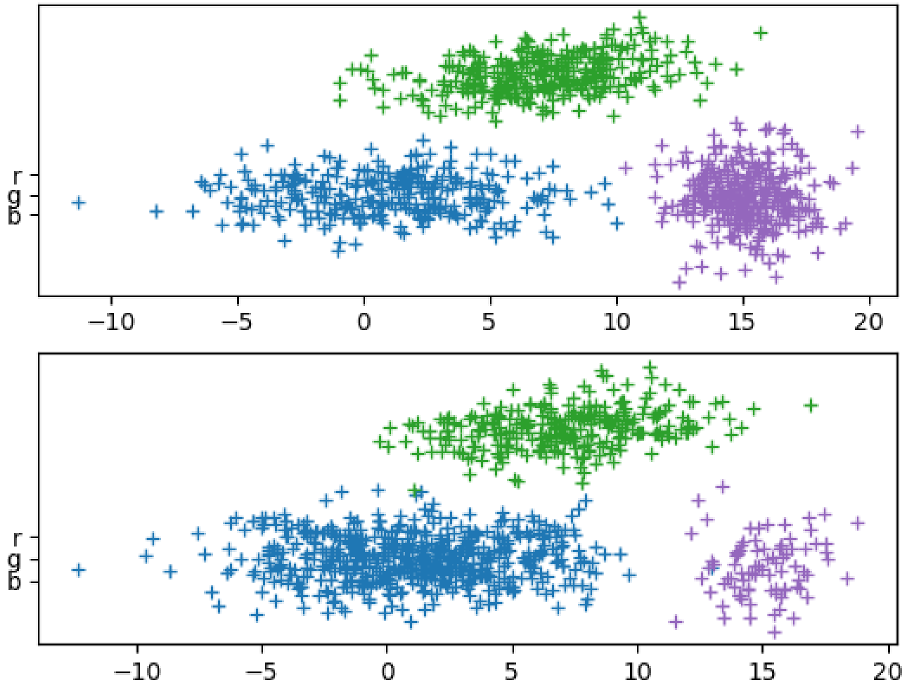
1.

得到散点图：



其中生成的点集的坐标在 1.txt 和 2.txt 中。

分别使用式 18,19 得到 MLE 估计的参数，用式 45,46 得到 BE 估计的参数，运行可得结果在 result.txt 中，可以看到，MLE 方法生成的参数比用 BE 得到的参数误差更大，尤其是在方差的估计上。当样本量减小时，两种方法的误差都增加了。这两者主要来源于 p(θ|D)的波形较宽（尤其是第一类和第二类），和样本数目不足带来的估计误差。

2.

17. (a) 由 $P(x|\theta)=\prod_{i=1}^{d}\theta_i^{x_i}(1-\theta_i)^{1-x_i}$ 可知.

$$P(x_1,x_2,\dots,x_n|\theta)=P(D|\theta)=\prod_{j=1}^{n}\prod_{i=1}^{d}\theta_i^{x_{ji}}(1-\theta_i)^{1-x_{ji}}$$

$$=\prod_{i=1}^{d}\theta_i^{\sum_{j=1}^{n}x_{ji}}(1-\theta_i)^{\sum_{j=1}^{n}(1-x_{ji})}$$

$$=\prod_{i=1}^{d}\theta_i^{s_i}(1-\theta_i)^{n-s_i}$$

(b) 有 $P(\theta|D)=\dfrac{P(D|\theta)P(\theta)}{P(D)}$，对于 $P(D|\theta)$ 有以上式

对于 $P(D)$，有 $P(D)=\int P(D|\theta)p(\theta)d\theta$ ∵ $p(\theta)$ 均匀

$$=\int_0^1\cdots\int_0^1\prod_{i=1}^{d}\theta_i^{s_i}(1-\theta_i)^{n-s_i}(d\theta_1 d\theta_2\cdots d\theta_d)$$

$$=\prod_{i=1}^{d}\int_0^1\theta_i^{s_i}(1-\theta_i)^{n-s_i}d\theta_i$$

使用恒等式：$P(D)=\prod_{i=1}^{d}\dfrac{s_i!(n-s_i)!}{(n+1)!}$

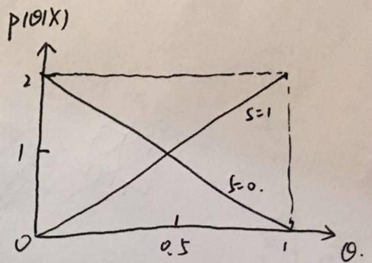
$$\therefore P(\theta|D)=\frac{\prod_{i=1}^{d}\theta_i^{s_i}(1-\theta_i)^{n-s_i}}{\prod_{i=1}^{d}\frac{s_i!(n-s_i)!}{(n+1)!}}=\prod_{i=1}^{d}\frac{(n+1)!}{s_i!(n-s_i)!}\theta_i^{s_i}(1-\theta_i)^{n-s_i}$$

$\prod_{i=1}^{d}\frac{s_i!(n-s_i)!}{(n+1)!}$

(c) $p(\theta|X)=\dfrac{2}{x!(1-x)!}\theta^{x}(1-\theta)^{1-x}$ 有

$x=0$ 时，$p(\theta|x)=2(1-\theta)$

$x=1$ 时，$p(\theta|x)=2\theta$.



