Anna Sedlackova | AXS1202 Homework 5

Q1:

a) $\int_0^1 p(y|\theta, n)\, d\theta = \frac{1}{1+n}$ derive posterior distribution

$$p(y|\theta, n) = \binom{n}{y} \theta^y (1-\theta)^{n-y}$$

$$\int_0^1 p(y|\theta, n)\, d\theta = \frac{1}{n+1} = p(y|n)$$

$p(\theta|n) = 1$ (uniform prior on θ)

posterior distribution for $\theta = p(\theta|y, n) =$

$$= \frac{p(y|\theta, n) \cdot p(\theta|n)}{p(y|n)} \quad \text{(using Baye's theorem)}$$

$$= \frac{\binom{n}{y} \theta^y (1-\theta)^{n-y} \cdot 1}{\frac{1}{n+1}} = \boxed{(n+1)\binom{n}{y} \theta^y (1-\theta)^{n-y}}$$

b) $n = 4$, $\theta = 3/4$. plot includes $y = 0$

$$p(y|\theta = 3/4, n = 4) = \binom{4}{y} (3/4)^y (1/4)^{4-y} \quad y \in \{0, \dots, 4\}$$

plot included at the end of the writeup

c) the four plots will be done as follows: (using the formulas from a)

$$p(\theta|y=1, n=1) = (1+1)\binom{1}{1} \theta^1 (1-\theta)^0 = \boxed{2\theta}$$

$$p(\theta|y=2, n=2) = (1+2)^1 \binom{2}{2} \theta^2 (1-\theta)^0 = \boxed{3\theta^2}$$

$$p(\theta|y=2, n=3) = (1+3)\binom{3}{2} \theta^2 (1-\theta)^1 = 4\frac{6}{2}\theta^2(1-\theta) =$$

$$= 12\theta^2(1-\theta) = \boxed{12\theta^2 - 12\theta^3}$$

$$p(\theta|y=3, n=4) = (1+4)\binom{4}{3} \theta^3 (1-\theta)^1 = 5\left(\frac{4 \times 3 \times 2}{3 \times 2}\right)\theta^3(1-\theta) =$$

$$= 20\theta^3(1-\theta) = \boxed{20\theta^3 - 20\theta^4}$$

4 plots included at the end of the writeup

Q2:

a) five plots were generated and are attached at the end of the writeup.

* The derivation for the eqn goes as follows for drawing the posterior probability plots:

$$\underbrace{p(h_i \mid \vec{d})}_{\text{posterior}} = \underbrace{\alpha}_{\substack{\text{norm}\\\text{constant}}} \underbrace{[P(\vec{d} \mid h_i)]}_{\text{likelyhood}} \underbrace{P(h_i)}_{\text{prior}}$$

from the textbook we know that:

prior:
$P(h_1) = 0.1$ $\quad$ $P(h_4) = 0.2$
$P(h_2) = 0.2$ $\quad$ $P(h_5) = 0.1$
$P(h_3) = 0.4$

likelyhood:
$$P(\vec{d} \mid h_1) = (1)^{n_c} (0)^{n_l}$$
$$P(\vec{d} \mid h_2) = (0.75)^{n_c} (0.25)^{n_l}$$
$$P(\vec{d} \mid h_3) = (0.5)^{n_c} (0.5)^{n_l}$$
$$P(\vec{d} \mid h_4) = (0.25)^{n_c} (0.75)^{n_l}$$
$$P(\vec{d} \mid h_5) = (0)^{n_c} (1)^{n_l}$$

where

$\underset{\text{cherries}}{n_c} + \underset{\text{limes}}{n_l} = N$ $\quad$ n_c number of cherries
$\quad$ n_l number of limes

we also know that

$$\sum_{i=1}^{5} p(h_i \mid d) = 1$$

so then $\alpha \{ (0.1)(1)^{n_c} + (0.2)(0.75)^{n_c}(0.25)^{n_l} + (0.4)(0.5)^{n_c}(0.5)^{n_l} + (0.2)(0.25)^{n_c} (0.75)^{n_l} + (0.1)(1)^{n_l} \} = 1$

from there we can derive the eqn to graph each of the posterior probability plots such as

$$h_4\ p(h_4 \mid \vec{d}) = \frac{\overbrace{0.25^{n_c}\ 0.75^{n_l}}^{\text{likelyhood}}\ \overbrace{0.2}^{\text{prior}}}{\underbrace{(0.1)(1)^{n_c}(0)^{n_l} + (0.2)(0.75)^{n_c}(0.25)^{n_l} + (0.4)(0.5)^{n_c}(0.5)^{n_l} + (0.2)(0.25)^{n_c}(0.75)^{n_l} + (0.1)(1)^{n_l}(0)^{n_c}}_{\alpha}}$$

similarity done for $p(h_1, \vec{d}), p(h_2, \vec{d}), p(h_3, \vec{d}), p(h_4, \vec{d})$.

* The derivation for the equations is as follows for the graphs of probability:

$$P(d_{N+1} = \text{lime} \mid \vec{d}) = \sum_{i=1}^{5} P(d_{N+1} = \text{lime} \mid h_i) \underbrace{P(h_i \mid \vec{d})}_{\text{posterior}}$$

$P(d_{N+1} = \text{lime} \mid h_1) = 0$
$P(d_{N+1} = \text{lime} \mid h_2) = 0.25$
$P(d_{N+1} = \text{lime} \mid h_3) = 0.5$
$P(d_{N+1} = \text{lime} \mid h_4) = 0.75$
$P(d_{N+1} = \text{lime} \mid h_5) = 1$

$$P(d_{N+1} = \text{lime} \mid \vec{d}) = (0)P(h_1 \mid \vec{d}) + 0.25P(h_2 \mid \vec{d}) + 0.5P(h_3 \mid \vec{d}) + 0.75P(h_4 \mid \vec{d}) + (1)P(h_5 \mid \vec{d})$$

From our results we can see that in the posterior probability graph, the probability of each bag increases with each candy pull, this is very apparent for h_1 and h_5 where after 20 candy pun the algorith is almost certain what type of bag it is. For h_2 and h_4 there are oscillations and cerbinty isn't established until upto the 80th candy pull. We can see the predictive probability increasing as across the five graphs as we are increasing the percentage of lime candy contents in the bags.

a) Plots are attached at the end of the write-up and labelled.

b) find N st $\max(P(h_1|\vec{d}), \dots, P(h_5|\vec{d})) \geq 0.9$
where $\vec{d} = (d_1, \dots, d_N)$

we know that $n_c + n_l = N$ $\quad n_c \geq 0$
however, it is $\quad n_l \geq 0$
difficult to simplify this and solve for (the problem is non linear).

continuing $\max\left(\frac{(1)^{n_c}(0)^{n_l}}{\alpha}, \frac{(0.75)^{n_c}(0.25)^{n_l}}{\alpha}, \frac{(0.5)^{n_c}(0.5)^{n_l}}{\alpha}, \frac{(0.25)^{n_c}(0.75)^{n_l}}{\alpha}, (0.1) \times \frac{(0)^{n_c}(1)^{n_l}}{\alpha}\right) \geq$

≥ 0.9

where $\alpha = (0.1)(1)^{n_c} + (0.2)(0.75)^{n_c}(0.25)^{n_l} + (0.4)(0.5)^{n_c}(0.5)^{n_l} + (0.2)(0.25)^{n_c}(0.75)^{n_l} + (0.1)(1)^{n_l}$

Q3: $C_1 \rightarrow$

a) $P(C_1) = 2/3 \quad P(C_2) = 1/3 \quad \mu_1 < \mu_2$

$P(x) = P(C_1)P(x|C_1) + P(C_2)P(x|C_2)$

$$\Rightarrow \boxed{P(x) = \left(\frac{2}{3}\right)\frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left(\frac{-(x-\mu_1)^2}{\sigma_1^2}\right) + \left(\frac{1}{3}\right)\frac{1}{\sqrt{2\pi\sigma_2^2}}\exp \; \exp\left(\frac{-(x-\mu_2)^2}{\sigma_2^2}\right)}$$

b) $$P(\text{error}) = \sum_{i=1}^{2}\int_R p(\vec{x}|C_i)\,p(C_i)\,d\vec{x}$$

$$= \int_{R12} p(x|C_1)P(C_1)dx + \int_{R21} P(x|C_2)P(C_2)dx$$

$R_{12} \rightarrow C_1$ classified as C_2

$R_{21} \rightarrow C_2$ classified as C_1

$$P(\text{error}) = \int_\theta^\infty P(x|C_1)P(C_1)\,dx \int_{-\infty}^{\theta} P(x|C_2)P(C_2)dx$$

$$= \boxed{\int_\theta^\infty \left(\frac{2}{3}\right)\left[\frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left(\frac{-(x-\mu_1)^2}{\sigma_1^2}\right)\right]dx + \int_{-\infty}^{\theta}\left(\frac{1}{3}\right)\times \left[\frac{1}{\sqrt{2\pi\sigma_2^2}}\exp\left(\frac{-(x-\mu_2)^2}{\sigma_2^2}\right)\right]dx}$$

c) to minimize misclassification of error, take derivativ of P(error) w.r.t θ and set to zero

$$P(\text{error}) = \int_\theta^\infty \left(\frac{2}{3}\right)\left[\frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left(-\frac{(x-\mu_1)^2}{\sigma_1^2}\right)\right]dx + \int_{-\infty}^{\theta}\left(\frac{1}{3}\right)\left[\frac{1}{\sqrt{2\pi\sigma_2^2}}\exp\left(-\frac{(x-\mu_2)^2}{\sigma_2^2}\right)\right]dx$$

$$0 = \frac{dP(\text{error})}{d\theta}$$

from 2nd fundamental thm of calculus

$$0 = \left[0 - \left(\frac{2}{3}\right)\frac{1}{\sqrt{2\pi\sigma_1}}\exp\left(-\frac{(\theta-\mu_1)^2}{\sigma_1^2}\right)\right] +$$ continued on 2nd page

$$\left[\left(\frac{1}{3}\right)\frac{1}{\sqrt{2\pi\sigma_2^2}}\exp\left(\frac{-(\theta-\mu_2)^2}{\sigma_2^2}\right)-0\right]$$

simplifications

$$\Rightarrow \left(\frac{2}{\cancel{3}}\right)\left(\cancel{\frac{1}{\sqrt{2\pi}}}\right)\left(\frac{1}{\sigma_1}\right)\exp\left(\frac{-(\theta-\mu_1)^2}{\sigma_1^2}\right)=\left(\frac{1}{\cancel{3}}\right)\left(\cancel{\frac{1}{\sqrt{2\pi}}}\right)\left(\frac{1}{\sigma_2}\right)\exp\left(\frac{-(\theta-\mu_2)^2}{\sigma_2^2}\right)$$

$$\Rightarrow \ln\left(\frac{2}{\sigma_1}\right)+\left[\frac{-(\theta-\mu_1)^2}{\sigma_1^2}\right]=\ln\left(\frac{1}{\sigma_2}\right)+\left[\frac{-(\theta-\mu_2)^2}{\sigma_2^2}\right]$$

$$\Rightarrow \ln\left(\frac{2\sigma_2}{\sigma_1}\right)+\left[\frac{-1}{\sigma_1^2}\left(\sigma^2-2\mu_1\theta+\mu_1^2\right)\right]=\left[\frac{-1}{\sigma_2^2}\left(\theta^2-2\mu_2\theta+\mu_2^2\right)\right]$$

$$\Rightarrow \ln\left(\frac{2\sigma_2}{\sigma_1}\right)+\left[\left(\frac{-1}{\sigma_1^2}\right)\theta^2+\left(\frac{2\mu_1}{\sigma_1^2}\right)\theta-\frac{\mu_1^2}{\sigma_1^2}\right]=\left[\left(\frac{-1}{\sigma_2^2}\right)\theta^2+\left(\frac{2\mu_2}{\sigma_2^2}\right)\theta-\frac{\mu_2^2}{\sigma_2^2}\right]$$

$$\Rightarrow \left(\frac{1}{\sigma_2^2}-\frac{1}{\sigma_1^2}\right)\theta^2+\left(\frac{2\mu_1}{\sigma_1^2}-\frac{2\mu_2}{\sigma_2^2}\right)\theta+\left\{\frac{\mu_2^2}{\sigma_2^2}-\frac{\mu_1^2}{\sigma_1^2}+\ln\left(\frac{2\sigma_2}{\sigma_1}\right)\right\}=$$

$$=0$$

$$\Rightarrow \theta^2+\left[\frac{\left(\frac{2\mu_1}{\sigma_1^2}-\frac{2\mu_2}{\sigma_2^2}\right)}{\left(\frac{1}{\sigma_2^2}-\frac{1}{\sigma_1^2}\right)}\right]\theta+\left[\frac{\left(\frac{\mu_2^2}{\sigma_2^2}-\frac{\mu_1^2}{\sigma_1^2}+\ln\left(\frac{2\sigma_2}{\sigma_1}\right)\right)}{\left(\frac{1}{\sigma_2^2}-\frac{1}{\sigma_1^2}\right)}\right]=0$$

$$\Rightarrow \theta_1^2+\frac{\left(\frac{2\mu_1}{\sigma_1^2}-\frac{2\mu_2}{\sigma_2^2}\right)}{\left(\frac{1}{\sigma_2^2}-\frac{1}{\sigma_1^2}\right)}\theta+\left[\frac{\frac{\mu_1}{\sigma_1^2}-\frac{\mu_2}{\sigma_2^2}}{\left(\frac{1}{\sigma_2^2}-\frac{1}{\sigma_1^2}\right)}\right]^2-\left[\frac{\left(\frac{\mu_1}{\sigma_1^2}-\frac{\mu_2}{\sigma_2^2}\right)}{\frac{1}{\sigma_2^2}-\frac{1}{\sigma_1^2}}\right]^2+$$

$$+\left[\frac{\left(\frac{\mu_2^2}{\sigma_2^2}-\frac{\mu_1^2}{\sigma_1^2}+\ln\left(\frac{2\sigma_2}{\sigma_1}\right)\right)}{\left(\frac{1}{\sigma_2^2}-\frac{1}{\sigma_1^2}\right)}\right]=0$$

