Q5 – Markov Chain

From textbook we can find that 1st order markov chain’s Goodness Of Fit is 3.12(df=8) and AIC is 64.39

> 1-pchisq(3.12,df=8)

[1] 0.9266026

And its p-value is 0.9266026

Proceed the 2nd order markov chain model

> setwd('C:\\Users\\ahn92\\Documents\\dev\\Sources\\02\_Lecture\_19\_1\\01\_Intermediate\_Statistics\\hw2\\Q5-Markov')

> dat = read.csv('illness.csv', encoding = 'latin1')

> head(dat)

i..no yes t.1 t.2 s t

1 261 13 0 0 0 10

2 22 4 0 1 0 10

3 18 6 1 0 0 10

4 12 14 1 1 0 10

5 125 9 0 0 1 10

6 15 3 0 1 1 10

> colnames(dat)=c("no" ,"yes" ,"t.1" ,"t.2" , "s" ,"t" )

> fit2 = glm(cbind(yes,no) ~ t.1 + t.2 + s+t ,family = binomial, data = dat)

> summary(fit2)

Call:

glm(formula = cbind(yes, no) ~ t.1 + t.2 + s + t, family = binomial,

data = dat)

Deviance Residuals:

Min 1Q Median 3Q Max

-0.68041 -0.25801 0.01553 0.27960 0.59014

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 1.3736 1.9006 0.723 0.4698

t.1 1.9452 0.2152 9.040 < 2e-16 \*\*\*

t.2 1.1481 0.2231 5.146 2.66e-07 \*\*\*

s 0.1741 0.2044 0.852 0.3944

t -0.4366 0.2017 -2.164 0.0304 \*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 173.6281 on 15 degrees of freedom

Residual deviance: 2.3424 on 11 degrees of freedom

AIC: 69.054

Number of Fisher Scoring iterations: 4

We can get the result and t.2 which is the 2nd order term has very low p-value(2.66e-07), AIC is 69.054 and its GOF is 2.3424(df=11) p-value is

> 1-pchisq(2.3424,df=11)

[1] 0.9968856

As 1st order markov chain and 2nd order markov chain’s p-value > 0.05, we cannot reject both model, but by AIC score 1st order markov chain is better than 2nd order markov chain model