CS6190: Probabilistic Modeling Homework 3 MCMC

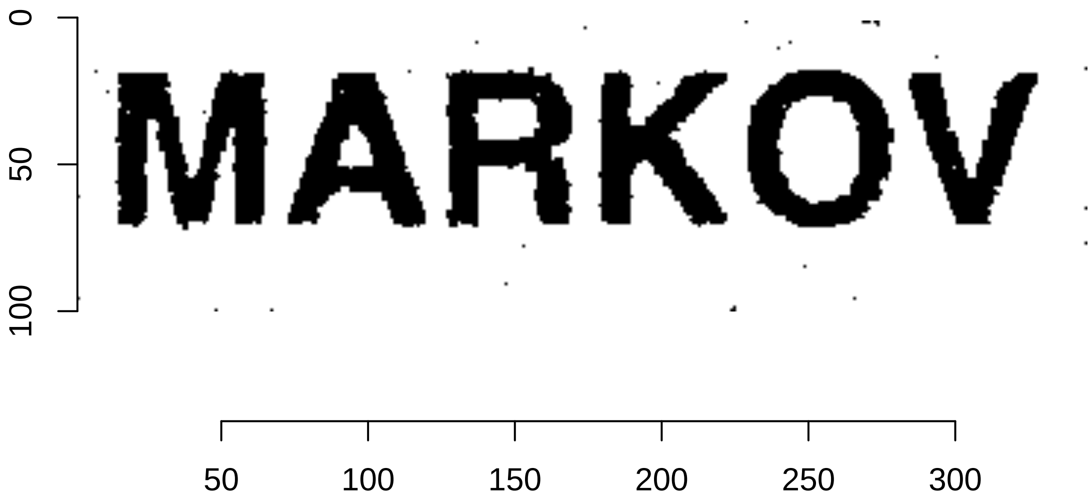
Gopal Menon

15 April, 2018

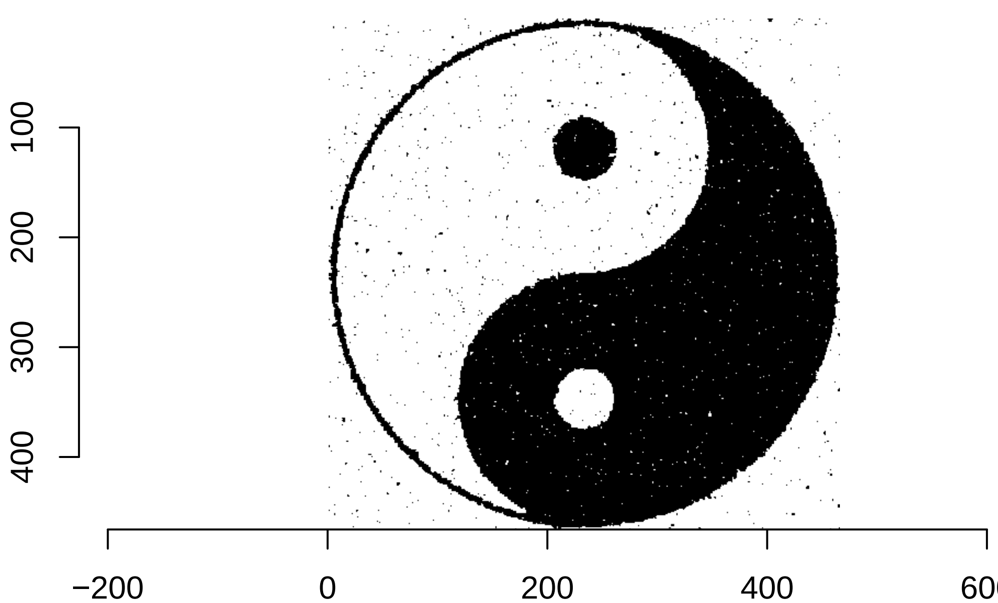
1(a) Gibbs Sampling of noisy image with Ising prior only.

The sampled images are shown for various values of α and β. The variance values shown should be disregarded as that part of the energy function is not used.

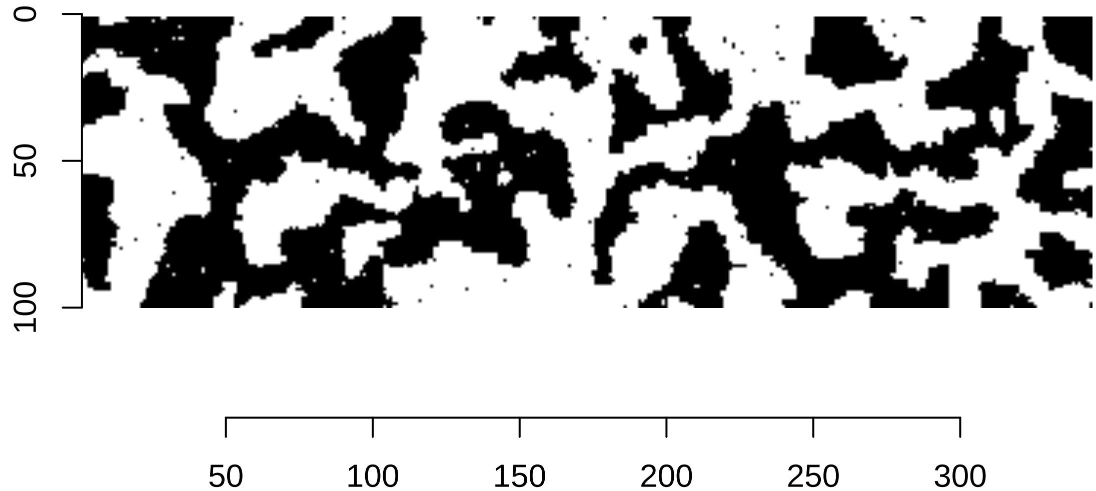
noisy−message.png , Alpha = 0 , Beta = 1 , Variance = 1



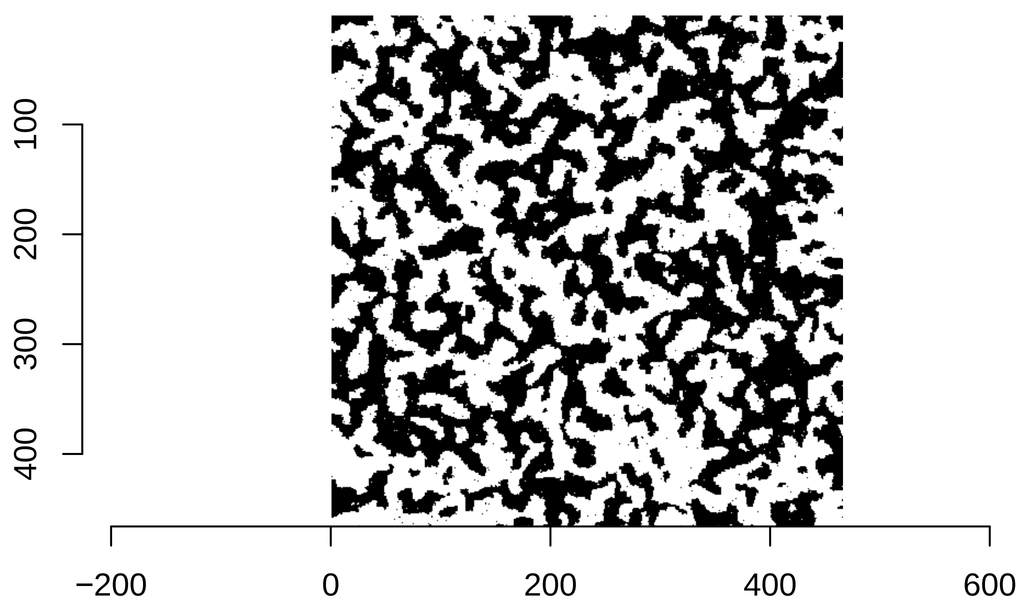
noisy−yinyang.png , Alpha = 0 , Beta = 1 , Variance = 1



