

# assignment6\_EEOB563

*Devin Molnau*

*April 2, 2019*

## Assignment 6 - EEOB 563 - Spring 2019

### PART 1:

#### Question 1:

Since the original situation has three doors that are equally likely, the prior probability can be set to  $1/3$  for Door 1, Door 2, and Door 3. When the third door is open, we now know that the likelihood that that has the car behind it is 0. Now one of the doors DOES have the car behind it which makes its likelihood 1 while the other door have a likelihood of  $1/2$ . Now, if we selected Door 1 originally, and Door 3 was opened, then there is  $1/2$  likelihood that the car is behind door number 1. To calculate the posterior probability it would be the prior \* likelihood / the sum of the prior and likelihoods. The sum of the likelihood would be  $(1/3 \times 1/2) + (1/3 \times 1)$  which will be the denominator. So the posterior probability of Door 1 is  $(1/2 \times 1/3) / (1/2 + 1/3) = 1/5$  and the posterior probability that the car is behind Door 2 is  $(1 \times 1/3) / (1/2 + 1/3) = 2/5$  and of course the posterior probability of Door number 3 is 0. So, it would be advantageous to switch doors and go with Door 2.

### PART 2:

#### Question 2:

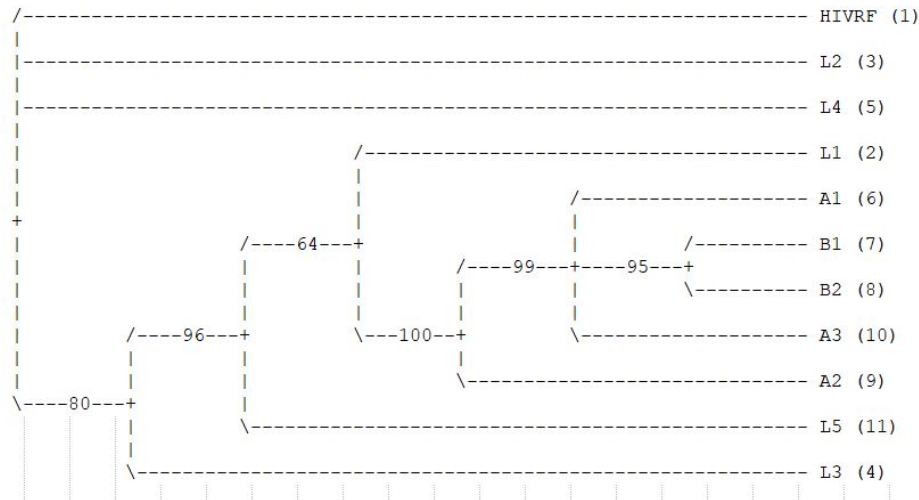
To determine the probability of defendant A and complainant B's HIV sequence forming a clade, I ran a MCMC with a mixed gamma model for 2,000,000 generations, with 4 chains and sampled every 100 generations.

```
module unload /opt/rit/spack-modules/lmod/linux-rhel7-x86_64/Core
module use /opt/rit/modules
```

```
screen -S mrbayes
module load mrbayes
mb
log start filename=hiv-rjmcmm-log.txt
execute hiv.nxs
outgroup HIVRF
lset nst=mixed rates=gamma
showmodel
mcmc data=no nruns=1 nchains=1
mcmc ngen=2000000 printfreq=100 samplefreq=100
mcmc filename=hiv-rjmcmm-prior
mcmc data=yes nruns=2 nchains=4 savebrlens=yes
mcmc ngen=2000000 printfreq=100 samplefreq=100
mcmc diagnfreq=1000 diagnstat=maxstddev
mcmc filename=hiv-rjmcmm
sump filename=hiv-rjmcmm
sumt filename=hiv-rjmcmm
quit
```

Based on the confidence tree generated after 2,000,000 iterations of MCMC, there is a 100% chance that the A and B form a clade.

Clade credibility values:



Calculating tree probabilities...

Credible sets of trees (496 trees sampled):

50 % credible set contains 10 trees  
 90 % credible set contains 40 trees  
 95 % credible set contains 90 trees  
 99 % credible set contains 257 trees

### Question 3:

Model probabilities above 0.050  
 Estimates saved to file "hiv-rjmc.mstat".

Model	Posterior Probability	Standard Deviation	Min. Probability	Max. Probability
gtrsubmodel[112212]	0.138	0.003	0.135	0.140
gtrsubmodel[112312]	0.086	0.002	0.084	0.087
gtrsubmodel[112313]	0.067	0.001	0.066	0.068
gtrsubmodel[123323]	0.062	0.000	0.062	0.063
gtrsubmodel[112213]	0.053	0.001	0.052	0.053

The most likely model is gtrsubmodel[112212] with a probability of 0.138 with a standard deviation of 0.003.

### Question 4:

To make sure that the MCMC simulations convergence to equilibrium, I look at the standard deviation of the mcmc after 2,000,000 iterations. The MaxStdDev after 2,000,000 iterations is 8.097e-3 which I would accept as reasonable.

Below are the Arithmetic and harmonic means of the 2 runs.

Run	Arithmetic mean	Harmonic mean
1	-1547.14	-1567.80
2	-1547.22	-1564.66
TOTAL	-1547.18	-1567.15

The length of the MCMC simulation that I ran was 2,000,000 generations, with a print and sampling frequency of 100 generations and 4 chains. Below are the paramters after the runs.

Model parameter summaries over the runs sampled in files  
 "hiv-rjmcmm.run1.p" and "hiv-rjmcmm.run2.p":  
 Summaries are based on a total of 30002 samples from 2 runs.  
 Each run produced 20001 samples of which 15001 samples were included.  
 Parameter summaries saved to file "hiv-rjmcmm.pstat".

Parameter	Mean	Variance	95% HPD Interval		Median	min ESS*	avg ESS	PSRF+
			Lower	Upper				
TL	0.257648	0.000775	0.205966	0.313751	0.255980	9944.57	10919.11	1.000
r(A<->C)	0.235618	0.001998	0.133110	0.296243	0.248393	1632.18	1912.04	1.000
r(A<->G)	0.267560	0.001156	0.199037	0.339739	0.268619	2973.97	3524.12	1.000
r(A<->T)	0.054816	0.000309	0.021079	0.088639	0.054370	3940.17	4470.05	1.000
r(C<->G)	0.096715	0.003008	0.036082	0.228729	0.075372	1008.77	1202.84	1.000
r(C<->T)	0.278719	0.001660	0.199976	0.370772	0.273837	2977.75	3751.90	1.000
r(G<->T)	0.066571	0.000769	0.021186	0.120713	0.061774	2492.24	2792.19	1.000
k_revmat	3.332178	0.712360	2.000000	5.000000	3.000000	3999.05	4145.56	1.000
pi(A)	0.401025	0.000321	0.365345	0.435052	0.400984	4200.06	4335.20	1.000
pi(C)	0.155612	0.000164	0.129876	0.180248	0.155308	4837.33	5003.02	1.000
pi(G)	0.212751	0.000218	0.184012	0.241704	0.212490	5193.93	5276.03	1.000
pi(T)	0.230612	0.000251	0.199667	0.261885	0.230443	4325.00	4398.91	1.000
alpha	0.719528	0.188382	0.185493	1.522572	0.609479	7791.34	7953.17	1.000