The logistic function: The law of population growth

Unit 7 Lecture 4

Jonathan Auerbach
STAT 489 Pre-Cap Prof Development
jauerba@gmu.edu

November 30, 2021

Learning Objectives

After this lecture, you will be able to:

1. Define exponential and logistic growth as stated by Verhulst.
2. Graph exponential and logistic growth curves using the `stat_function` and `geom_textpath` functions from the `R` packages `ggplot2` and `geomtextpath`.
3. Fit exponential and logistic growth curves to data with Non-Linear Least Squares using the `nls` function. See Appendix for details.
4. Derive the exponential and logistic functions from their respective differential equations.

These slides use the following `R` packages

Setup:

```
library("knitr")
library("tidyverse")
library("geomtextpath")
theme_set(theme_bw(base_size = 20))
```

The logistic function: The law of population growth

- Malthus argued in *The Principle of Population* (1798) that when resources are abundant, populations grow at an exponential rate.
  - For example, the population of the Thirteen American Colonies doubled every 23 years between 1610 and 1780.
  - But exponential growth cannot continue indefinitely. Limited resources and other "checks" eventually constrain growth.

- Verhulst derived mathematical equations that describe Malthus' findings in his book *Mathematical Investigations on the Law of Population Growth* (1845).
  - Euler—for whom the constant e is named—had already studied the exponential growth of populations in 1748.
  - Verhulst added a capacity constraint to account for limited resources, calling his equation logistique (logistic) growth.
  - Today, the logistic function is used to describe a wide range of phenomenon from the onset of rare diseases to the choices made by economic agents.

Pierre Verhulst, *Mathematical Investigations* (1845)

RECHERCHES MATHÉMATIQUES

SUR

LA LOI D'ACCROISSEMENT DE LA POPULATION.

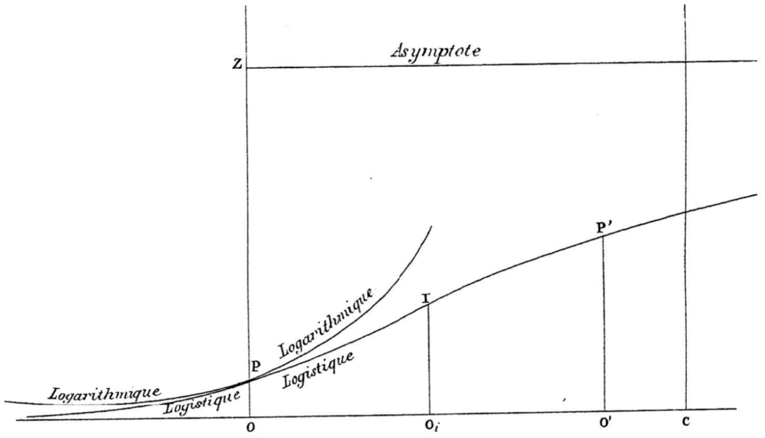
THÉORIE GÉNÉRALE.

§ 1. De tous les problèmes que l'économie politique offre aux méditations des philosophes, l'un des plus intéressants est, sans contredit, la connaissance de la loi qui règle les progrès de la population. Pour le résoudre avec exactitude, il faudrait pouvoir apprécier l'influence des causes nombreuses qui empêchent ou favorisent la multiplication de l'espèce humaine. Et comme plusieurs de ces causes sont variables par leur nature et par leur mode d'action, le problème considéré dans toute sa généralité, est visiblement insoluble.

Il faut observer cependant, qu'à mesure que la civilisation se perfectionne, l'influence des causes purement perturbatrices s'affaiblit de plus en plus, pour laisser dominer les causes constantes; de manière qu'à une certaine époque, il devient permis de faire abstraction des premières, sauf à considérer les données du problème comme soumises à de légères variations.