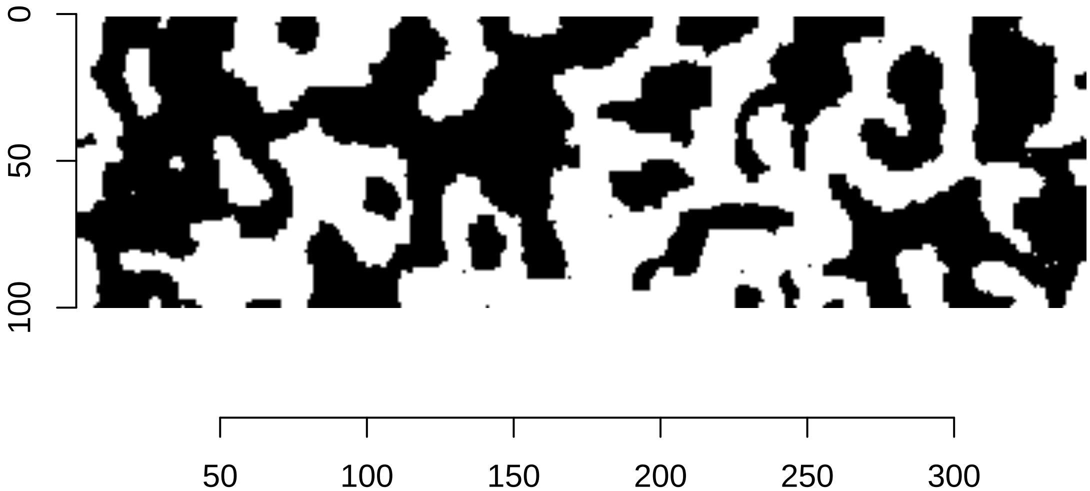
noisy−message.png , Alpha = 1e−07 , Beta = 0.75 , Variance = 1



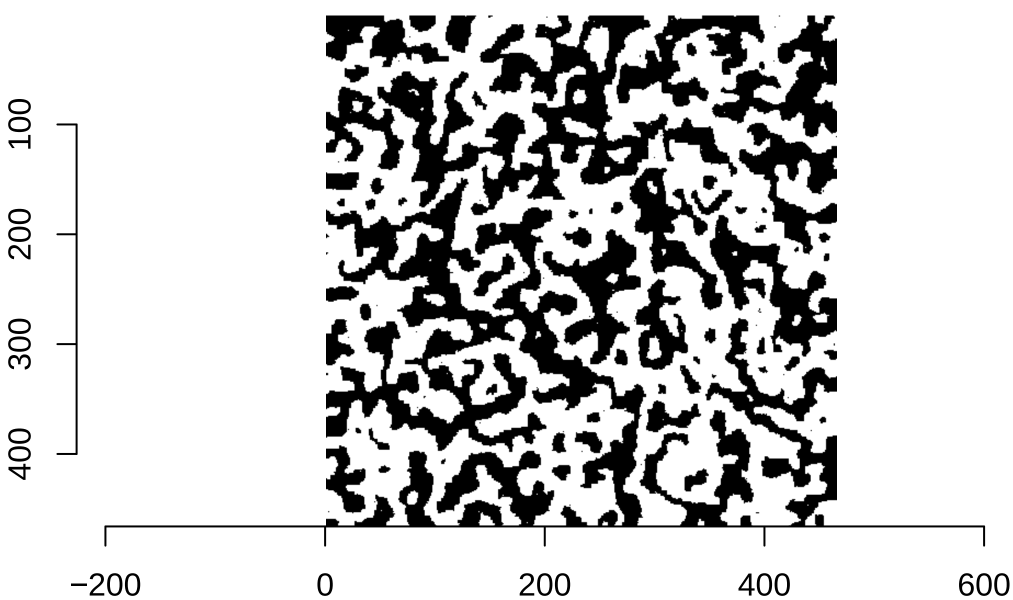
noisy−yinyang.png , Alpha = 1e−07 , Beta = 0.75 , Variance = 1



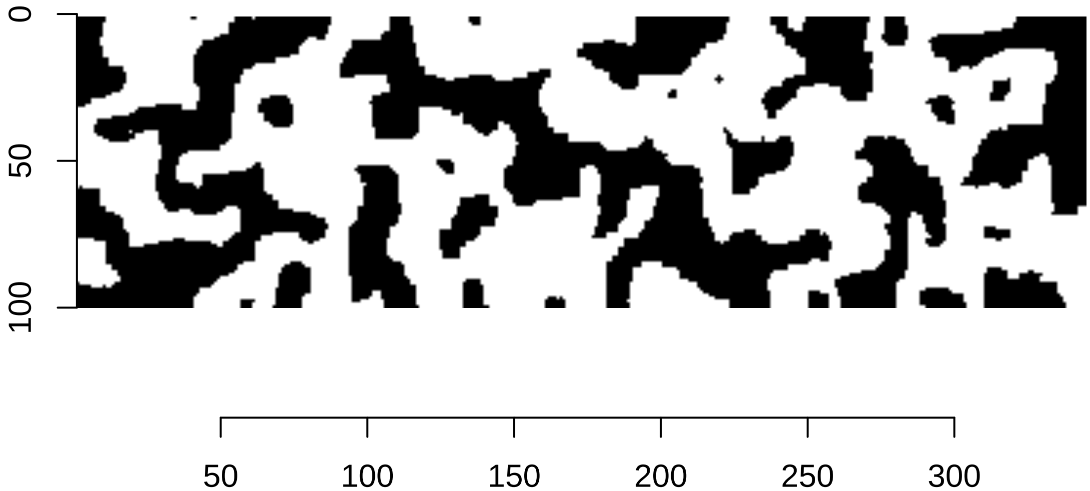
noisy−message.png , Alpha = 1e−07 , Beta = 1 , Variance = 1



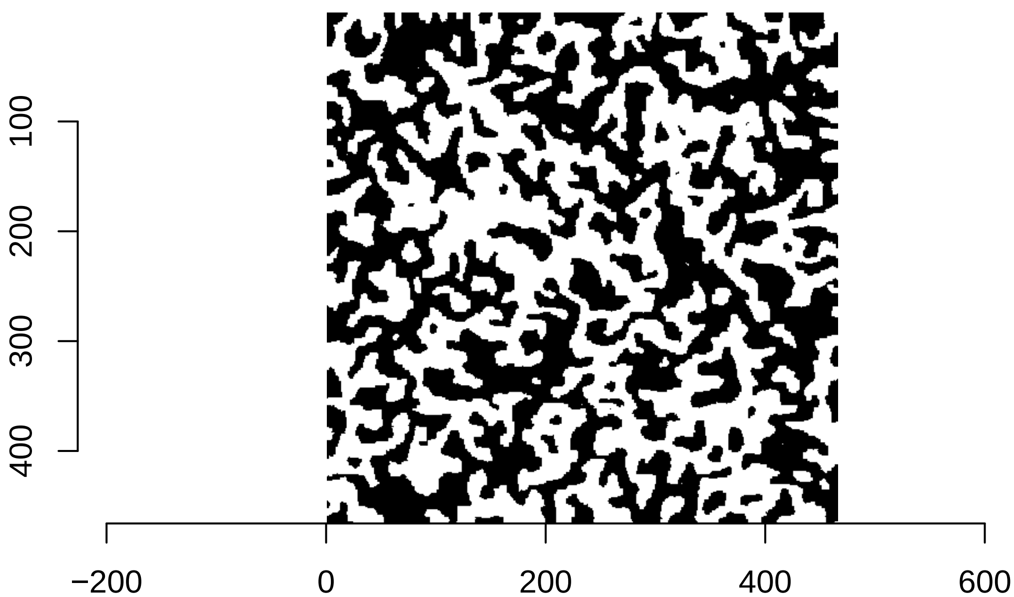
noisy−yinyang.png , Alpha = 1e−07 , Beta = 1 , Variance = 1