$\Rightarrow$ on the other page

$$\Rightarrow \left[\theta + \frac{\left(\frac{\mu_1}{\sigma_1^2} - \frac{\mu_2}{\sigma_2^2}\right)}{\left(\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}\right)}\right]^2 - \left[\frac{\left(\frac{\mu_1}{\sigma_1^2} - \frac{\mu_2}{\sigma_2^2}\right)^2}{\left(\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}\right)^2}\right] + \frac{\left(\frac{\mu_2^2}{\sigma_2^2} - \frac{\mu_1^2}{\sigma_1^2} + \ln\left(\frac{2\sigma_2}{\sigma_1}\right)\right)}{\left(\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}\right)}$$

simplify

$$= \frac{\left(\frac{\mu_1}{\sigma_1^2} - \frac{\mu_2}{\sigma_2^2}\right)}{\left(\frac{\sigma_1^2}{\sigma_2^2\sigma_1^2} - \frac{\sigma_2^2}{\sigma_2^2\sigma_1^2}\right)} = \frac{\sigma_1^2\sigma_2^2\left(\frac{\mu_1}{\sigma_1^2} - \frac{\mu_2}{\sigma_2^2}\right)}{\sigma_1^2 - \sigma_2^2} = \frac{\mu_1\sigma_2^2 - \mu_2\sigma_1^2}{\sigma_1^2 - \sigma_2^2}$$

simplify

$$= \frac{\frac{\mu_1^2}{\sigma_1^4} - \frac{2\mu_1\mu_2}{\sigma_1^2\sigma_2^2} + \frac{\mu_2^2}{\sigma_2^4}}{\left(\frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^2\sigma_2^2}\right)} = \frac{\sigma_1^4\sigma_2^4\left(\frac{\mu_1^2}{\sigma_1^4} - \frac{2\mu_1\mu_2}{\sigma_1^2\sigma_2^2} + \frac{\mu_2^2}{\sigma_2^4}\right)}{(\sigma_1^2 - \sigma_2^2)^2} = \frac{\mu_1^2\sigma_2^4 - 2\mu_1\mu_2\sigma_1^2\sigma_2^2 + \mu_2^2\sigma_1^4}{(\sigma_1^2 - \sigma_2^2)^2}$$

simplify

$$= \frac{\sigma_1^2\sigma_2^2\left(\frac{\mu_2^2}{\sigma_2^2} - \frac{\mu_1^2}{\sigma_1^2} + \ln\left(\frac{2\sigma_2}{\sigma_1}\right)\right)}{(\sigma_1^2 - \sigma_2^2)} =$$

$$\frac{\left(\mu_2^2\sigma_1^2 - \mu_1^2\sigma_2^2 + \sigma_1^2\sigma_2^2\ln\left(\frac{2\sigma_2}{\sigma_1}\right)\right)}{\sigma_1^2 - \sigma_2^2}$$

$$= \frac{(\sigma_1^2 - \sigma_2^2)\left(\mu_2^2\sigma_1^2 - \mu_1^2\sigma_2^2 + \sigma_1^2\sigma_2^2\ln\left(\frac{2\sigma_2}{\sigma_1}\right)\right)}{(\sigma_1^2 - \sigma_2^2)^2}$$

$$= \frac{1}{(\sigma_1^2 - \sigma_2^2)^2}\left[\mu_2^2\sigma_1^4 - \mu_1^2\sigma_1^2\sigma_2^2 + \sigma_1^4\sigma_2^2\ln\left(\frac{2\sigma_2}{\sigma_1}\right) \;\mu_2^2\sigma_1^2\sigma_2^2 + \mu_1^2\sigma_2^4 - \sigma_1^2\sigma_2^4\ln\left(\frac{2\sigma_2}{\sigma_1}\right)\right]$$

Putting all this together:

$$\Rightarrow \left[\theta + \left(\frac{\mu_1\sigma_2^2 - \mu_2\sigma_1^2}{\sigma_1^2 - \sigma_2^2}\right)\right]^2 + \frac{1}{(\sigma_1^2 - \sigma_2^2)^2}\left\{\begin{aligned} &-\mu_1\sigma_2^4 + 2\mu_1\mu_2\sigma_1^2\sigma_2^2 - \mu_2^2\sigma_1^4 + \\ &+\mu_2^2\sigma_1^4 - \mu_1^2\sigma_1^2\sigma_2^2 + \sigma_1^4\sigma_2^2\ln\left(\frac{2\sigma_2}{\sigma_1}\right) - \\ &-\mu_2^2\sigma_1^2\sigma_2^2 + \mu_1^2\sigma_2^4 - \sigma_1^2\sigma_2^4\left(\ln\left(\frac{2\sigma_2}{\sigma_1}\right)\right)\end{aligned}\right\} = 0$$

$$\Rightarrow \left[\theta + \left(\frac{\mu_1\sigma_2^2 - \mu_2\sigma_1^2}{\sigma_1^2 - \sigma_2^2}\right)\right]^2 + \frac{1}{(\sigma_1^2 - \sigma_2^2)^2}\left\{(-\sigma_1^2\sigma_2^2)(\mu_1^2 - 2\mu_1\mu_2 + \mu_2^2) + (\sigma_1^2 - \sigma_2^2)\sigma_1^2\sigma_2^2\ln\left(\frac{2\sigma_2}{\sigma_1}\right)\right\} = 0$$

$$\Rightarrow \left[\theta + \left(\frac{\mu_1\sigma_2^2 - \mu_2\sigma_1^2}{\sigma_1^2 - \sigma_2^2}\right)\right]^2 = \frac{1}{(\sigma_1^2 - \sigma_2^2)^2}\left\{\sigma_1^2\sigma_2^2(\mu_1 - \mu_2)^2 - (\sigma_1^2 - \sigma_2^2)\sigma_1^2\sigma_2^2\ln\left(\frac{2\sigma_2}{\sigma_1}\right)\right\}$$

try positive root

$$\theta + \frac{\mu_1\sigma_2^2 - \mu_2\sigma_1^2}{\sigma_1^2 - \sigma_2^2} = \frac{\sigma_1\sigma_2}{\sigma_1^2 - \sigma_2^2}\sqrt{(\mu_1 - \mu_2)^2 - (\sigma_1^2 - \sigma_2^2)\ln\left(\frac{2\sigma_2}{\sigma_1}\right)}$$

$$\Rightarrow \boxed{\theta = \left(\frac{1}{\sigma_1^2 - \sigma_2^2}\right)\left(\mu_2\sigma_1^2 - \mu_1\sigma_2^2 + \sigma_1\sigma_2\sqrt{(\mu_1 - \mu_2)^2 - (\sigma_1^2 - \sigma_2^2)\ln\left(\frac{2\sigma_2}{\sigma_1}\right)}\right)}$$

Q4:

a) derive update rule for μ_k using the objective fn

$$D = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{n,k} \|x_n - \mu_k\|^2$$

solving value where gradient is zero
express answer in scalar form and vector form

The objective fn is $D = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{n,k} \|x_n - \vec{\mu}_k\|^2$,

differentiate w.r.t to the mean (the parameter we want to estimate)

$$\frac{\partial D}{\partial \mu_k} = 2 \sum_{n=1}^{N} r_{n,k} (\vec{x}_n - \vec{\mu}_k)$$

solve for when the gradient is zero:

$$\frac{\partial D}{\partial \vec{\mu}_k} = 2 \sum_{n=1}^{N} r_{n,k} (\vec{x}_n - \vec{\mu}_k) = 0$$

$$2 \sum r_{n,k} \vec{x}_n - 2 \sum r_{n,k} \vec{\mu}_k \Rightarrow$$

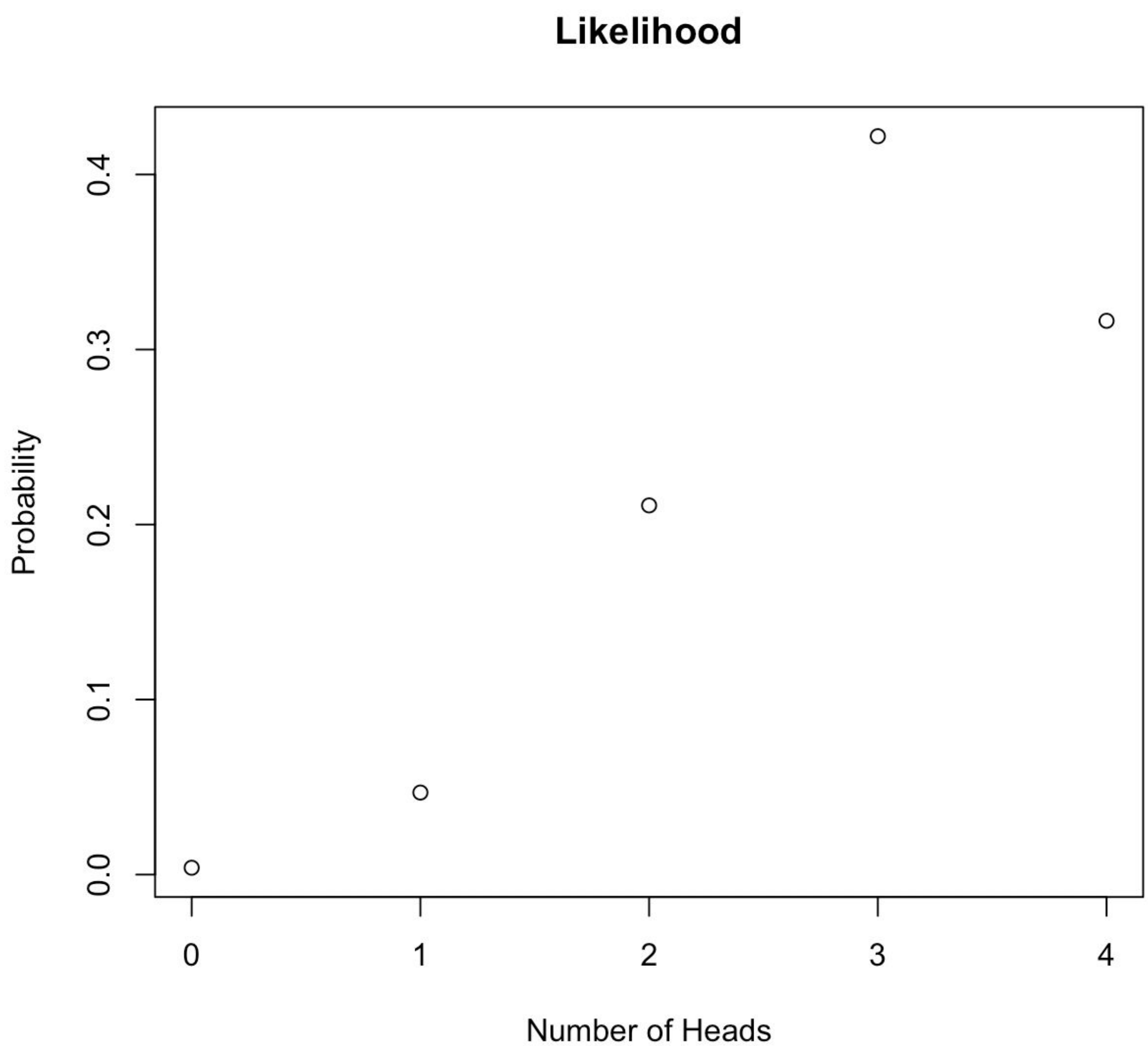
$$\Rightarrow \quad \vec{\mu}_k = \frac{\cancel{2} \sum r_{n,k} \vec{x}_n}{\cancel{2} \sum r_{n,k}} = \boxed{\frac{\sum_n r_{n,k} \vec{x}_n}{\sum_n r_{n,k}}}$$

b) code on canvas
plots attached at the end of the writeup

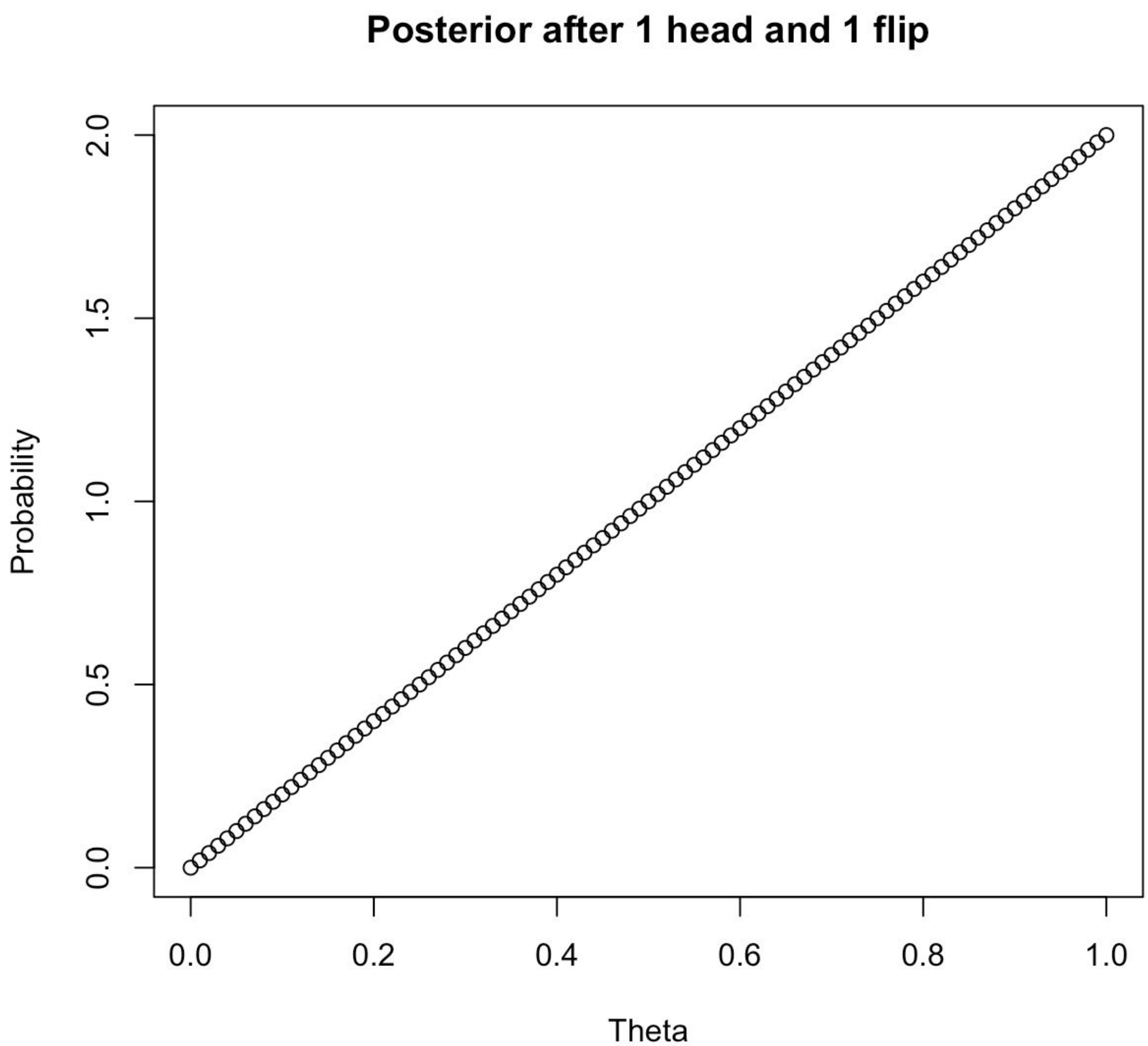
scalar derivation:

