

Gompertz Mixture fit of lifespan of WT BY4742, BY4741 genetic backgrounds and single deletion mutants

Emine Guven

10 May 2021

```
library('varhandle')
library('flexsurv')
```

```
## Loading required package: survival
```

```
library('stringr')
library("base")
library("fitmix")
```

load the lifespan data previously generated from “rls 2016-11-29.csv”

cooked lifespans previously using “rls 2016-08-26.csv” and write outputs to the “update-dRLS_05052021.csv”

```
lifespan_data<-read.csv("updatedRLS_05052021.csv")
names(lifespan_data)
```

```
## [1] "X"           "genotype"    "mat"         "media"
## [5] "temp"        "n"           "avgLS"       "StddevLS"
## [9] "medianLS"    "single_lifespan"
```

mixture plot function

```
plot.gompertz.components <- function(mixture,component.number,...) {
  curve(mixture$lambda[component.number] *
    dgomperz(x,shape=mixture$shape[component.number],
      rate=mixture$rate[component.number]), add=TRUE, col="darkred")}
```

BY4742

```
#BY4742 plots
WT.BY4742<- lifespan_data[lifespan_data$genotype=="BY4742",]
#no need to filter temp==30 since I did filtering in the beginning
#WT.BY4742.temp<-WT.BY4742[WT.BY4742$temp==30,]
WT.BY4742.media<-WT.BY4742[WT.BY4742$media=="YPD",]

WT.BY4742.media<-WT.BY4742.media[WT.BY4742.media$mat=="MATalpha",]
```

```
WT.BY4742.media= WT.BY4742.media[!is.na(WT.BY4742.media[,1]), ]
dim(WT.BY4742.media)
```

```
## [1] 6 10
```

```
class(WT.BY4742.media$single_lifespan)
```

```
## [1] "factor"
```

```
str(WT.BY4742.media$single_lifespan)
```

```
## Factor w/ 20222 levels "", "1", "1", "1", "1", "1", "1", "1", "...: 5955 17058 3397 2648 16177 6000
```

BY4742 individuals

```
BY4742_lifespan_list=list()
```

```
for (k in c(1:length(WT.BY4742.media$single_lifespan))){
  # k=c(94:95)
  fk<-WT.BY4742.media$single_lifespan[k]
  #f<-unfactor(fk[k])
  BY4742_lifespan_single<-as.numeric(unlist(str_split(fk, ",")))
  #BY4742_lifespan_list<-mean(ref_lifespan_single)
  BY4742_lifespan_list[[length(BY4742_lifespan_list)+1]]<-BY4742_lifespan_single
}
```

```
length(BY4742_lifespan_list)
```

```
## [1] 6
```

```
cumulative_BY4742_lifespan<-unlist(BY4742_lifespan_list)
length(cumulative_BY4742_lifespan)
```

```
## [1] 410
```

```
cumulative_BY4742_vector<-na.omit(cumulative_BY4742_lifespan)
length(cumulative_BY4742_vector)
```

```
## [1] 410
```

mixture BY4742

```
cumulative_BY4742_mixGompertz<-fitmixEM(cumulative_BY4742_vector,"gompertz", 2, initial=FALSE)
```

```
## original BY4742
```

```
print(cumulative_BY4742_mixGompertz$estimate)
```

```
##          weight      alpha      beta
## [1,] 0.4879878 0.1711092 0.0044341969
## [2,] 0.5120122 0.1342279 0.0005399596
```

```
print(cumulative_BY4742_mixGompertz$measures)
```

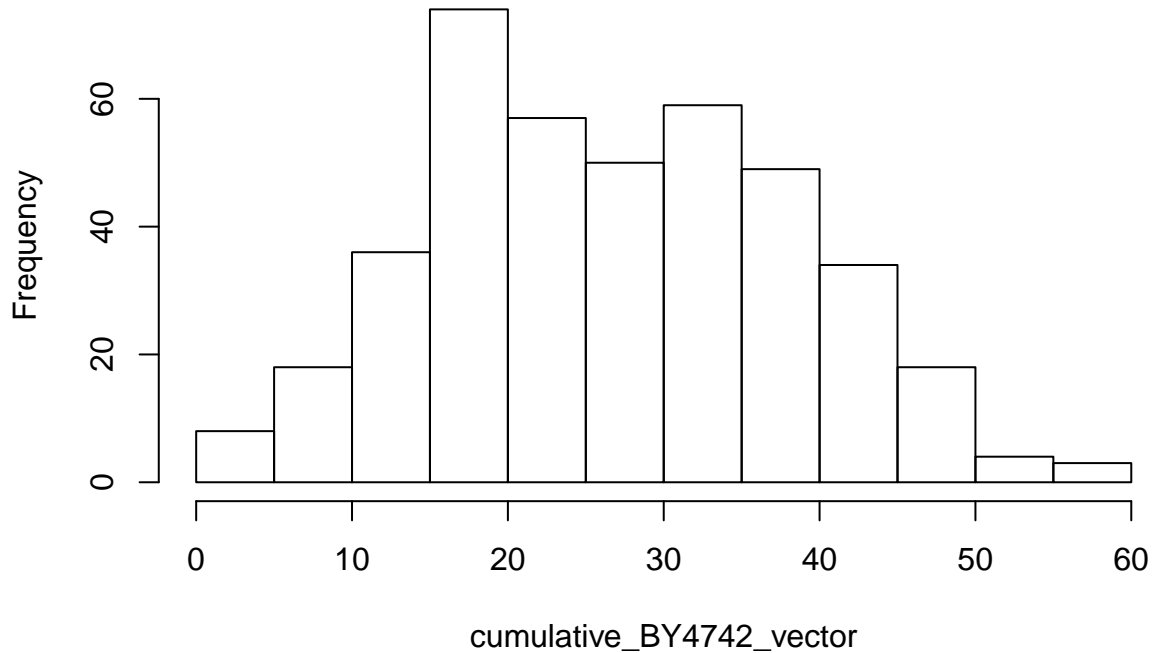
```
##          AIC      BIC      KS log.likelihood
## [1,] 3199.059 3219.14 0.09490343      -1594.53
```

BY4742 mixture plot

```
dfBY4742=data.frame(lambda=c(cumulative_BY4742_mixGompertz$estimate[,1]),
  shape=c(cumulative_BY4742_mixGompertz$estimate[,2]),
  rate=c(cumulative_BY4742_mixGompertz$estimate[,3]))

plot(hist(cumulative_BY4742_vector),col="grey",border="white",freq=FALSE,ylim=c(0,0.08))
```