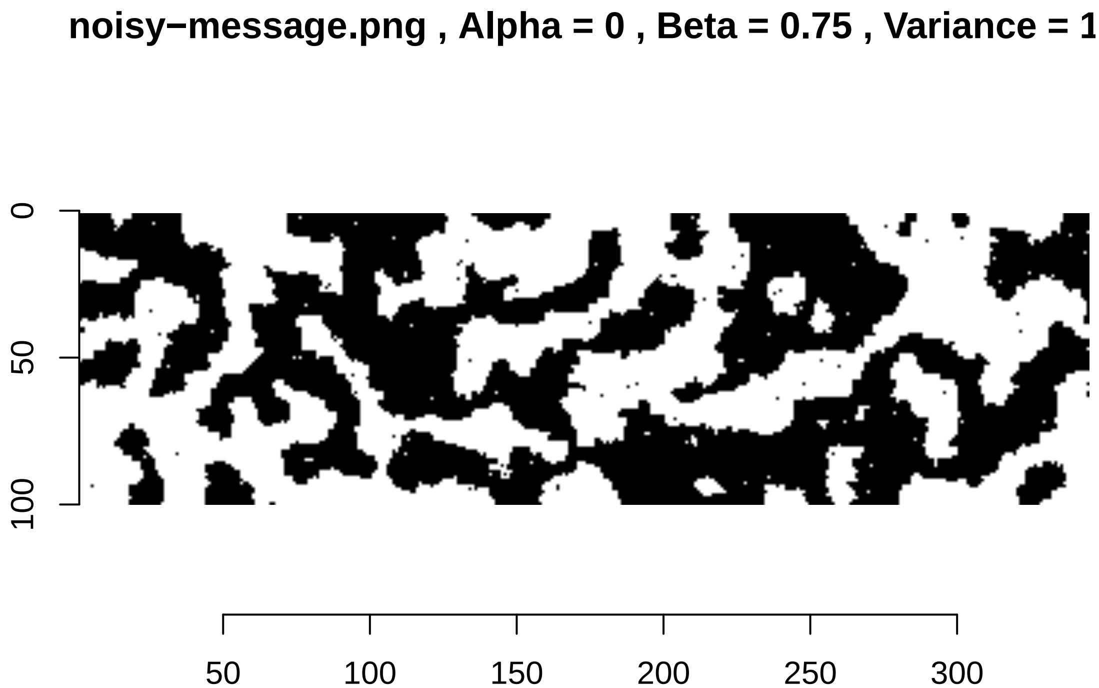


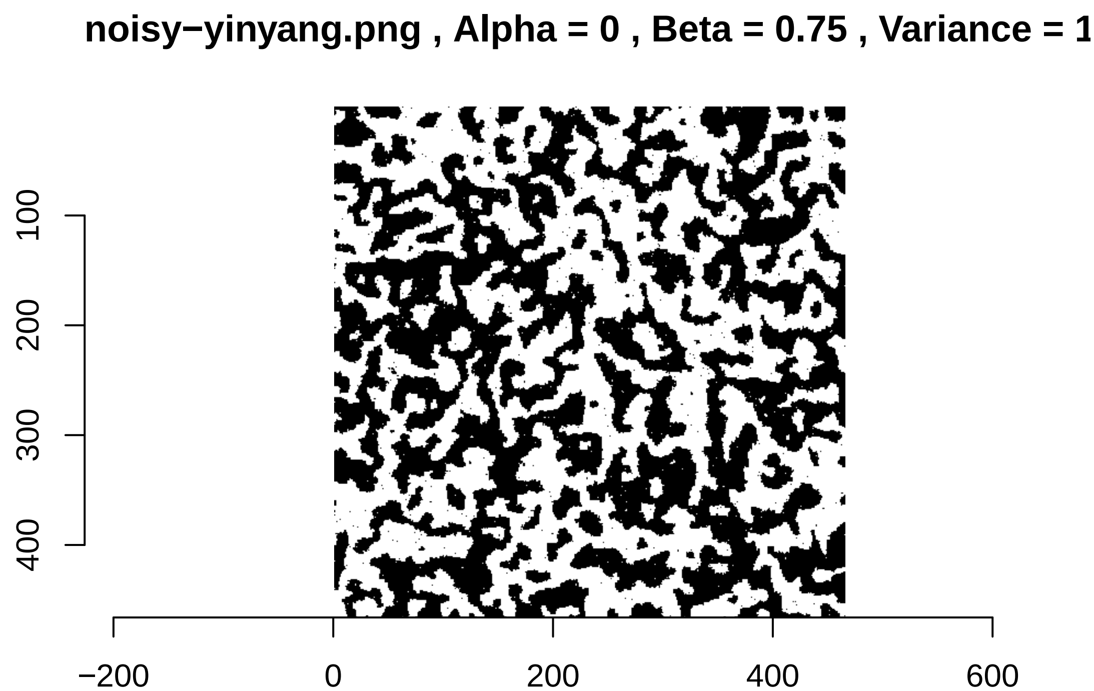
noisy−message.png , Alpha = 1e−07 , Beta = 1.25 , Variance = 1



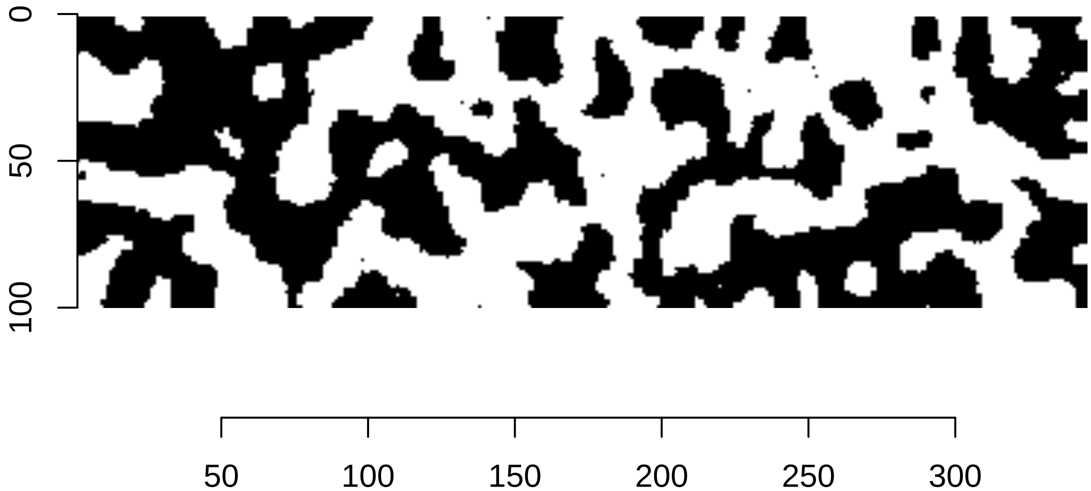
noisy−yinyang.png , Alpha = 1e−07 , Beta = 1.25 , Variance = 1



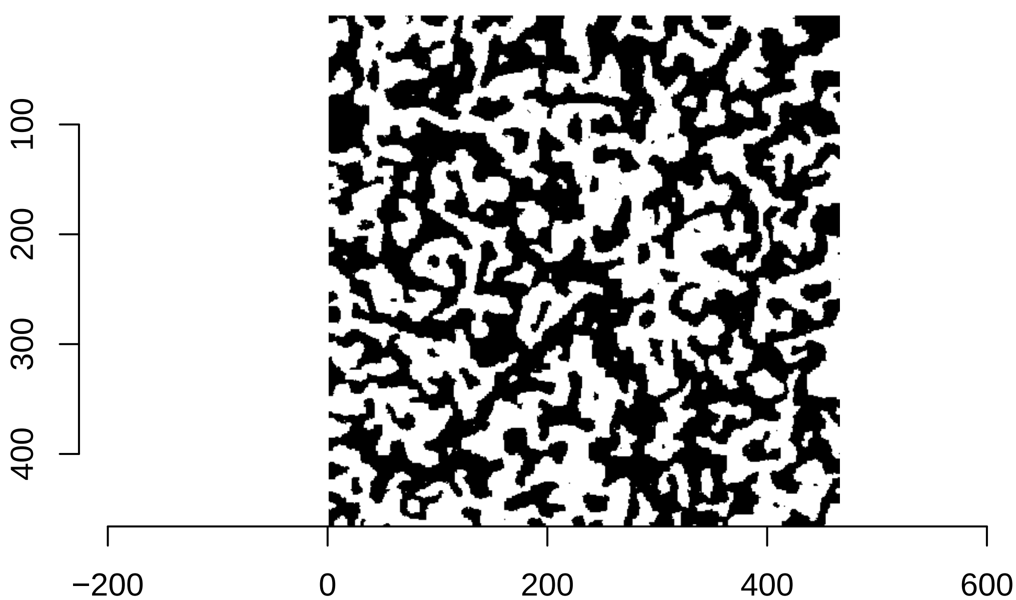

noisy−message.png , Alpha = 0 , Beta = 0.75 , Variance = 1
0
50
100
50
100
150
200
250
300

