = 6, y = 6 \Rightarrow X_1 \in \{2, 3, 4, 5\} \)

So the 3 factors of the likelihood, given these data, are:

\[
i = 1: \sum_{x_1=2}^{5} C_{x_1}^{5} C_{7-x_1}^{5} \theta_1^{x_1} (1 - \theta_1)^{5-x_1} \theta_2^{7-x_1} (1 - \theta_2)^{-2+x_1}
\]

\[
i = 2: \sum_{x_1=1}^{5} C_{x_1}^{6} C_{5-x_1}^{4} \theta_1^{x_1} (1 - \theta_1)^{6-x_1} \theta_2^{5-x_1} (1 - \theta_2)^{-1+x_1}
\]

\[
i = 3: \sum_{x_1=0}^{4} C_{x_1}^{4} C_{6-x_1}^{6} \theta_1^{x_1} (1 - \theta_1)^{4-x_1} \theta_2^{6-x_1} (1 - \theta_2)^{x_1}
\]
After plugging in the values of $X_1$, for $i = 1, 2, 3$, and calculating the sums, omitting the gory details, we get for the likelihood:

$$L(\theta_1, \theta_2) = \{10 \theta_1^2 (1-\theta_1)^3 \theta_2^5 + 50 \theta_1^3 (1-\theta_1)^2 \theta_2^4 (1-\theta_2) + 50 \theta_1^4 (1-\theta_1) \theta_2^3 (1-\theta_2)^2 + 10 \theta_1^5 \theta_2^2 (1-\theta_2)^3 \} \times$$

$$\{6 \theta_1 (1-\theta_1)^5 \theta_2^4 + 60 \theta_1^2 (1-\theta_1)^4 \theta_2^3 (1-\theta_2) +$$

$$120 \theta_1^3 (1-\theta_1)^3 \theta_2^2 (1-\theta_2)^2 + 60 \theta_1^4 (1-\theta_1)^2 \theta_2 (1-\theta_2)^3 + 6 \theta_1^5 (1-\theta_1)(1-\theta_2)^4 \} \times$$

$$\{(1-\theta_1)^4 \theta_2^5 + 24 \theta_1 (1-\theta_1)^3 \theta_2^4 (1-\theta_2) +$$

$$90 \theta_1^2 (1-\theta_1)^2 \theta_2^3 (1-\theta_2)^2 + 80 \theta_1^3 (1-\theta_1) \theta_2^2 (1-\theta_2)^3 + 15 \theta_1^4 \theta_2 (1-\theta_2)^4 \}$$

(Part 2) With a uniform prior $\pi(\theta_1, \theta_2) = 1$ on $(\theta_1, \theta_2)$, the posterior is

$$\pi(\theta_1, \theta_2 | y) \propto L(\theta_1, \theta_2) \pi(\theta_1, \theta_2) = L(\theta_1, \theta_2).$$

The normalizing constant is

$$\int_0^1 \int_0^1 \pi(\theta_1, \theta_2 | y) d\theta_1 d\theta_2$$

$$= \int_0^1 \int_0^1 L(\theta_1, \theta_2) d\theta_1 d\theta_2$$

which can be found via Maple to be $29993/7927920 = .00378321$.

So the posterior density $\pi(\theta_1, \theta_2 | y)$ is $\frac{29993}{7927920} L(\theta_1, \theta_2)$. It is clear from the form of $L(\theta_1, \theta_2)$ that the conditional density of $\theta_1$, $f_1(\theta_1 | \theta_2, y)$ is a mixture of beta distributions. Similarly, the conditional density of $\theta_2$ is also a mixture of beta distributions. So the Gibbs sampler algorithm is:

- Start with arbitrary $(\theta_1^{(0)}, \theta_2^{(0)})$, say, $(0.5, 0.5)$.

- At step $t+1$, generate $\theta_1^{(t+1)} \sim f_1(\theta_1^{(t)} | \theta_2^{(t)}, y)$.

- Generate $\theta_2^{(t+1)} \sim f_2(\theta_2^{(t)} | \theta_1^{(t+1)}, y)$.

- Repeat previous two steps for $t + 2, t + 3, \ldots$
(Part 3) The transformation of the parameters which may be considered is the logit transformation, so that the parameters are \( \log \frac{\theta_1}{1-\theta_1} \) and \( \log \frac{\theta_2}{1-\theta_2} \). This may help convergence. The posterior could then be expressed as \( \pi(\theta_1, \theta_2 \mid y) \propto \exp(a \log \frac{\theta_1}{1-\theta_1} + b \log \frac{\theta_2}{1-\theta_2}) \), where \( a \) and \( b \) were constants. However, this would require that in its most simplified form, the likelihood would contain exponents that “matched”, i.e., that \( \theta_i \) and \( 1 - \theta_i \) had equal exponents for \( i = 1, 2 \). This requirement seems to be heavily dependent on the observed data, so this transformation may not work in all cases.

Whether a Metropolis-Hastings algorithm would speed up convergence would depend on whether the posterior distribution were bounded. If so, we could use an independent Metropolis-Hastings algorithm that would achieve uniform ergodicity and beat the Gibbs sampler. However, a 3-D plot of the posterior using Maple shows that over the region \( \theta_1 \in [0, 1], \theta_2 \in [0, 1] \), the posterior increases without bound at one of the boundaries. So uniform ergodicity cannot be achieved.