Histogram of cumulative_BY4742_vector

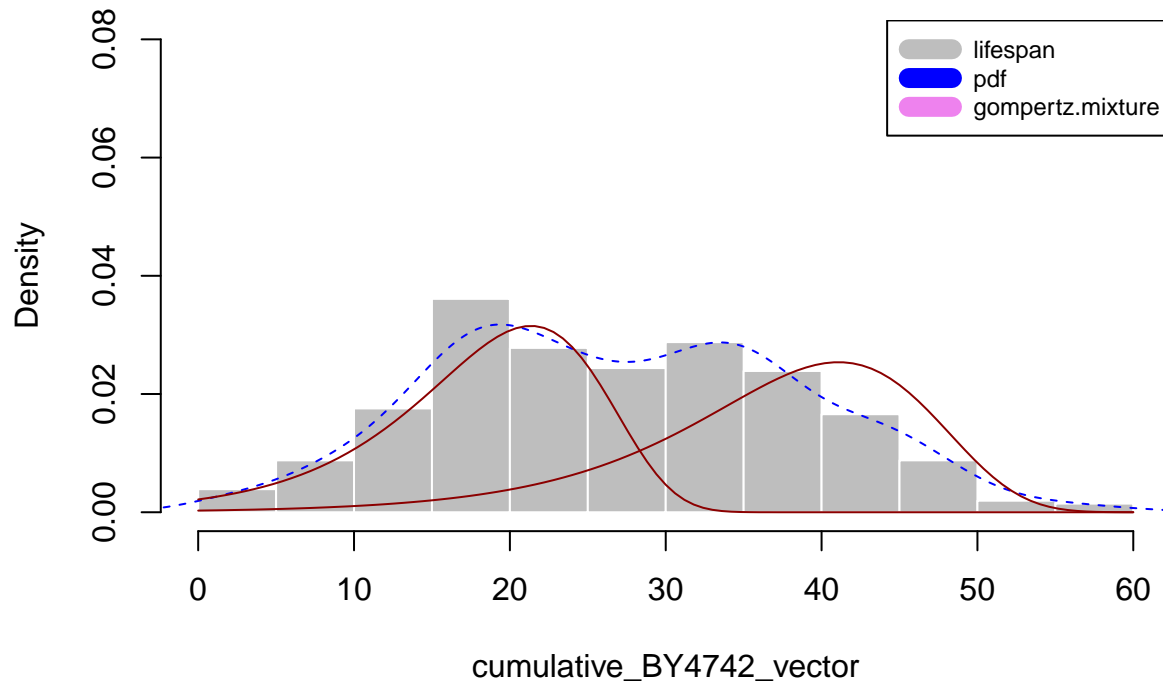


```
#pdf("plots/mixture_fits.pdf")
lines(density(cumulative_BY4742_vector),lty=2,col="blue")
sapply(1:2,plot.gompertz.components,mixture=dfBY4742)

##      [,1]      [,2]
## x Numeric,101 Numeric,101
## y Numeric,101 Numeric,101

legend("topright", c("lifespan", "pdf","gompertz.mixture"),
  col=c("grey","blue","violet"), lwd=10,cex=0.75)
```

Histogram of cumulative_BY4742_vector



BY4741

```
WT.BY4741<- lifespan_data[lifespan_data$genotype=="BY4741",]
WT.BY4741.media<-WT.BY4741[WT.BY4741$media=="YPD",]
WT.BY4741.media<-WT.BY4741.media[WT.BY4741.media$mat=="MATa",]
WT.BY4741.media= WT.BY4741.media[!is.na(WT.BY4741.media[,1]), ]
dim(WT.BY4741.media)

## [1] 1 10

class(WT.BY4741.media$single_lifespan)

## [1] "factor"

str(WT.BY4741.media$single_lifespan)

## Factor w/ 20222 levels "", "1, 1", "1, 1, 1, 1, 1, 1", ...: 11412

### BY4741 individuals
BY4741_lifespan_list=list()

for (k in c(1:length(WT.BY4741.media$single_lifespan))){
  # k=c(94:95)
  fk<-WT.BY4741.media$single_lifespan[k]
  #f<-unfactor(fk[k])
  BY4741_lifespan_single<-as.numeric(unlist(str_split(fk, ",")))
  #BY4741_lifespan_list<-mean(ref_lifespan_single)
  BY4741_lifespan_list[[length(BY4741_lifespan_list)+1]]<-BY4741_lifespan_single
}
```

```
length(BY4741_lifespan_list)
```

```
## [1] 1
```

```
cumulative_BY4741_lifespan<-unlist(BY4741_lifespan_list)  
length(cumulative_BY4741_lifespan)
```

```
## [1] 60
```

```
cumulative_BY4741_vector<-na.omit(cumulative_BY4741_lifespan)  
length(cumulative_BY4741_vector)
```

```
## [1] 60
```

mixture BY4741

```
#no need to do sampling cause it is already small enough  
cumulative_BY4741_mixGompertz<-fitmixEM(cumulative_BY4741_vector,"gompertz", 2, initial=FALSE)
```

```
print(cumulative_BY4741_mixGompertz$estimate)
```

```
##           weight      alpha      beta  
## [1,] 0.4645235 0.2646889 0.0011761887  
## [2,] 0.5354765 0.1978986 0.0001788796
```

```
print(cumulative_BY4741_mixGompertz$measures)
```

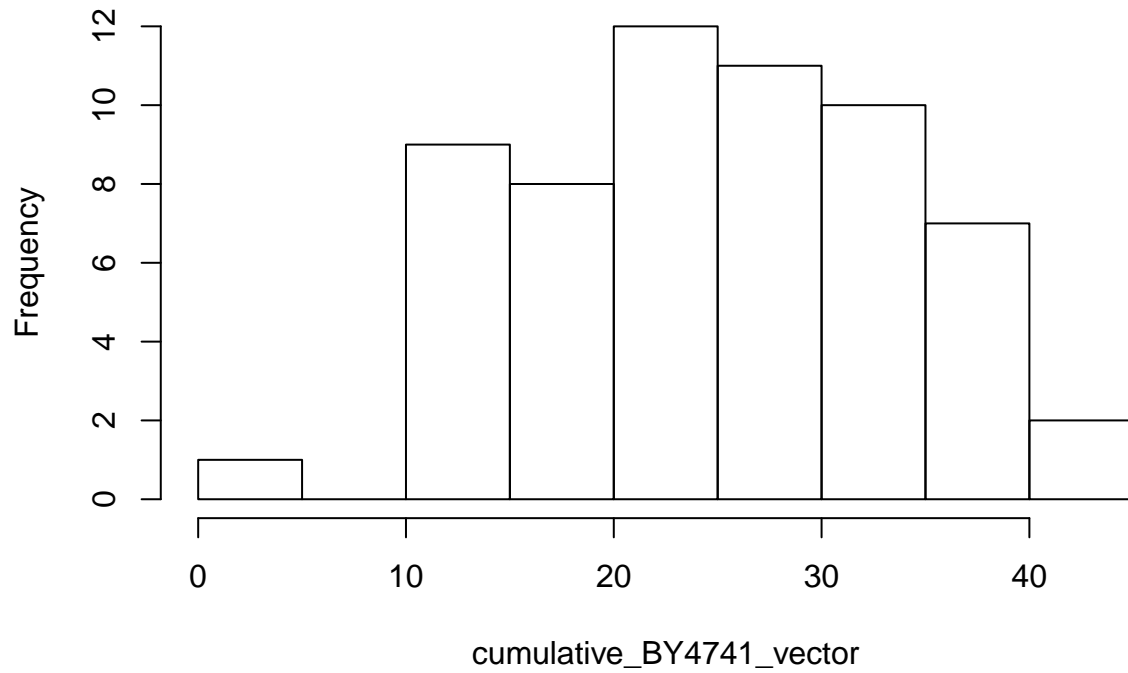
```
##           AIC      BIC      KS log.likelihood  
## [1,] 442.9331 453.4048 0.08643545      -216.4665
```

mixture plot of BY4741

```
dfBY4741=data.frame(lambda=c(cumulative_BY4741_mixGompertz$estimate[,1]),  
                      shape=c(cumulative_BY4741_mixGompertz$estimate[,2]),  
                      rate=c(cumulative_BY4741_mixGompertz$estimate[,3]))
```

```
plot(hist(cumulative_BY4741_vector),xlab="lifespan,divisions",col="grey",border="white"
```