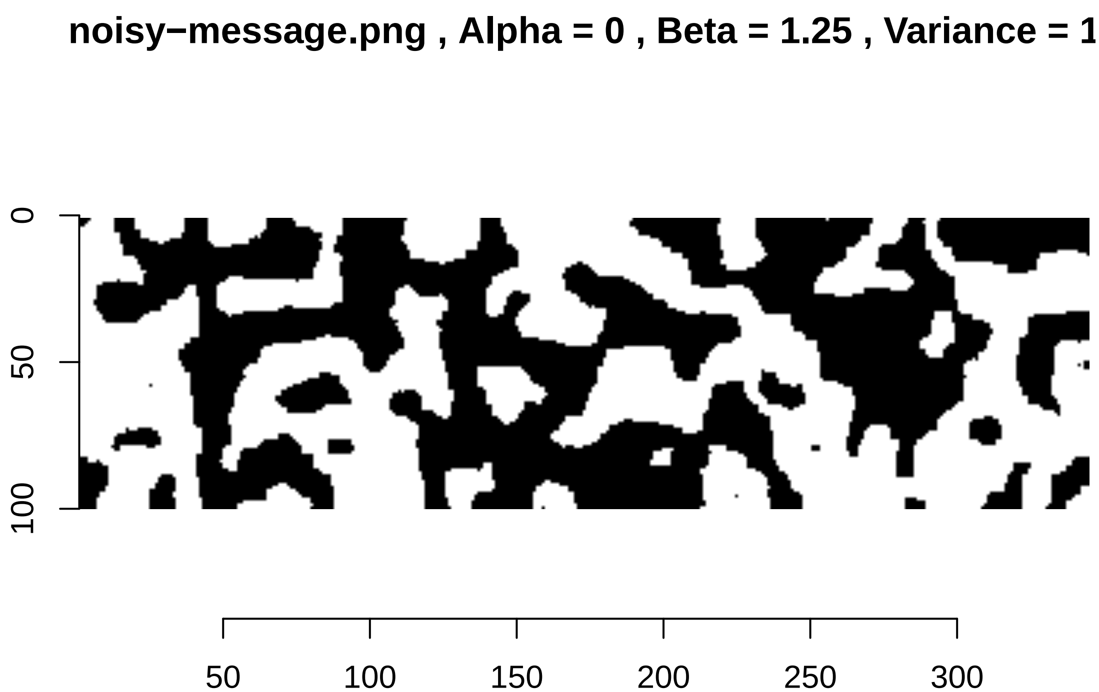

noisy−yinyang.png , Alpha = 0 , Beta = 0.75 , Variance = 1
100
200
300
400
−200
0
200
400
600

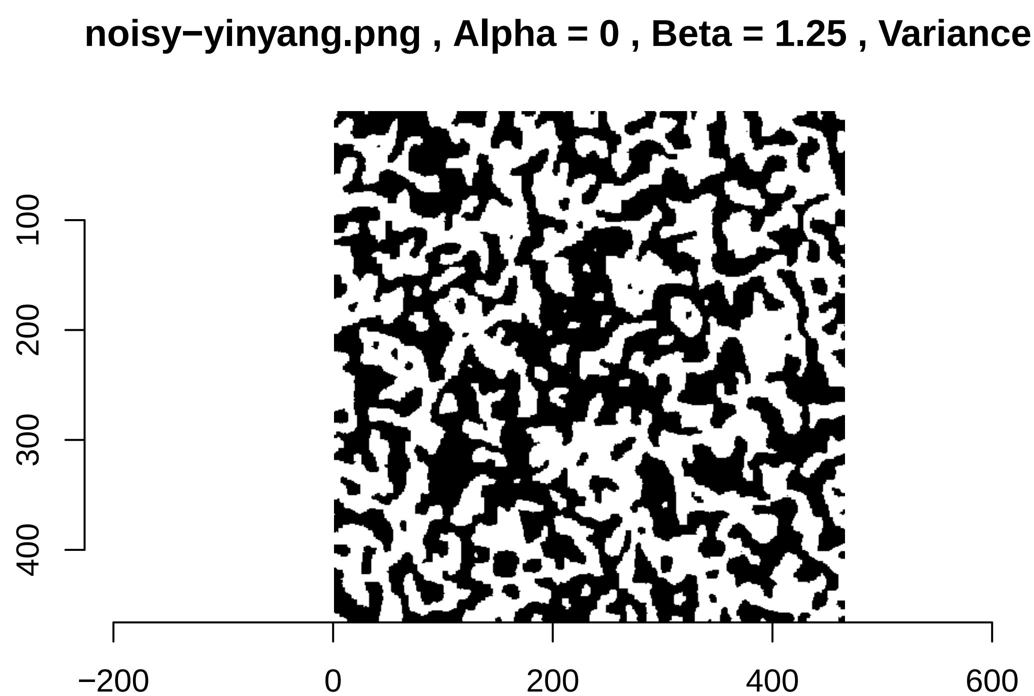
noisy−message.png , Alpha = 0 , Beta = 1 , Variance = 1



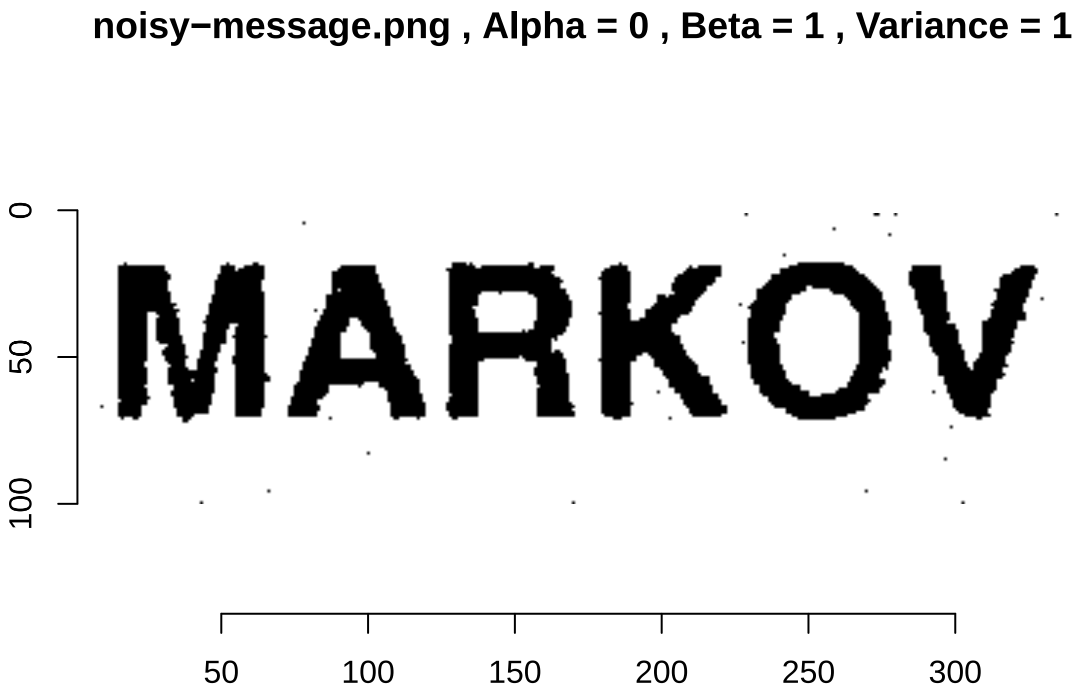
noisy−yinyang.png , Alpha = 0 , Beta = 1 , Variance = 1

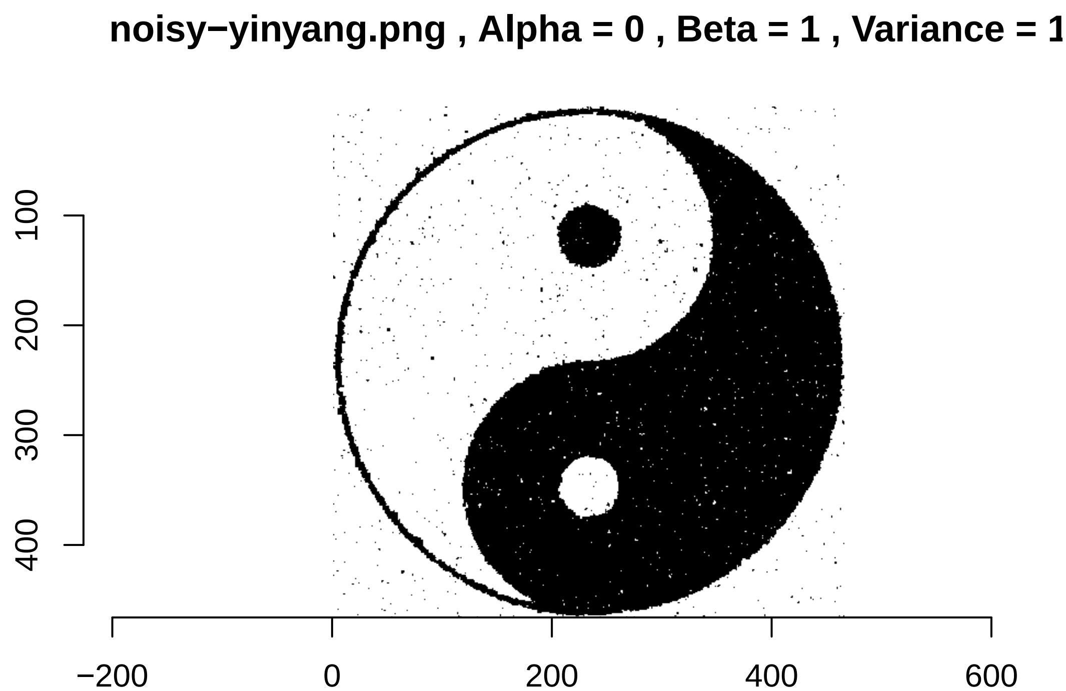






1(b) Here are the posterior images for 10 burn in and 20 denoising iterations. The parameter values I used were $\alpha = 0$, $\beta = 1$ and $\sigma^2 = 1$.