(d) $$\int P(x|\theta)P(\theta|D)d\theta=\int_0^1\cdots\int_0^1\prod_{i=1}^{d}\theta_i^{x_i}(1-\theta_i)^{1-x_i}\prod_{i=1}^{d}\frac{(n+1)!}{s_i!(n-s_i)!}\theta_i^{s_i}(1-\theta_i)^{n-s_i}(d\theta_1 d\theta_2\cdots d\theta_d)$$

$$=\prod_{i=1}^{d}\int_0^1\theta_i^{x_i+s_i}(1-\theta_i)^{n+1-s_i-x_i}\frac{(n+1)!}{s_i!(n-s_i)!}d\theta_i$$

$$=\prod_{i=1}^{d}\frac{(x_i+s_i)!(n+1-s_i-x_i)!}{(n+2)s_i!(n-s_i)!}=\prod_{i=1}^{d}f(i)$$

当 $x_i=0$ 时. $f(i)=1-\dfrac{s_i+1}{n+2}$ 当 $x_i=1$ 时. $f(i)=\dfrac{s_i+1}{n+2}$

$$\therefore \text{原式}=\prod_{i=1}^{d}\left(1-\frac{s_i+1}{n+2}\right)^{1-x_i}\left(\frac{s_i+1}{n+2}\right)^{x_i}$$

(e) 则 $\hat{\theta}=\dfrac{s_i+1}{n+2}$

3.

因为课本的HMM学习模型没有涉及多序列的训练，我参考了网上的训练方式，结合课本进行训练。我设置了4个隐状态，5个可见状态（包括终止状态）：

model1:

a:

[[5.11153030e-01 3.99091110e-01 8.97558599e-02 0.00000000e+00]

[4.84562083e-02 6.82177035e-01 2.69366757e-01 2.64947772e-53]

[1.93322411e-48 1.22711973e-13 4.81777124e-01 5.18222876e-01]

[0.00000000e+00 0.00000000e+00 0.00000000e+00 1.00000000e+00]]

b:

[[7.27599845e-01 2.72400155e-01 1.65851872e-29 5.50861830e-41 0.00000000e+00]

[6.44927062e-07 4.64760235e-01 4.41612459e-01 9.36266618e-02 0.00000000e+00]

[1.97991544e-80 6.14585955e-51 2.09549239e-01 7.90450761e-01 0.00000000e+00]

[0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00 1.00000000e+00]]

pi:

[1.0, 1.117977272357278e-95, 0.0, 0.0]

model2:

a:

[[6.91264444e-01 3.08733491e-01 2.70950567e-19 2.06483037e-06]

[3.91302779e-02 5.82779810e-01 3.78089912e-01 4.77902797e-32]

[3.04281063e-16 2.32316017e-06 4.05081946e-01 5.94915731e-01]

[0.00000000e+00 0.00000000e+00 0.00000000e+00 1.00000000e+00]]

b:

[[9.85196529e-02 1.19409128e-01 1.38575354e-01 6.43495864e-01 0.00000000e+00]

[1.38756478e-01 5.19184865e-01 3.42058655e-01 9.07262443e-10 0.00000000e+00]

[9.40494512e-01 1.77601038e-05 4.48538229e-22 5.94877274e-02 0.00000000e+00]

[0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00 1.00000000e+00]]

pi:

[0.999999998368456, 1.6315439359589093e-09, 5.08155973217321e-121, 0.0] evaluation:

0.0001776704185648031 3.1165150426908877e-09

type: 1

8.21681048066042e-105 2.7214006288007157e-05

type: 2

9.345360777484206e-93 1.8092950792981618e-05

type: 2

4.212158972514054e-05 2.3815495026759947e-06

type: 1

8.95529592177064e-65 1.5279118498656896e-06

Etype: 2

需要使 p(lamda1): 1.0 p(lamda2): 0.0 才能使二者后验概率相差不大。

发现以上结果不够理想，我尝试训练了包含 3 个隐状态，4 个可见状态（不包括终止状态）的模型：

1. Learning:

Model1:

a:

[[9.51425535e-001 4.85744646e-002 2.55948131e-175]

[4.69127158e-001 4.55980945e-053 5.30872842e-001]

[4.30002659e-001 3.34799152e-001 2.35198189e-001]]

b:

[[2.46951533e-127 2.32152177e-001 3.72954657e-001 3.94893166e-001]

[1.76980243e-004 9.99823020e-001 3.16080731e-127 5.76094451e-117]

[1.00000000e+000 1.59550838e-079 1.64124391e-109 3.36605533e-162]]

pi:

[0.0, 0.20003169580035002, 0.7999683041996499]

Model2:

a:

[[7.32865749e-01 2.67134251e-01 1.09846842e-17]

[1.13965712e-14 4.09617997e-01 5.90382003e-01]

[6.30611312e-02 2.21555867e-23 9.36938869e-01]]

b:

[[1.14915196e-01 6.63174119e-02 1.53741998e-01 6.65025394e-01]

[3.89133406e-05 4.99552176e-01 5.00408910e-01 4.84861492e-10]

[7.24479182e-01 2.75520818e-01 1.68019727e-11 8.45941838e-33]]

pi:

[0.9373051449770543, 0.06269485500490389, 1.804190207704293e-11]

2. Evaluation:

0.0002235850160520472 1.630870257147244e-06

type: 1

5.474604534876615e-122 0.00012322631972470354

type: 2

3.62450959780954e-113 0.0004125752367839571

type: 2

0.00026091922755094605 3.3460673331124657e-06

type: 1

可见模型可以较显著地把一个序列分类。

3. Classify:

1.5172357134266758e-05 2.4710995110576897e-06

Etype: 1

p(lamda1): 0.14005756132330996 p(lamda2): 0.8599424386766901

需要使 p(lamda1): 0.14 p(lamda2): 0.86 才能使二者后验概率相差不大。

结果较好。