Histogram of cumulative_BY4741_vector

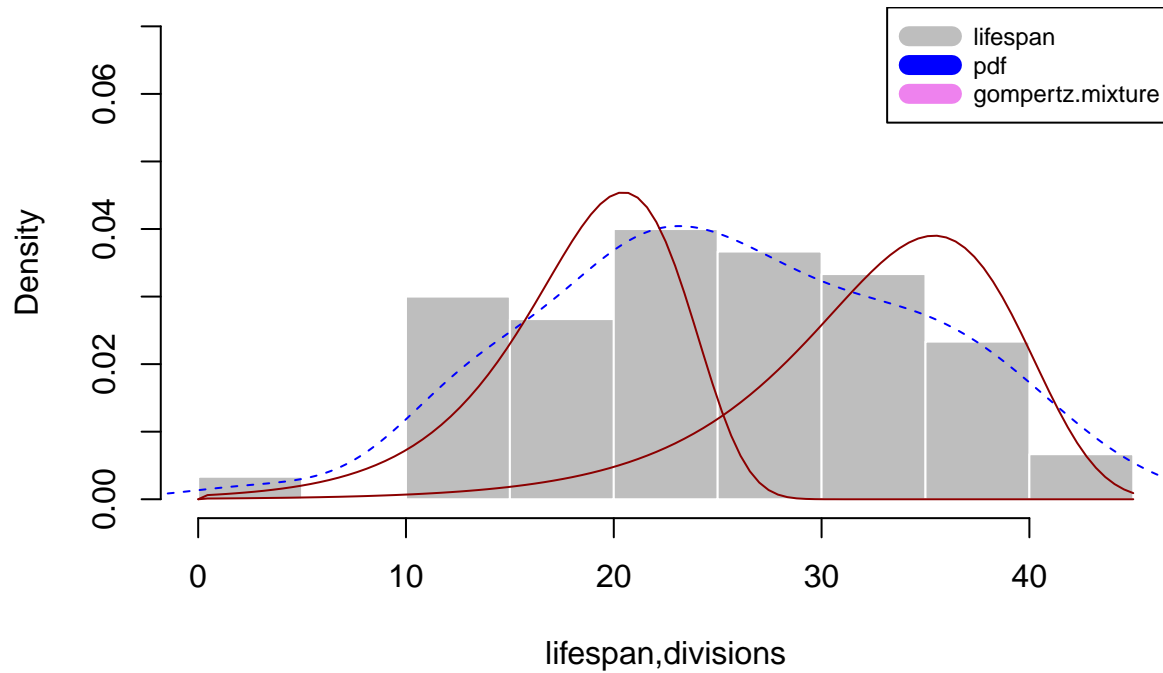


```
#pdf("plots/mixture_fitsBY4741.pdf")
lines(density(cumulative_BY4741_vector),lty=2,col="blue")
sapply(1:2,plot.gompertz.components,mixture=dfBY4741)

##      [,1]      [,2]
## x Numeric,101 Numeric,101
## y Numeric,101 Numeric,101

legend("topright", c("lifespan", "pdf","gompertz.mixture"),
      col=c("grey","blue","violet"),
      lwd=10,cex=0.75)
```

Histogram of cumulative_BY4741_vector



```
#dev.off()
```

```
tor1
```

```
tor1<- lifespan_data[lifespan_data$genotype=="tor1",]
#WT.BY4741.temp<-WT.BY4741[WT.BY4741$temp==30,]
#there is no need to this since I did filtering in the beginning
tor1.media<-tor1[tor1$media=="YPD",]
tor1.media<-tor1.media[tor1.media$mat=="MATalpha",]
tor1.media= tor1.media[!is.na(tor1.media[,1]), ]
dim(tor1.media)
```

```
## [1] 1 10
```

```
class(tor1.media$single_lifespan)
```

```
## [1] "factor"
```

```
str(tor1.media$single_lifespan)
```

```
## Factor w/ 20222 levels "", "1, 1", "1, 1, 1, 1, 1, 1", ...: 17428
```

```
### tor1 individuals
```

```
tor1_lifespan_list=list()
```

```
for (k in c(1:length(tor1.media$single_lifespan))){
  # k=c(94:95)
  fk<-tor1.media$single_lifespan[k]
  #f<-unfactor(fk[k])
  tor1_lifespan_single<-as.numeric(unlist(str_split(fk, ",")))
}
```

```

#tor1_lifespan_list<-mean(ref_lifespan_single)
tor1_lifespan_list[[length(tor1_lifespan_list)+1]]<-tor1_lifespan_single
}

```

```
length(tor1_lifespan_list)
```

```
## [1] 1
```

```

cumulative_tor1_lifespan<-unlist(tor1_lifespan_list)
length(cumulative_tor1_lifespan)

```

```
## [1] 39
```

```

cumulative_tor1_vector<-na.omit(cumulative_tor1_lifespan)
length(cumulative_tor1_vector)

```

```
## [1] 39
```

mixture tor1

```

#no need to do sampling cause it is already small enough
cumulative_tor1_mixGompertz<-fitmixEM(cumulative_tor1_vector,"gompertz", 2, initial=FALSE)

```

```
cumulative_tor1_mixGompertz$estimate
```

```

##          weight      alpha      beta
## [1,] 0.5819315 0.1185385 0.0002599403
## [2,] 0.4180685 0.1731600 0.0011806545

```

```
cumulative_tor1_mixGompertz$measures
```

```

##          AIC      BIC      KS log.likelihood
## [1,] 319.7924 328.1102 0.1285071      -154.8962