1(c) Final estimates for $\sigma^2 = 1$.

```
## [1] "Standard deviation estimate for noisy message: 1.188"
## [1] "Standard deviation estimate for noisy yinyang: 1.5856694784"
```

2. Say you are given data (X, Y), with $X \in \mathbb{R}^d$ and $Y \in \{0, 1\}$. The goal is to train a classifier that will predict an unknown class label $\tilde{y}$ from a new data point $\tilde{x}$. Consider the following model:

$$
\begin{aligned}
&Y \sim Ber\left(\frac{1}{1 + e^{-X^T\beta}}\right), \\
&\beta \sim N(0, \sigma^2 I).
\end{aligned}
$$

This is a **Bayesian logistic regression** model. Your goal is to derive and implement a Hamiltonian Monte Carlo sampler for doing Bayesian inference on β.

(a) Write down the formula for the unormalized posterior of $\beta|Y$, i.e.,

$$
\begin{aligned}
p(\beta|y;x,\sigma) &\propto \prod_{i=1}^{n} p(y_i|\beta;x_i)p(\beta;\sigma) \\
&= \prod_{i=1}^{n}\left(\left[\frac{1}{1+e^{-x_i^T\beta}}\right]^{y_i}\left[1-\frac{1}{1+e^{-x_i^T\beta}}\right]^{1-y_i}\right)\frac{1}{\sqrt{(2\pi)^k|\sigma^2 I|}}e^{\left(-\frac{1}{2}\beta^T(\sigma^2 I)^{-1}\beta\right)}
\end{aligned}
$$

(b) Show that this posterior is proportional to $exp(-U(\beta))$, where

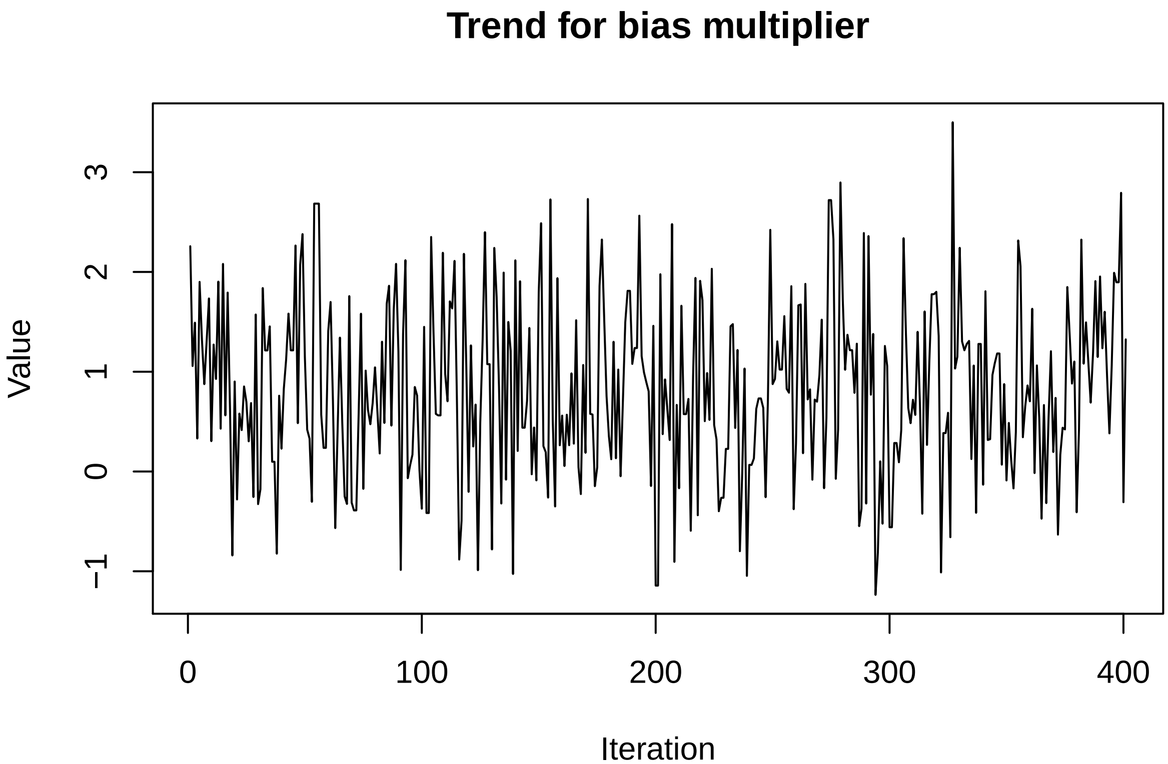
$$
U(\beta) = \sum_{i=1}^{n}(1-y_i)x_i^T\beta + \log(1+e^{-x_i^T\beta}) + \frac{1}{2\sigma^2}\|\beta\|^2
$$

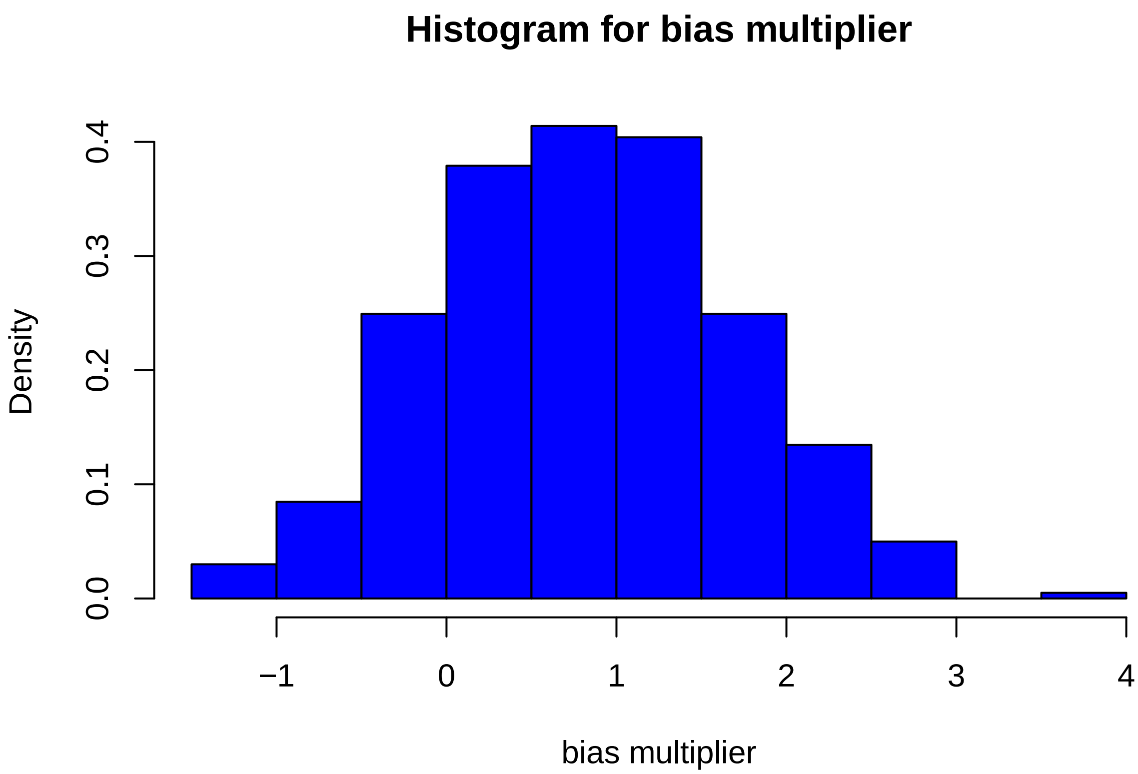
$$
\begin{aligned}
p(\beta|y;x,\sigma) &\propto \prod_{i=1}^{n}\left(\left[\frac{1}{1+e^{-x_i^T\beta}}\right]^{y_i}\left[1-\frac{1}{1+e^{-x_i^T\beta}}\right]^{1-y_i}\right)\frac{1}{\sqrt{(2\pi)^k|\sigma^2 I|}}e^{\left(-\frac{1}{2}\beta^T(\sigma^2 I)^{-1}\beta\right)} \\
&= \prod_{i=1}^{n}\left(\left[\frac{1}{1+e^{-x_i^T\beta}}\right]^{y_i}\left[\frac{e^{-x_i^T\beta}}{1+e^{-x_i^T\beta}}\right]^{1-y_i}\right)\frac{1}{\sqrt{(2\pi)^k|\sigma^2 I|}}e^{\left(-\frac{\|\beta\|^2}{2\sigma^2}\right)} \\
\log\left(p(\beta|y;x,\sigma)\right) &\propto -\sum_{i=1}^{n} y_i\log\left(1+e^{-x_i^T\beta}\right) + \sum_{i=1}^{n}(1-y_i).-x_i^T\beta - \sum_{i=1}^{n}(1-y_i)\log\left(1+e^{-x_i^T\beta}\right) - \frac{1}{2\sigma^2}\|\beta\|^2 \\
&= -\sum_{i=1}^{n} y_i\log\left(1+e^{-x_i^T\beta}\right) - \sum_{i=1}^{n}(1-y_i)x_i^T\beta - \sum_{i=1}^{n}\log\left(1+e^{-x_i^T\beta}\right) + \sum_{i=1}^{n} y_i\log\left(1+e^{-x_i^T\beta}\right) - \frac{\|\beta\|^2}{2\sigma^2} \\
&= -\sum_{i=1}^{n}(1-y_i)x_i^T\beta - \sum_{i=1}^{n}\log\left(1+e^{-x_i^T\beta}\right) - \frac{1}{2\sigma^2}\|\beta\|^2
\end{aligned}
$$