Source: https://en.wikipedia.org/wiki/Pierre_Fran%C3%A7ois_Verhulst
https://gdz.sub.uni-goettingen.de/id/PPN129323640_0018

Growth curves, *Mathematical Investigations* (1845)



Source: https://gdz.sub.uni-goettingen.de/id/PPN129323640_0018

Exponential vs. logistic growth

- Let y_x denote the population at time x with $0 \leq y_0 \leq K$. Exponential growth refers to the model:

$$y_x = e^{rx} \, y_0$$

- Verhulst modified exponential growth, calling it logistic growth:

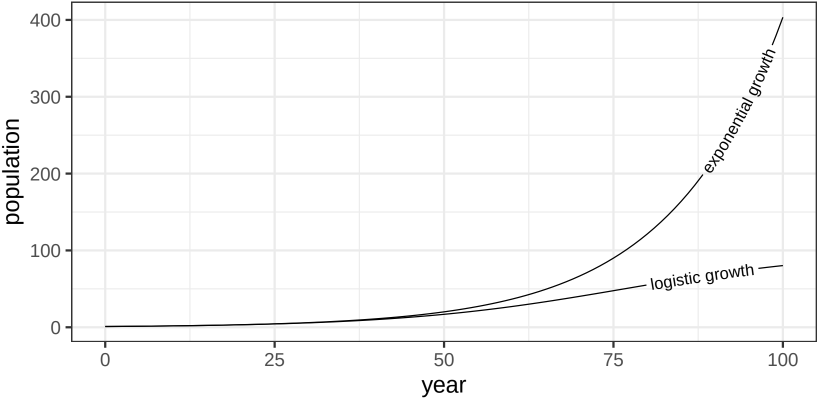
$$y_x = K \left(1 + \frac{K - y_0}{y_0} \, e^{-rx} \right)^{-1}$$

  - r is called the growth rate, and K is called the capacity constraint.

- When K is large and x and y_0 are small, the two functions are nearly identical.
  - As x increases, exponential growth increases without bound.
  - In contrast, e^{-rx} goes to 0 as x increases. It follows that logistic growth essentially stops as the population approaches capacity K.

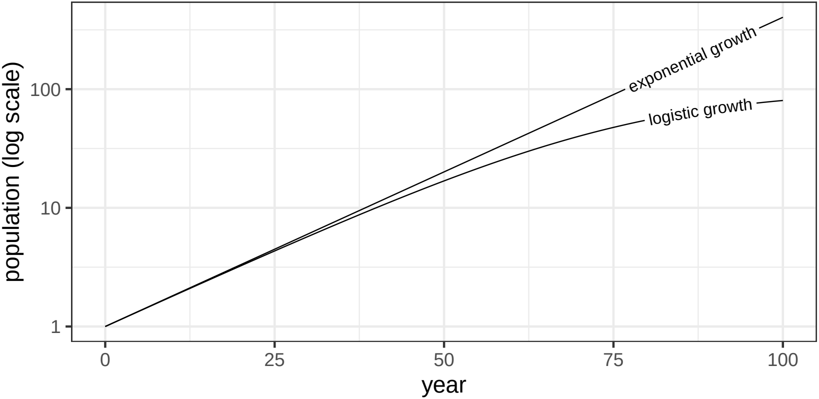
Exponential vs. logistic growth (r = .06)

```
ggplot(tibble(x = 1)) + xlim(0, 100) +
 labs(x = "year", y = "population") +
 geom_textpath(fun = ~ exp(.06 *.x), hjust = .95, size = 5,
   stat = "function", label = "exponential growth") +
 geom_textpath(stat = "function", size = 5,
   fun = ~ 100 / (1 + (100 - 1) / 1 * exp(-.06 * .x)),
   label = "logistic growth", hjust = .95)
```



Exponential vs. logistic growth (r = .06)

```
ggplot(tibble(x = 1)) + xlim(0, 100) + scale_y_log10() +
 labs(x = "year", y = "population (log scale)") +
 geom_textpath(fun = ~ exp(.06 *.x), hjust = .95, size = 5,
   stat = "function", label = "exponential growth") +
 geom_textpath(stat = "function", size = 5,
   fun = ~ 100 / (1 + (100 - 1) / 1 * exp(-.06 * .x)),
   label = "logistic growth", hjust = .95)
```



