

```

[1] "# of discriminatory genes = 25"

[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 19396.2
[2,]    1  0.3  0.2 19396.2
[3,]    1  0.3  0.3 19396.2
[4,]    1  0.3  0.4 19396.2
[5,]    1  0.4  0.1 19396.2
[6,]    1  0.4  0.2 19396.2
[7,]    1  0.4  0.3 19396.2
[8,]    1  0.5  0.1 19396.2
[9,]    1  0.5  0.2 19396.2
[10,]   1  0.6  0.1 19396.2

```

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

```

[1] "Mean pi: 0.603913043478261" "Mean pi: 0.396086956521739"

[1] "First 3 genes:"

      [,1]      [,2]
[1,] 3.367745 4.350240
[2,] 4.593873 5.589917
[3,] 4.118598 5.104477

[1] "Last 3 genes:"

      [,1]      [,2]
[98,] 6.313384 6.483693
[99,] 5.868213 6.032464
[100,] 6.291623 6.461208

```

```
[1] "Mean % of nondiscriminatory genes: 0.7489"  
[1] "Mean ARI: 1"  
[1] "Mean sensitivity: 1"  
[1] "Mean false positive rate: 0.00146666666666667"
```