LAB 1 200A

Ioana Anghel

October 10, 2017

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

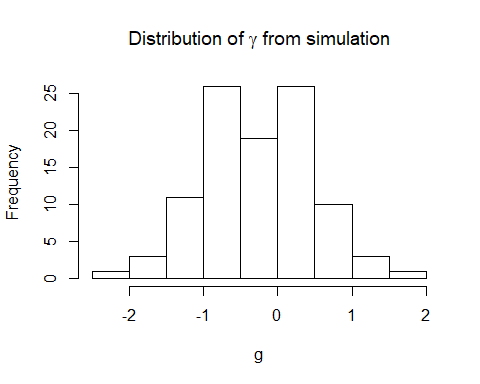
## 1.

library(phytools)

## Loading required package: ape

## Loading required package: maps

setwd("~/GitHub/200A/")  
  
snake.tree<-read.tree("homalops.phy")  
  
h<-max(nodeHeights(snake.tree))  
x<-seq(0,h,by=h/100)  
b<-(log(Ntip(snake.tree))-log(2))/h  
  
trees<-pbtree(b=b,n=Ntip(snake.tree),t=h,nsim=100,method="direct",  
 quiet=TRUE)  
g<-sapply(trees,function(x) ltt(x,plot=FALSE)$gamma)  
hist(g,main=expression(paste("Distribution of ",gamma," from simulation")))



mean(g)

## [1] -0.2621326

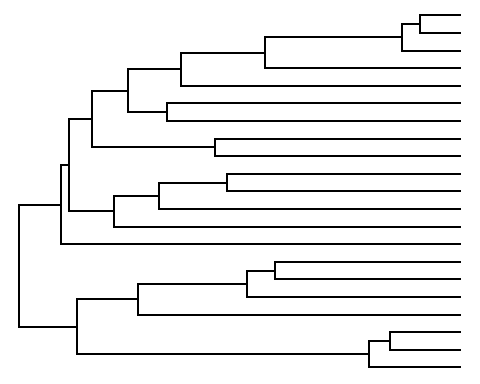
var(g)

## [1] 0.5440225

#We can test hypotheses about ??. This is done automatically with ltt. For example  
  
obj<-ltt(snake.tree,plot=FALSE)  
print(obj)

## Object of class "ltt" containing:  
##   
## (1) A phylogenetic tree with 21 tips and 20 internal nodes.  
##   
## (2) Vectors containing the number of lineages (ltt) and branching times (times) on the tree.  
##   
## (3) A value for Pybus & Harvey's "gamma" statistic of -3.2411, p-value = 0.0012.

plotTree(snake.tree,ftype="off")

 #2. Given this gamma value, what would you conclude about the tempo of speciation in this clade?

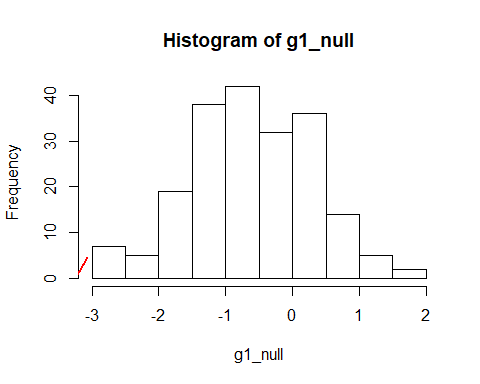
This gamma value is negative and indicates that speciation events are closer to the roots of the tree than expected if the speciation rate is constant through time. Perhaps early history diversifcation was rapid and subsequently slowed down.

# 3.

library(geiger)  
age <- 22  
richness <- 34  
snakebirth <- (log(richness) - log(2))/age  
snakebirth

## [1] 0.1287824

richness <- 34  
missing <- 13  
#this simulates gamma values when trees are undersampled.  
#we will grow trees with n=34 and prune them down to 21 taxa  
  
snake.gamma <- obj$gamma  
  
num\_simulations<-200 #number of simulations  
g1\_null<-numeric(num\_simulations) #g1\_null will hold the simulated gamma values  
for(i in 1:num\_simulations) {  
 sim.bdtree(snakebirth, d=0, stop = "taxa", n=richness)->sim\_tree   
 drop.random(sim\_tree, missing)->prune # prune down to the # of taxa in the phylogeny  
 gammaStat(prune)->g1\_null[i]  
}  
# create a histogram of the null distribution  
hist(g1\_null)  
  
#arrow indicates where the observed gamma falls in the null you just generated  
arrows(snake.gamma, 40, snake.gamma, 0, col="red", lwd=2)



# Which of the null values are smaller (more negative) than the data?  
smallerNull<-g1\_null<=snake.gamma  
# How many TRUEs are there?  
count<-sum(smallerNull)  
  
# finally, what is the p-value?  
mccr\_pval<-(count+1)/(num\_simulations+1)  
mccr\_pval

## [1] 0.004975124

For MCCR test to suggest rapid initial diversification, we need P < 0.001. The p-value was higher, therefore we conclude that apparent rapid initial diversification may have been due to incomplete sampling. The gamma value does not fall in the tails of the distribution, so the tree could have been produced by a tree evolving under a constant rate that was incompletey sampled.