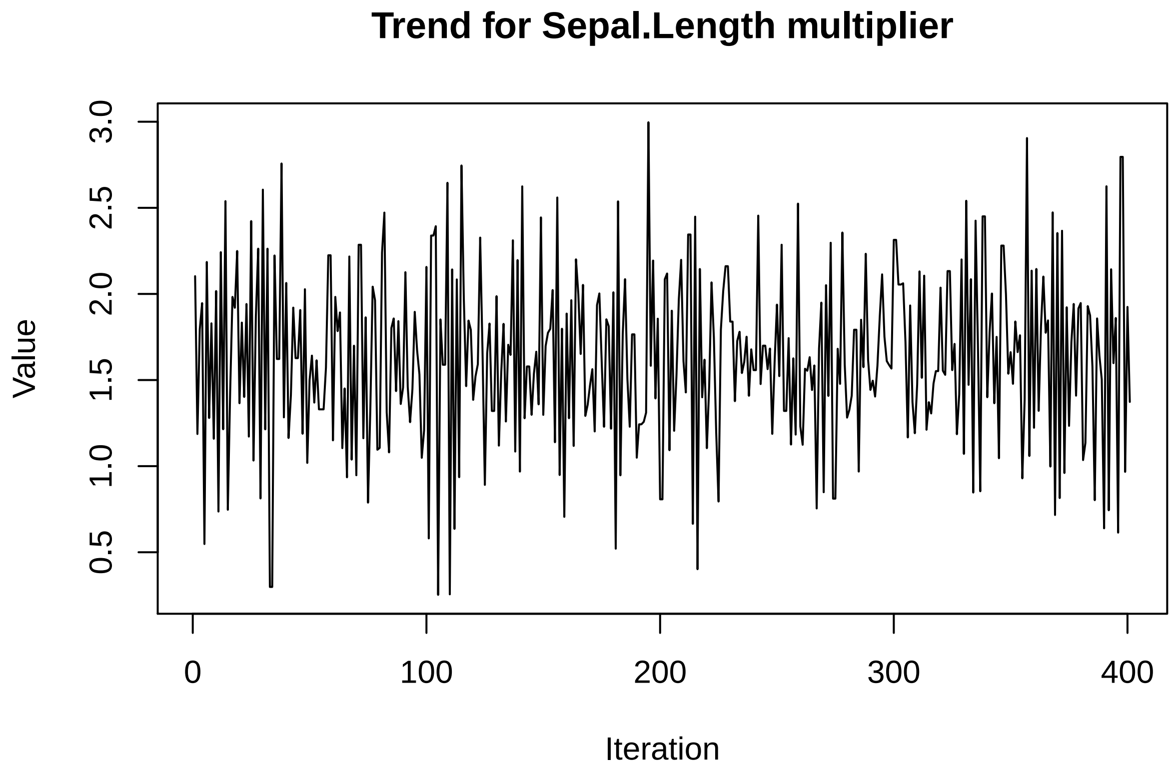
$\Rightarrow p(\beta|y;x,\sigma) \propto exp\left(-U(\beta)\right)$, where

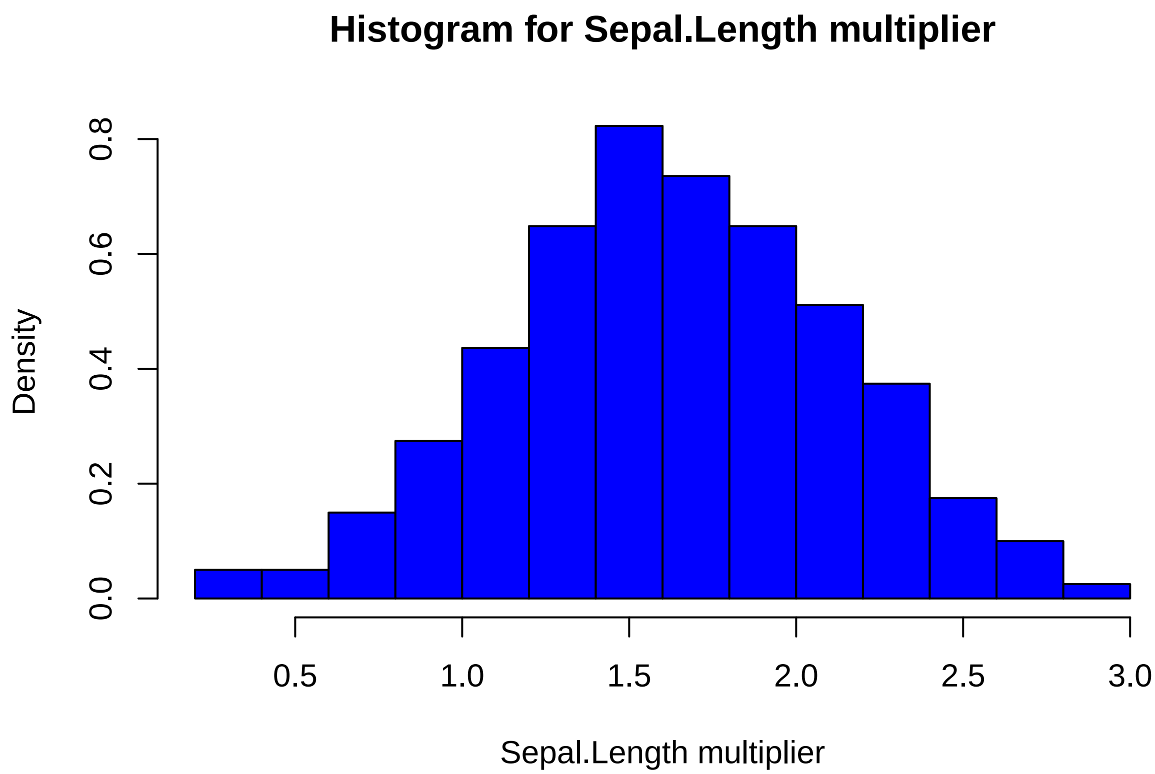
$$
U(\beta) = \sum_{i=1}^{n}(1-y_i)x_i^T\beta + \sum_{i=1}^{n}\log\left(1+e^{-x_i^T\beta}\right) + \frac{1}{2\sigma^2}\|\beta\|^2
$$