$$\mu_{k,i} = \frac{r_{i,k} x_i}{\sum_n r_{n,k}}$$

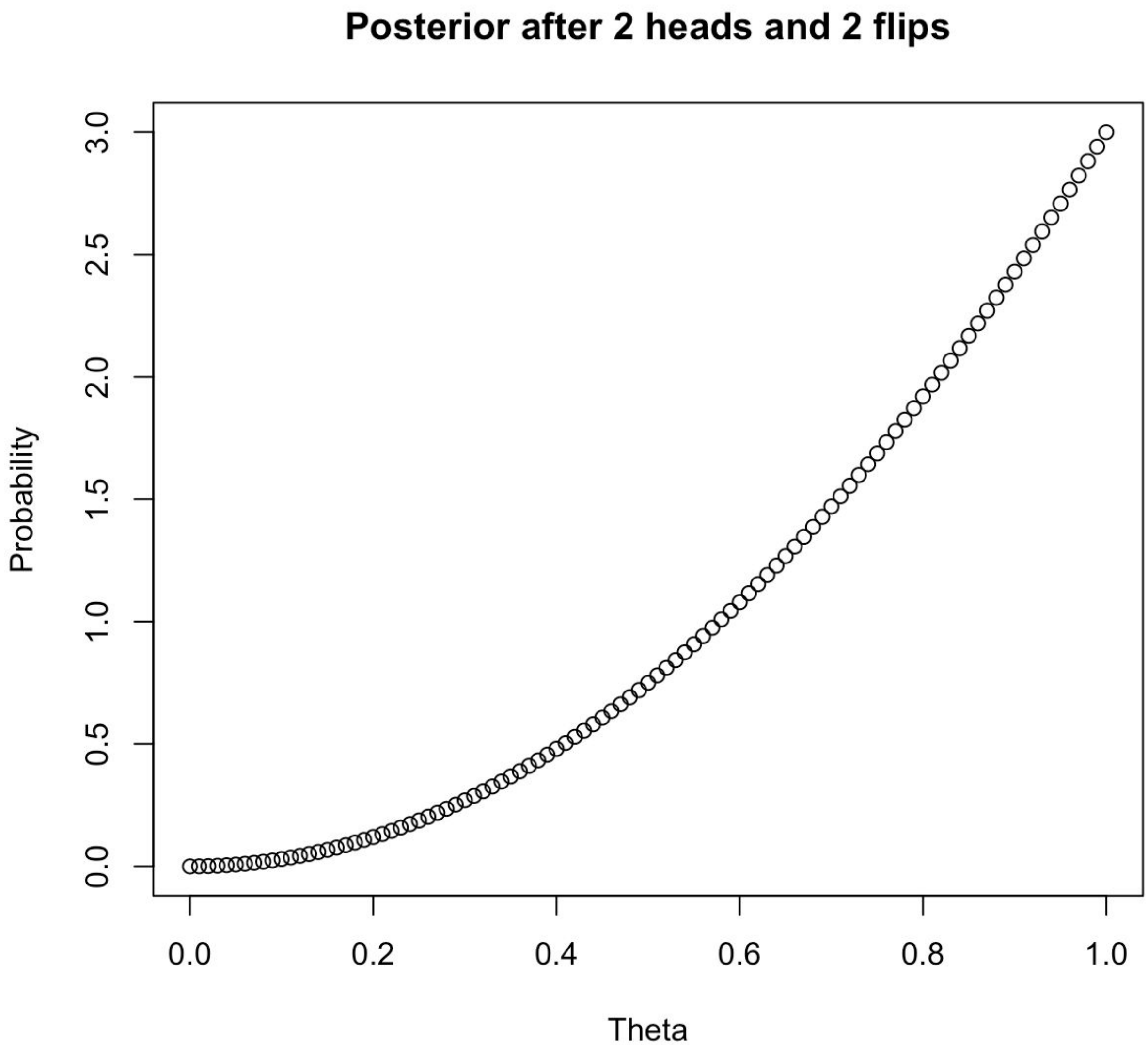
Question 1, part b



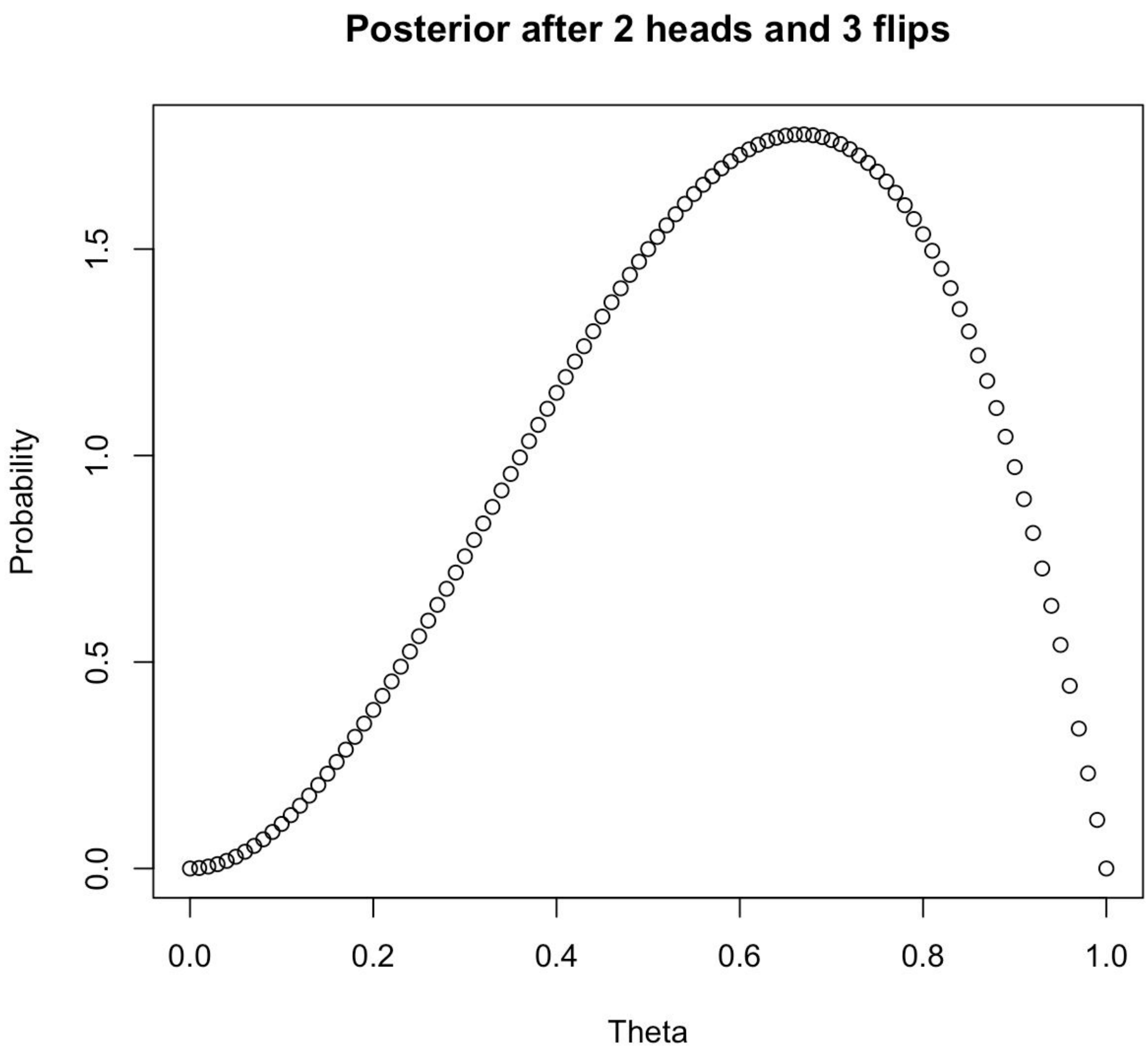
Question 1, part c



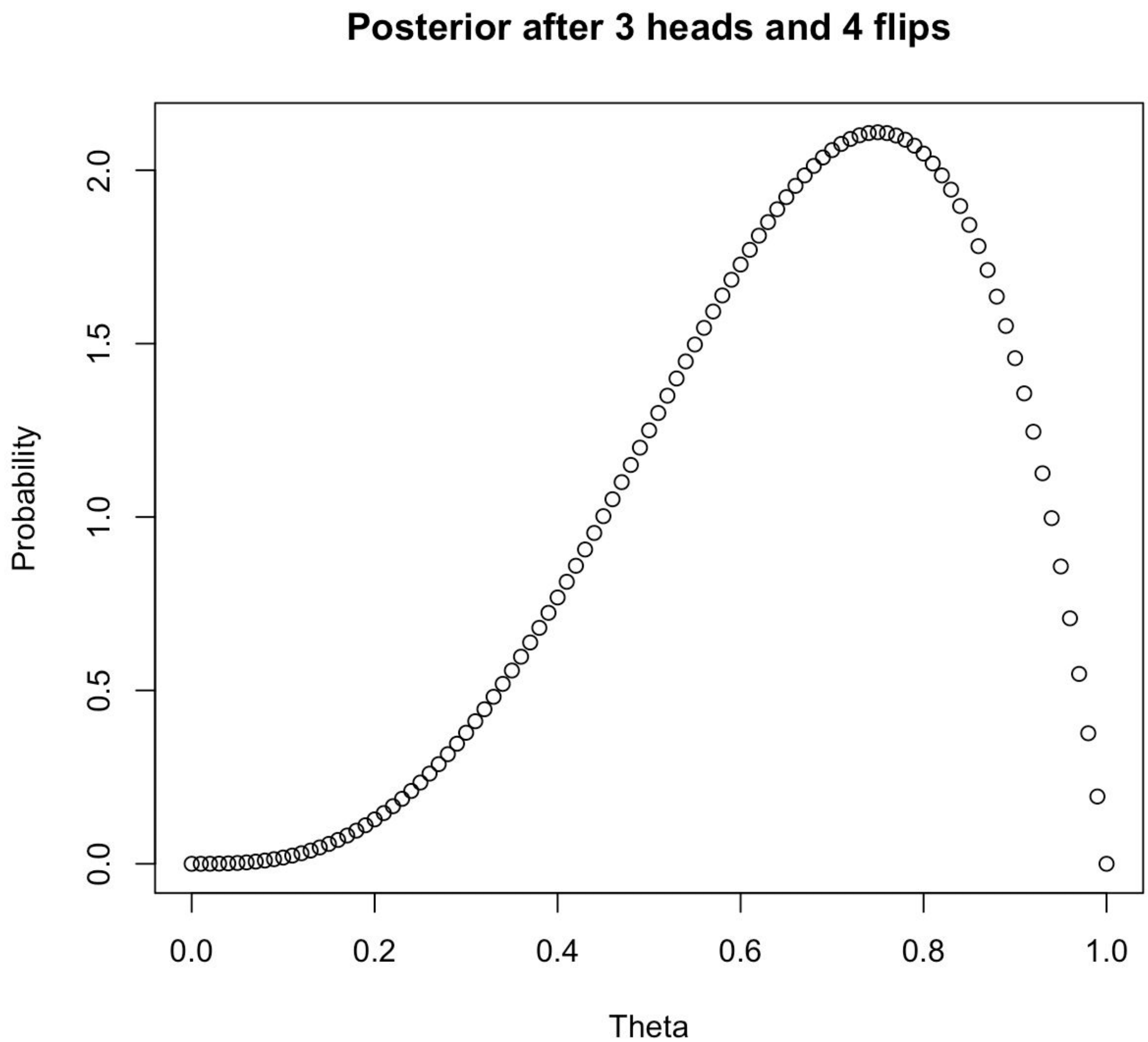
Question 1, part c



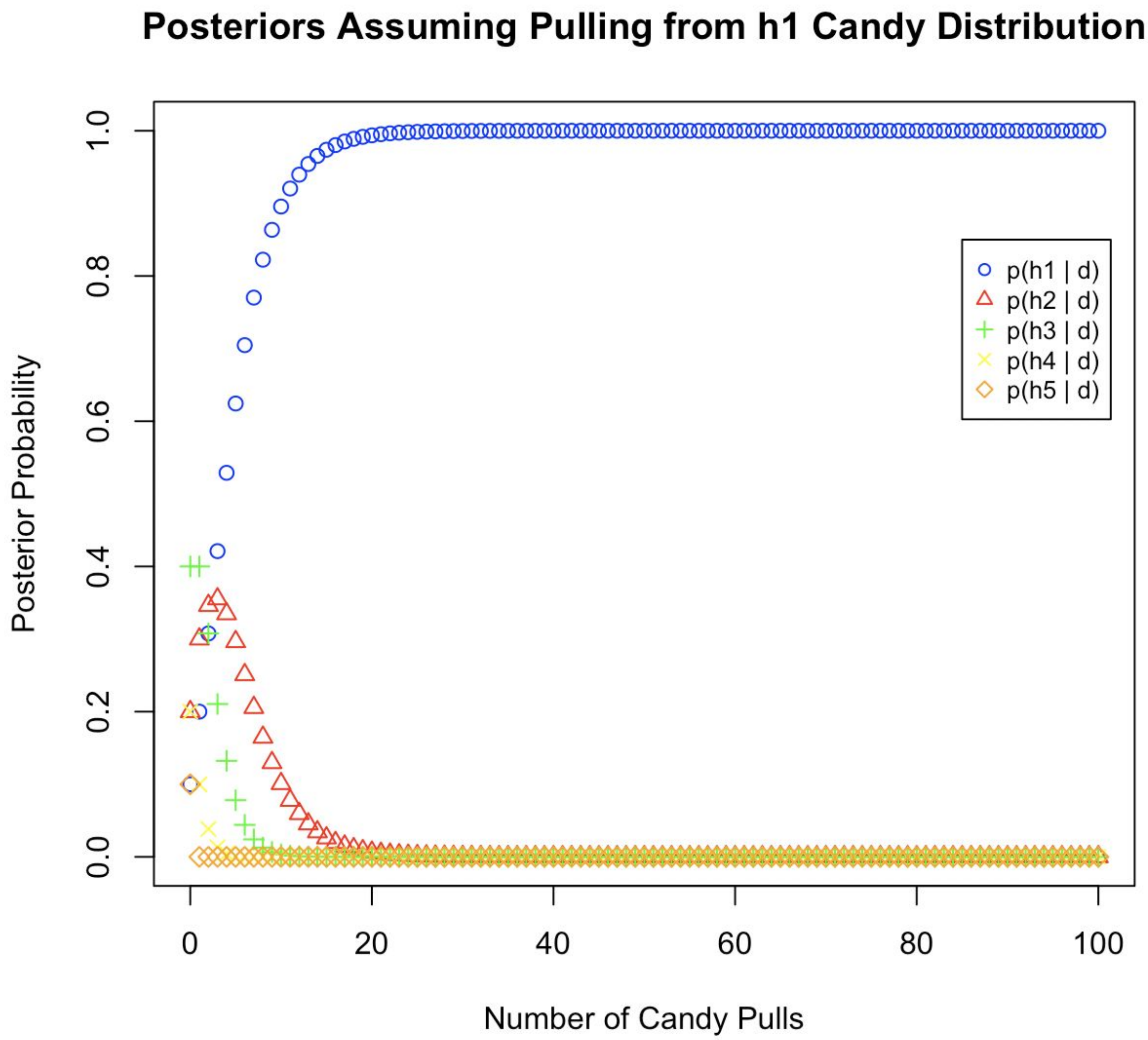
