

Biological sequence analysis

Probabilistic models
of proteins and
nucleic acids

R. Durbin
S. Eddy
A. Krogh
G. Mitchison



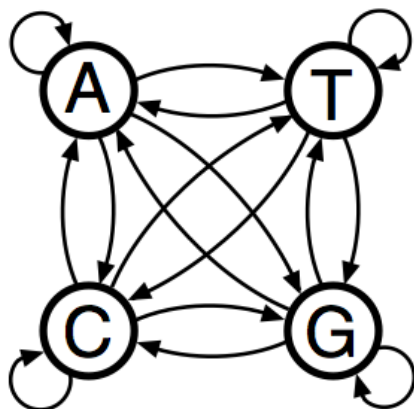
Markov Chains

M+

M-

	A	C	G	T
A	0.180	0.274	0.426	0.120
C	0.171	0.368	0.274	0.188
G	0.161	0.339	0.375	0.125
T	0.079	0.355	0.384	0.182

	A	C	G	T
A	0.300	0.205	0.285	0.210
C	0.322	0.298	0.078	0.302
G	0.248	0.246	0.298	0.208
T	0.177	0.239	0.292	0.292



β	A	C	G	T
A	-0.740	0.419	0.580	-0.803
C	-0.913	0.302	1.812	-0.685
G	-0.624	0.461	0.331	-0.730
T	-1.169	0.573	0.393	-0.679

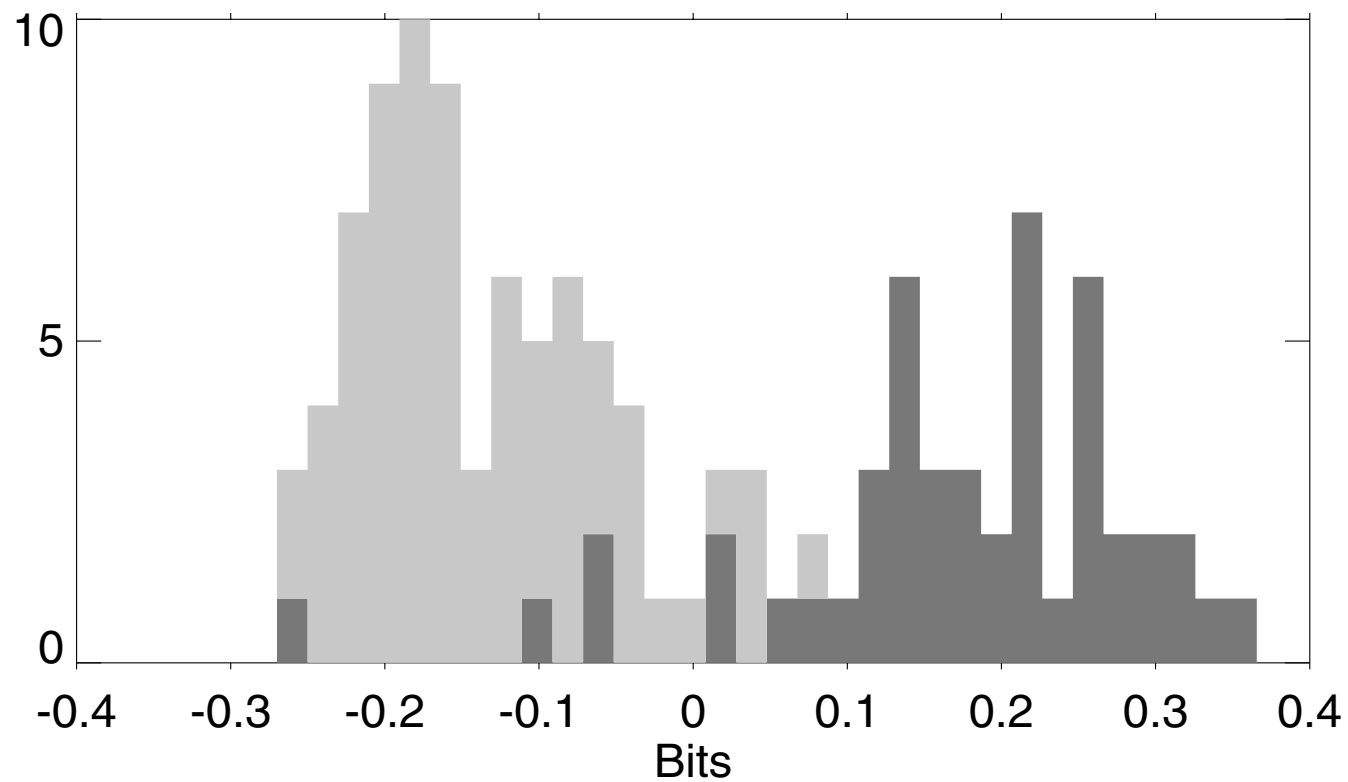


Figure 3.2 *The histogram of the length-normalised scores for all the sequences. CpG islands are shown with dark grey and non-CpG with light grey.*

Loaded/Fair Coin Hidden Markov Models (HMM)

000--0--00--0000000000000000-0---00--0--0-00-00--0-000-000000--0-----

-00--0000-00-0-00000000000000-0000-0000000000000000000000000000-00000-0-00

-----o-----ooooooooooooooooooooooooooooooooooooo-----oooooooooooooo--o---o-----oooooooooooooooooooooooo-----

