

```

[1] "# of discriminatory genes = 1"

[1] "fold change for discriminatory genes: 0.5"

[1] "fold change for nondiscriminatory genes: 0.1"

choose_k
  2
100

```

The optimal number of clusters K is set as the most frequently found K from the 100 simulations tabulated above. It is found by using the BIC criterion after running the unpenalized EM algorithm on K spanning from 2 to 8.

Using the last set of simulated counts, I ran a grid search across varying tuning parameters. As done in Pan et al, I fixed  $\lambda_1 = 1$ , and searched over  $\lambda_2 = (0.1, 0.2, \dots, 2)$  and  $\tau = (0.1, 0.2, \dots, 2)$ :

```

[1] "lambda1, lambda2, tau, BIC:"

      [,1] [,2] [,3]      [,4]
[1,]    1  0.3  0.1 19247.54
[2,]    1  0.3  0.2 19247.54
[3,]    1  0.3  0.3 19247.54
[4,]    1  0.3  0.4 19247.54
[5,]    1  0.3  0.5 19247.54
[6,]    1  0.3  0.6 19247.54
[7,]    1  0.4  0.1 19247.54
[8,]    1  0.4  0.2 19247.54
[9,]    1  0.4  0.3 19247.54
[10,]   1  0.4  0.4 19247.54
[11,]   1  0.4  0.5 19247.54
[12,]   1  0.5  0.1 19247.54
[13,]   1  0.5  0.2 19247.54
[14,]   1  0.5  0.3 19247.54
[15,]   1  0.5  0.4 19247.54
[16,]   1  0.6  0.1 19247.54
[17,]   1  0.6  0.2 19247.54
[18,]   1  0.6  0.3 19247.54
[19,]   1  0.7  0.1 19247.54
[20,]   1  0.7  0.2 19247.54
[21,]   1  0.8  0.1 19247.54

```

The results of the final run based on optimal tuning parameters are below:  
Below are the summary of results:

```

[1] "Mean pi: 0.585217391304348" "Mean pi: 0.414782608695652"

[1] "First 3 genes:"

```

```

      [,1]      [,2]
[1,] 3.407622 4.333236
[2,] 5.020155 5.159425
[3,] 4.535741 4.672607

[1] "Last 3 genes:"

      [,1]      [,2]
[98,] 6.318445 6.481589
[99,] 5.872330 6.027641
[100,] 6.297777 6.456920

[1] "Mean % of nondiscriminatory genes: 0.9872"

[1] "Mean ARI: 1"

[1] "Mean sensitivity: 1"

[1] "Mean false positive rate: 0.00282828282828283"

```