Question 1, part c



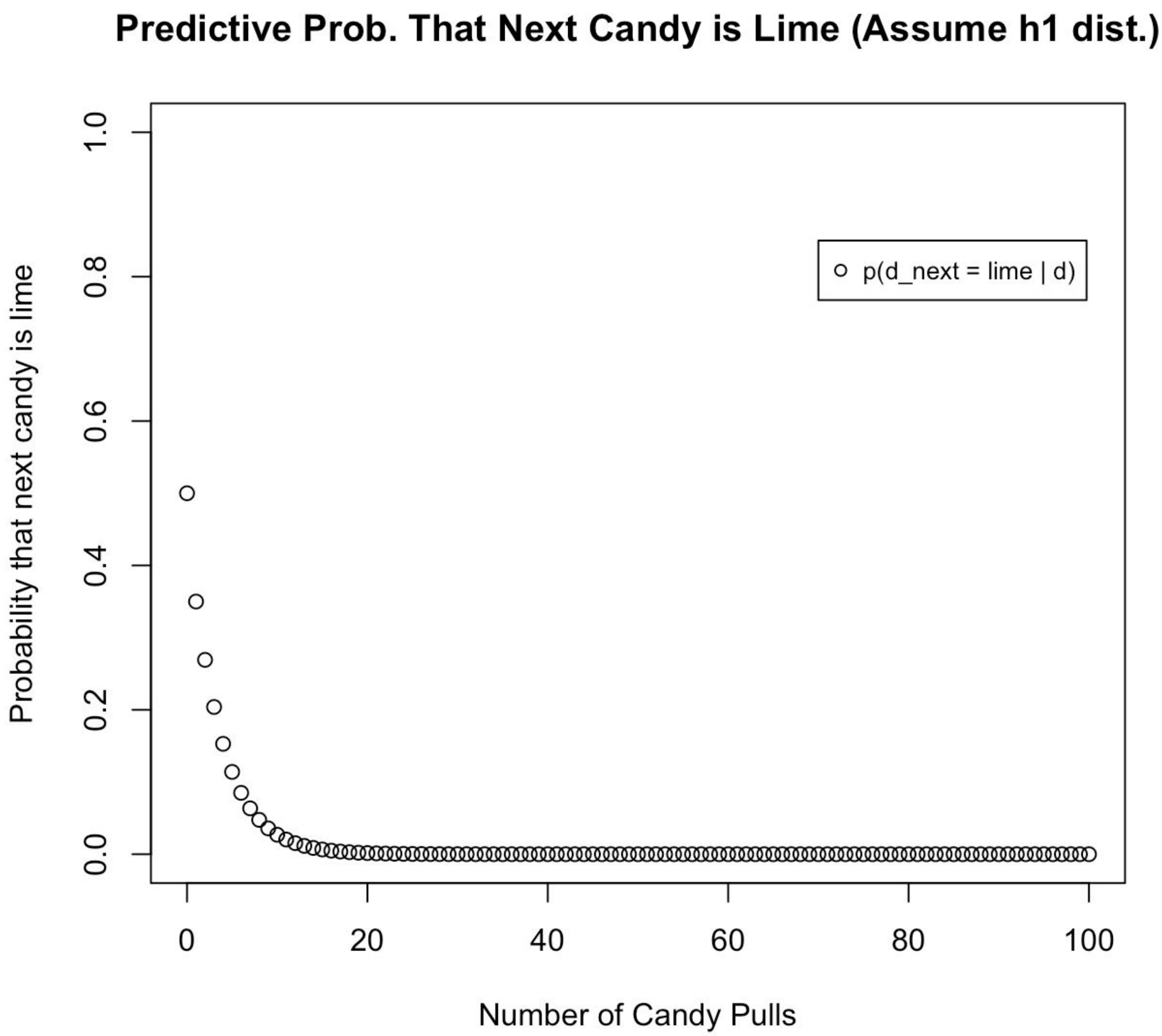
Question 1, part c



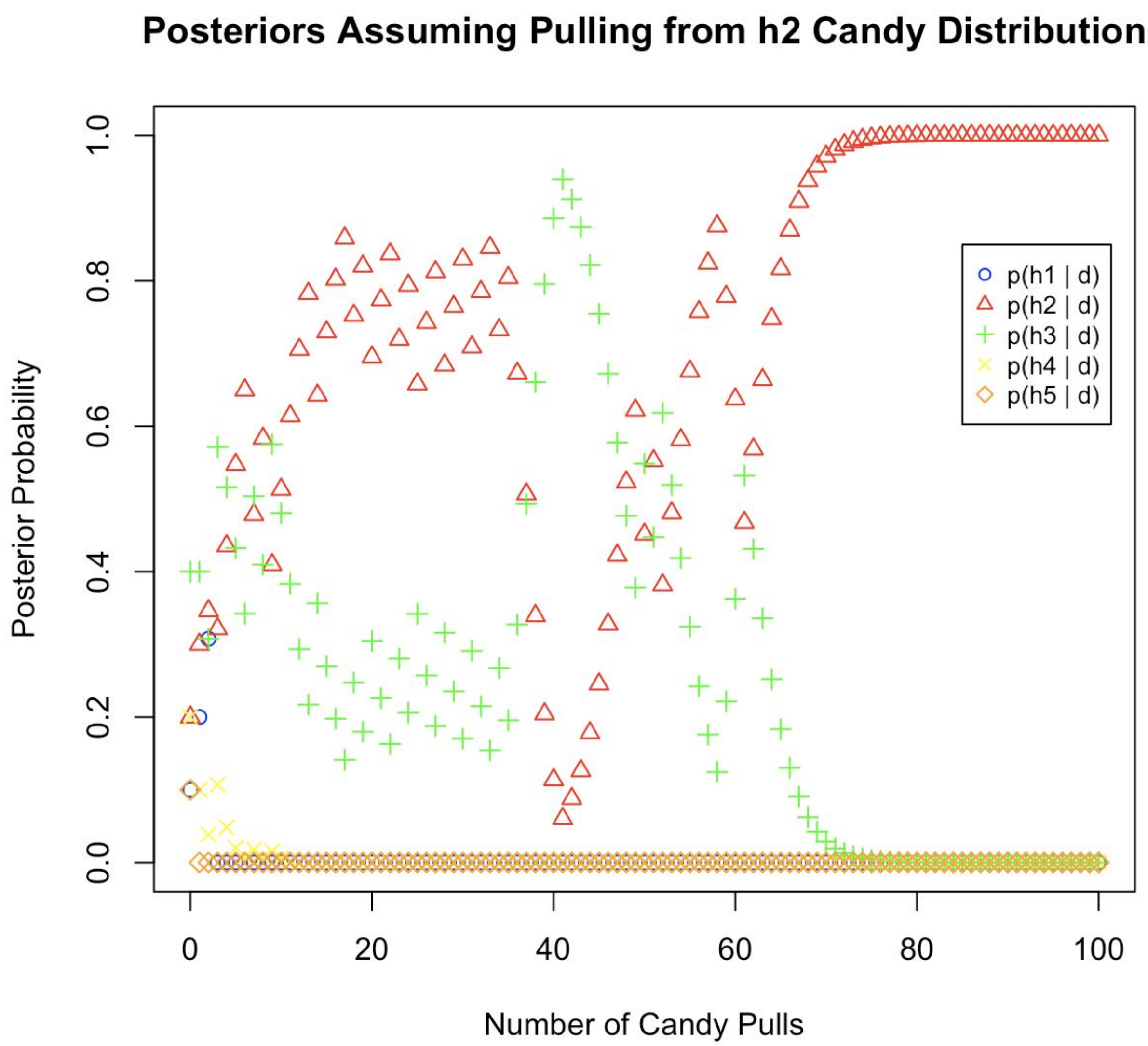
Question 2, part a



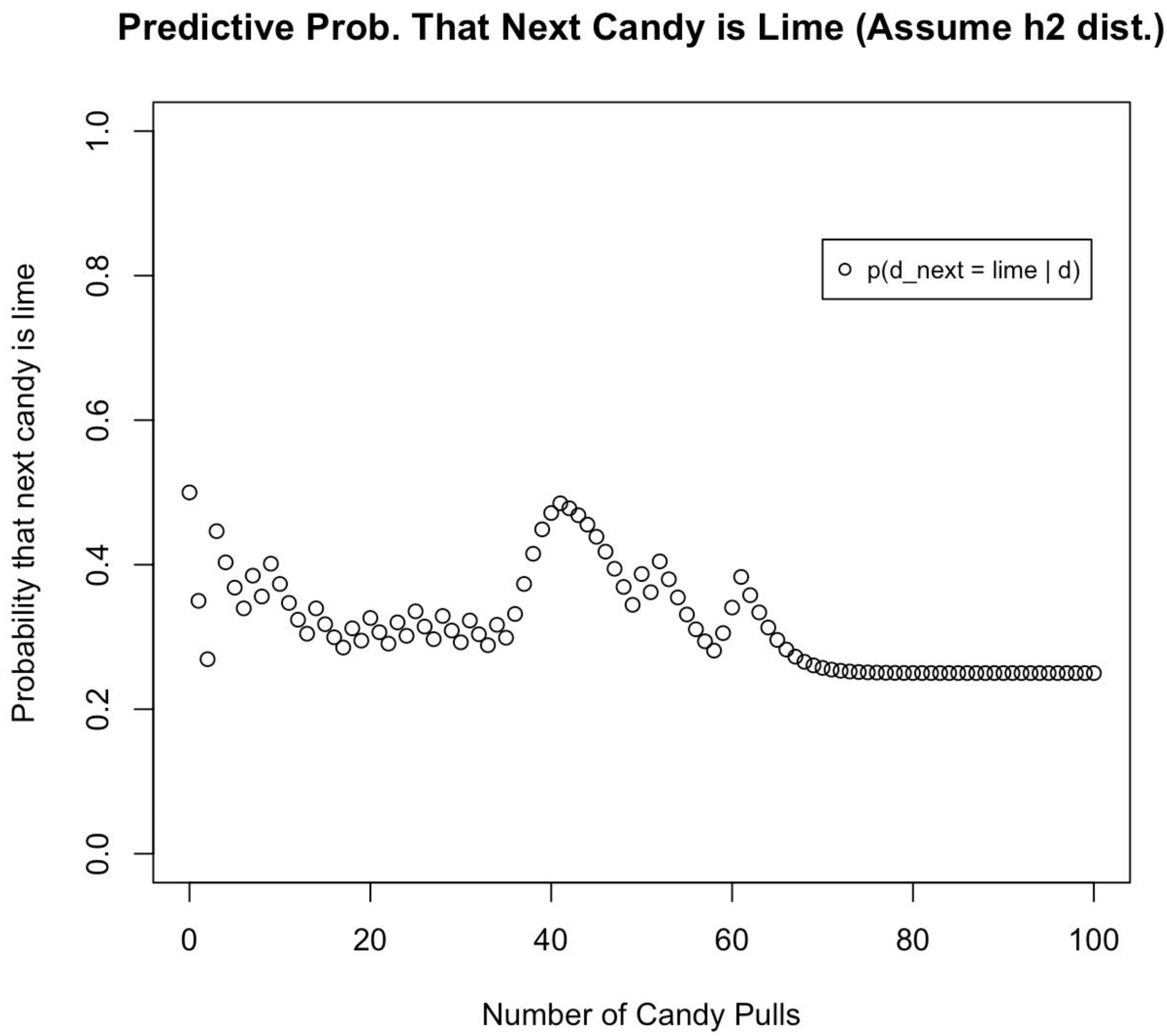
Question 2, part a



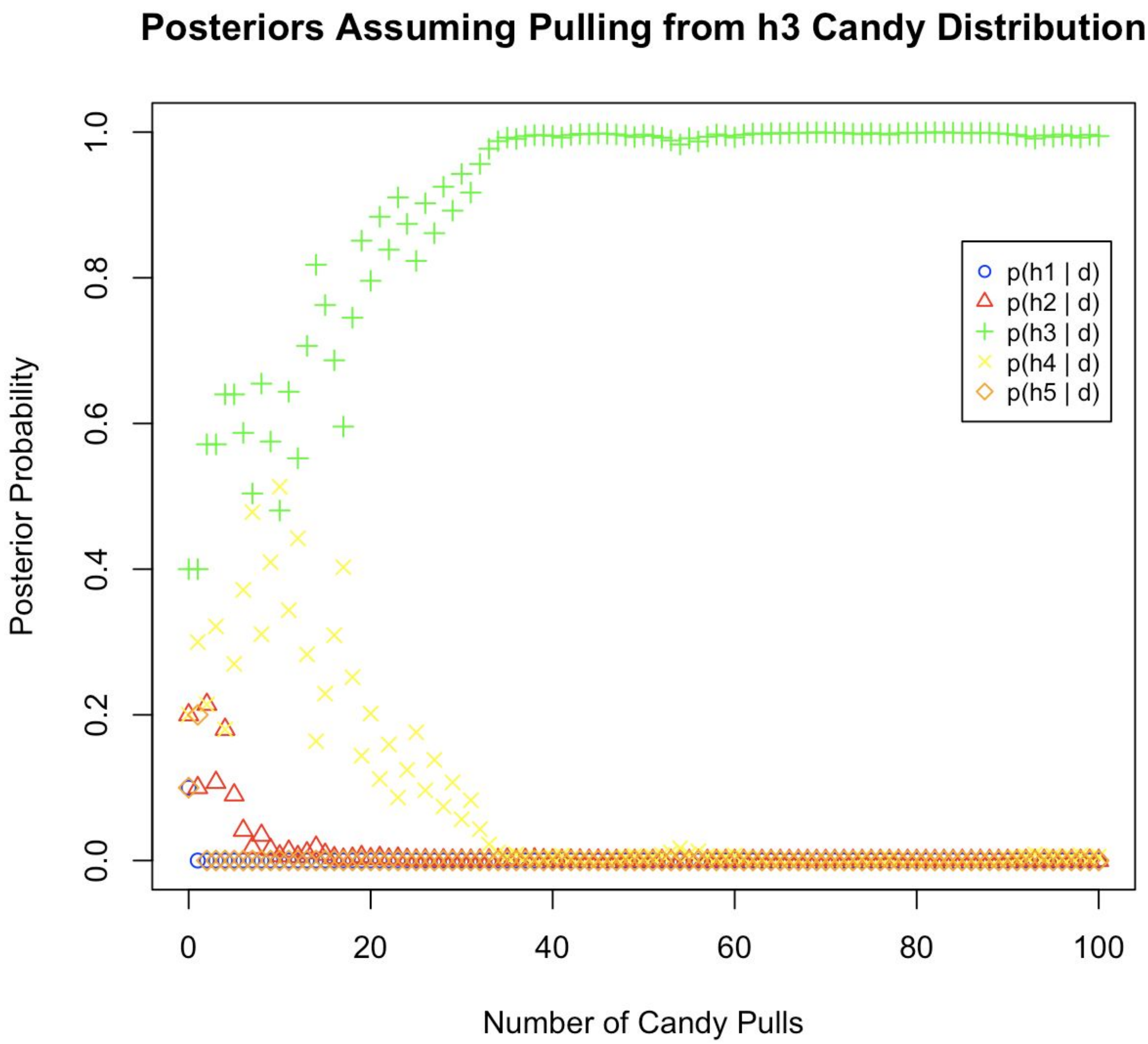
