

Improvement to nls() : Solution to Tests(Medium)

Arkajyoti Bhattacharjee

Theory :

The logistic sigmoid growth curve, here, is given by -

$$y = f(t) = \frac{a}{1 + be^{-ct}} \dots (\Delta)$$

Here, the growth $y = f(t)$ is a function of time t .

b : the scale parameter of the sigmoid.

a : the curve's maximum value.

c : the logistic growth rate or steepness of the curve.

Now, given the test data, we need to estimate a logistic sigmoid growth curve of the form as in Eq. Δ .

R Code :

```
### "Improvements to nls()" - GSOC'21
## Medium

# test data

time <- c( 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16,
  17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29,
  30, 31, 32, 33, 34, 35)
y <- c( 0.0074203, 0.3188325, 0.2815891, -0.3171173, -0.0305409, 0.2266773,
  -0.0216102, 0.2319695, -0.1082007, 0.2246899, 0.6144181, 1.1655192,
  1.8038330, 2.7644418, 4.1104270, 5.0470456, 6.1896092, 6.4128618,
  7.2974793, 7.8965245, 8.4364991, 8.8252770, 8.9836204, 9.6607736,
  9.1746182, 9.5348823, 10.0421165, 9.8477874, 9.2886090, 9.3169916,
  9.6270209 )

# Plot observed y vs. time
plot(time,y,xlab="t",ylab=expression(y[t]),
main="Plot of Observed and Fitted y against Time",
lwd=2,col="red",type="o")
```

```
# clean the data
time <- time[y>=0]
y <- y[y>=0]

## Approximate estimation of a,b,c
# approximate value of a
a<-max(y)

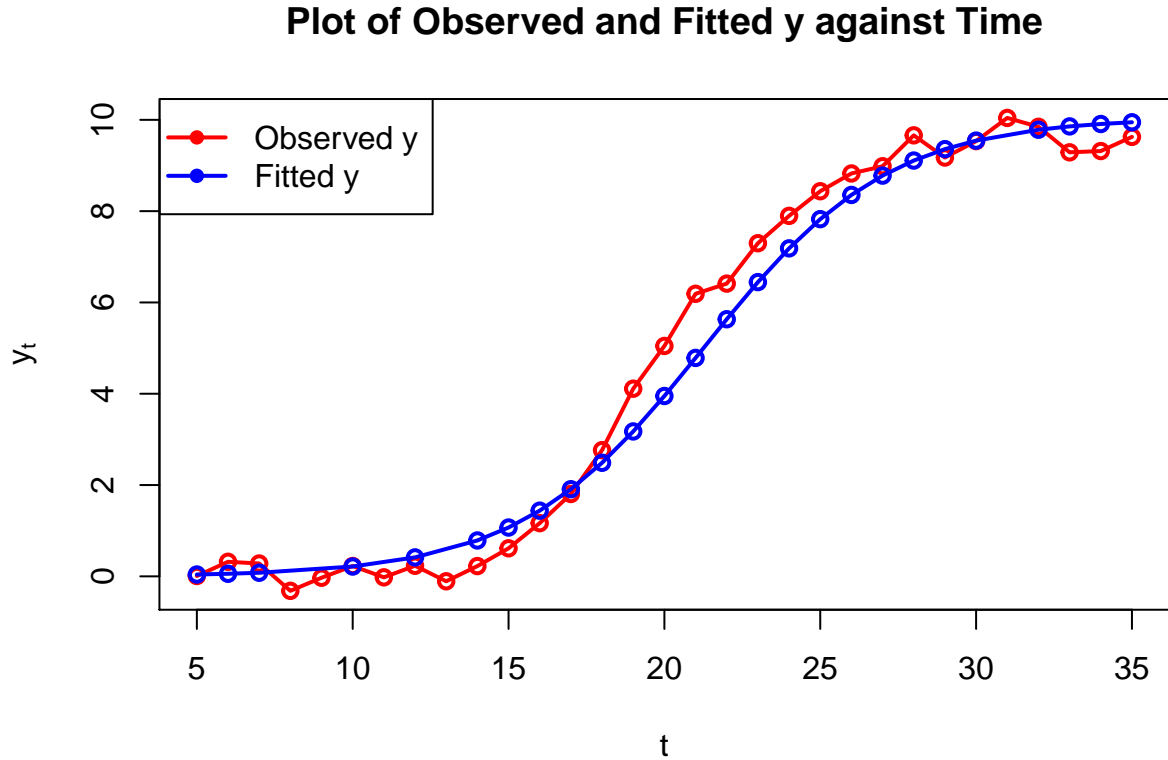
# approximate value of b
b <- a/y[1]-1

# approximate value of c
time<-time[y!=a]
y<-y[y!=a]
c<-log( (b*y)/(a-y) )/time
c<-mean(c)

# plot of fitted y values
lines(time, a/(1+b*exp(-c*time)),type="o",lwd=2,col="blue")
legend("topleft",legend=c("Observed y","Fitted y"),
col=c("red","blue"),lwd=2,pch=16)

# print a,b,c
cat("The estimated value of a is : ",a,"\n")
cat("The estimated value of b is : ",b,"\n")
cat("The estimated value of c is : ",c,"\n")
```

Output :



```
## The estimated value of a is : 10.04212
## The estimated value of b is : 1352.33
## The estimated value of c is : 0.3388275
```

Explanation :

I first plot the graph to know how the given y 's look as a function of time. Since, the logistic function is strictly positive, I first discard the negative y values from the data and the corresponding time points.

Now, a being the maximum value of the growth curve, I take a to be the maximum of the given y 's.

Here, the scale parameter b is actually $b = e^{cx_0}$ (on comparing with the form given in Eq Δ of the "Easy" file). Unlike x_0 , b itself involves two parameters c and x_0 ; so, estimating b becomes difficult as we cannot estimate c and x_0 explicitly using Eq. Δ (Note that this Eq. Δ does not contain any x_0 term).

So, I note that $f(0) = \frac{a}{1+b}$ and the logistic growth function is an increasing function. Hence, I estimate b by $b = a/y_{\min} - 1$.

Finally, given y , $time$ and estimated a , b , I compute c 's by plugging these values in Eq. Δ and then take the mean of the resulting vector of c 's. In this process, I compute c by $c = \frac{\log(by/(a-y))}{time}$; so, I discard that value of y and corresponding value of $time$ for which $a = y$.

Using the above procedure, my estimated logistic growth function is -

$$\hat{g} = \hat{f}(t) = \frac{10.04212}{1 + 1352.33e^{-0.3388275t}}$$