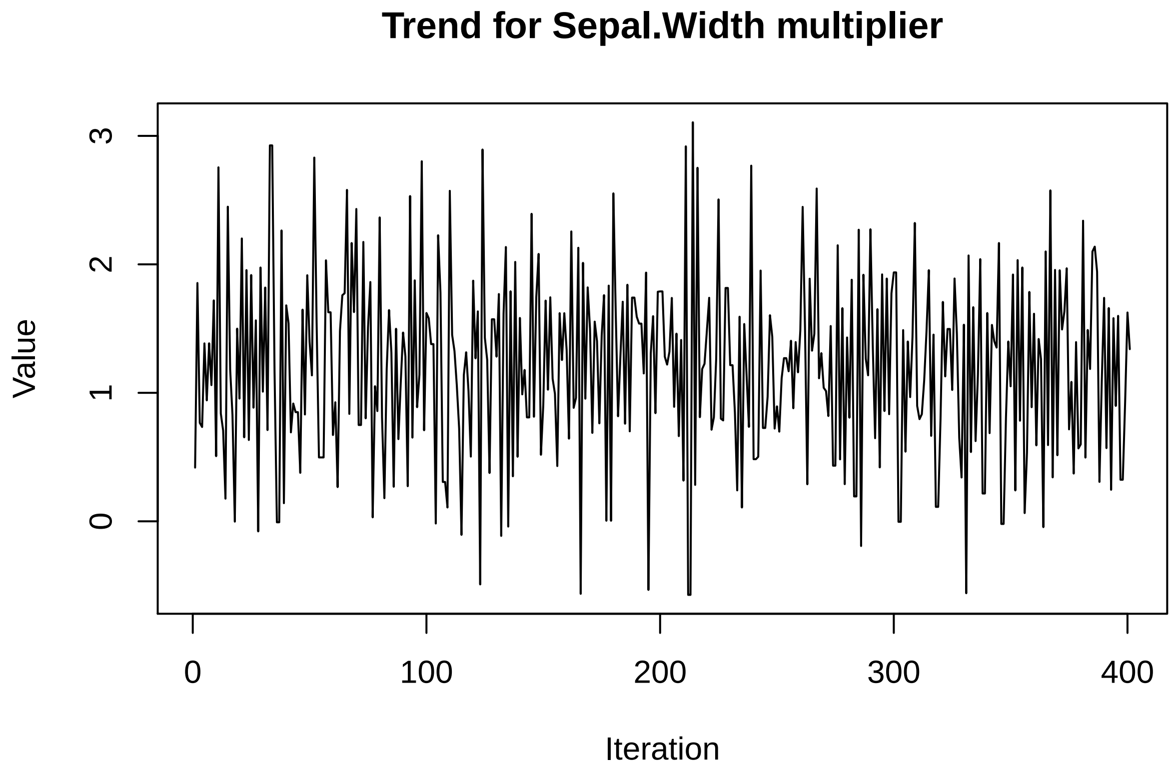
2(e) Trace plots of the β sequence and histograms are shown below.

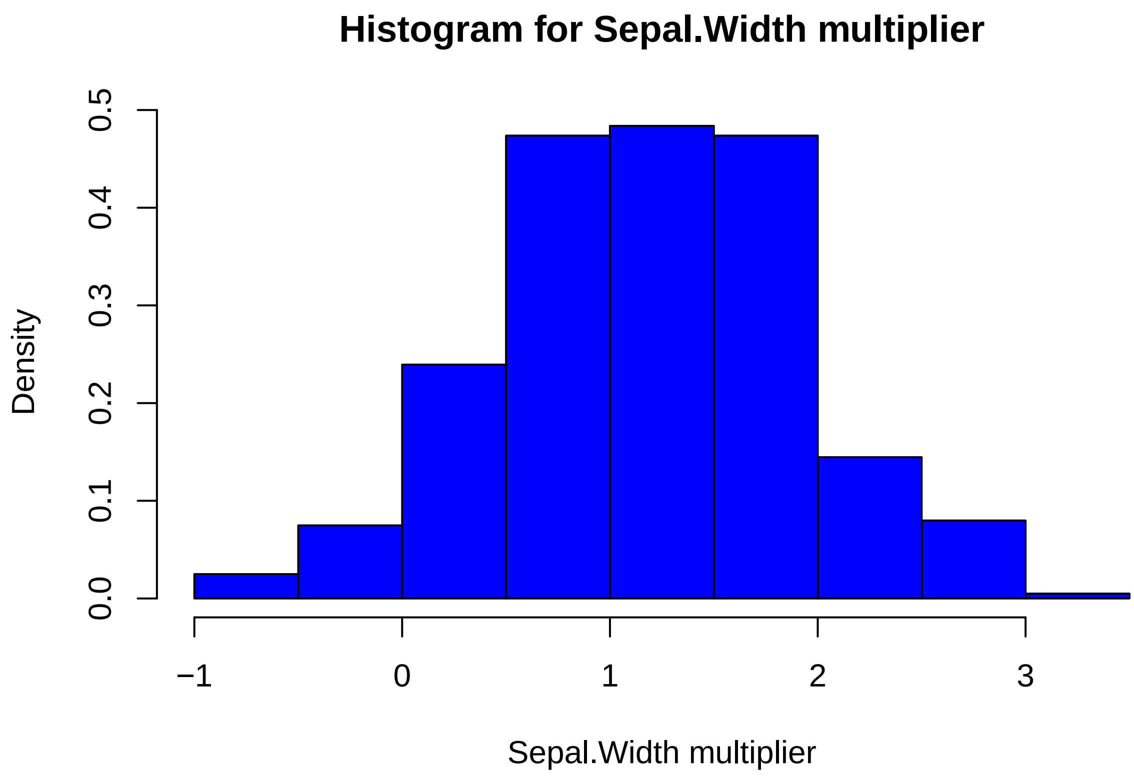

Trend for bias multiplier
Value
−1
0
1
2
3
0
100
200
300
400
Iteration

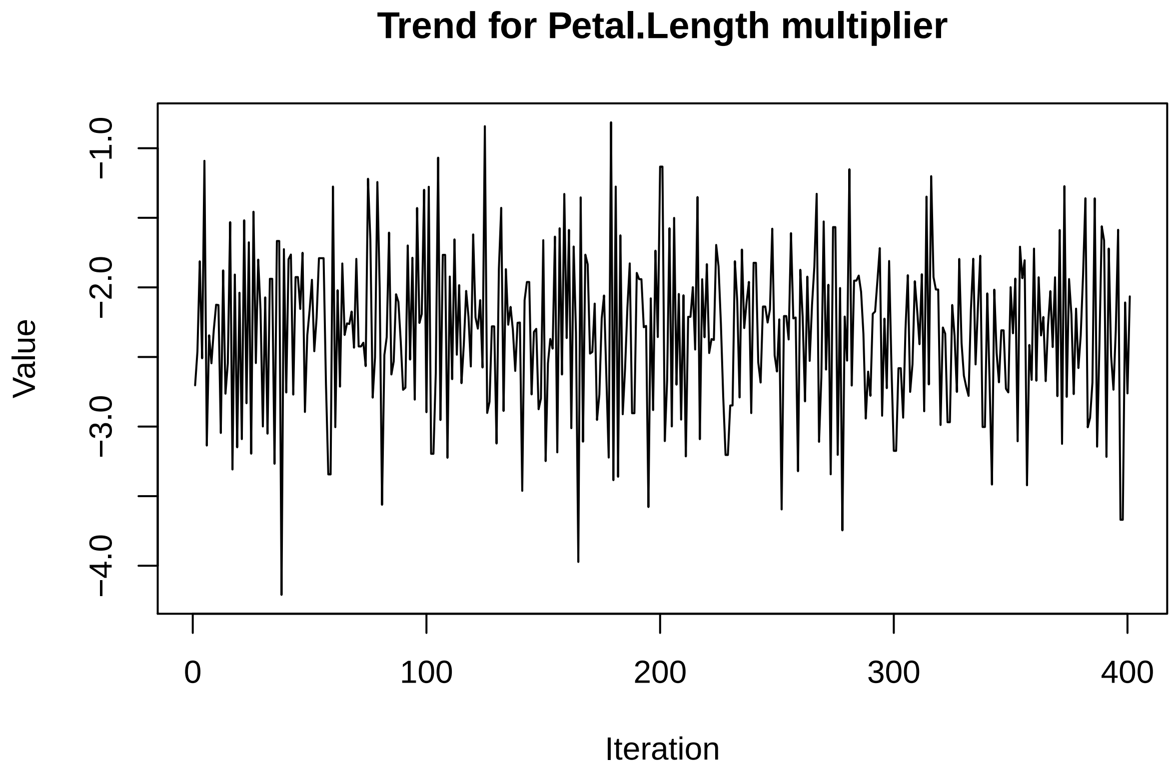

Histogram for bias multiplier
Density
0.0
0.1
0.2
0.3
0.4
−1
0
1
2
3
4
bias multiplier