Loaded/Fair Coin Hidden Markov Models (HMM)

```

FFFFFFFFFFFFFFFFLLLLLLLLLLLLFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
000--0--00--0000000000000000--0---00--0--0-00-00--0-000-000000--0-----

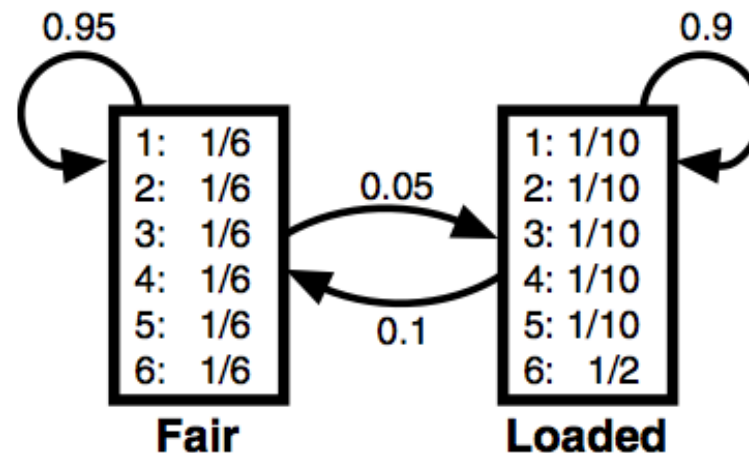
```

```
FLLFFFFFFFFFFFFFFFFLLLLLLLLLLLLFFFFFFFFLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLL
-00--0000-00-0-0000000000000-0000-00000000000000000000000000000000000000000-00000-0-00
```

[illegible]

Occasionally Dishonest Casino

2265154453325541636245212164633212434633644145452554564641642462532533

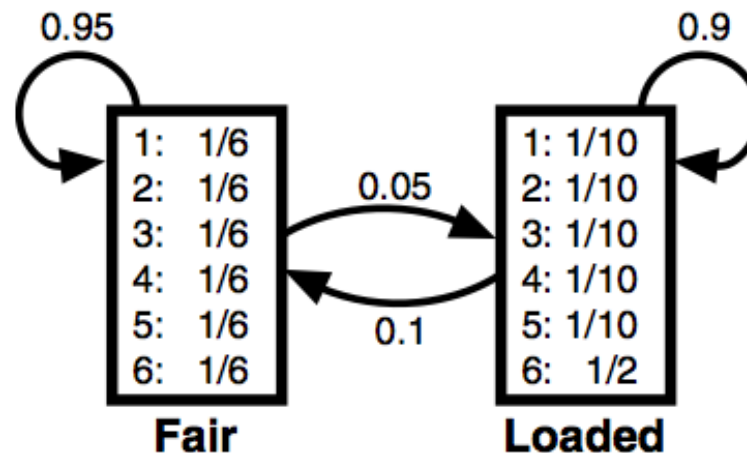




Andrew Viterbi, 1935 -

Occasionally Dishonest Casino

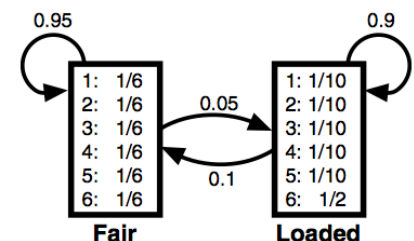
FFFFFFFFFFFFFFFFLLFFFFFFFFFFF
2265154453325541636245212164633212434633644145452554564641642462532533



Occasionally Dishonest Casino

Rolls	315116246446644245311321631164152133625144543631656626566666
Die	FFL
Viterbi	FFL
Rolls	651166453132651245636664631636663162326455236266666625151631
Die	LLLLLLFF
Viterbi	LLLLLLFF
Rolls	222555441666566563564324364131513465146353411126414626253356
Die	FFLL
Viterbi	FFL
Rolls	366163666466232534413661661163252562462255265252266435353336
Die	LLLLLLLLFF
Viterbi	LLLLLLLLLLLLLFFF
Rolls	233121625364414432335163243633665562466662632666612355245242
Die	FF
Viterbi	FF

Figure 3.5 The numbers show 300 rolls of a die as described in the example. Below is shown which die was actually used for that roll (F for fair and L for loaded). Under that the prediction by the Viterbi algorithm is shown.



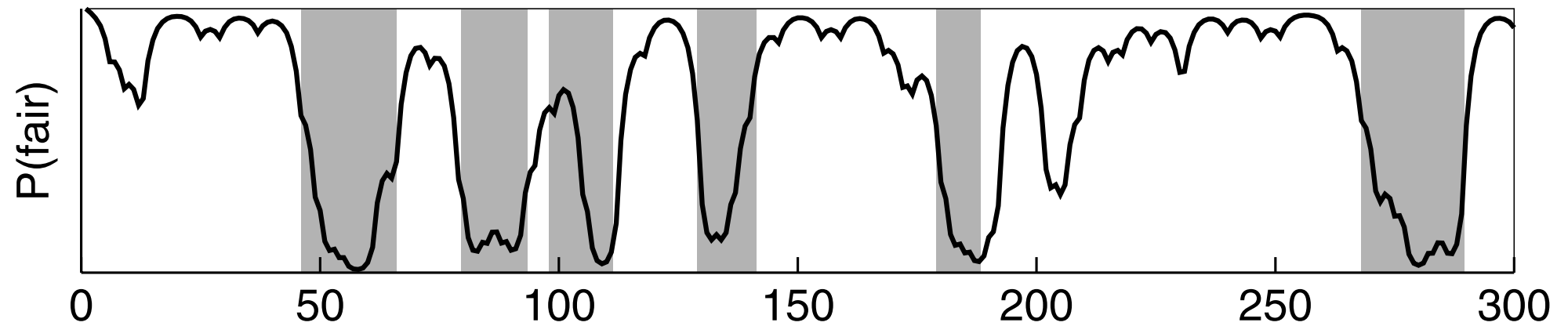


Figure 3.6 *The posterior probability of being in the state corresponding to the fair die in the casino example. The x axis shows the number of the roll. The shaded areas show when the roll was generated by the loaded die.*