# 4

library(phytools)  
fitbd <- birthdeath(snake.tree)  
fitbd

##   
## Estimation of Speciation and Extinction Rates  
## with Birth-Death Models  
##   
## Phylogenetic tree: snake.tree   
## Number of tips: 21   
## Deviance: 55.26213   
## Log-likelihood: -27.63107   
## Parameter estimates:  
## d / b = 0 StdErr = 0   
## b - d = 0.06839495 StdErr = 0.01111134   
## (b: speciation rate, d: extinction rate)  
## Profile likelihood 95% confidence intervals:  
## d / b: [-1.977961, 0.3422611]  
## b - d: [0.04207165, 0.1039128]

bd(fitbd)

## b d   
## 0.06839495 0.00000000

# 5

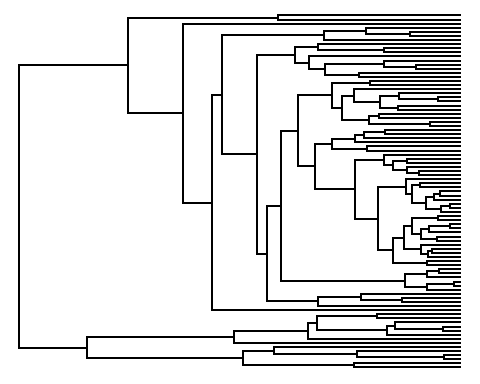
plant.tree<-read.nexus("T70460.nex")  
  
obj<-ltt(plant.tree,plot=FALSE)  
print(obj)

## Object of class "ltt" containing:  
##   
## (1) A phylogenetic tree with 87 tips and 86 internal nodes.  
##   
## (2) Vectors containing the number of lineages (ltt) and branching times (times) on the tree.  
##   
## (3) A value for Pybus & Harvey's "gamma" statistic of -1.7379, p-value = 0.0822.

# 5.(1)

# The study used 53 species from Salvia subgenus Calosphace and 17 outgroups. There are 87 tips in this phylogeny (Benitez-Vieyra et al., 2014 - <http://rspb.royalsocietypublishing.org/content/281/1782/20132934#ref-24>) The Calosphace subgenus has a total of 468 species. (<https://link.springer.com/content/pdf/10.1007/s00606-017-1445-4.pdf>)

plotTree(plant.tree,ftype="off")



# 5.(2)

library(phytools)  
fitbd <- birthdeath(plant.tree)  
fitbd

##   
## Estimation of Speciation and Extinction Rates  
## with Birth-Death Models  
##   
## Phylogenetic tree: plant.tree   
## Number of tips: 87   
## Deviance: -1219.352   
## Log-likelihood: 609.6758   
## Parameter estimates:  
## d / b = 0 StdErr = 0   
## b - d = 103.6111 StdErr = 7.947405   
## (b: speciation rate, d: extinction rate)  
## Profile likelihood 95% confidence intervals:  
## d / b: [-0.634676, 0.2130442]  
## b - d: [83.12069, 127.2206]

bd(fitbd)

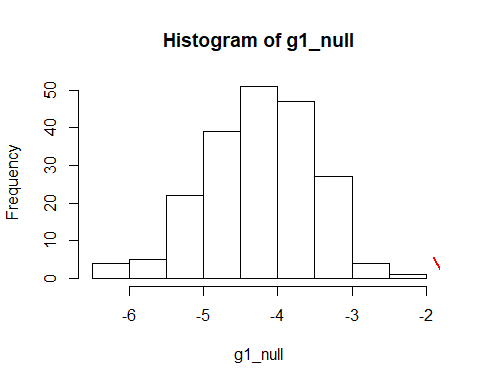
## b d   
## 103.6111 0.0000

# 5.(3)

library(geiger)  
age <- 12  
richness <- 468  
plantbirth <- (log(richness) - log(2))/age  
plantbirth

## [1] 0.4546101

richness <- 468  
missing <- 398  
  
plant.gamma <- obj$gamma  
  
num\_simulations<-200 #number of simulations  
g1\_null<-numeric(num\_simulations) #g1\_null will hold the simulated gamma values  
for(i in 1:num\_simulations) {  
 sim.bdtree(plantbirth, d=0, stop = "taxa", n=richness)->sim\_tree   
 drop.random(sim\_tree, missing)->prune # prune down to the # of taxa in the phylogeny  
 gammaStat(prune)->g1\_null[i]  
}  
# create a histogram of the null distribution  
hist(g1\_null)  
  
#arrow indicates where the observed gamma falls in the null you just generated  
arrows(plant.gamma, 40, plant.gamma, 0, col="red", lwd=2)



smallerNull<-g1\_null<=plant.gamma  
# How many TRUEs are there?  
count<-sum(smallerNull)  
  
# finally, what is the p-value?  
mccr\_pval<-(count+1)/(num\_simulations+1)  
mccr\_pval

## [1] 1

p-value is extremely high. Though the gamma is in the range of possibilities, gamma is not significant when taking into account the incomplete sampling. This was to be expected because the study had such few species.