# CS228: Probabilistic Graphical Models

## Homework 4

Luke Jaffe

Due: 03/03/2017
Submitted: 03/03/2017

**Problem 4: Programming Assignment**

a) **TODO:** rewrite and scan

b)

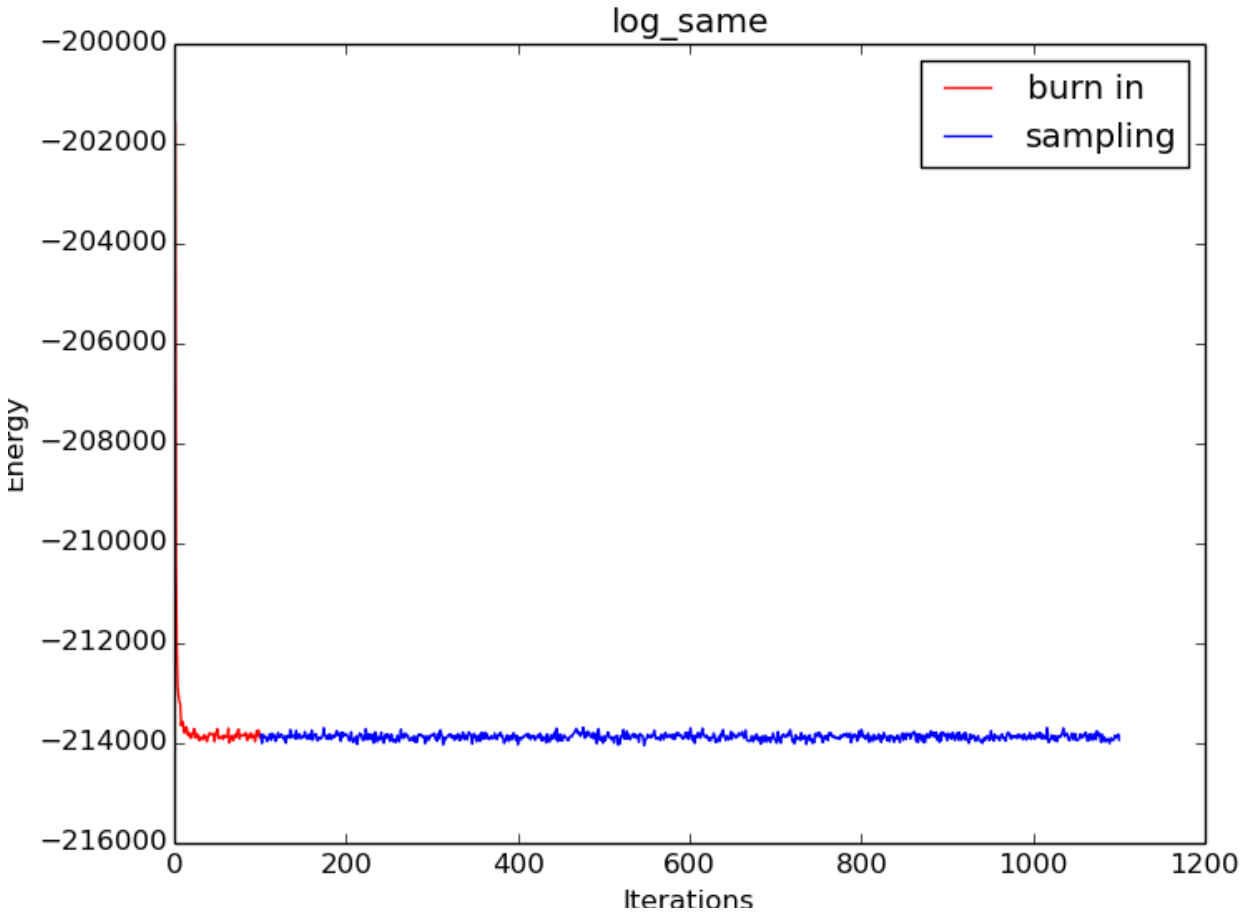
i. Outline a Gibbs sampling algorithm (in pseudocode) that iterates over the pixels in the image and samples each $y_{ij}$ given its Markov Blanket.