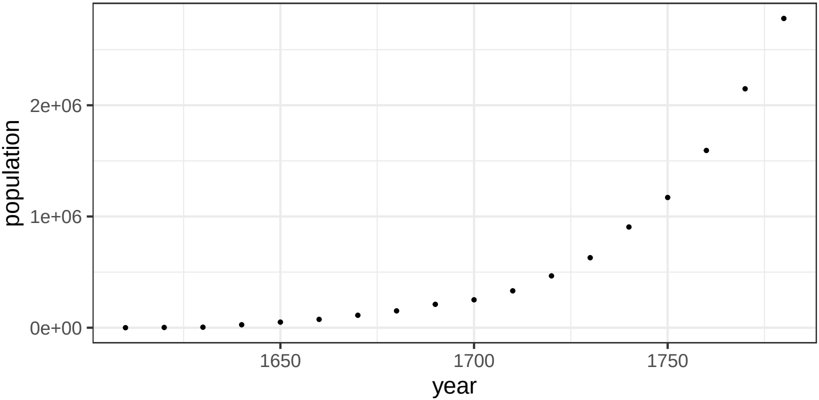
Population of American Colonies 1610 - 1790

```r
USA_pre1790 <-
  tibble(x = seq(1610, 1780, 10),
         y = c(350, 2302, 4646, 26634, 50368, 75058,
               111935, 151507, 210372, 250888, 331711,
               466185, 629445, 905563, 1170760, 1593625,
               2148076, 2780369))

USA_pre1790 %>%
  top_n(5) %>%
  pivot_wider(names_from = x,
              values_from = y) %>%
  kable(format.args = list(big.mark = ","))
```

1740	1750	1760	1770	1780
905,563	1,170,760	1,593,625	2,148,076	2,780,369

Population of American Colonies 1610 - 1790

```
pop_plot <-
  USA_pre1790 %>%
  ggplot() +
  aes(x, y) +
  labs(x = "year", y = "population")

pop_plot + geom_point()
```



Exponential curve fit to Colonies population

```
exponential = "y ~ exp(x * r) * y_0"
start <- list(y_0 = 1, r = log(2) / 24)
nls(formula = exponential, data = USA_pre1790,
    start = start) %>%
  coefficients() %>%
  t() %>%
  as_tibble() %>%
  kable(col.names = c("$y_0$", "r"),
        digits = 2)
```

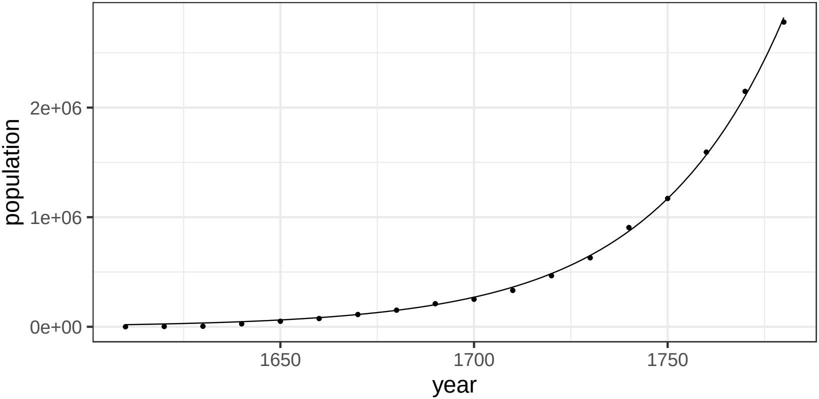
y_0	r
0	0.03

Interpretation:

1. Population increases $\approx$ 3 percent a year.
2. Population will double approximate every log(2) / 0.03 $\approx$ 23 years.

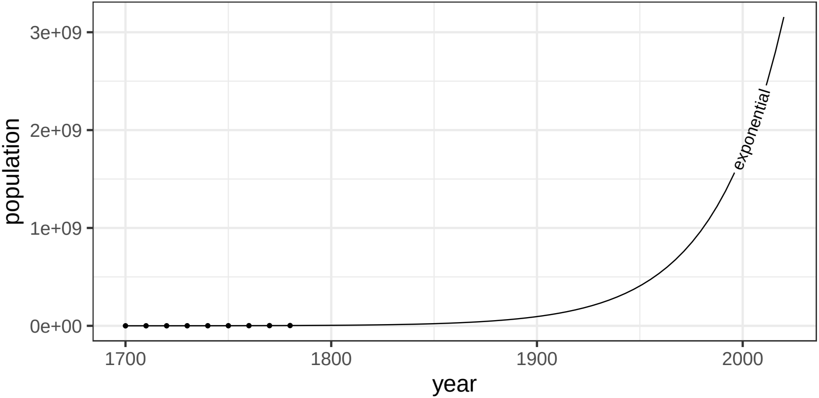
If Colonies continued to grow at ≈ 3% a year…

```
pop_plot + geom_point() +
  geom_smooth(method = "nls", se = FALSE,
              color = "black", linewidth = .5,
              formula = exponential,
              method.args = list(start = start))
```



