Political Methodology III: Model Based Inference

Justin Grimmer

Professor
Department of Political Science
Stanford University

April 9th, 2019

Statistical Inference

- Model based inference:
- Assume: data generated via distributional process
- Thursday: Derived a likelihood for single/bivariate problems
- Today: Interested in conditional relationships
- Today: Build Linear Model Using Likelihood

Bivariate $\rightsquigarrow$ Multivariate $\rightsquigarrow$ Computational Model

Regression Model$\rightsquigarrow$ Time for Change

Model incumbent vote share, Y_i

1) X_{i1} =Incumbent Presidential Popularity (Net Approval/Disapproval)
2) X_{i2} =GDP Growth
3) X_{i3} =First-term Incumbent In Race

$\boldsymbol{X}_i = (X_{i1}, X_{i2}, X_{i3})$

Bivariate Regression Model via Maximum Likelihood

Linear regression model:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{i1} + \epsilon_i \\ \epsilon_i &\sim \textsf{Normal}(0, \sigma^2) \end{aligned}$$

Bivariate Regression Model via Maximum Likelihood

Linear regression model:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{i1} + \epsilon_i \\ \epsilon_i &\sim \text{Normal}(0, \sigma^2) \end{aligned}$$

Equivalently:

$$\begin{aligned} Y_i &\sim \text{Normal}(\mu_i, \sigma^2) \\ E[Y_i|X_{i1}] = \mu_i &= \beta_0 + \beta_1 X_{i1} \end{aligned}$$

Bivariate Regression Model via Maximum Likelihood

Linear regression model:

$$
\begin{aligned}
Y_i &= \beta_0 + \beta_1 X_{i1} + \epsilon_i \\
\epsilon_i &\sim \mathsf{Normal}(0, \sigma^2)
\end{aligned}
$$

Equivalently:

$$
\begin{aligned}
Y_i &\sim \mathsf{Normal}(\mu_i, \sigma^2) \\
E[Y_i|X_{i1}] = \mu_i &= \beta_0 + \beta_1 X_{i1}
\end{aligned}
$$

Parameters:

$$
\begin{aligned}
\boldsymbol{\beta} &= (\beta_0, \beta_1) \\
\sigma^2 &
\end{aligned}
$$

Bivariate Regression Model via Maximum Likelihood

$$L(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{Y}, \boldsymbol{X}) = f(\boldsymbol{Y} | \boldsymbol{\beta}, \sigma^2, \boldsymbol{X})$$

Bivariate Regression Model via Maximum Likelihood

$$\begin{aligned} L(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{Y}, \boldsymbol{X}) &= f(\boldsymbol{Y} | \boldsymbol{\beta}, \sigma^2, \boldsymbol{X}) \\ &= \prod_{i=1}^{N} f(Y_i | \boldsymbol{\beta}, \sigma^2, X_{i1}) \end{aligned}$$

Bivariate Regression Model via Maximum Likelihood

$$
\begin{aligned}
L(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{Y}, \boldsymbol{X}) & = f(\boldsymbol{Y} | \boldsymbol{\beta}, \sigma^2, \boldsymbol{X}) \\
& = \prod_{i=1}^{N} f(Y_i | \boldsymbol{\beta}, \sigma^2, X_{i1}) \\
& = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(Y_i - \beta_0 - \beta_1 X_{i1})^2}{2\sigma^2} \right)
\end{aligned}
$$

Bivariate Regression Model via Maximum Likelihood

$$
\begin{aligned}
L(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{Y}, \boldsymbol{X}) &= f(\boldsymbol{Y}|\boldsymbol{\beta}, \sigma^2, \boldsymbol{X}) \\
&= \prod_{i=1}^{N} f(Y_i|\boldsymbol{\beta}, \sigma^2, X_{i1}) \\
&= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(Y_i - \beta_0 - \beta_1 X_{i1})^2}{2\sigma^2}\right) \\
&= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\sum_{i=1}^{N} \frac{(Y_i - \beta_0 - \beta_1 X_{i1})^2}{2\sigma^2}\right)
\end{aligned}
$$

Bivariate Regression Model via Maximum Likelihood

$$L(\boldsymbol{\beta}, \sigma^2|\boldsymbol{Y}, \boldsymbol{X}) \quad = \quad \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\sum_{i=1}^{N} \frac{(Y_i - \beta_0 - \beta_1 X_{i1})^2}{2\sigma^2}\right)$$

Bivariate Regression Model via Maximum Likelihood

$$
\begin{aligned}
L(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{Y}, \boldsymbol{X}) &= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(- \sum_{i=1}^{N} \frac{(Y_i - \beta_0 - \beta_1 X_{i1})^2}{2\sigma^2} \right) \\
\log L(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{Y}, \boldsymbol{X}) &= \underbrace{-\frac{N}{2} \log(2\pi)}_{\text{constant}} - \frac{N}{2} \log(\sigma^2) - \sum_{i=1}^{N} \frac{(Y_i - \beta_0 - \beta_1 X_{i1})^2}{2\sigma^2}
\end{aligned}
$$