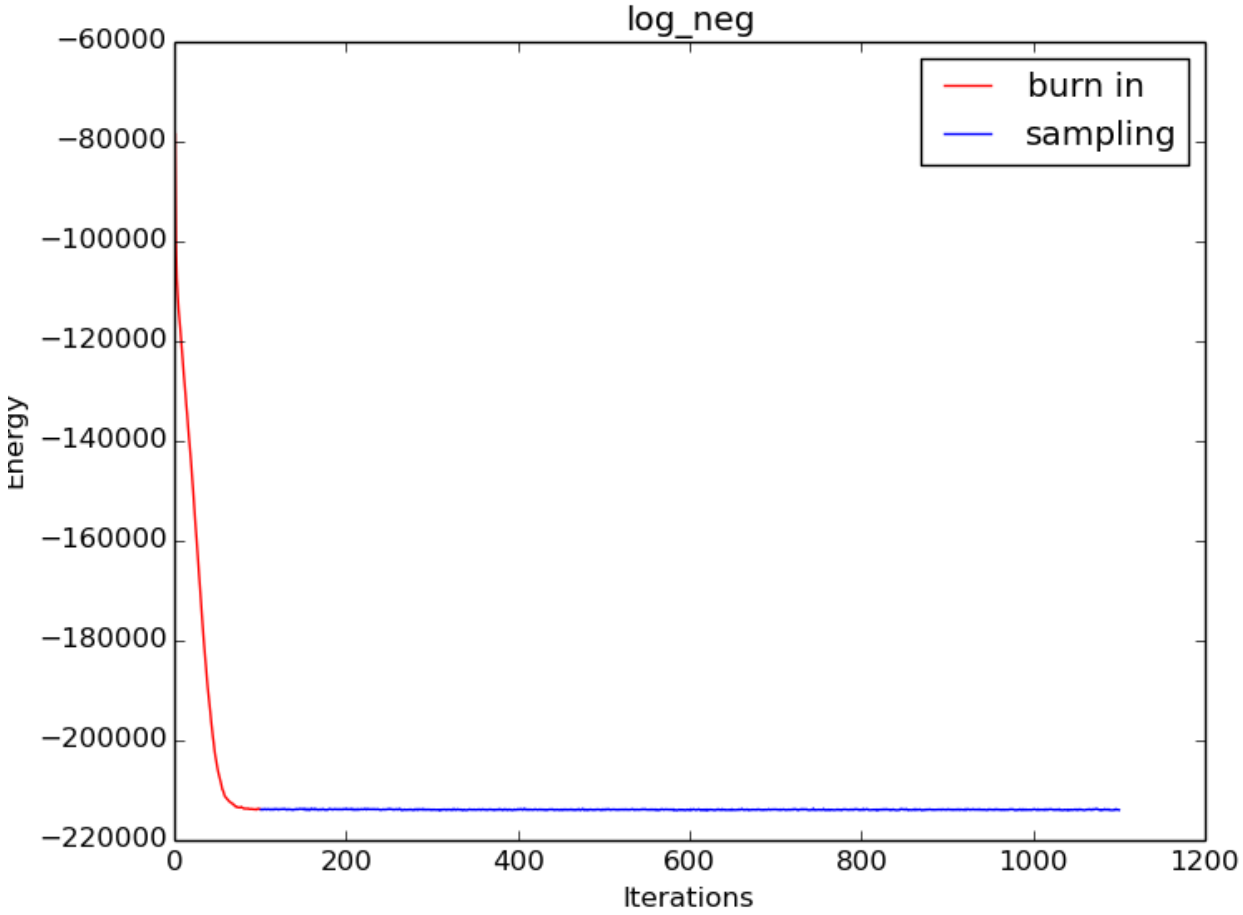
```
given constants eta, beta

function sample(Y, X):
    initialize matrix X to the noisy image
    initialize matrix Y randomly (same size as X)
    for I = 1 to N:
        for j = 1 to M:
            x_term = eta*X[i][j]
            y_term = 0
            for y in markov_blanket(Y, i, j):
                y_term += y
            y_term *= beta
            a = x_term + y_term
            s = sigmoid(2*a)
            Y[i][j] = s

function gibbs(Y, X, B, N):
    for t = 1 to B:
        sample(Y, X)
    for t = 1 to N
        sample(Y, X)
    return Y
```

