

## Exercise: from sites to matrices

### Sequences

Seq1 AAAAACCG  
 Seq2 AAAACCGG  
 Seq3 AATTGGGG  
 Seq4 ATTGTGGG

### Alignment

Position	1	2	3	4	5	6	7	8
Seq1	A	A	A	A	A	C	C	G
Seq2	A	A	A	A	C	C	G	G
Seq3	A	A	T	T	G	G	G	G
Seq4	A	T	T	G	T	G	G	G

### Count matrix

Position	1	2	3	4	5	6	7	8
A	4	3	2	2	1	0	0	0
C	0	0	0	0	1	2	1	0
G	0	0	0	1	1	2	3	4
T	0	1	2	1	1	0	0	0
Total	4	4	4	4	4	4	4	4

### Equal priors

Residue	Prior
A	0.25
C	0.25
G	0.25
T	0.25

### Frequency matrices with equal priors

Pseudo-weight 0

Position	1	2	3	4	5	6	7	8
A	1	0.75	0.5	0.5	0.25	0	0	0
C	0	0	0	0	0.25	0.5	0.25	0
G	0	0	0	0.25	0.25	0.5	0.75	1
T	0	0.25	0.5	0.25	0.25	0	0	0
Total	1	1	1	1	1	1	1	1

Pseudo-weight 1

Position	1	2	3	4	5	6	7	8
A	0.85	0.65	0.45	0.45	0.25	0.05	0.05	0.05
C	0.05	0.05	0.05	0.05	0.25	0.45	0.25	0.05
G	0.05	0.05	0.05	0.25	0.25	0.45	0.65	0.85
T	0.05	0.25	0.45	0.25	0.25	0.05	0.05	0.05
Total	1	1	1	1	1	1	1	1

Pseudo-weight 10

Position	1	2	3	4	5	6	7	8
A	0.464	0.393	0.321	0.321	0.250	0.179	0.179	0.179
C	0.179	0.179	0.179	0.179	0.250	0.321	0.250	0.179
G	0.179	0.179	0.179	0.250	0.250	0.321	0.393	0.464
T	0.179	0.250	0.321	0.250	0.250	0.179	0.179	0.179
Total	1	1	1	1	1	1	1	1

Pseudo-weight 100

Position	1	2	3	4	5	6	7	8
A	0.279	0.269	0.260	0.260	0.250	0.240	0.240	0.240
C	0.240	0.240	0.240	0.240	0.250	0.260	0.250	0.240
G	0.240	0.240	0.240	0.250	0.250	0.260	0.269	0.279
T	0.240	0.250	0.260	0.250	0.250	0.240	0.240	0.240
Total	1	1	1	1	1	1	1	1

### Weight matrices with equal priors

Pseudo-weight 0

Position	1	2	3	4	5	6	7	8
A	1.39	1.10	0.69	0.69	0.00	-Inf	-Inf	-Inf
C	-Inf	-Inf	-Inf	-Inf	0.00	0.69	0.00	-Inf
G	-Inf	-Inf	-Inf	0.00	0.00	0.69	1.10	1.39
T	-Inf	0.00	0.69	0.00	0.00	-Inf	-Inf	-Inf
Total	1.386	1.099	1.386	0.693	0	1.386	1.099	1.386

Pseudo-weight 1

Position	1	2	3	4	5	6	7	8
A	1.22	0.96	0.59	0.59	0.00	-1.61	-1.61	-1.61
C	-1.61	-1.61	-1.61	-1.61	0.00	0.59	0.00	-1.61
G	-1.61	-1.61	-1.61	0.00	0.00	0.59	0.96	1.22
T	-1.61	0.00	0.59	0.00	0.00	-1.61	-1.61	-1.61
Total	-3.605	-2.263	-2.043	-1.022	0	-2.043	-2.263	-3.605

Pseudo-weight 10

Position	1	2	3	4	5	6	7	8
A	0.62	0.45	0.25	0.25	0.00	-0.34	-0.34	-0.34
C	-0.34	-0.34	-0.34	-0.34	0.00	0.25	0.00	-0.34
G	-0.34	-0.34	-0.34	0.00	0.00	0.25	0.45	0.62
T	-0.34	0.00	0.25	0.00	0.00	-0.34	-0.34	-0.34
Total	-0.39	-0.221	-0.17	-0.085	0	-0.17	-0.221	-0.39

