

Prediction of CpG island

We're going to calculate the probability that the hidden state is CpG island (I) at every position of the given DNA sequence **TTACGCGCGCGCGATATTT**.

```
library(HMM)

dnaSeq <- unlist(strsplit(x="TTACGCGCGCGCGATATTT", ""))
nucleotides <- c("A", "C", "G", "T")

# I is for CpG-island
# N is for non-CpG-island
states <- c("I", "N")
transProbs <- matrix(c(0.8, 0.2, 0.3, 0.7), c(length(states), length(states)),
                     byrow=TRUE)
emissionProbs <- matrix(c(0.1, 0.4, 0.4, 0.1, 0.3, 0.2, 0.2, 0.3),
                        c(length(states), length(nucleotides)), byrow=TRUE)
hmm = initHMM(states, nucleotides, startProbs=c(0.4,0.6), transProbs=transProbs,
              emissionProbs=emissionProbs)

p <- posterior(hmm, dnaSeq)
p <- as.matrix(cbind(dnaSeq, p[1,]))
colnames(p) <- c("State", "P(I)")
p
```

```
##      State P(I)
## 1  "T"      "0.102902143356443"
## 2  "T"      "0.138290083344347"
## 3  "A"      "0.263987824663092"
## 4  "C"      "0.681949738364792"
## 5  "G"      "0.823554781079921"
## 6  "C"      "0.871473457910994"
## 7  "G"      "0.887520861970078"
## 8  "C"      "0.892398144493408"
## 9  "G"      "0.892396999936075"
## 10 "C"      "0.88751595181912"
## 11 "G"      "0.871458448071672"
## 12 "C"      "0.823510308859526"
## 13 "G"      "0.681818434599012"
## 14 "A"      "0.263600307494064"
## 15 "T"      "0.137512486212712"
## 16 "A"      "0.100973871229589"
## 17 "T"      "0.0952790878051435"
## 18 "T"      "0.111278517237432"
## 19 "T"      "0.174677909480996"
```