… population would have hit 3 billion in 2020

```r
pop_plot + geom_point() + xlim(1700, 2020) +
  geom_textsmooth(label = "exponential", method = "nls",
                  formula = exponential, hjust = .9,
                  fullrange = T, se = F, size = 5,
                  method.args = list(start = start))
```



Logistic curve fit to Colonies population

```
logistic <- "y ~ SSlogis(x, Asym, xmid, scal)"
nls(logistic, data = USA_pre1790) %>%
  coefficients() %>%
  t() %>%
  as_tibble() %>%
  transmute(`$y_0$` = Asym / (1 + exp(xmid / scal)),
            K = Asym,
            r = 1/scal) %>%
  kable(digits = 2,
        format.args = list(big.mark = ","))
```

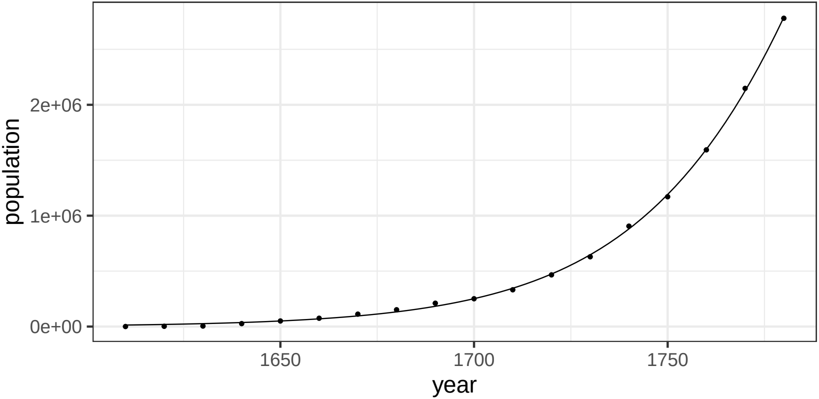
y_0	K	r
0	15,100,062	0.03

Interpretation:

1. Population increases first at $\approx$ 3 percent a year, slowing thereafter.
2. Population will reach a maximum capacity at 15 million.

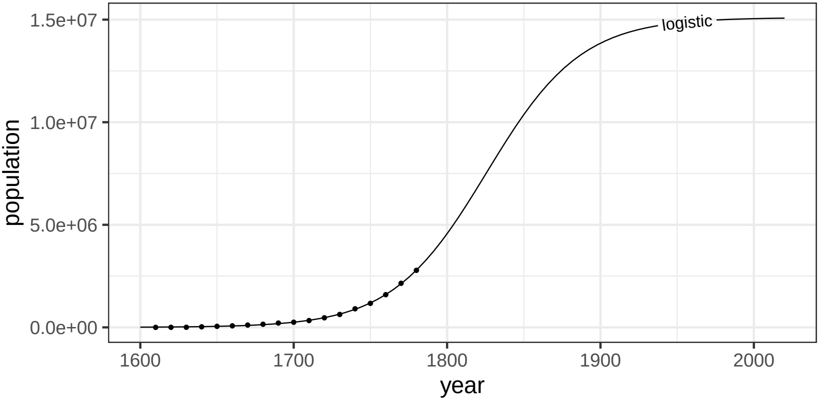
If $r \approx 3\%$ growth with $\approx$ 15,000,000 capacity...

```
pop_plot + geom_point() +
  geom_smooth(method = "nls", formula = logistic,
              se = FALSE, color = "black", linewidth = .5)
```



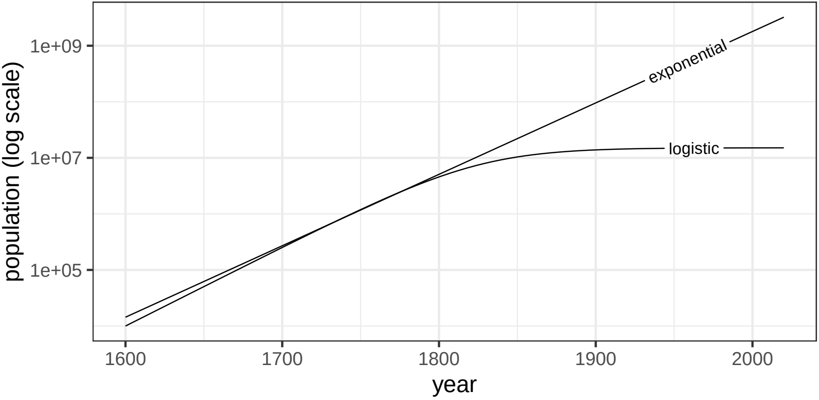
… population would be close to capacity by 1920

```
pop_plot <- pop_plot + xlim(1600, 2020) +
  geom_textsmooth(label = "logistic", formula = logistic,
                  hjust = .9, fullrange = T, size = 5,
                  method = "nls", se = F)

pop_plot + geom_point()
```