```

mixture plot of tor1

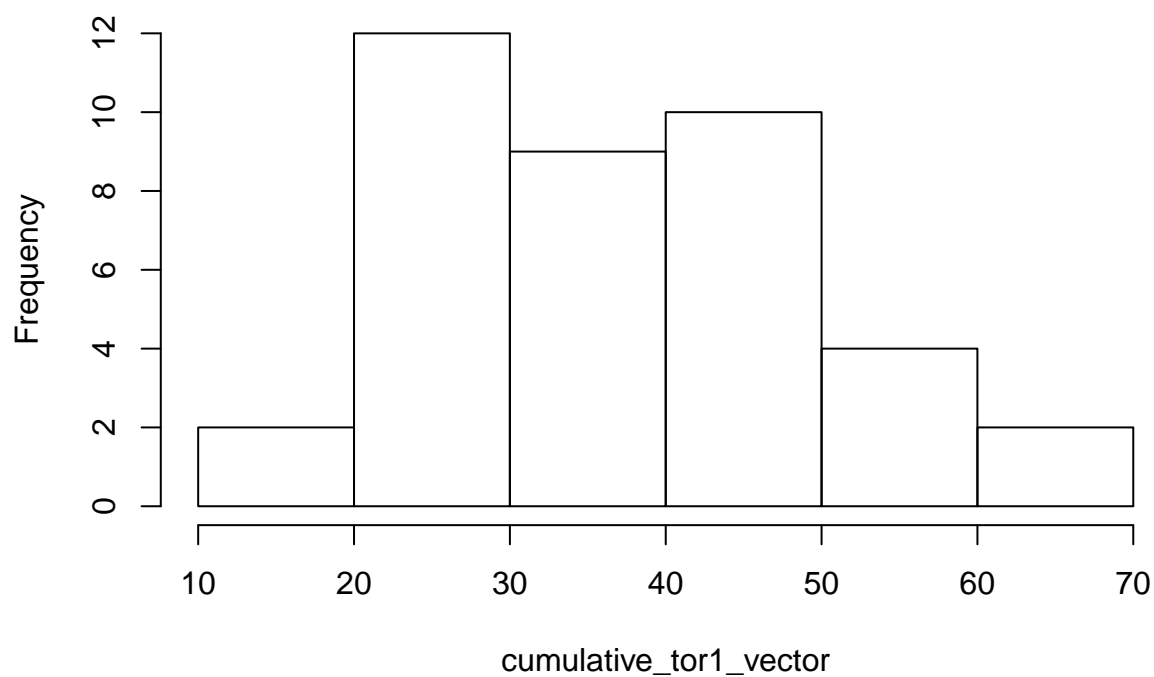
```

dftor1=data.frame(lambda=c(cumulative_tor1_mixGompertz$estimate[,1]),
                    shape=c(cumulative_tor1_mixGompertz$estimate[,2]),
                    rate=c(cumulative_tor1_mixGompertz$estimate[,3]))

```

```
plot(hist(cumulative_tor1_vector),xlab="lifespan,divisions", col="grey",border="white",
```


Histogram of cumulative_tor1_vector

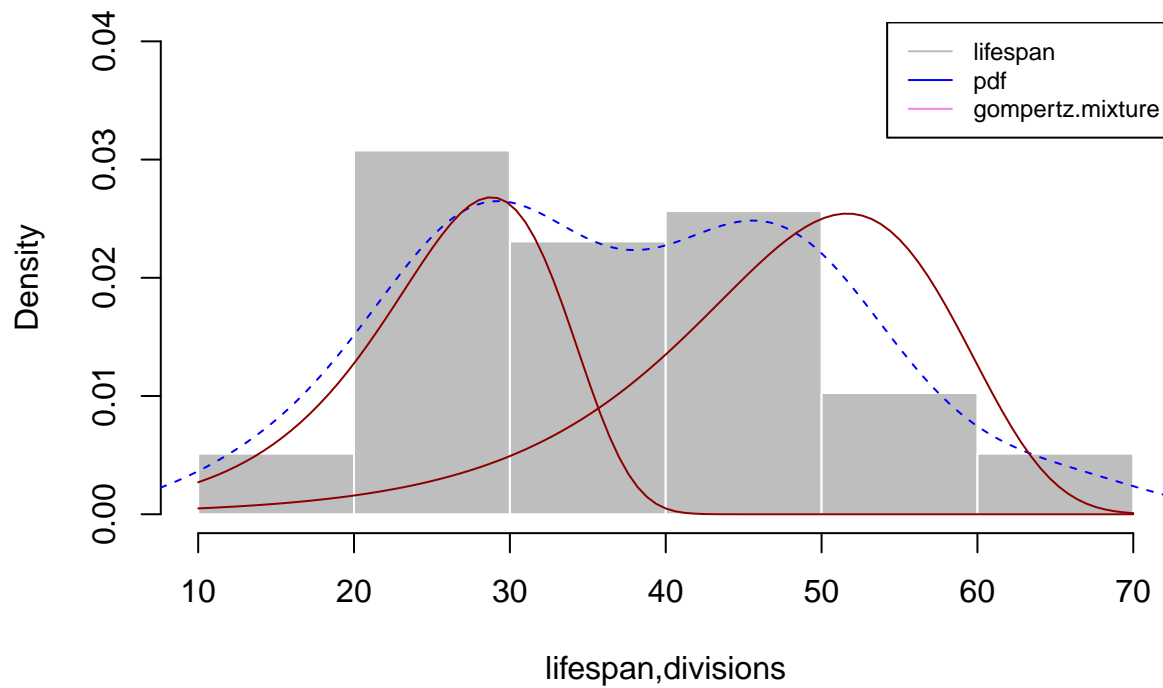


```
#pdf("plots/mixture_fitstor1.pdf")
lines(density(cumulative_tor1_vector),lty=2,col="blue")
sapply(1:2,plot.gompertz.components,mixture=dftor1)

##      [,1]      [,2]
## x Numeric,101 Numeric,101
## y Numeric,101 Numeric,101

legend("topright", c("lifespan", "pdf","gompertz.mixture"),
      col=c("grey","blue","violet"),lwd=1,cex=0.75)
```

Histogram of cumulative_tor1_vector



```
#dev.off()
```

fob1

```
fob1<- lifespan_data[lifespan_data$genotype=="fob1",]

#WT.BY4741.temp<-WT.BY4741[WT.BY4741$temp==30,]
#there is no need to this since I did filtering in the beginning

fob1.media<-fob1[fob1$media=="YPD",]

fob1.media<-fob1.media[fob1.media$mat=="MATalpha",]

fob1.media= fob1.media[!is.na(fob1.media[,1]), ]
dim(fob1.media)

## [1] 1 10

class(fob1.media$single_lifespan)

## [1] "factor"

str(fob1.media$single_lifespan)

## Factor w/ 20222 levels "", "1, 1", "1, 1, 1, 1, 1, 1", ...: 17504
### fob1 individuals
fob1_lifespan_list=list()

for (k in c(1:length(fob1.media$single_lifespan))){
```

```

# k=c(94:95)
fk<-fob1.media$single_lifespan[k]
#f<-unfactor(fk[k])
fob1_lifespan_single<-as.numeric(unlist(str_split(fk, ",")))
#fob1_lifespan_list<-mean(ref_lifespan_single)
fob1_lifespan_list[[length(fob1_lifespan_list)+1]]<-fob1_lifespan_single
}

length(fob1_lifespan_list)

```

```

## [1] 1
cumulative_fob1_lifespan<-unlist(fob1_lifespan_list)
length(cumulative_fob1_lifespan)

```

```

## [1] 39
cumulative_fob1_vector<-na.omit(cumulative_fob1_lifespan)
length(cumulative_fob1_vector)

```

```
## [1] 39
```

mixture fob1

```

#no need to do sampling cause it is already small enough
cumulative_fob1_mixGompertz<-fitmixEM(cumulative_fob1_vector,"gompertz", 2, initial=FALSE)

cumulative_fob1_mixGompertz$estimate

##          weight      alpha      beta
## [1,] 0.4829328 0.1952019 0.0017610300
## [2,] 0.5170672 0.1156508 0.0009052184

cumulative_fob1_mixGompertz$measures

##          AIC      BIC      KS log.likelihood
## [1,] 309.0855 317.4033 0.1481874 -149.5428