Pseudo-weight 100

Position	1	2	3	4	5	6	7	8
A	0.11	0.07	0.04	0.04	0.00	-0.04	-0.04	-0.04
C	-0.04	-0.04	-0.04	-0.04	0.00	0.04	0.00	-0.04
G	-0.04	-0.04	-0.04	0.00	0.00	0.04	0.07	0.11
T	-0.04	0.00	0.04	0.00	0.00	-0.04	-0.04	-0.04
Total	-0.008	-0.004	-0.003	-0.001	0	-0.003	-0.004	-0.008

## Unequal priors

Residue	Prior
A	0.3
C	0.2
G	0.2
T	0.3

## Frequency matrices with unequal priors

Pseudo-weight 0

Position	1	2	3	4	5	6	7	8
A	1	0.75	0.5	0.5	0.25	0	0	0
C	0	0	0	0	0.25	0.5	0.25	0
G	0	0	0	0.25	0.25	0.5	0.75	1
T	0	0.25	0.5	0.25	0.25	0	0	0
Total	1	1	1	1	1	1	1	1

Pseudo-weight 1

Position	1	2	3	4	5	6	7	8
A	0.86	0.66	0.46	0.46	0.26	0.06	0.06	0.06
C	0.04	0.04	0.04	0.04	0.24	0.44	0.24	0.04
G	0.04	0.04	0.04	0.24	0.24	0.44	0.64	0.84
T	0.06	0.26	0.46	0.26	0.26	0.06	0.06	0.06
Total	1	1	1	1	1	1	1	1

Pseudo-weight 10

Position	1	2	3	4	5	6	7	8
A	0.500	0.429	0.357	0.357	0.286	0.214	0.214	0.214
C	0.143	0.143	0.143	0.143	0.214	0.286	0.214	0.143
G	0.143	0.143	0.143	0.214	0.214	0.286	0.357	0.429
T	0.214	0.286	0.357	0.286	0.286	0.214	0.214	0.214
Total	1	1	1	1	1	1	1	1

Pseudo-weight 100

Position	1	2	3	4	5	6	7	8
A	0.327	0.317	0.308	0.308	0.298	0.288	0.288	0.288
C	0.192	0.192	0.192	0.192	0.202	0.212	0.202	0.192
G	0.192	0.192	0.192	0.202	0.202	0.212	0.221	0.231
T	0.288	0.298	0.308	0.298	0.298	0.288	0.288	0.288
Total	1	1	1	1	1	1	1	1

## Weight matrices with unequal priors

Pseudo-weight 0

Position	1	2	3	4	5	6	7	8
A	1.20	0.92	0.51	0.51	-0.18	-Inf	-Inf	-Inf
C	-Inf	-Inf	-Inf	-Inf	0.22	0.92	0.22	-Inf
G	-Inf	-Inf	-Inf	0.22	0.22	0.92	1.32	1.61
T	-Inf	-0.18	0.51	-0.18	-0.18	-Inf	-Inf	-Inf
Total	1.204	0.734	1.022	0.552	0.082	1.833	1.545	1.609

Pseudo-weight 1

Position	1	2	3	4	5	6	7	8
A	1.05	0.79	0.43	0.43	-0.14	-1.61	-1.61	-1.61
C	-1.61	-1.61	-1.61	-1.61	0.18	0.79	0.18	-1.61
G	-1.61	-1.61	-1.61	0.18	0.18	0.79	1.16	1.44
T	-1.61	-0.14	0.43	-0.14	-0.14	-1.61	-1.61	-1.61
Total	-3.775	-2.574	-2.364	-1.143	0.078	-1.642	-1.873	-3.393

Pseudo-weight 10

Position	1	2	3	4	5	6	7	8
A	0.51	0.36	0.17	0.17	-0.05	-0.34	-0.34	-0.34
C	-0.34	-0.34	-0.34	-0.34	0.07	0.36	0.07	-0.34
G	-0.34	-0.34	-0.34	0.07	0.07	0.36	0.58	0.76
T	-0.34	-0.05	0.17	-0.05	-0.05	-0.34	-0.34	-0.34
Total	-0.499	-0.365	-0.324	-0.142	0.04	0.04	-0.024	-0.247

Pseudo-weight 100

Position	1	2	3	4	5	6	7	8
A	0.09	0.06	0.03	0.03	-0.01	-0.04	-0.04	-0.04
C	-0.04	-0.04	-0.04	-0.04	0.01	0.06	0.01	-0.04
G	-0.04	-0.04	-0.04	0.01	0.01	0.06	0.10	0.14
T	-0.04	-0.01	0.03	-0.01	-0.01	-0.04	-0.04	-0.04
Total	-0.032	-0.029	-0.028	-0.011	0.006	0.034	0.032	0.025