Colonies population: exponential vs. logistic growth

```
(pop_plot <- pop_plot + coord_trans(y = "log10") +
   geom_textsmooth(label = "exponential", method = "nls",
     formula = exponential, hjust = .9, fullrange = T,
     size = 5, se = F, method.args = list(start = start)) +
   scale_y_continuous(name = "population (log scale)",
     br = 10^c(5,7,9), mi = 10^c(4,6,8)) + xlim(1600,2020))
```



Derivation of exponential growth

- Exponential growth $y_x = e^{rx}\ y_0$ is the solution to the separable differential equation
$$\frac{dy}{dx} = ry_x \qquad (1)$$

- This differential equation can be solved by separation of variables: Equation (1) is "rewritten" as $y_x^{-1} dy = r dx$

- Integrating the left side with respect to y and the right side with respect to x yields $\log(y_x) = rx + c$. Solving for y_x results in $y_x = e^{rx+c}$

- Since $y_0 = e^c$, $c = \log(y_0)$, and thus $y_x = e^{rx}\ y_0$

Derivation of logistic growth

- Logistic growth $y_x = K\left(1 + \frac{K - y_0}{y_0}\ e^{-rx}\right)^{-1}$ is the solution to the seperable differential equation

$$\frac{dy}{dx} = ry_x\left(1 - \frac{y_x}{K}\right) \quad (2)$$

- This differential equation can be solved by separation of variables: Equation (2) is "rewritten" as $\left[y_x\left(1 - \frac{y_x}{K}\right)\right]^{-1} dy = rdx$
  - Note that the left side is equal to $\left[y_x^{-1} - (y_x - K)^{-1}\right] dy$

- Integrating the left side with respect to y and the right side with respect to x yields $\log(y_x) - \log(y_x - K) = rx + c$. Solving for y_x results in $y_x = K\left(1 - e^{-(rx+c)}\right)^{-1}$

- Since $y_0 = K(1 - e^{-c})^{-1}$, $c = -\log\left(\frac{y_0 - K}{y_0}\right)$, and thus $y_x = K\left(1 + \frac{K - y_0}{y_0}\ e^{-rx}\right)^{-1}$

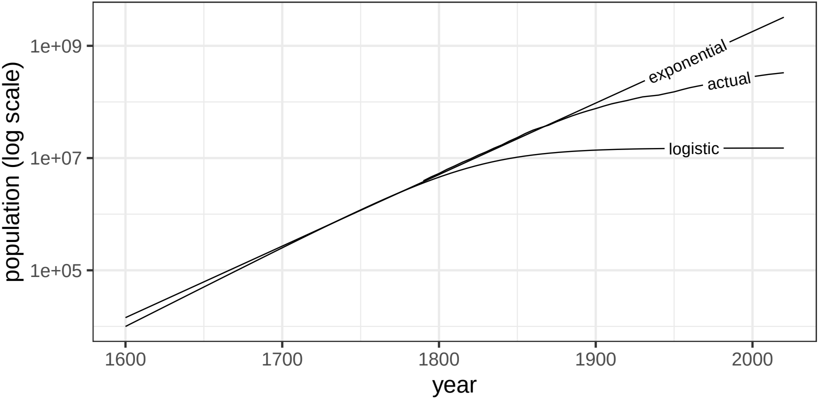
U.S. population 1790 - 2020 (Decennial Census)

```
USA <-
  tibble(
  x = seq(1790, 2020, 10),
  y = c(3929214, 5308483, 7239881, 9638453,
        12866020, 17069453, 23191876, 31443321,
        38558371, 50189209, 62979766, 76212168,
        92228496, 106021537, 123202624, 132164569,
        151325798, 179323175, 203211926, 226545805,
        248709873, 281421906, 308745538, 331449281))

USA %>%
  top_n(5) %>%
  pivot_wider(names_from = x, values_from = y) %>%
  kable(format.args = list(big.mark = ","))
```

1980	1990	2000	2010	2020
226,545,805	248,709,873	281,421,906	308,745,538	331,449,281