```

mixture plot of fob1

```

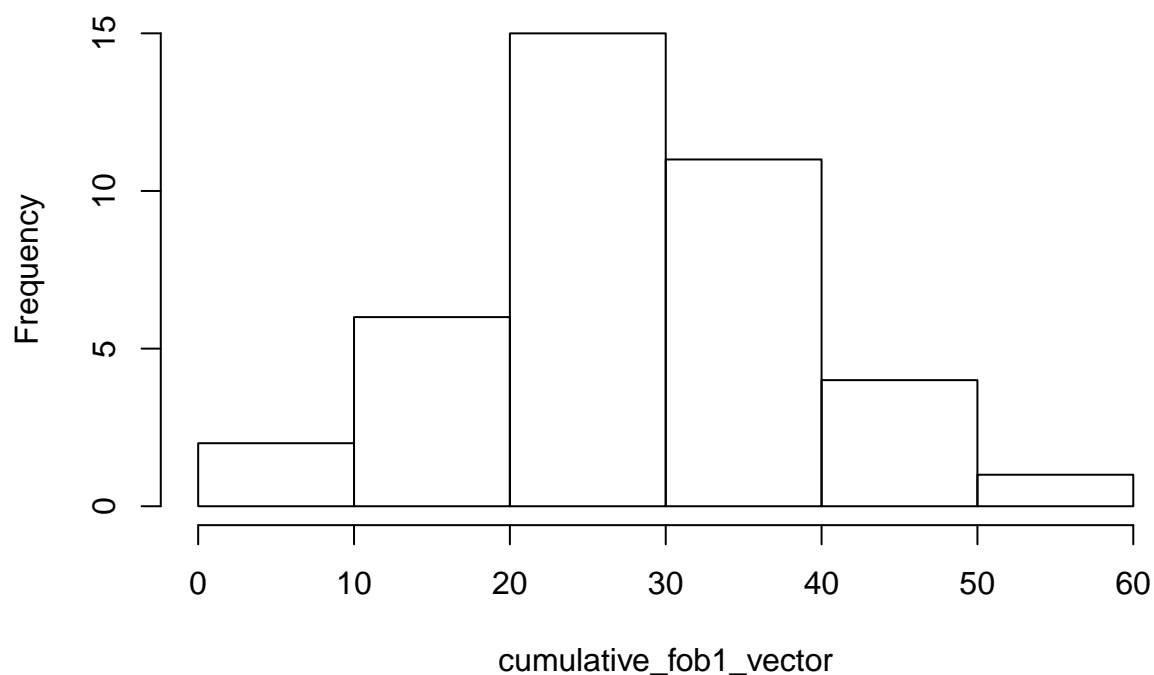
dffob1=data.frame(lambda=c(cumulative_fob1_mixGompertz$estimate[,1]),
                    shape=c(cumulative_fob1_mixGompertz$estimate[,2]),
                    rate=c(cumulative_fob1_mixGompertz$estimate[,3]))

plot(hist(cumulative_fob1_vector),xlab="lifespan,divisions",

```

```
col="grey",border="white",
```

Histogram of cumulative_fob1_vector



```
#pdf("plots/mixture_fitsfob1.pdf")
lines(density(cumulative_fob1_vector),lty=2,col="blue")
sapply(1:2,plot.gompertz.components,mixture=dffob1)

##      [,1]      [,2]
## x Numeric,101 Numeric,101
## y Numeric,101 Numeric,101

legend("topright", c("lifespan", "pdf","gompertz.mixture"),
      col=c("grey","blue","violet"),lwd=1,cex=0.75)
```

sod2 lifespan

13

```

for (k in c(1:length(sod2.media$single_lifespan))){
  # k=c(94:95)
  fk<-sod2.media$single_lifespan[k]
  #f<-unfactor(fk[k])
  sod2_lifespan_single<-as.numeric(unlist(str_split(fk, ",")))
  #sod2_lifespan_list<-mean(ref_lifespan_single)
  sod2_lifespan_list[[length(sod2_lifespan_list)+1]]<-sod2_lifespan_single
}

```

```
length(sod2_lifespan_list)
```

```
## [1] 1
```

```

cumulative_sod2_lifespan<-unlist(sod2_lifespan_list)
length(cumulative_sod2_lifespan)

```

```
## [1] 40
```

```

cumulative_sod2_vector<-na.omit(cumulative_sod2_lifespan)
length(cumulative_sod2_vector)

```

```
## [1] 40
```

mixture sod2

```

#no need to do sampling cause it is already small enough
cumulative_sod2_mixGompertz<-fitmixEM(cumulative_sod2_vector,"gompertz", 2, initial=FALSE)

cumulative_sod2_mixGompertz$estimate

```

```

##          weight      alpha      beta
## [1,] 0.5012542 0.1792547 0.0035327856
## [2,] 0.4987458 0.1735168 0.0001001064

```

```
cumulative_sod2_mixGompertz$measures
```

```

##          AIC      BIC      KS log.likelihood
## [1,] 319.5947 328.0391 0.1024277      -154.7973

```

mixture plot of sod2

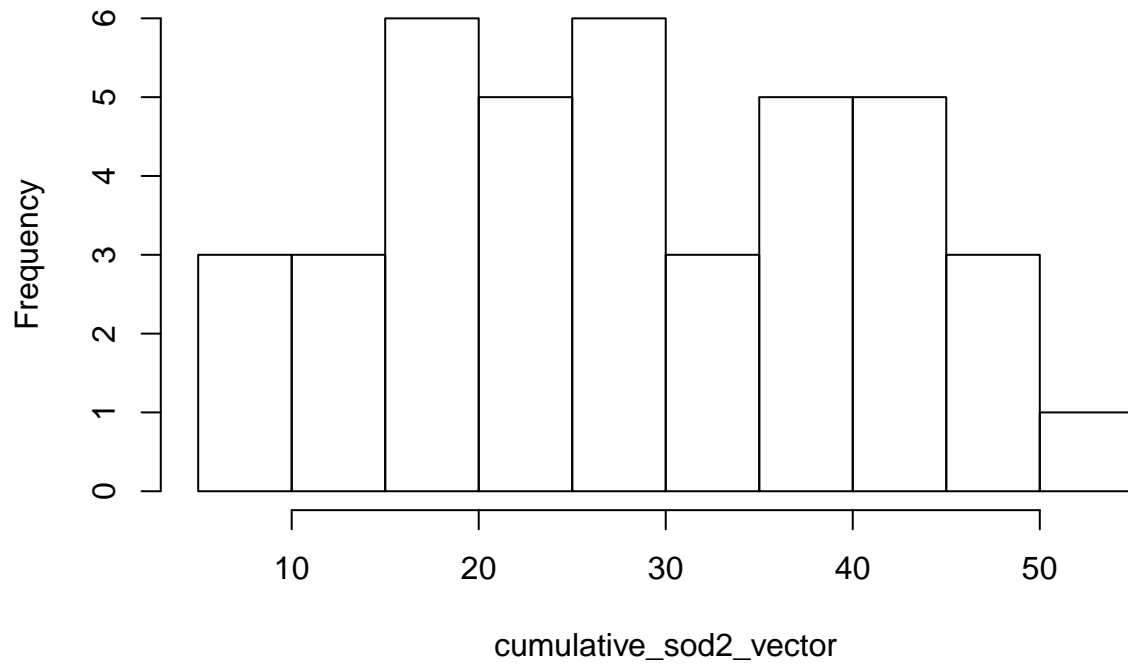
```

dfsod2=data.frame(lambda=c(cumulative_sod2_mixGompertz$estimate[,1]),
  shape=c(cumulative_sod2_mixGompertz$estimate[,2]),
  rate=c(cumulative_sod2_mixGompertz$estimate[,3]))