Question 2, part a



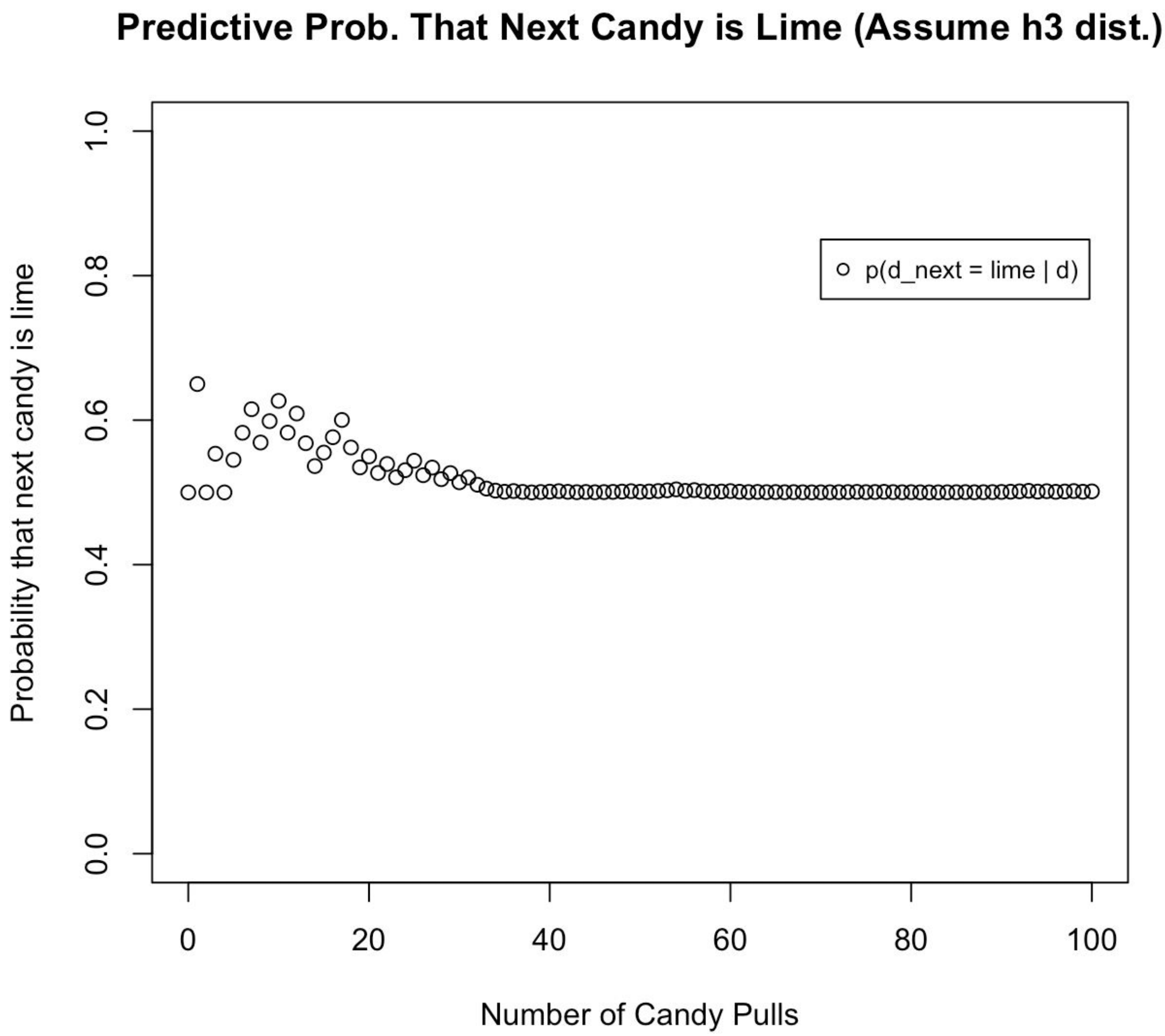
Question 2, part a



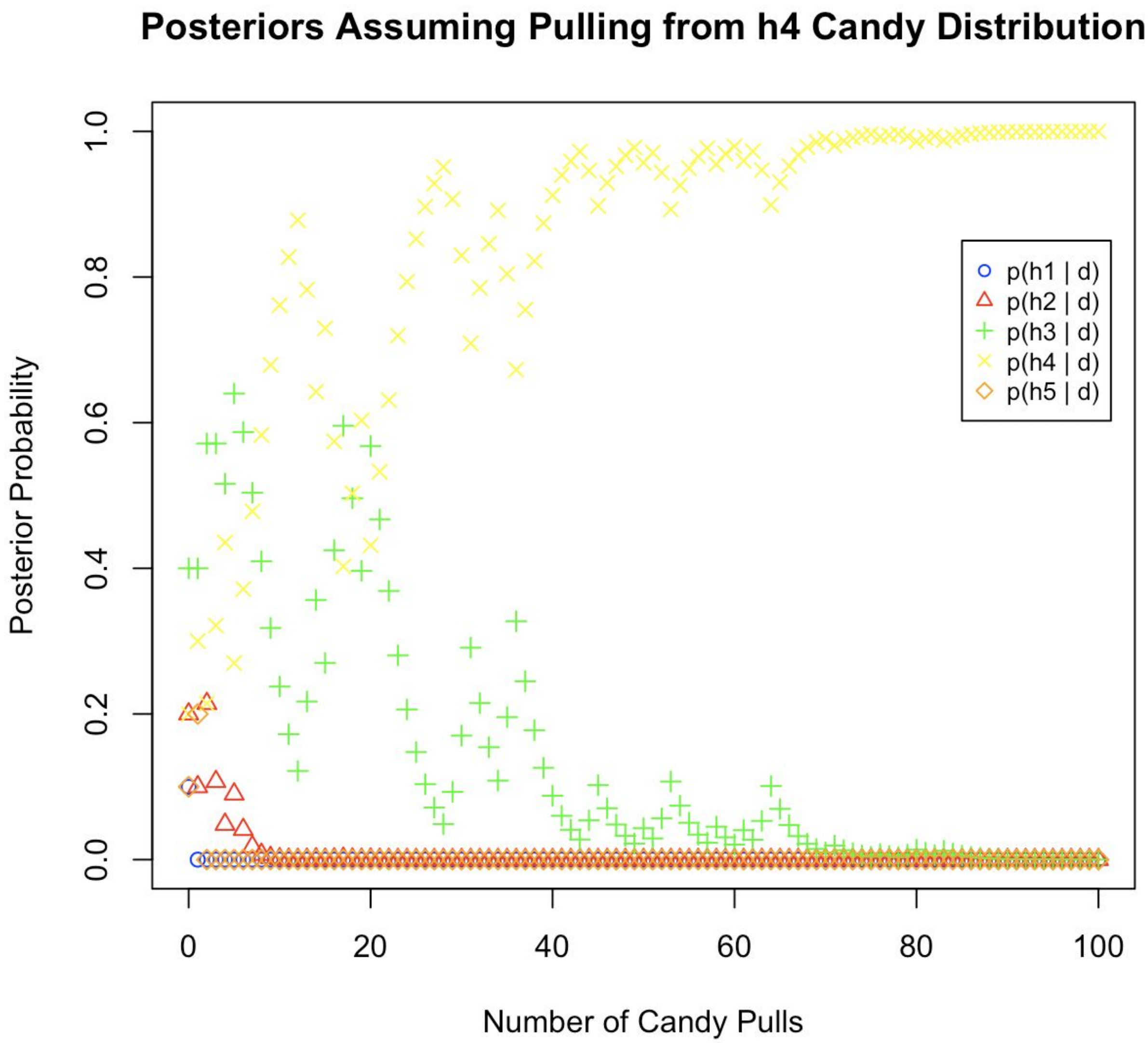
Question 2, part a



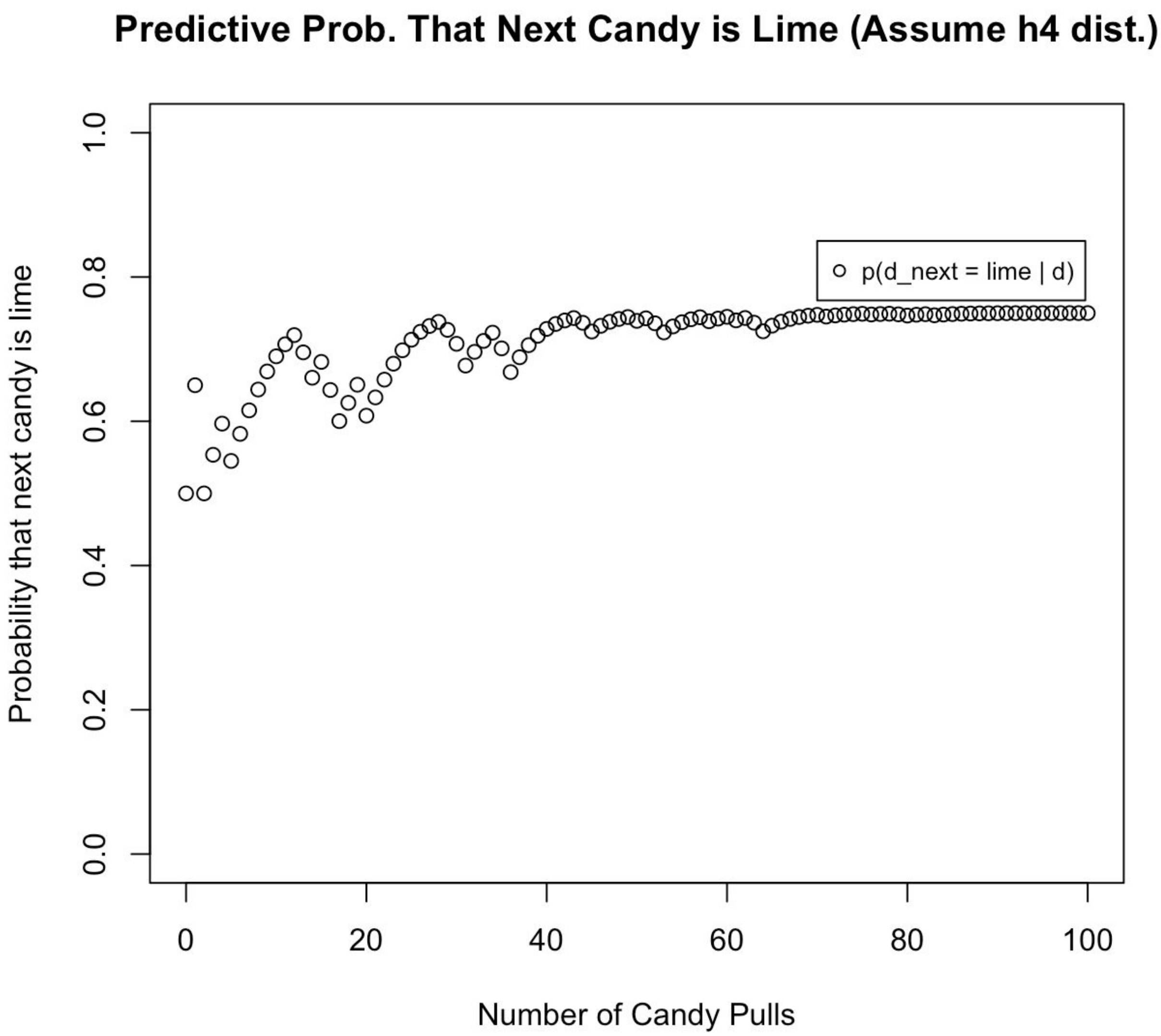
Question 2, part a



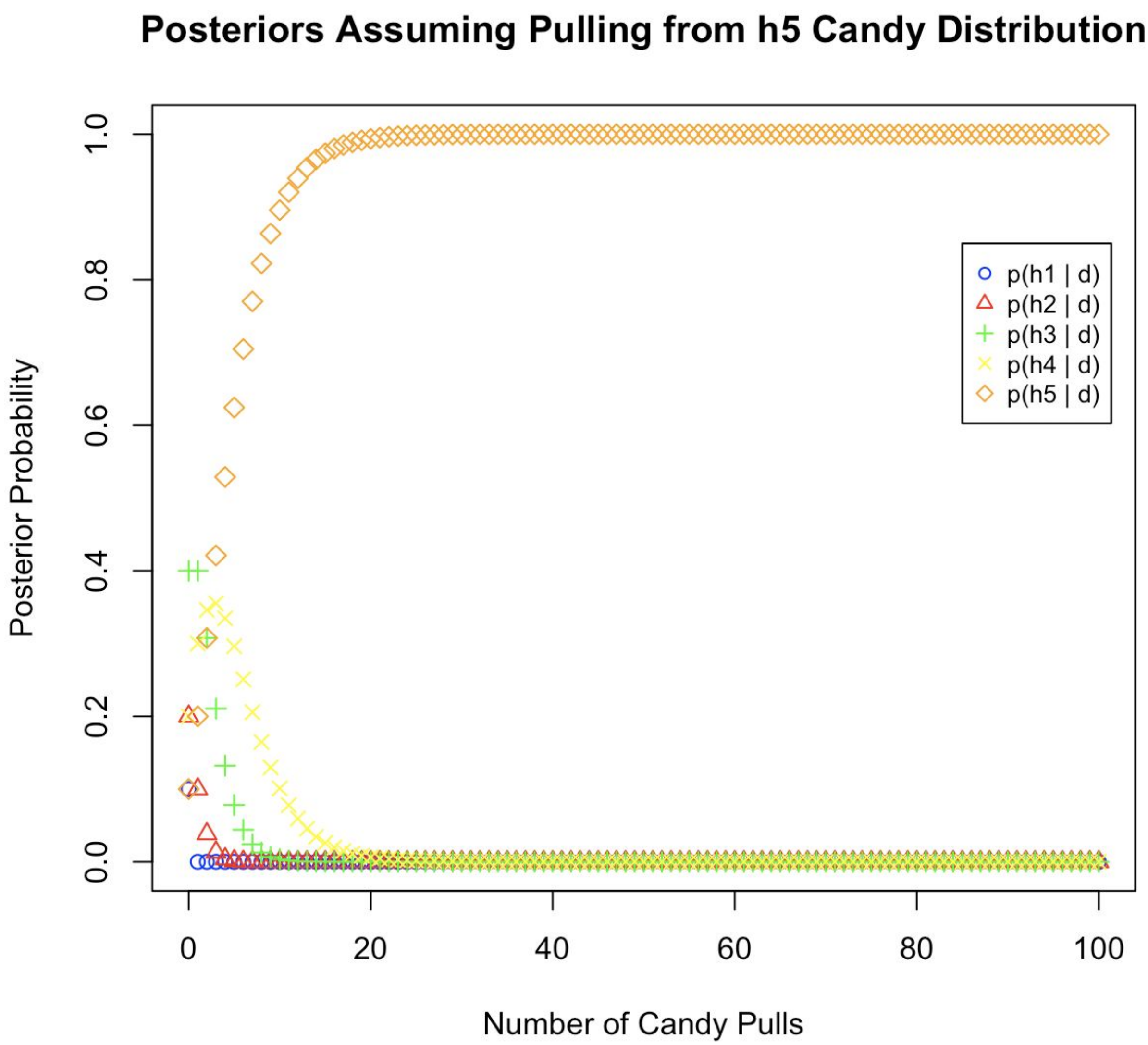
Question 2, part a