U.S. population: growth historically closer to exponential but appears to be reaching capacity

```
pop_plot +
  geom_textline(label = "actual",
                data = USA,
                hjust = .9, size = 5)
```



U.S. population 1790 - 2020 (Decennial Census)

```
logistic <- "y ~ SSlogis(x, Asym, xmid, scal)"
nls(logistic, data = USA) %>%
  coefficients() %>%
  t() %>%
  as_tibble() %>%
  transmute(`$y_0$` = Asym / (1 + exp(xmid / scal)),
            K = Asym,
            r = 1/scal) %>%
  kable(digits = 2,
        format.args = list(big.mark = ","))
```

y_0	K	r
0	499,344,363	0.02

Interpretation:

1. Population increases first at $\approx$ 2 percent a year, slowing thereafter.
2. Population will reach a maximum capacity at 500 million.

Why didn't the U.S. reach capacity as predicted?

- Malthus believed limited resources constrain population growth.
  - He argued overpopulation inevitably leads to war or famine.
  - Though not the first to make this argument, the timing of his work after the American and French revolutions made it extremely popular
- Many have resurrected Malthus' argument over the years.
  - Pearl and Reed (1920) warned that agriculture production was not keeping pace with population growth. Initially unaware of Verhulst, they rediscovered the logistic growth curve.
  - More recently, Erlich predicted in *The Population Bomb* (1968) that hundreds of millions of people would starve to death in the 1970s.
- Doomsday predictions such as these have not (yet) come to fruition.
  - Largely due to industrialization and expansion in the 1800s and the Green Revolution that increased crop harvests in the 1970s.
  - While population growth does stress environments, public concern is often elevated during periods of xenophobia.

Verhulst's legacy and the logistic function today

- Verhulst died in 1849 at the age of 45, four years after the publication of *Mathematical Investigations on the Law of Population Growth*.
  - The logistic function was rediscovered several times before Verhulst's work finally reached a wide audience.
- The frequent rediscovery of the logistic function is likely due to the regularity with which it occurs in the study of statistics.
  - For example, the logistic function arises naturally in the study of the binomial distribution and is commonly used in logistic regression and artificial neural networks.
  - It also has a variety of special properties that make it ideal for studying a wide range of phenomenon, from the onset of rare diseases to the choices made by economic agents.

References

1. Bacaër, N. A short history of mathematical population dynamics. Vol. 618. Springer, 2011.

2. Erlich, Paul R. The Population Bomb. 1971.

3. Malthus, T. R. An essay on the principle of population. 1798.

4. Pearl, R., and L. J. Reed. "On the rate of growth of the population of the United States since 1790 and its mathematical representation." PNAS. 6.6 (1920): 275-288.

5. US Census Bureau. "Historical Statistics of the United States: Colonial Times to 1970." 1975. https://www.census.gov/library/publications/1975/compendia/hist_stats_colonial-1970.html

6. Verhulst, P. F. (1845). "Mathematical Investigations on the Law of Population Growth." In David, H. and A.W.F. Edwards Eds. Annotated Readings in the History of Statistics. (2001): 69-75.

Appendix: Non-Linear Least Squares

- Exponential and logistic growth are examples of nonlinear models.
  - Let $y_x = f_x(\theta)$ denote population growth.
  - For example, if $f_x = e^{rx}\ y_0$, then $\theta = \{r, y_0\}$
  - Non-Linear Least Squares: $\min_{\theta}\ \sum_i [y_{x_i} - f_{x_i}(\theta)]^2$

- Just like Least Squares, we solve $\frac{\partial}{\partial\theta} \sum_i [y_{x_i} - f_{x_i}(\theta)]^2 \overset{set}{=} 0$.

- But unlike Least Squares, θ cannot usually be expressed as a function of y_{x_i} and x_i in closed form. Instead, θ is often calculated using the Gauss-Newton algorithm:
  1. We make an initial guess of the parameter values, $\theta^{(0)}$
  2. We repeatedly update our guess using the formula:

$$\theta^{(t+1)} = \theta^{(t)} + (J^T J)^{-1} J^T r$$

where residual vector $r = [y_{x_1} - f_{x_1}(\theta),\ y_{x_2} - f_{x_2}(\theta),\ \ldots\ ]^T$ and Jacobian matrix $J_{ij} = \frac{\partial}{\partial\theta_j}[y_{x_i} - f_{x_i}(\theta_j)]^2$ are evaluated at $\theta^{(t)}$.