Analysis of HMM-state Models

Ricky Lim

May 1, 2013

```
Filename: Analysis_StateModel.Rnw
Working directory: /TEMP_DDN/users/gfilion/rlim/H1_ColorChromatin/ColorStates
```

1 Goal

- Compare the stability of 4state and 5state-HMM models in the first five random runs

2 Load datasets

```r
LogLik_4state <- read.delim("outputLogLik_4states/loglik_4states.ls")
LogLik_5state <- read.delim("outputLogLik_5states/loglik_5states.ls")

# get the first 5 runs
state4_top5 <- as.numeric(LogLik_4state)[1:5]
state5_top5 <- as.numeric(LogLik_5state)[1:5]

box_state <- rbind(state4_top5, state5_top5)
box_state <- t(box_state)

# draw the histograms
h <- hist(box_state, plot = FALSE)

# color each bin according to the box_state(4-state or 5-state)
latest.ob_4state <- box_state[, 1]
latest.ob_5state <- box_state[, 2]
bin_4state <- cut(latest.ob_4state, h$breaks)
bin_5state <- cut(latest.ob_5state, h$breaks)
clr <- rep("white", length(h$counts))
clr[bin_4state] <- "green"
clr[bin_5state] <- "red"

# producing the figure
pdf("figs/log_likelihood.pdf", useDingbats = FALSE)
plot(h, col = clr, xlab = "", main = "Profile Log-Likelihood")
legend("topright", fill = c("green", "red"), c("Four-state HMM", "Five-state HMM"), bg = "transparent",
    box.lwd = 0, box.col = "transparent")
dev.off()

## pdf
##   2
```

2.1 Comparison

3 Compare State Assignment

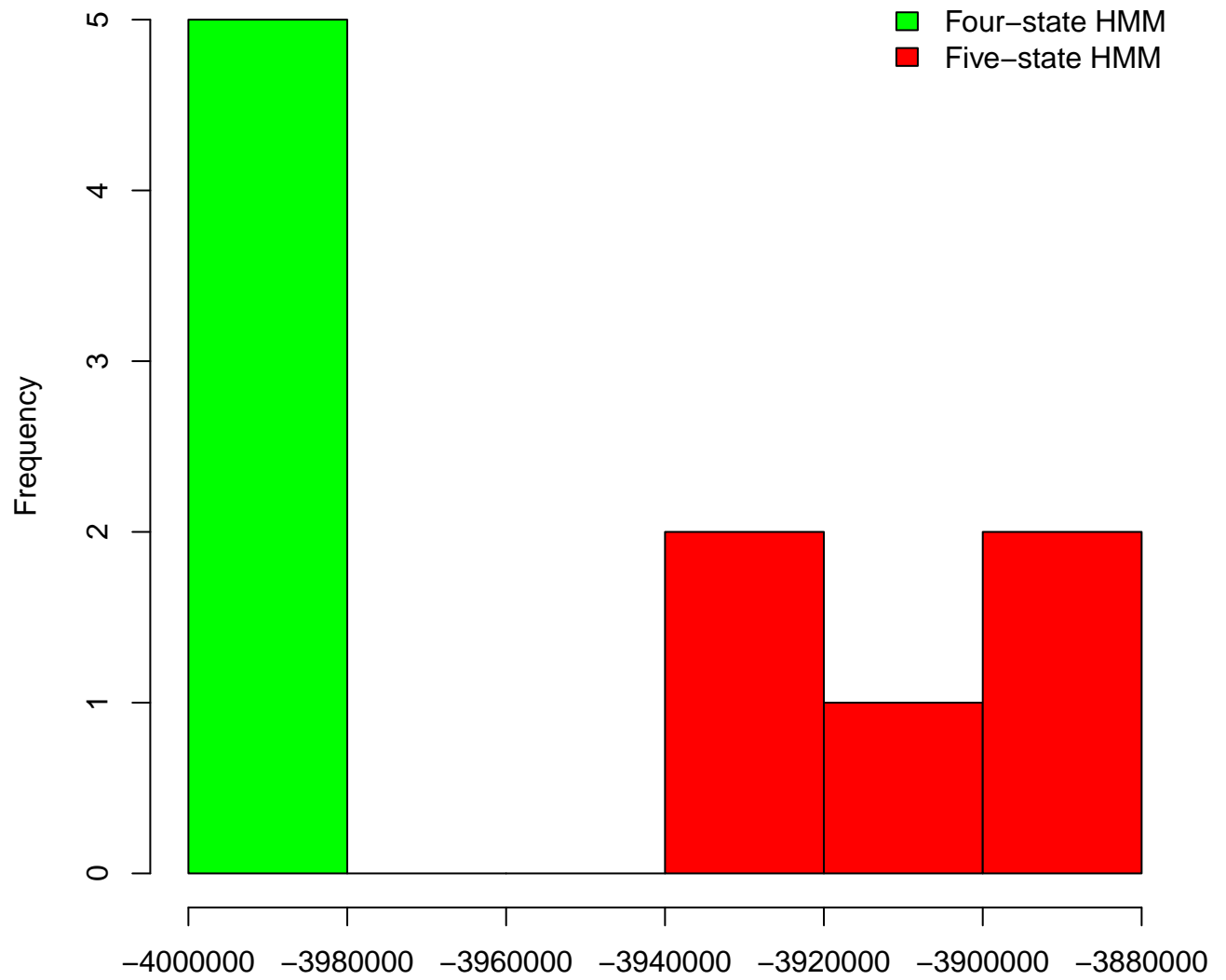## 3.1 State Assignment in Four-state HMM