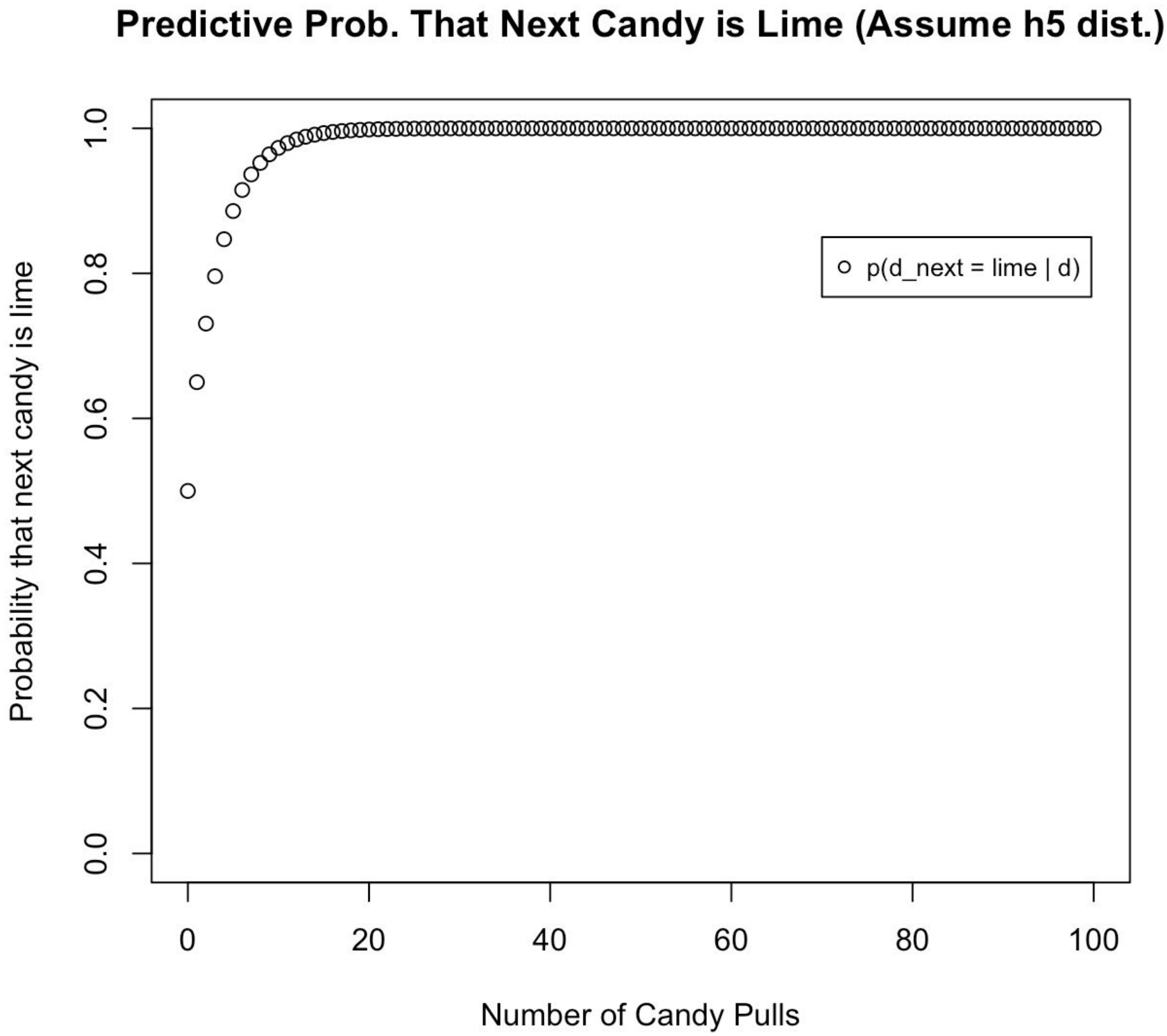
Question 2, part a



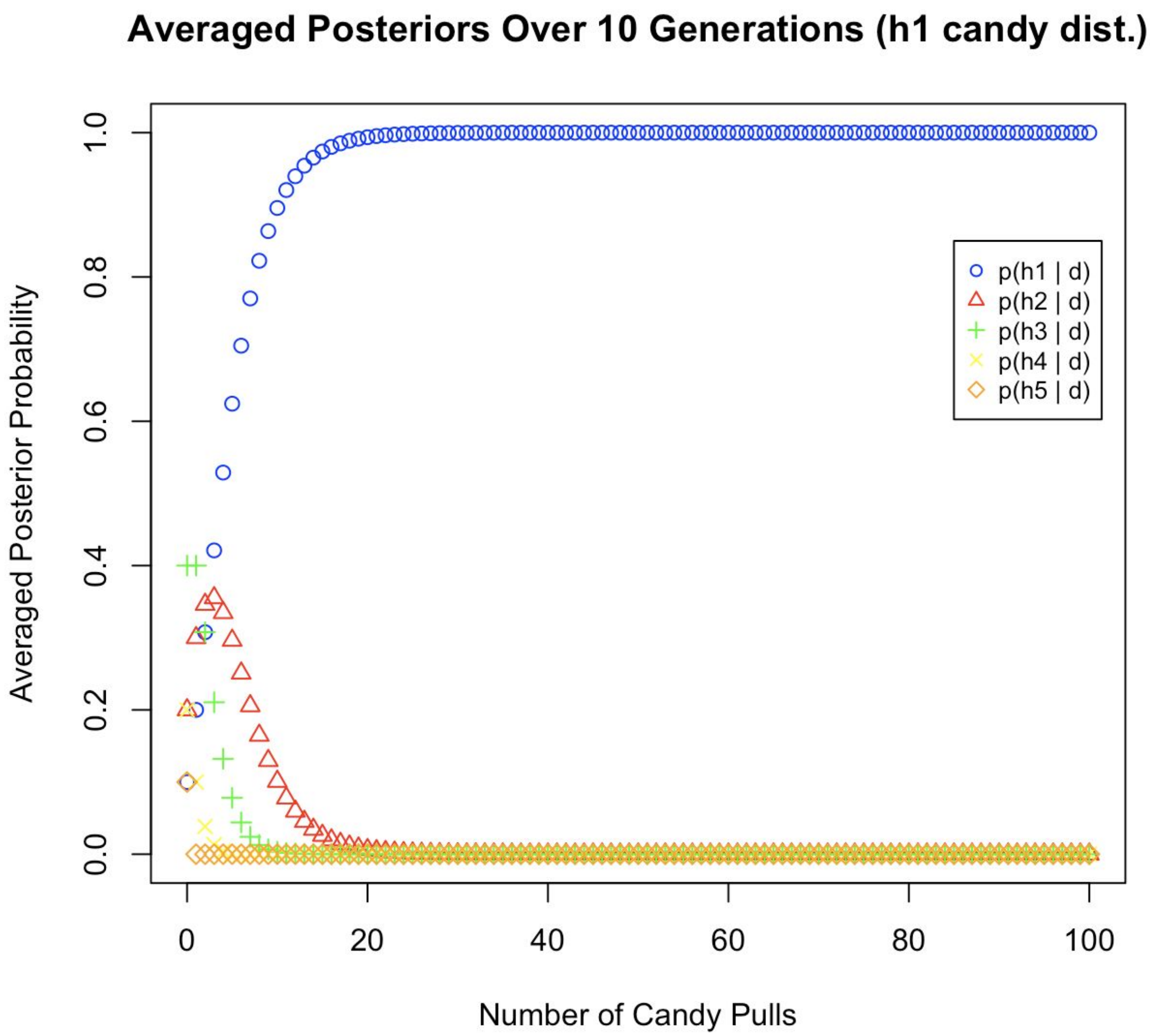
Question 2, part a



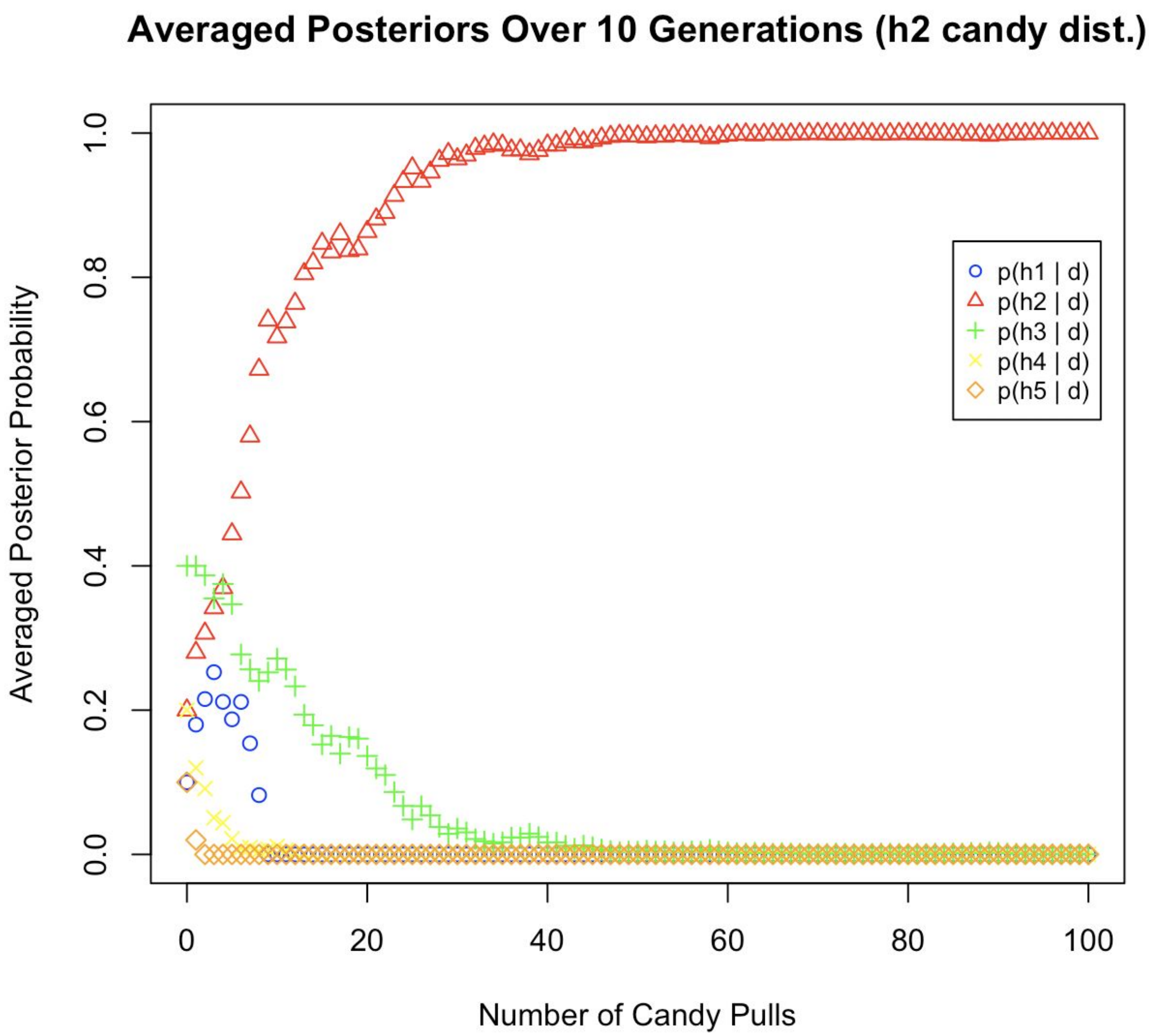
Question 2, part a



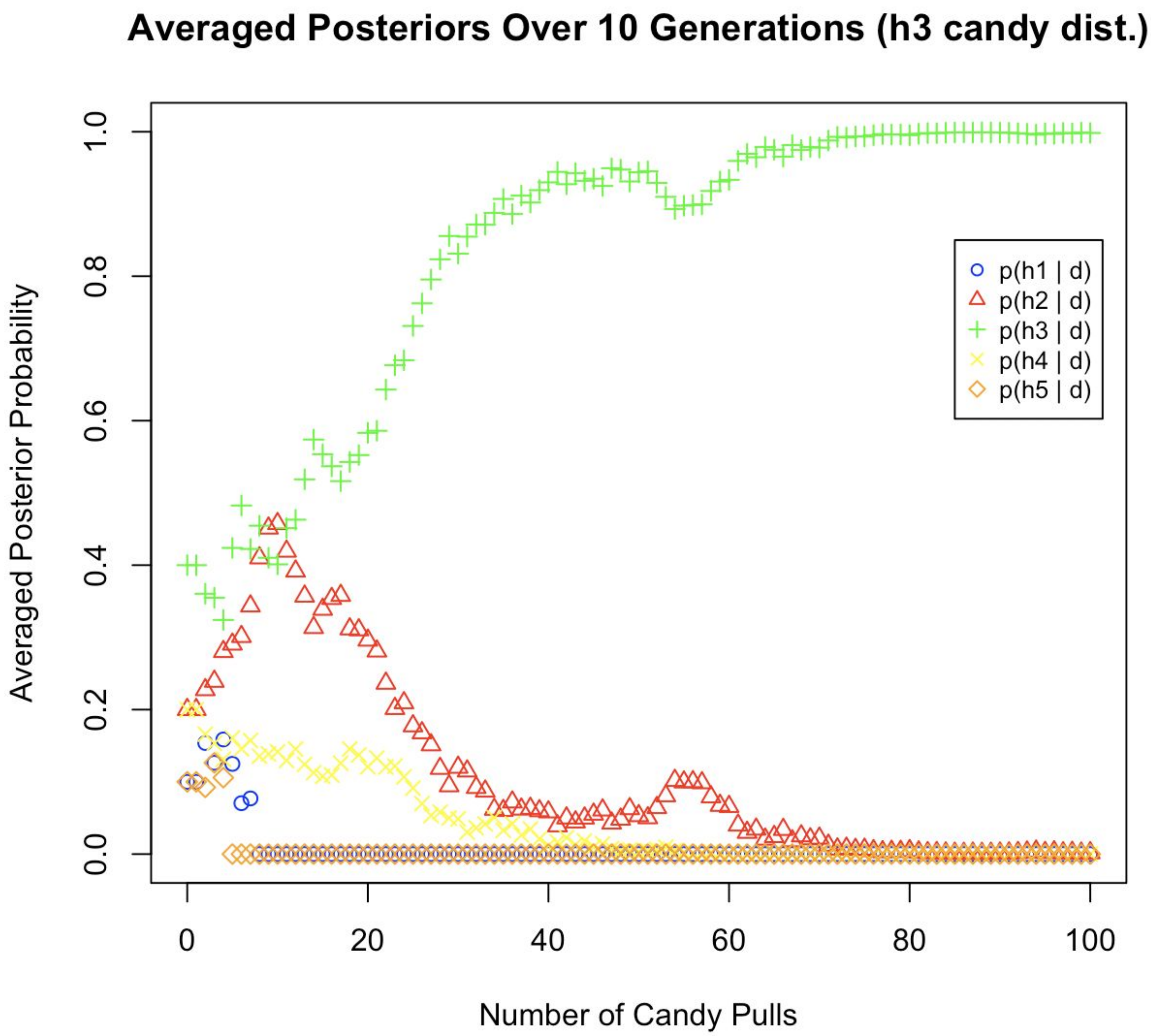
Question 2, part a



