Statistics 520 - Assignment 6

Sam Olson

The Model

Define random variables $Y_1, \dots, Y_n$ connected with the growth of some organism, observed at times $t_1, \dots, t_n$. Assume the response variables are independent and follow the model:

$$Y_i = \mu_i + \sigma \varepsilon_i,$$

$$\mu_i = B + A \exp\Big[- \exp\{-k(t_i - T)\}\Big],$$

$$\varepsilon_i \overset{\text{iid}}{\sim} N(0,1).$$

This model describes a sigmoidal curve with both upper and lower asymptotes, but is not constrained to be symmetric about its inflection point in the way a logistic curve is.

Parameters

- A: distance between upper and lower asymptotes
- B: lower asymptote
- k: growth rate (slope of the linear portion of the curve)
- T: the time at which the inflection point occurs

Data

On the course web page in the Data module is a file `growthdat.txt` that contains data simulated from the Gompertz regression model. The variable names in this file are `x` and `y`, where:

- `x` = time of observation
- `y` = growth in appropriate units

Q1

Find generalized least squares estimates of the parameters A, B, k, T, and the associated moment-based estimate of σ^2. Describe how you determined effective starting values for the estimation procedure. Present a scatterplot along with the fitted expectation function.

Answer

We fit the four-parameter Gompertz model

$$\mu(t) = B + A\exp(-\exp\{-k(t-T)\})$$

with random component

$$Y_i = \mu(t_i) + \sigma\varepsilon_i, \quad \varepsilon_i \overset{\text{iid}}{\sim} N(0,1).$$

Here:

- (A): distance between the upper and lower asymptotes,
- (B): lower asymptote,
- (k): growth rate (slope at the inflection point),
- (T): time of inflection,
- (σ^2): error variance.

Effective starting values were intuited inductively, i.e., based upon visual inspection/descriptive statistics of the dataset. Since there is no reasonable basis to suppose numbers must be discrete, the observed values were largely used as starting values (treated the observations as a continuous R.V.).

- Lower asymptote ((B)): The min observed y-value is 0.948, so set ($B \approx 0.95$).
- Upper asymptote ((A + B)): The max observed y-value is 233.37. So, to get A we take the (observed) range ($A \approx 232.42$).
- Inflection time ((T)): From visual inspection, the curve rises steeply from t = 10 to t = 20; taking the midpoint of this period aas an initial starting point for "inflection" ($T \approx 15$).
- Growth rate ((k)): The sharpest rise in the data occurs around t = 12 to 18. For example, between t=13 ($y \approx 64$)) and (t=17) (($y \approx 140$)), the increase is about 76 units over 4 time units, or roughly 19 units per unit time. This provides an empirical sense of the "steepness" of the curve. However, in the Gompertz formulation the parameter (k) does not equal the raw slope; instead, the actual slope at the inflection point is ($\frac{kA}{e}$). Using ($A \approx 232$ and equating ($\frac{kA}{e}$) with the observed slope (~19) yields ($k \approx 0.22$). To keep the starting value moderate and ensure convergence, we round slightly upward and set ($k \approx 0.3$).