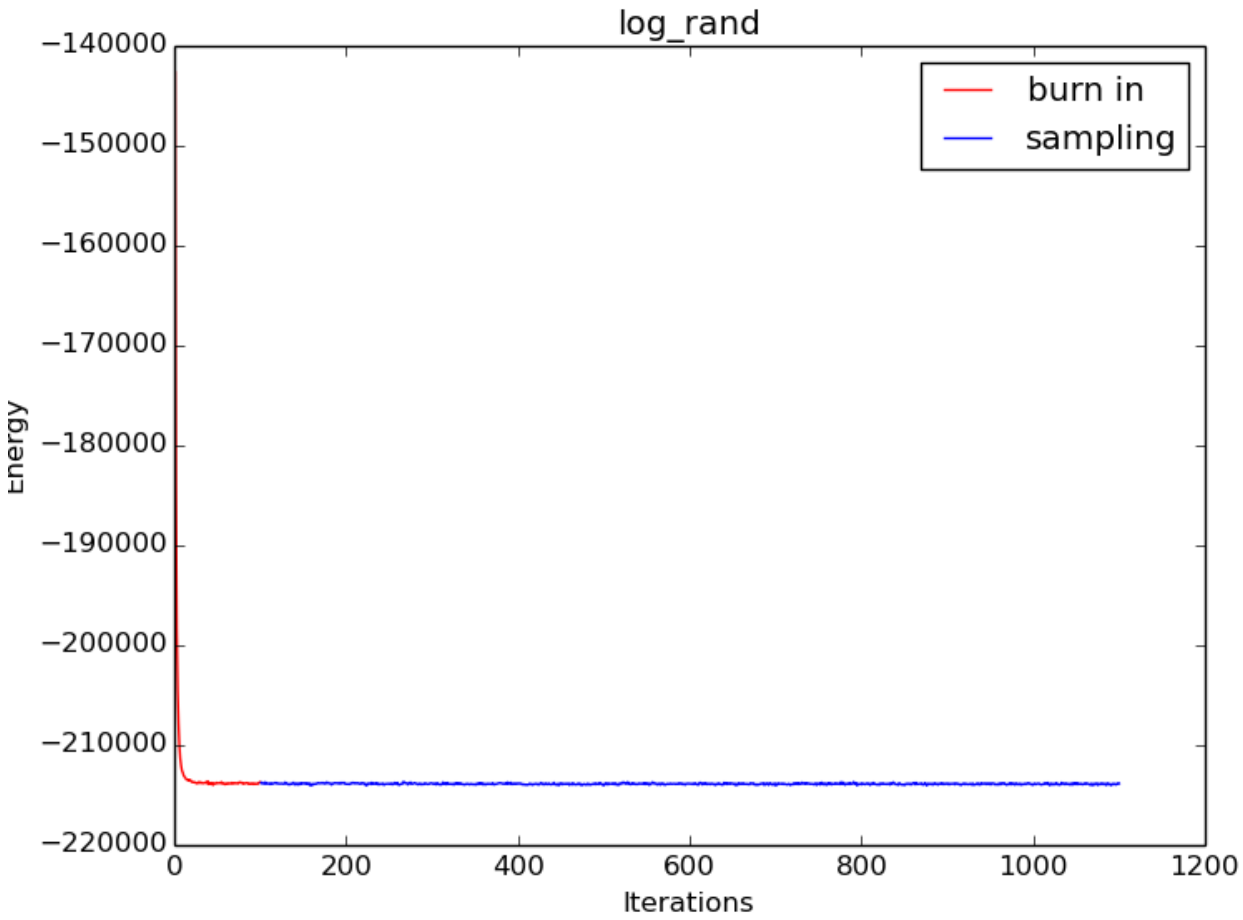
ii. How can we show in our case that the equilibrium distribution is in fact the posterior distribution p(y|x)?

We can show this using the Monte Carol equation.

c)


log_same
burn in
sampling
−200000
−202000
−204000
−206000
−208000
−210000
−212000
−214000
−216000
Energy
0
200
400
600
800
1000
1200
Iterations


log_neg
burn in
sampling
−60000
−80000
−100000
−120000
−140000
−160000
−180000
−200000
−220000
Energy
0
200
400
600
800
1000
1200
Iterations




i. Do all three seem to be converging to the same general region of the posterior, or are some obviously suboptimal?

While all three methods do not converge in exactly the same way, they all converge to the same general region of the posterior after some iterations. The negative initialization may be considered suboptimal since it takes longer to converge than the same or random initialization methods.

ii. Does the burn-in seem to be adequate in length?

Yes, the burn-in is more than sufficient in length. The same and random initializations seem to converge in less than 20 iterations, while the negative initialization needs 80 or so.

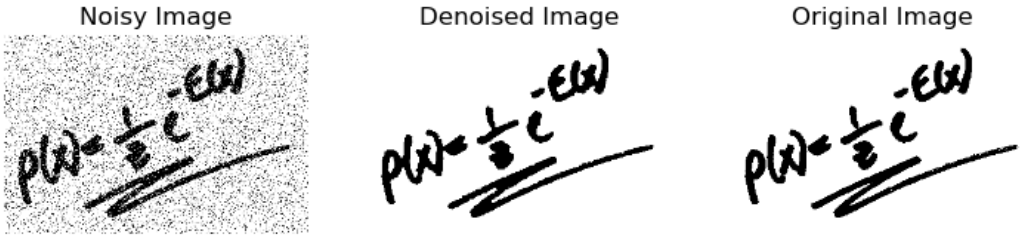
iii. Is there substantial fluctuation from iteration to iteration, indicating that the chain is mixing well, or does it become stuck at particular energies for several iterations at a time?