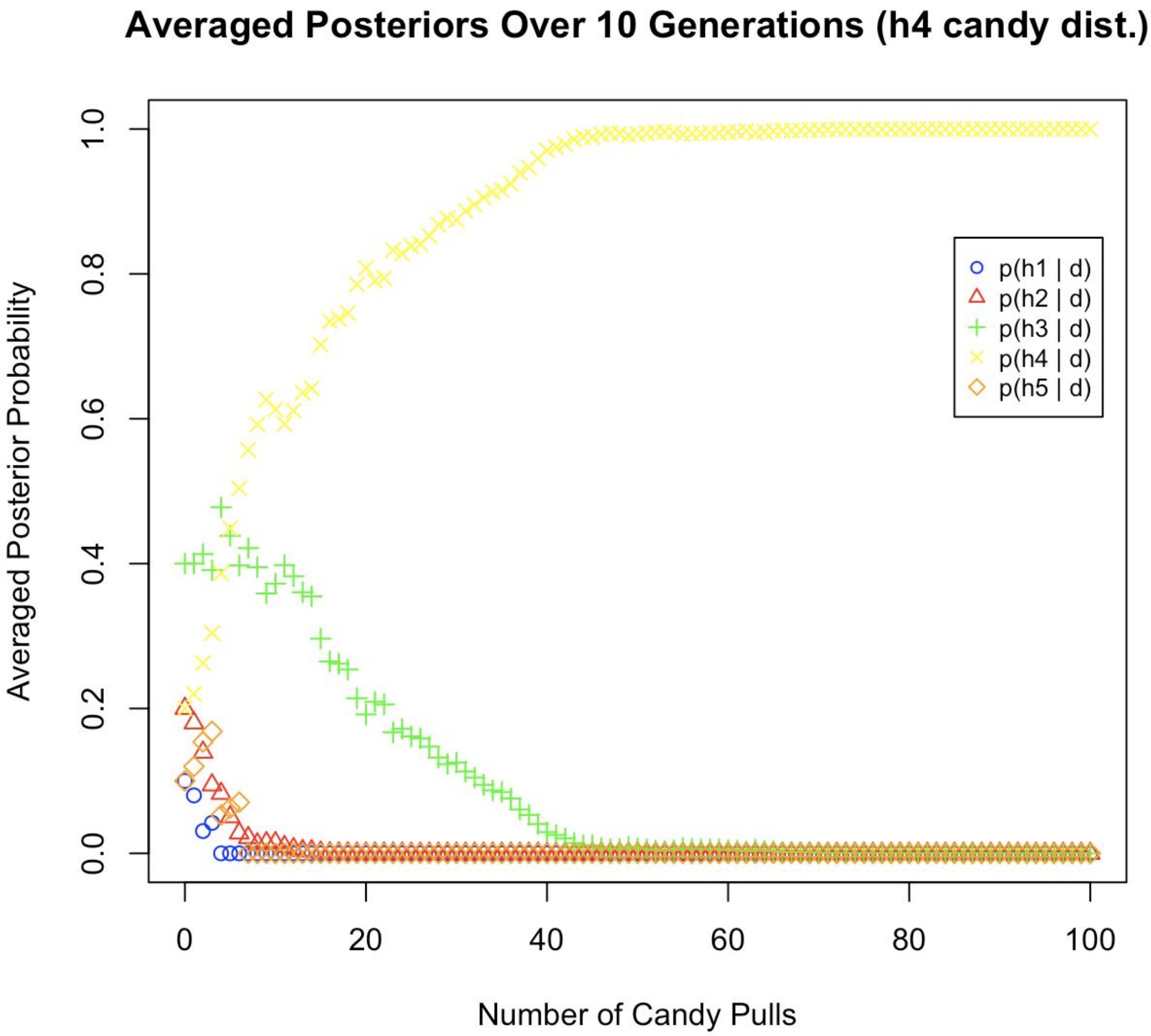
Question 2, part a



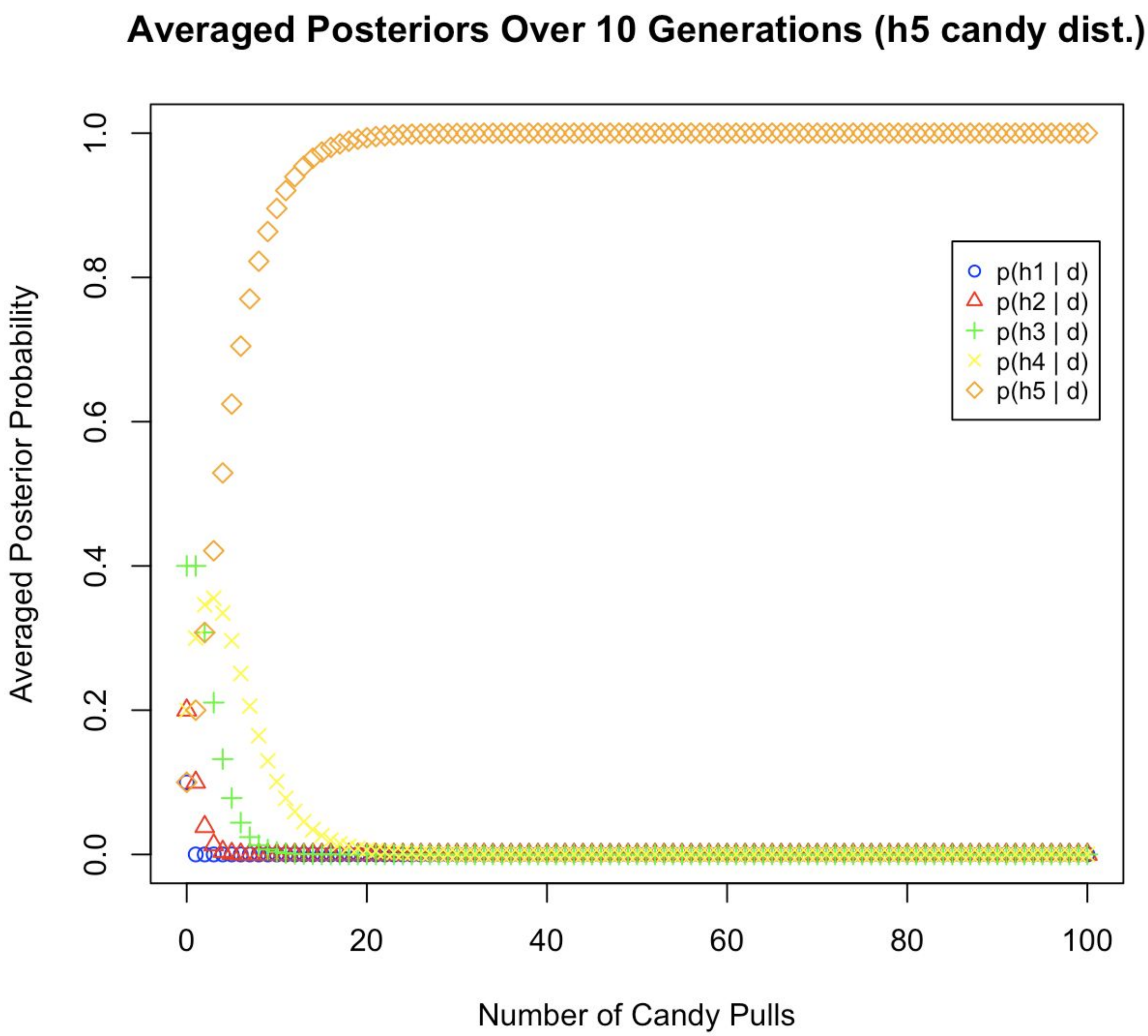
Question 2, part a

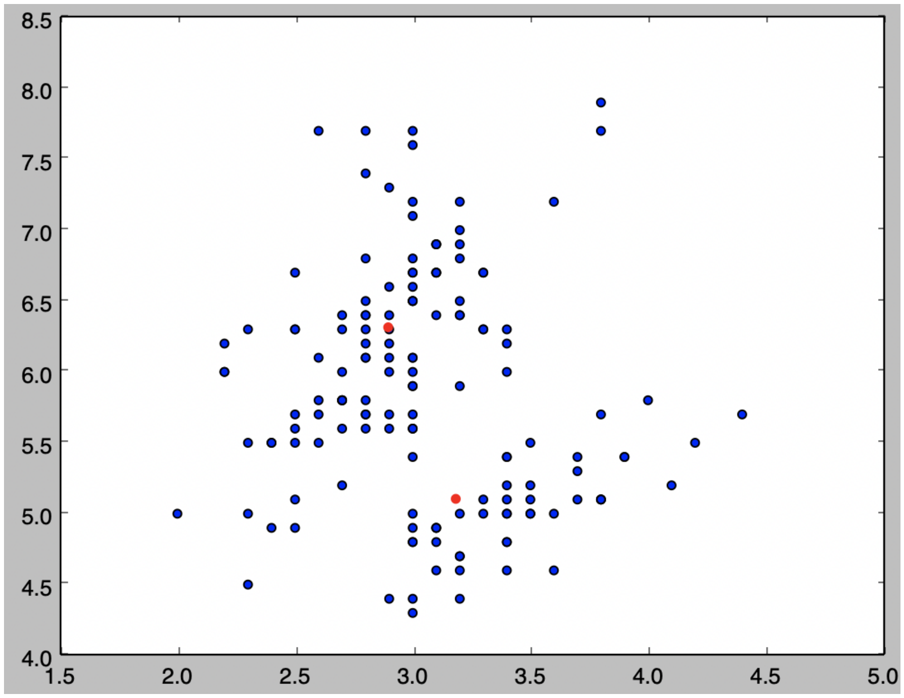


Question 2, part a

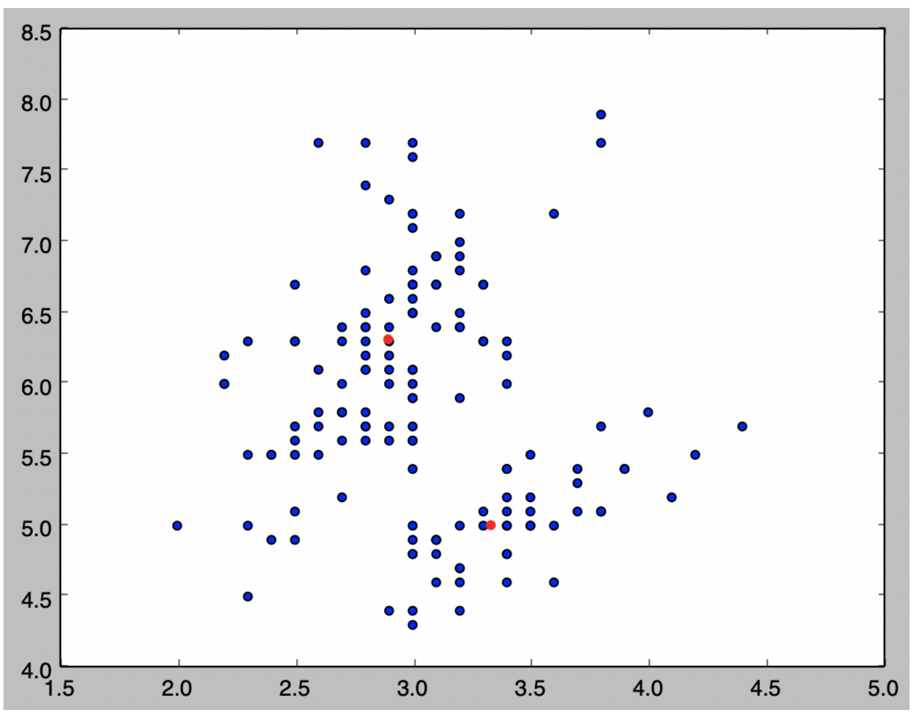


Question 2, part a