Thus the starting parameter vector is

$$\theta^{(0)} = (A = 232.42, B = 0.95, k = 0.3, T = 15)$$

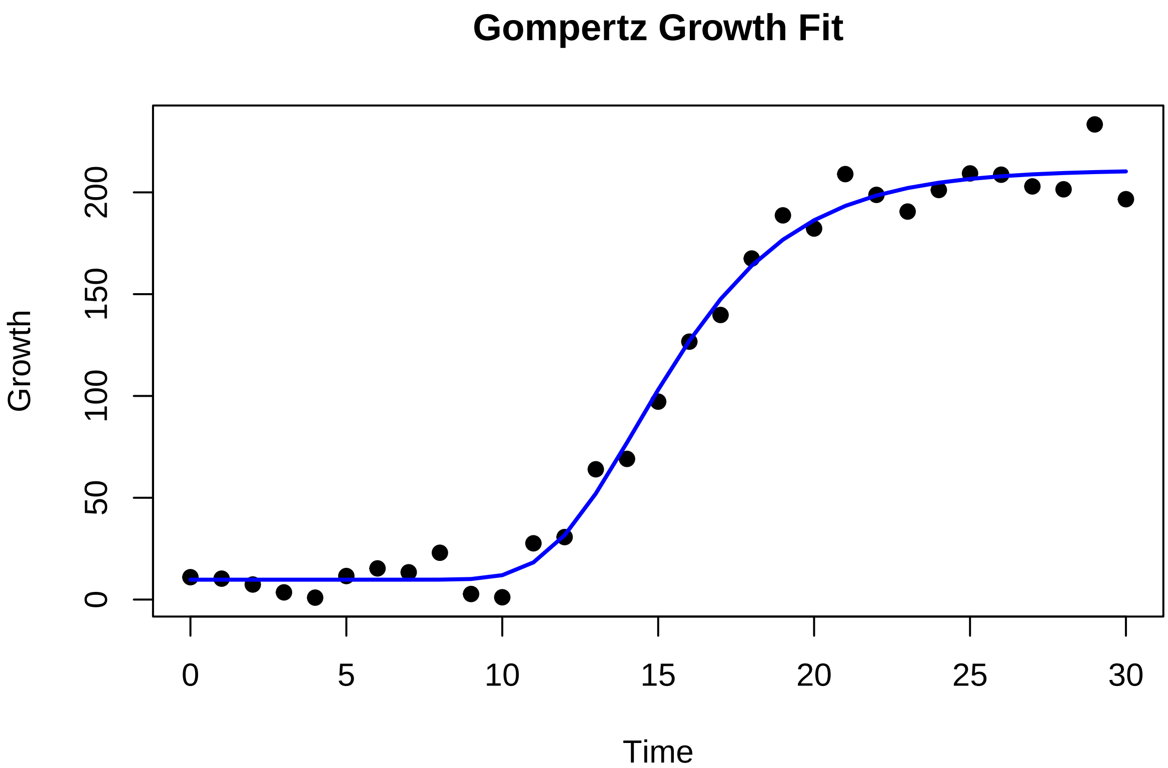
Actual Computation yields:

```
## New Estimates at Iteration  1  :
## 199.3338 9.76253 0.342326 14.24739
## New Estimates at Iteration  2  :
## 201.165 9.774896 0.3546216 14.25486
## New Estimates at Iteration  3  :
## 201.3052 9.761514 0.353611 14.25717
## New Estimates at Iteration  4  :
## 201.295 9.762608 0.3537187 14.25706
## New Estimates at Iteration  5  :
## 201.296 9.762523 0.3537084 14.25708
## New Estimates at Iteration  6  :
## 201.2959 9.762532 0.3537094 14.25707
## New Estimates at Iteration  7  :
## 201.2959 9.762532 0.3537093 14.25707
## New Estimates at Iteration  8  :
## 201.2959 9.762532 0.3537093 14.25707
## New Estimates at Iteration  9  :
## 201.2959 9.762532 0.3537093 14.25707
## New Estimates at Iteration  10  :
## 201.2959 9.762532 0.3537093 14.25707
## Convergence Criterion of  1e-08  met
## Final Estimates:  201.2959 9.762532 0.3537093 14.25707
```

```
##           [,1]
## dA 201.2958914
## dB   9.7625316
## dk   0.3537093
## dT  14.2570743
```

```
## [1] 85.56958
```

Plotting the observed vs. estimated values:


Gompertz Growth Fit
Growth
0
50
100
150
200
0
5
10
15
20
25
30
Time

Q2

Compute 95% approximate confidence intervals for the parameters of the expectation function (A, B, k, T).