Yes, there is substantial fluctuation between iterations for all methods. In the plots above, the “same” method may appear to fluctuate the most, but this is just because of the y-axis scale.

d)

Denoised 10% error: 0.0059
Denoised 20% error: 0.0104

denoised_10% plots

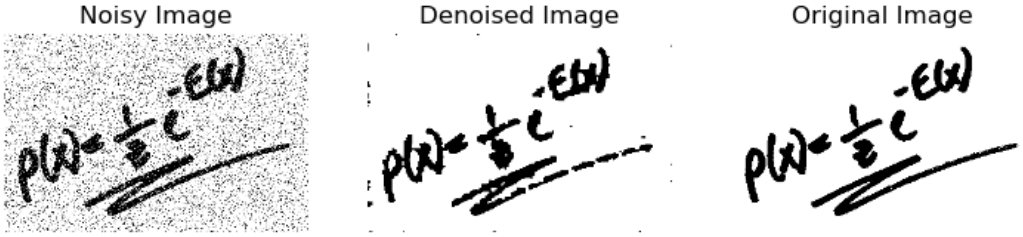



denoised_20% plots

e)

Denoised dumb 10% error: 0.0213
Denoised dumb 20% error: 0.0588

denoised_dumb_10% plots

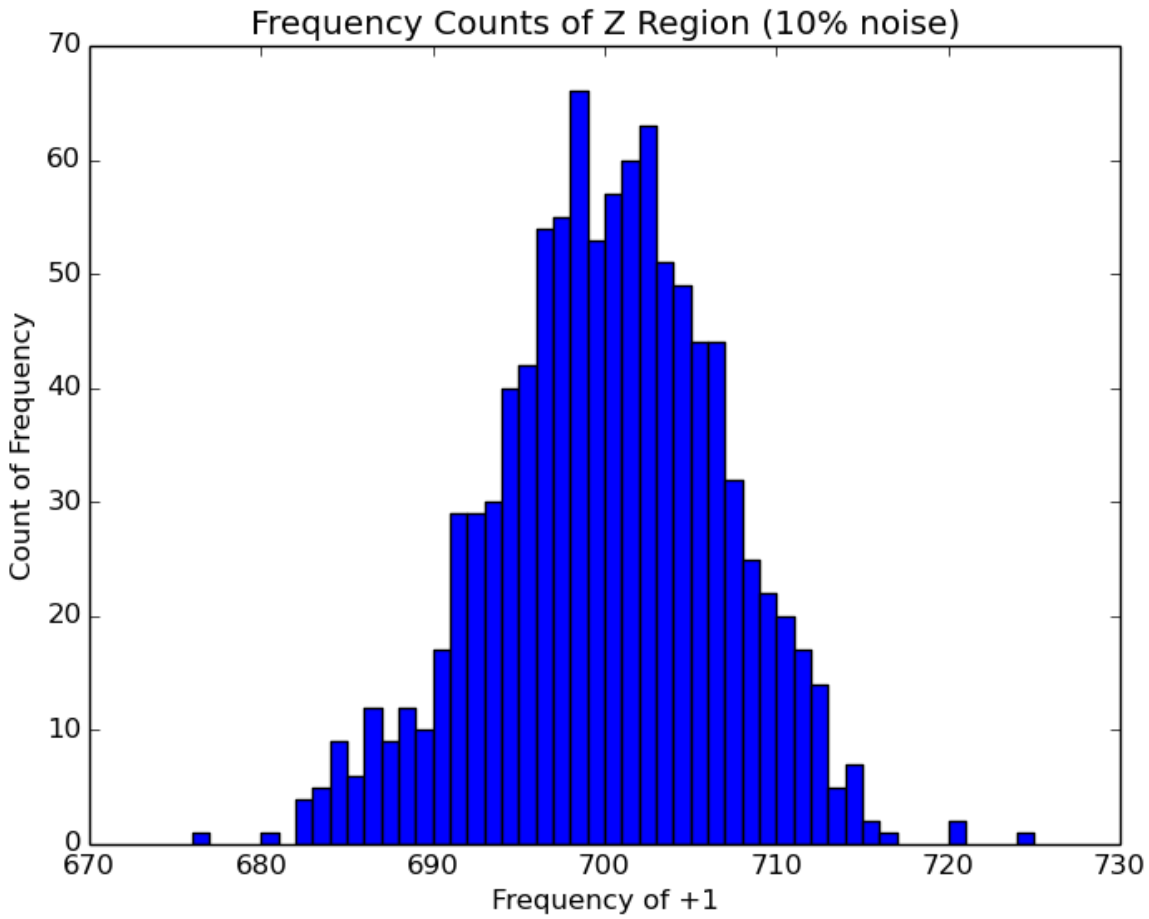


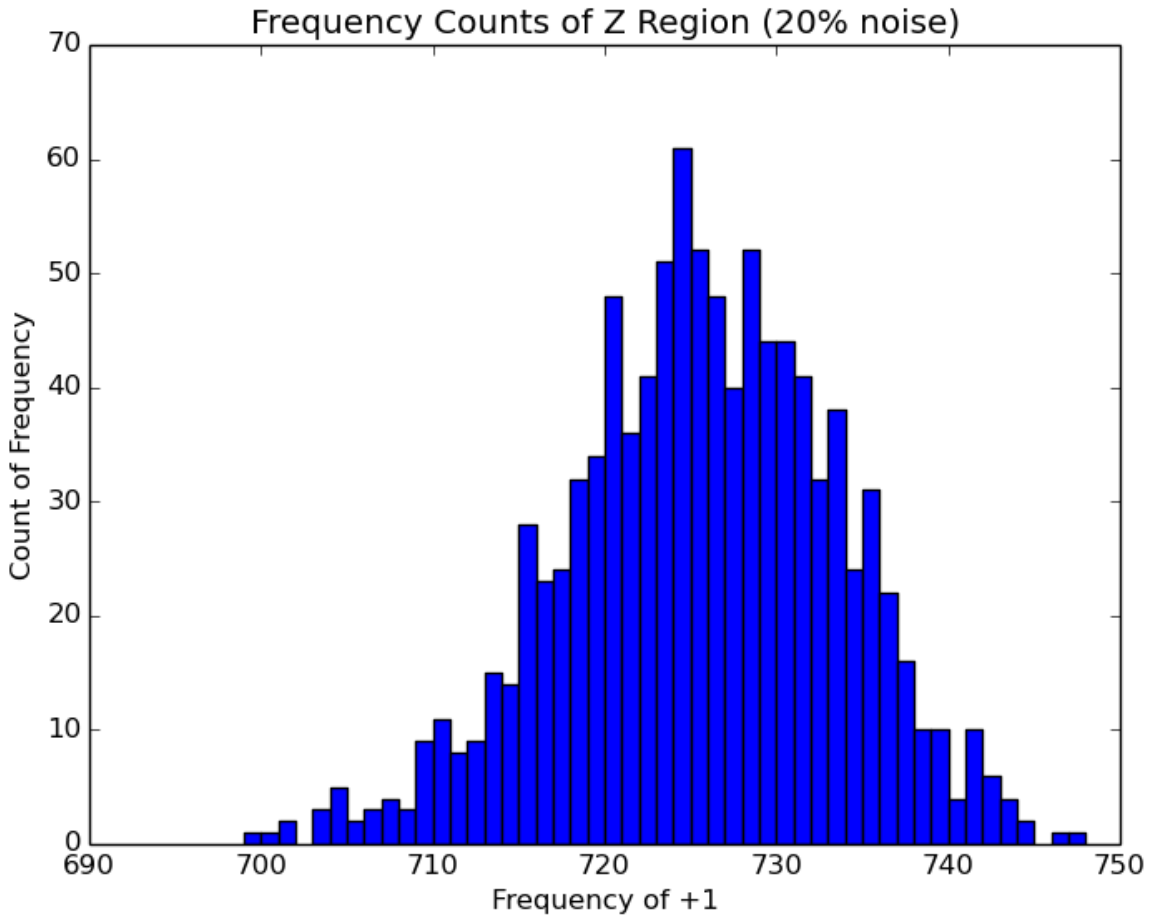

denoised_dumb_20% plots

i. Does the Gibbs sampler do better than the trivial algorithm? Why or why not?

Yes, the Gibbs sampler does better than the trivial algorithm. The error is lower in the 10% noise case (0.5% v. 2%) and the 20% noise case (1% v. 6%). The images for the Gibbs sampler are also visibly better denoised. This is because the Gibbs sampler is built on a model which to some degree correctly codifies the spatial locality relationships of the actual imagery, whereas the other method is naive.

f)






i. The distribution of frequencies for the noisier case is slightly wider (greater variance), and has a greater mean (~725 v. ~700).