```

```
plot(hist(cumulative_sod2_vector),xlab="lifespan,divisions", col="grey",border="white",
```

Histogram of cumulative_sod2_vector

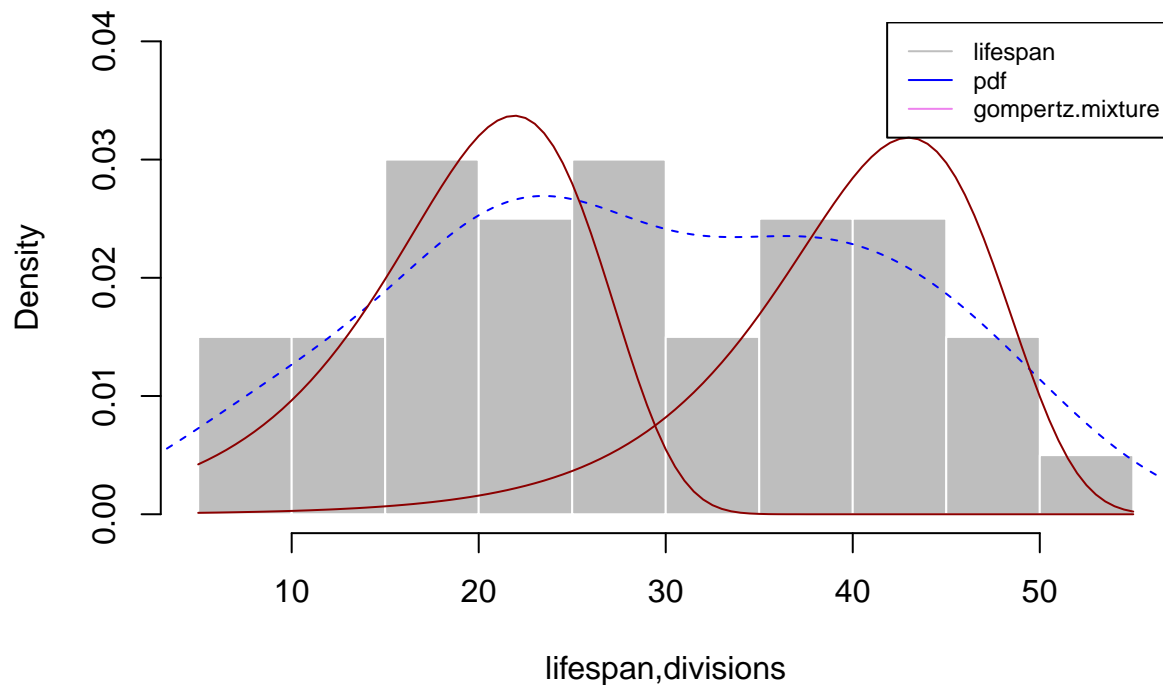


```
#pdf("plots/mixture_fitssod2.pdf")
lines(density(cumulative_sod2_vector),lty=2,col="blue")
sapply(1:2,plot.gompertz.components,mixture=dfsod2)

##      [,1]      [,2]
## x Numeric,101 Numeric,101
## y Numeric,101 Numeric,101

legend("topright", c("lifespan", "pdf","gompertz.mixture"),
      col=c("grey","blue","violet"),lwd=1,cex=0.75)
```

Histogram of cumulative_sod2_vector



```
#dev.off()
```

sir2 lifespan

```
sir2<- lifespan_data[lifespan_data$genotype=="sir2",]
dim(sir2)
```

```
## [1] 55 10
```

```
#WT.BY4741.temp<-WT.BY4741[WT.BY4741$temp==30,]
#there is no need to this since I did filtering in the beginning
```

```
sir2.media<-sir2[sir2$media=="YPD",]
dim(sir2.media)
```

```
## [1] 4 10
```

```
sir2.media<-sir2.media[sir2.media$mat=="MATalpha",]
dim(sir2.media)
```

```
## [1] 1 10
```

```
sir2.media= sir2.media[!is.na(sir2.media[,1]), ]
dim(sir2.media)
```

```
## [1] 1 10
```


sch9 lifespan

```
sch9<- lifespan_data[lifespan_data$genotype=="sch9",]

#WT.BY4741.temp<-WT.BY4741[WT.BY4741$temp==30,]
#there is no need to this since I did filtering in the beginning

sch9.media<-sch9[sch9$media=="YPD",]

sch9.media<-sch9.media[sch9.media$mat=="MATalpha",]

sch9.media= sch9.media[!is.na(sch9.media[,1]), ]
dim(sch9.media)

## [1] 1 10

class(sch9.media$single_lifespan)

## [1] "factor"

str(sch9.media$single_lifespan)

## Factor w/ 20222 levels "", "1, 1", "1, 1, 1, 1, 1, 1", ...: 1182
""

### sch9 individuals
sch9_lifespan_list=list()

for (k in c(1:length(sch9.media$single_lifespan))){
  # k=c(94:95)
  fk<-sch9.media$single_lifespan[k]
  #f<-unfactor(fk[k])
  sch9_lifespan_single<-as.numeric(unlist(str_split(fk, ",")))
  #sch9_lifespan_list<-mean(ref_lifespan_single)
  sch9_lifespan_list[[length(sch9_lifespan_list)+1]]<-sch9_lifespan_single
}

length(sch9_lifespan_list)

## [1] 1

cumulative_sch9_lifespan<-unlist(sch9_lifespan_list)
length(cumulative_sch9_lifespan)

## [1] 40

cumulative_sch9_vector<-na.omit(cumulative_sch9_lifespan)
length(cumulative_sch9_vector)

## [1] 40
```

mixture sch9

```
#no need to do sampling cause it is already small enough
cumulative_sch9_mixGompertz<-fitmixEM(cumulative_sch9_vector,"gompertz", 2, initial=FALSE)
```

```
cumulative_sch9_mixGompertz$estimate
```

```
##          weight      alpha      beta  
## [1,] 0.6743064 0.2354858 0.0018870361  
## [2,] 0.3256936 0.1651988 0.0003707972
```

```
cumulative_sch9_mixGompertz$measures
```