Answer

```
##   param    estimate         SE        CI_L        CI_U
## 1     A 201.2958914 4.97428547 191.5464710 211.0453118
## 2     B   9.7625316 2.80872640   4.2575290  15.2675342
## 3     k   0.3537093 0.03229084   0.2904204   0.4169982
## 4     T  14.2570743 0.19481081  13.8752522  14.6388965
```

```
## [1] "A CI"
```

```
## [1] 191.7788 211.2853
```

```
## [1] "B CI"
```

```
## [1]  5.554791 17.157624
```

```
## [1] "k CI"
```

```
## [1] 0.2957594 0.4230137
```

```
## [1] "T CI"
```

```
## [1] 13.88032 14.64406
```

Q3

Compute pairwise correlations between $\hat{A}, \hat{B}, \hat{k}, \hat{T}$.

Answer

The `nonlin()` output includes the estimated covariance matrix of the parameter estimates (`out$covb`). From this matrix we can derive the correlations:

$$\mathrm{Corr}(\hat{\theta}_i, \hat{\theta}_j) = \frac{\widehat{\mathrm{Cov}}(\hat{\theta}_i, \hat{\theta}_j)}{\sqrt{\widehat{\mathrm{Var}}(\hat{\theta}_i)}; \sqrt{\widehat{\mathrm{Var}}(\hat{\theta}_j)}}$$

for parameters ($\theta_i, \theta_j \in \{A, B, k, T\}$).

Computing these quantities:

```
##             A          B          k           T
## A  1.00000000 -0.6387723 -0.6322836 -0.02943357
## B -0.63877229  1.0000000  0.2430056  0.43917161
## k -0.63228357  0.2430056  1.0000000  0.10361238
## T -0.02943357  0.4391716  0.1036124  1.00000000
```

Q4

Two quantities of interest to scientists are called the maximum relative growth rate and the maximum absolute growth rate. These quantities are related to the slope of the growth curve at the inflection point and give the per time unit increase in growth relative to the upper asymptote and in absolute scale at that time point. The maximum absolute growth rate is defined as,

$$k_{\text{abs}} = \frac{k(A+B)}{\exp(1)}$$

A plug-in estimate of k_{abs} is then

$$\hat{k}_{\text{abs}} = \frac{\hat{k}(\hat{A}+\hat{B})}{\exp(1)}.$$

and note that k_{abs} is an absolute function of its components. Given this, compute a 90% approximate confidence interval for k_{abs} (note a 90% interval – I get tired of using 95% all the time). Outline the procedure you used to calculate the quantities needed.

Answer

As defined,

$$k_{\text{abs}} = \frac{k(A+B)}{e}.$$

Given $(\hat{A}, \hat{B}, \hat{k}, \hat{T})$ from `nonlin` and the estimated covariance matrix $\widehat{\text{Cov}}(\hat{\theta})$ for $\hat{\theta} = (\hat{A}, \hat{B}, \hat{k}, \hat{T})$,

use a Wald (delta-method) interval.

Let $g(A, B, k, T) = k(A+B)/e$. Its gradient at $\hat{\theta}$ is

$$\nabla g(\hat{\theta}) = \left(\frac{\partial g}{\partial A}, \frac{\partial g}{\partial B}, \frac{\partial g}{\partial k}, \frac{\partial g}{\partial T}\right)^\top = \left(\frac{\hat{k}}{e}, \ \frac{\hat{k}}{e}, \ \frac{\hat{A}+\hat{B}}{e}, \ 0\right)^\top.$$

Then

$$\widehat{\text{Var}}(\hat{k}_{\text{abs}}) = \nabla g(\hat{\theta})^\top \ \widehat{\text{Cov}}(\hat{\theta}) \ \nabla g(\hat{\theta}), \qquad \text{SE}(\hat{k}_{\text{abs}}) = \sqrt{\widehat{\text{Var}}(\hat{k}_{\text{abs}})}.$$

A two-sided 90% CI uses $z_{0.95} = 1.645$:

$$\hat{k}_{\text{abs}} \ \pm \ 1.645 \times \text{SE}(\hat{k}_{\text{abs}}).$$

Computing

```
## [1] 27.46343
```

```
## [1] 23.81144 31.11541
```