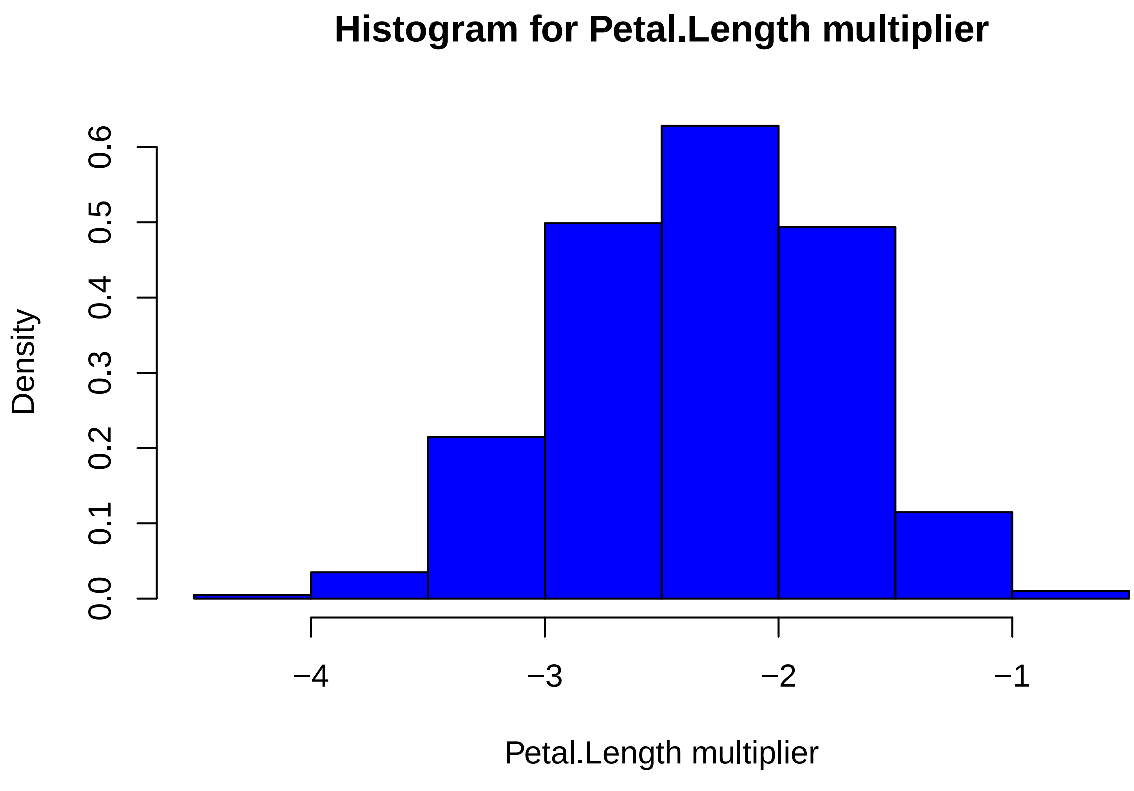

Trend for Sepal.Length multiplier
Value
3.0
2.5
2.0
1.5
1.0
0.5
0
100
200
300
400
Iteration

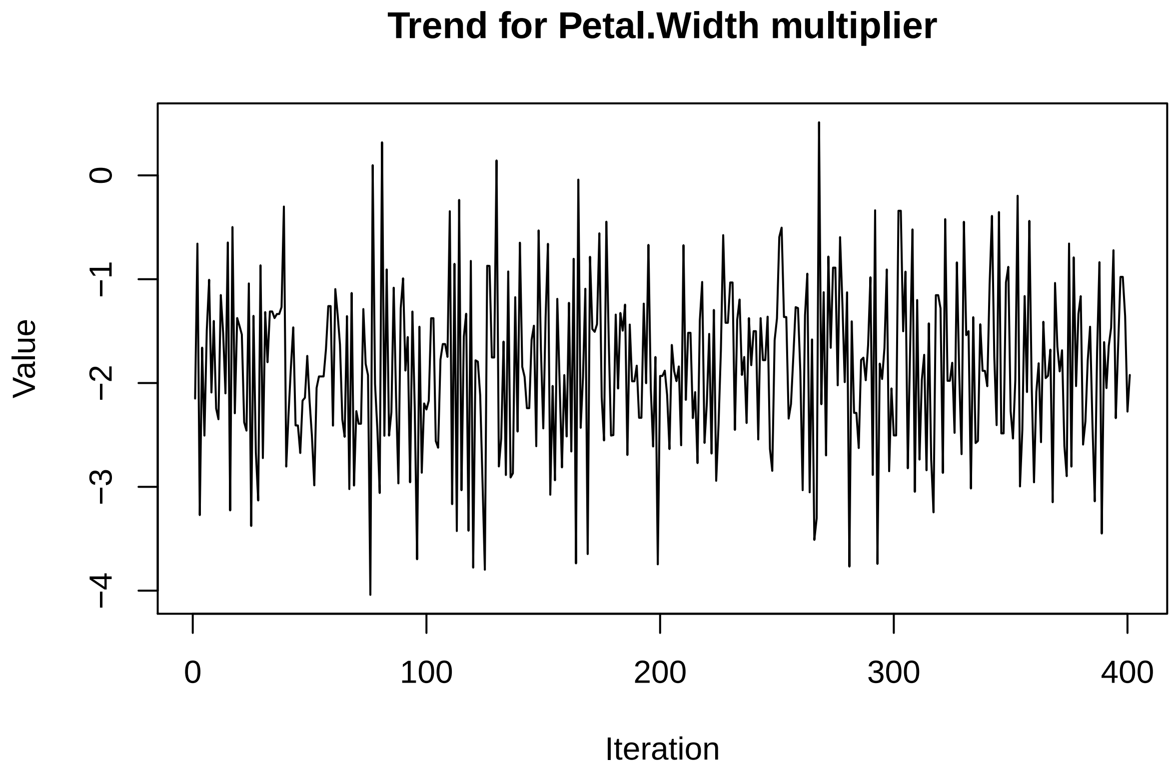

Histogram for Sepal.Length multiplier
Density
0.8
0.6
0.4
0.2
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Sepal.Length multiplier

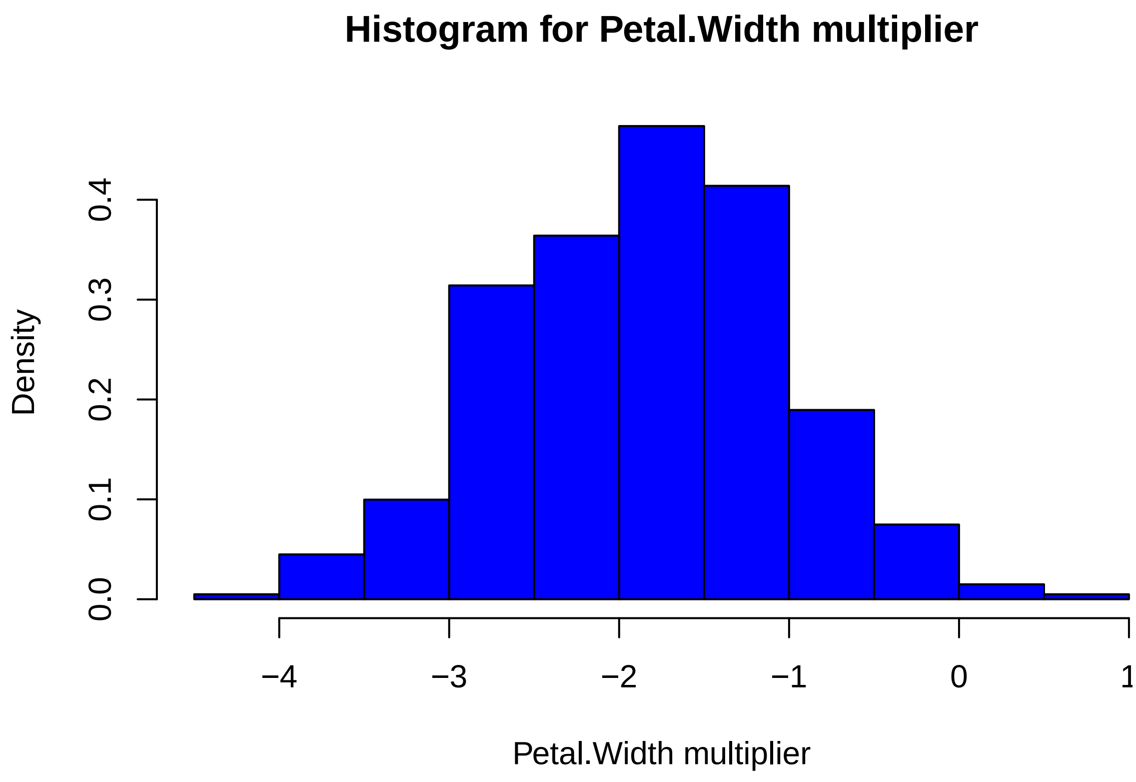

Trend for Sepal.Width multiplier
Value
0
1
2
3
0
100
200
300
400
Iteration


Histogram for Sepal.Width multiplier
Density
0.0
0.1
0.2
0.3
0.4
0.5
−1
0
1
2
3
Sepal.Width multiplier


Trend for Petal.Length multiplier
Value
−1.0
−2.0
−3.0
−4.0
0
100
200
300
400
Iteration


Histogram for Petal.Length multiplier
Density
0.0
0.1
0.2
0.3
0.4
0.5
0.6
−4
−3
−2
−1
Petal.Length multiplier





2(f) A variance σ^2 value of 1 was used, with a step size ϵ of 0.04 and number of steps L of 40.

```
## [1] "The zero-one loss was 2"
```