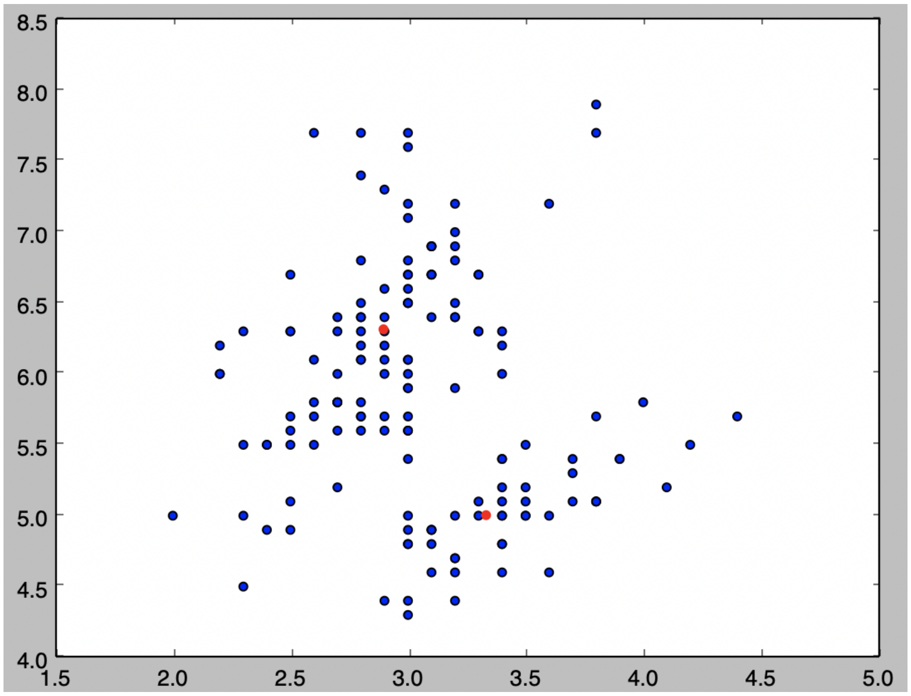


2 clusters step 1

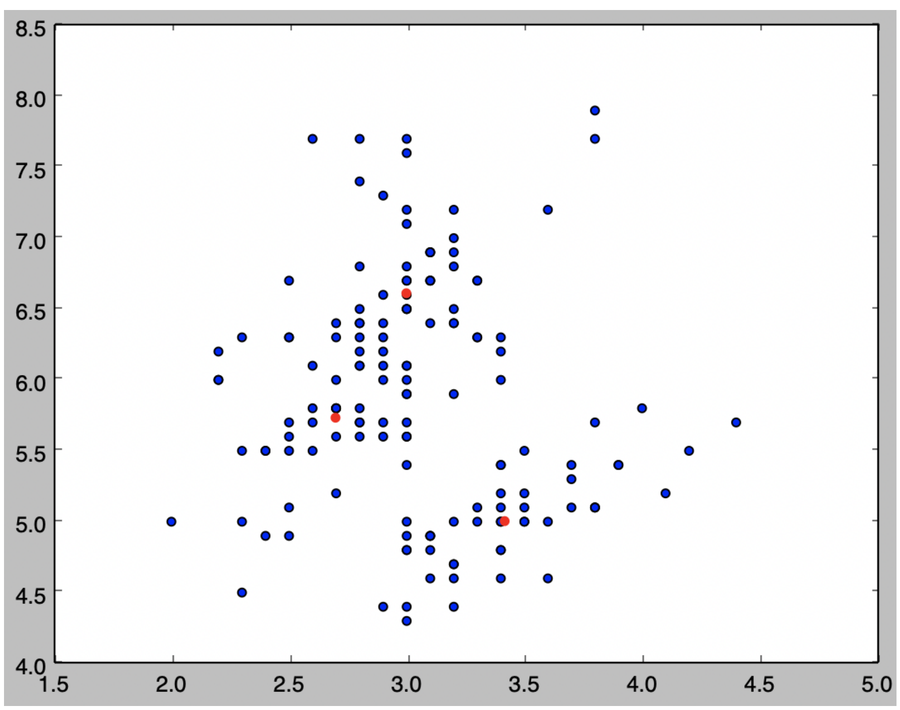


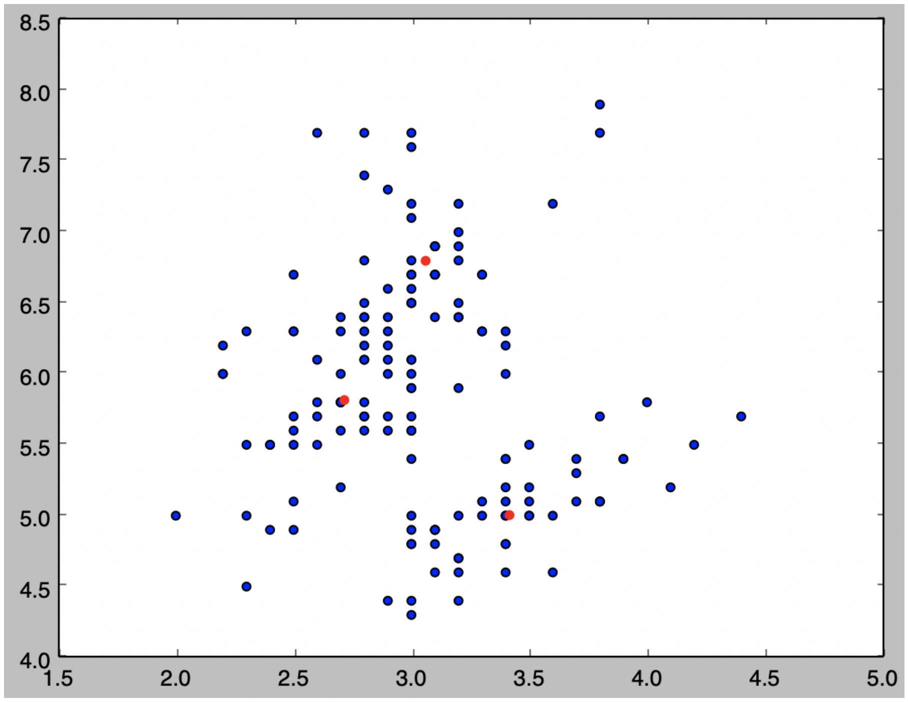
2 clusters step 2

2 clusters step 3

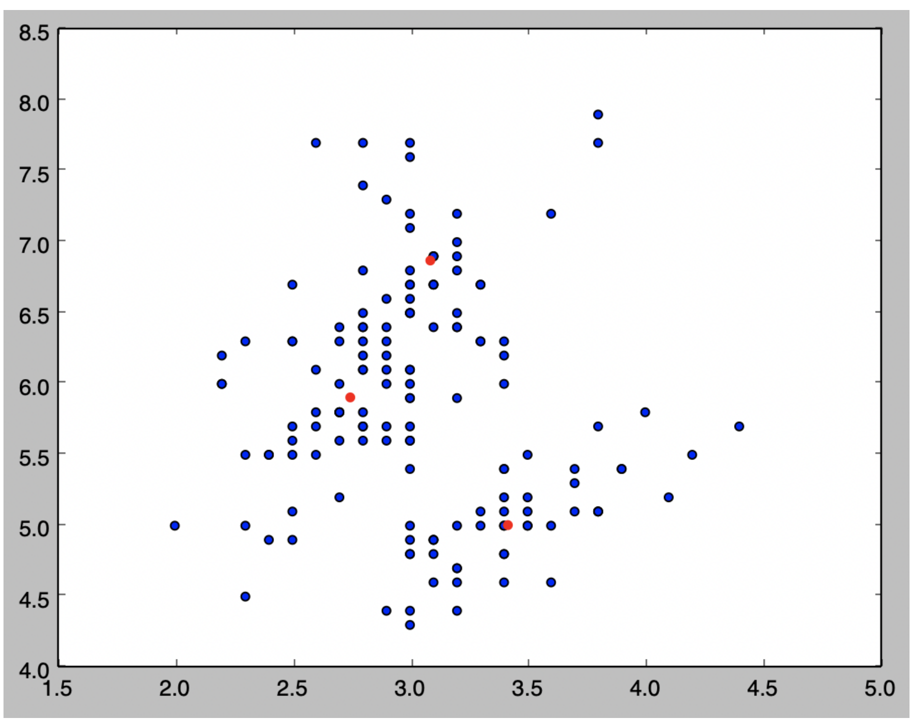


3 clusters step1





3 clusters step 2



3 clusters step 3