	fit41366702441	fit41366702570	fit41366702728	fit41366702873	fit41366702987
0	130214	71342	25029	805264	71369
1	25029	25029	71354	25029	805276
2	805296	805295	130232	71359	25039
3	71345	130218	805269	130232	130200

Table 1: Four-state HMM

	fit41366703076	fit41366703192	fit41366703253
0	189195	25029	71354
1	33869	130232	130232
2	773274	71359	25029
3	35546	805264	805269

Table 2: Four-state HMM

Profile Log−Likelihood

3.2 State Assignment in Five-state HMM

	fit51366732443	fit51366733163	fit51366738904	fit51366739827	fit51366740470
0	58055	22846	70528	33095	88321
1	798014	798562	79293	758764	33030
2	70762	53177	797769	131594	18897
3	26215	27718	26204	18944	760016
4	78838	129581	58090	89487	131620

Table 3: Five-state HMM

	fit51366741858	fit51366741922	fit51366742221	fit51366742307	fit51366742484
0	131620	760016	18899	760016	88321
1	88321	131620	88319	131620	18898
2	18897	18897	131620	88321	131620
3	33030	33030	760016	33029	760016
4	760016	88321	33030	18898	33029

Table 4: Five-state HMM

	fit51366742656	fit51366743106
0	88321	18898
1	33030	131620
2	760016	760016
3	131620	33029
4	18897	88321

Table 5: Five-state HMM

3.3   State Assignment Comparison for Top Five

	fit41366702441	fit41366702570	fit41366702728	fit41366702873	fit41366702987
0	130214	71342	25029	805264	71369
1	25029	25029	71354	25029	805276
2	805296	805295	130232	71359	25039
3	71345	130218	805269	130232	130200

	fit51366732443	fit51366733163	fit51366738904	fit51366739827	fit51366740470
0	58055	22846	70528	33095	88321
1	798014	798562	79293	758764	33030
2	70762	53177	797769	131594	18897
3	26215	27718	26204	18944	760016
4	78838	129581	58090	89487	131620

Figure above (Profile Log-Likelihood) indicates that five-state model yields higher log-likelihood values than four-state model although these values vary over the course of random runs. This variation is reflected by the state assignment of the five-state. As shown in tables above, the number of each state assignment in five-state model reflects fluctuations over different random runs. These, altogether, suggest that five-state HMM model produces less-stable state assignment. For this reason, we selected four-state HMM model that yields more stable state assignment in segmenting H1-hESC genome.

4 Metainfo

```
sessionInfo()
```

```
## R version 2.15.0 (2012-03-30)
## Platform: x86_64-unknown-linux-gnu (64-bit)
##
## locale:
## [1] C
##
## attached base packages:
## [1] stats     graphics  grDevices datasets  utils     methods   base
##
## other attached packages:
## [1] codetools_0.2-8  knitr_1.2        xtable_1.7-0      vimcom_0.9-8
## [5] setwidth_1.0-3   cacheSweave_0.6-1 stashR_0.3-5     filehash_2.2-1
##
## loaded via a namespace (and not attached):
## [1] digest_0.5.2   evaluate_0.4.3 formatR_0.7   plyr_1.7.1    stringr_0.6
## [6] tools_2.15.0
```