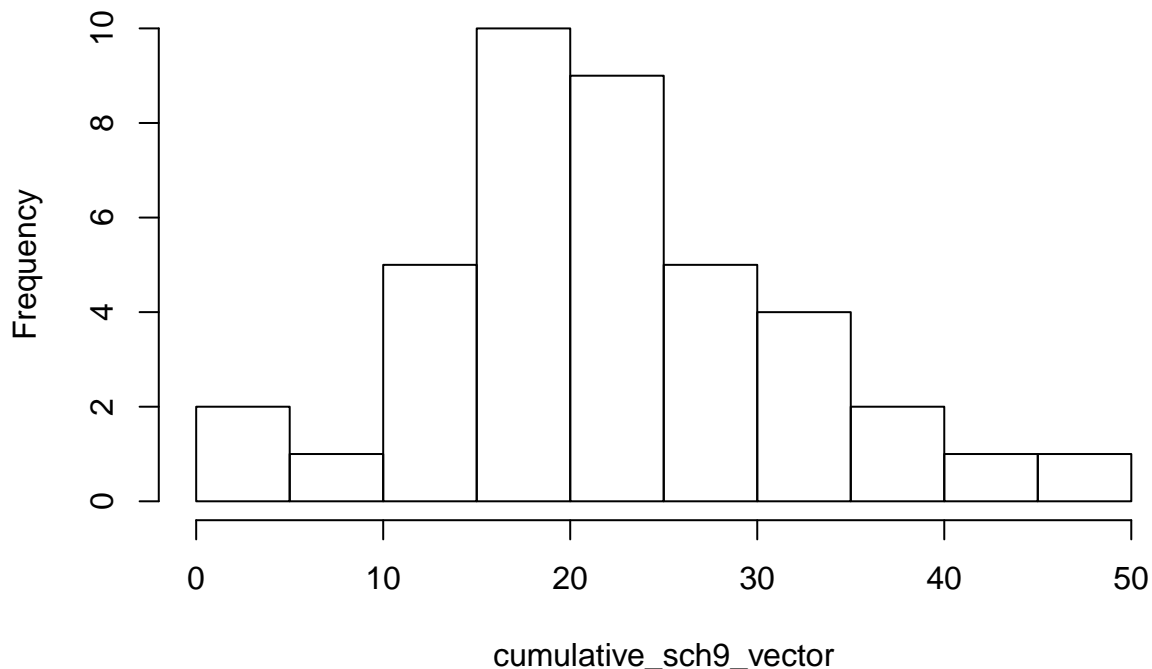
```
##          AIC      BIC      KS log.likelihood  
## [1,] 298.7891 307.2335 0.09757428      -144.3945
```

mixture plot of sch9

```
dfs9=data.frame(lambda=c(cumulative_sch9_mixGompertz$estimate[,1]),  
                  shape=c(cumulative_sch9_mixGompertz$estimate[,2]),  
                  rate=c(cumulative_sch9_mixGompertz$estimate[,3]))
```

```
plot(hist(cumulative_sch9_vector),xlab="lifespan,divisions", col="grey",border="white",
```

Histogram of cumulative_sch9_vector

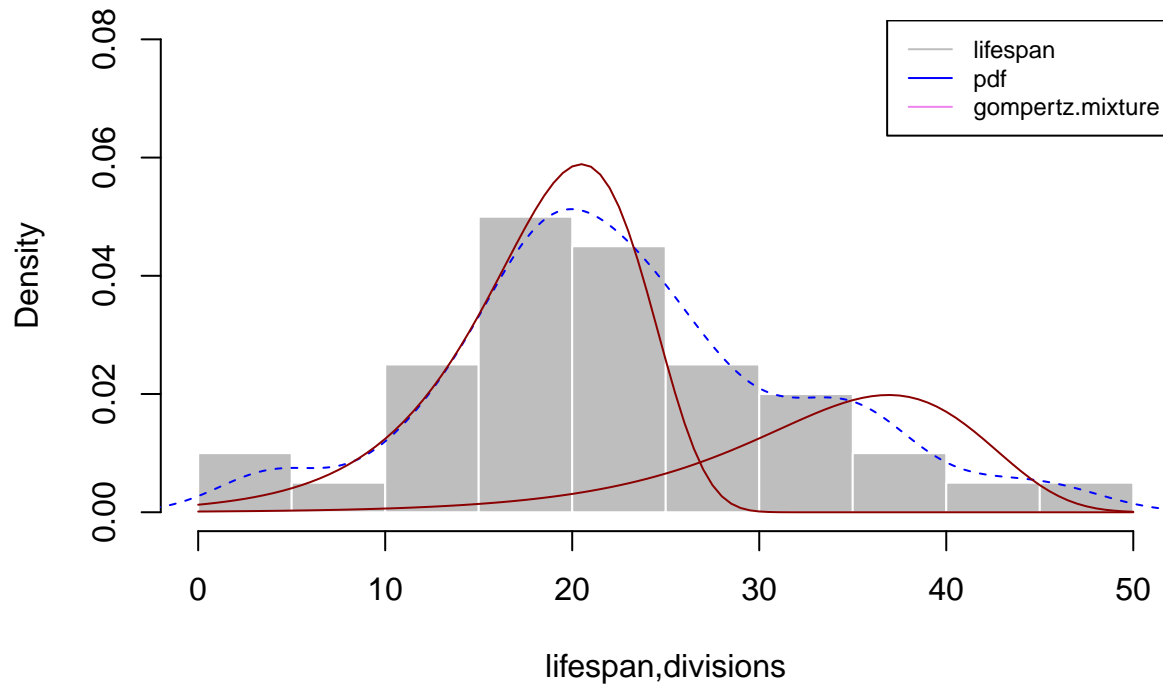


```
#pdf("plots/mixture_fitssch9.pdf")  
lines(density(cumulative_sch9_vector),lty=2,col="blue")  
sapply(1:2,plot.gompertz.components,mixture=dfs9)
```

```
##      [,1]      [,2]  
## x Numeric,101 Numeric,101  
## y Numeric,101 Numeric,101
```

```
legend("topright", c("lifespan", "pdf","gompertz.mixture"),  
      col=c("grey","blue","violet"),lwd=1,cex=0.75)
```

Histogram of cumulative_sch9_vector



```
#dev.off()
```

```
sir2<- lifespan_data[lifespan_data$genotype=="sir2",]
```

```
sir2.media<-sir2[sir2$media=="YPD",]
```

```
sir2.media<-sir2.media[sir2.media$mat=="MATalpha",]
```

```
sir2.media= sir2.media[!is.na(sir2.media[,1]), ]
dim(sir2.media)
```

```
## [1] 1 10
```

```
class(sir2.media$single_lifespan)
```

```
## [1] "factor"
```

```
str(sir2.media$single_lifespan)
```

```
## Factor w/ 20222 levels "", "1, 1", "1, 1, 1, 1, 1, 1", ...: 18307
```

```
### sir2 individuals
```

```
sir2_lifespan_list=list()
```

```
for (k in c(1:length(sir2.media$single_lifespan))){
  # k=c(94:95)
  fk<-sir2.media$single_lifespan[k]
  #f<-unfactor(fk[k])
  sir2_lifespan_single<-as.numeric(unlist(str_split(fk, ",")))
  #sir2_lifespan_list<-mean(ref_lifespan_single)
  sir2_lifespan_list[[length(sir2_lifespan_list)+1]]<-sir2_lifespan_single
}
```

```

length(sir2_lifespan_list)

## [1] 1
cumulative_sir2_lifespan<-unlist(sir2_lifespan_list)
length(cumulative_sir2_lifespan)

## [1] 35
cumulative_sir2_vector<-na.omit(cumulative_sir2_lifespan)
length(cumulative_sir2_vector)

## [1] 35
sir2<-cumulative_sir2_vector

### mixture sir2

#no need to do sampling cause it is already small enough
cumulative_sir2_mixGompertz<-fitmixEM(cumulative_sir2_vector,"gompertz", 2, initial=FALSE)

fitmixEM(sir2,"gompertz", 2, initial=FALSE)

## $estimate
##      weight      alpha      beta
## [1,] 0.4176364 0.1404278 0.002153822
## [2,] 0.5823636 0.1514844 0.227055631
##
## $measures
##      AIC      BIC      KS log.likelihood
## [1,] 237.0737 244.8505 0.1275805      -113.5369
##
## $cluster
## [1] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 2 1 2 1 1 1 1 1 1 1 2 2 1 1 1 1
cumulative_sir2_mixGompertz$estimate

##      weight      alpha      beta
## [1,] 0.4176364 0.1404278 0.002153822
## [2,] 0.5823636 0.1514844 0.227055631
cumulative_sir2_mixGompertz$measures

##      AIC      BIC      KS log.likelihood
## [1,] 237.0737 244.8505 0.1275805      -113.5369
cumulative_sir2_mixGompertz$cluster

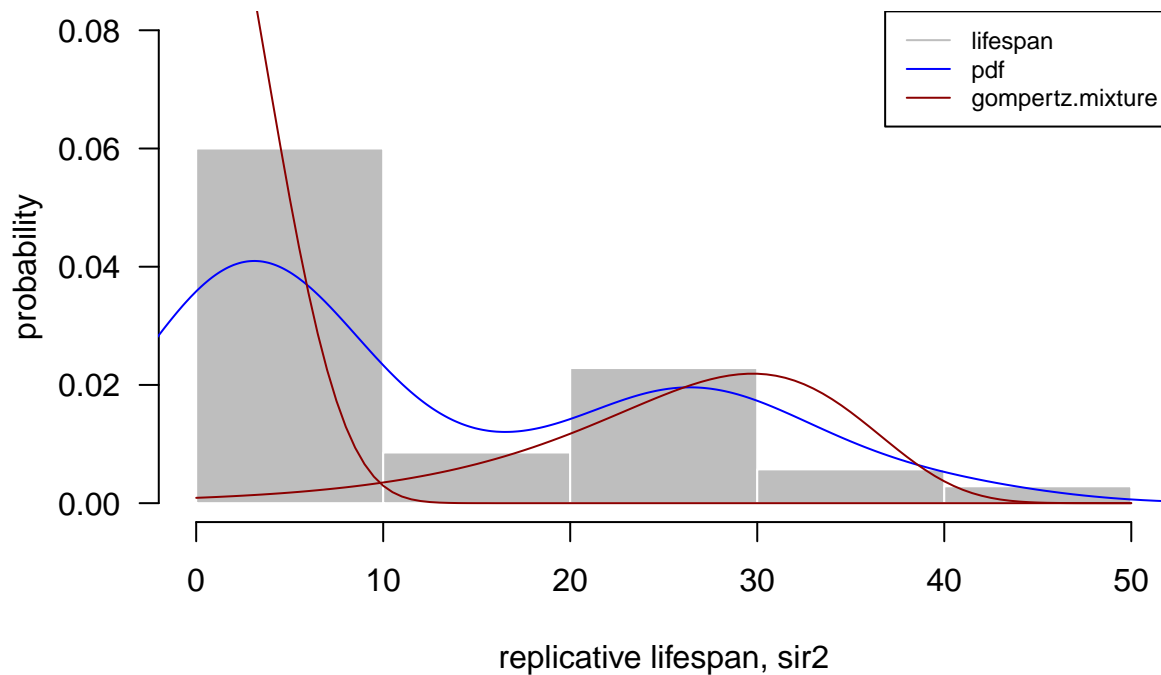
## [1] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 2 1 2 1 1 1 1 1 1 1 2 2 1 1 1 1
###mixture plot of sch9

dfsir2=data.frame(lambda=c(cumulative_sir2_mixGompertz$estimate[,1]),
                    shape=c(cumulative_sir2_mixGompertz$estimate[,2]),
                    rate=c(cumulative_sir2_mixGompertz$estimate[,3]))

```

```
hist(sir2,xlab="replicative lifespan, sir2",col="grey",border="white",breaks=5,las=1,
     freq=FALSE,ylim=c(0,0.08),ylab="probability")
lines(density(sir2),las=1,col="blue")
x<-sapply(1:2,plot.gompertz.components,mixture=dfsir2)
legend("topright", c("lifespan", "pdf", "gompertz.mixture"),
      col=c("grey","blue","darkred"),lwd=1,cex=0.75)
```

Histogram of sir2

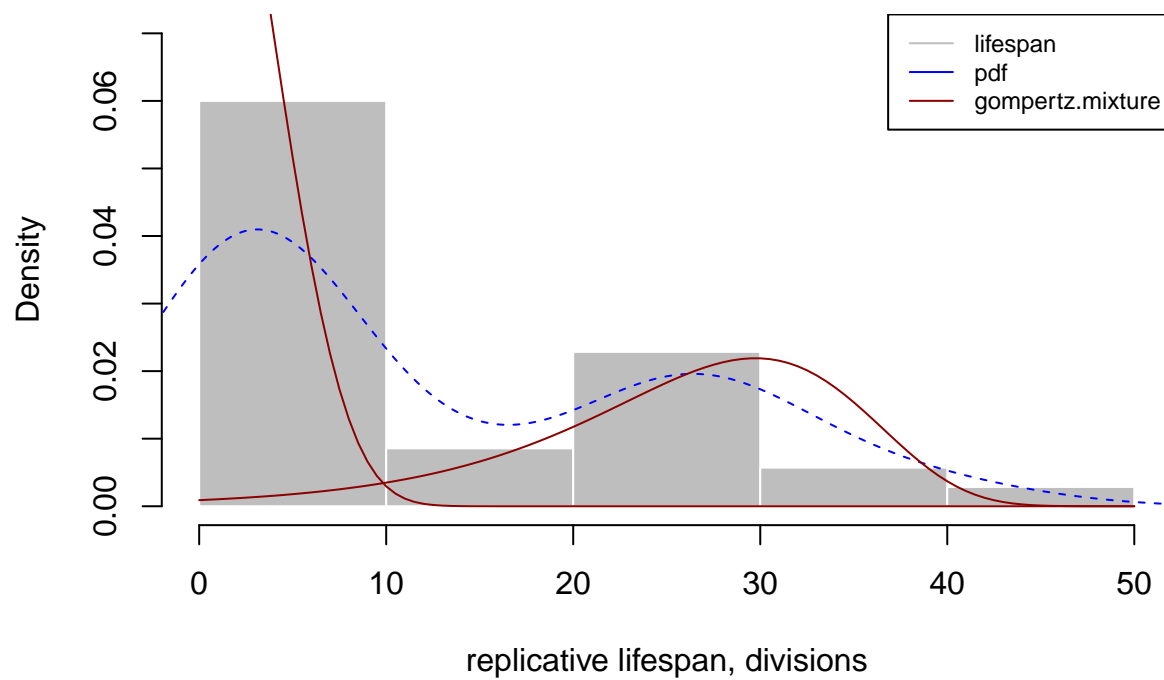


```
hist(cumulative_sir2_vector,breaks=5, xlab="replicative lifespan, divisions",col="grey",border="white",
     #pdf("plots/mixture_fitssir2.pdf")
     lines(density(cumulative_sir2_vector),lty=2,col="blue")
     sapply(1:2,plot.gompertz.components,mixture=dfsir2)
```

```
##      [,1]      [,2]
## x Numeric,101 Numeric,101
## y Numeric,101 Numeric,101
```

```
legend("topright", c("lifespan", "pdf","gompertz.mixture"),
      col=c("grey","blue","darkred"),lwd=1,cex=0.75)
```

Histogram of cumulative_sir2_vector



#dev.off()