Bivariate Regression Model via Maximum Likelihood

$$
\begin{aligned}
L(\boldsymbol{\beta}, \sigma^2|\boldsymbol{Y}, \boldsymbol{X}) &= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\sum_{i=1}^{N} \frac{(Y_i - \beta_0 - \beta_1 X_{i1})^2}{2\sigma^2}\right) \\
\log L(\boldsymbol{\beta}, \sigma^2|\boldsymbol{Y}, \boldsymbol{X}) &= \underbrace{-\frac{N}{2}\log(2\pi)}_{\text{constant}} - \frac{N}{2}\log(\sigma^2) - \sum_{i=1}^{N} \frac{(Y_i - \beta_0 - \beta_1 X_{i1})^2}{2\sigma^2} \\
l(\boldsymbol{\beta}, \sigma^2|\boldsymbol{Y}, \boldsymbol{X}) &= -\frac{N}{2}\log(\sigma^2) - \sum_{i=1}^{N} \frac{(Y_i - \beta_0 - \beta_1 X_{i1})^2}{2\sigma^2} + c
\end{aligned}
$$

Bivariate Regression Model via Maximum Likelihood

$$
\begin{aligned}
L(\boldsymbol{\beta}, \sigma^2|\boldsymbol{Y}, \boldsymbol{X}) &= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(- \sum_{i=1}^{N} \frac{(Y_i - \beta_0 - \beta_1 X_{i1})^2}{2\sigma^2}\right) \\
\log L(\boldsymbol{\beta}, \sigma^2|\boldsymbol{Y}, \boldsymbol{X}) &= \underbrace{-\frac{N}{2}\log(2\pi)}_{\text{constant}} - \frac{N}{2}\log(\sigma^2) - \sum_{i=1}^{N} \frac{(Y_i - \beta_0 - \beta_1 X_{i1})^2}{2\sigma^2} \\
l(\boldsymbol{\beta}, \sigma^2|\boldsymbol{Y}, \boldsymbol{X}) &= -\frac{N}{2}\log(\sigma^2) - \sum_{i=1}^{N} \frac{(Y_i - \beta_0 - \beta_1 X_{i1})^2}{2\sigma^2} + c
\end{aligned}
$$

Optimize with respect to:

- $\boldsymbol{\beta}$
- σ^2

Bivariate Regression Model via Maximum Likelihood

$$
\begin{aligned}
L(\boldsymbol{\beta}, \sigma^2|\boldsymbol{Y}, \boldsymbol{X}) &= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\sum_{i=1}^{N}\frac{(Y_i - \beta_0 - \beta_1 X_{i1})^2}{2\sigma^2}\right) \\
\log L(\boldsymbol{\beta}, \sigma^2|\boldsymbol{Y}, \boldsymbol{X}) &= \underbrace{-\frac{N}{2}\log(2\pi)}_{\text{constant}} - \frac{N}{2}\log(\sigma^2) - \sum_{i=1}^{N}\frac{(Y_i - \beta_0 - \beta_1 X_{i1})^2}{2\sigma^2} \\
l(\boldsymbol{\beta}, \sigma^2|\boldsymbol{Y}, \boldsymbol{X}) &= -\frac{N}{2}\log(\sigma^2) - \sum_{i=1}^{N}\frac{(Y_i - \beta_0 - \beta_1 X_{i1})^2}{2\sigma^2} + c
\end{aligned}
$$

Optimize with respect to:

- $\boldsymbol{\beta}$
- σ^2

Bivariate Regression Model via Maximum Likelihood

$$\begin{aligned}
\frac{\partial l(\boldsymbol{\beta}, \sigma^2|\boldsymbol{Y}, \boldsymbol{X})}{\partial \beta_0} &= \sum_{i=1}^{N} \frac{Y_i - \beta_0 - \beta_1 X_{i1}}{\sigma^2} \\
\frac{\partial l(\boldsymbol{\beta}, \sigma^2|\boldsymbol{Y}, \boldsymbol{X})}{\partial \beta_1} &= \sum_{i=1}^{N} \frac{Y_i - \beta_0 - \beta_1 X_{i1}}{\sigma^2} X_{i1} \\
\frac{\partial l(\boldsymbol{\beta}, \sigma^2|\boldsymbol{Y}, \boldsymbol{X})}{\partial \sigma^2} &= -\frac{N}{2\sigma^2} + \sum_{i=1}^{N} \frac{(Y_i - \beta_0 - \beta_1 X_{i1})^2}{2(\sigma^2)^2}
\end{aligned}$$

Bivariate Regression Model via Maximum Likelihood

$$\frac{\partial l(\boldsymbol{\beta}, \sigma^2|\boldsymbol{Y}, \boldsymbol{X})}{\partial \beta_0} = \sum_{i=1}^{N} \frac{Y_i - \beta_0 - \beta_1 X_{i1}}{\sigma^2}$$

$$\frac{\partial l(\boldsymbol{\beta}, \sigma^2|\boldsymbol{Y}, \boldsymbol{X})}{\partial \beta_1} = \sum_{i=1}^{N} \frac{Y_i - \beta_0 - \beta_1 X_{i1}}{\sigma^2} X_{i1}$$

$$\frac{\partial l(\boldsymbol{\beta}, \sigma^2|\boldsymbol{Y}, \boldsymbol{X})}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \sum_{i=1}^{N} \frac{(Y_i - \beta_0 - \beta_1 X_{i1})^2}{2(\sigma^2)^2}$$

Score function: $s_i(\boldsymbol{\beta}, \sigma^2) = \nabla l(\boldsymbol{\beta}, \sigma^2|Y_i, X_i) =$

$$\left(\frac{Y_i - \beta_0 - \beta_1 X_{i1}}{\sigma^2}, \frac{Y_i - \beta_0 - \beta_1 X_{i1}}{\sigma^2} X_{i1}, -\frac{1}{2\sigma^2} + \frac{(Y_i - \beta_0 - \beta_1 X_{i1})^2}{2(\sigma^2)^2}\right)$$

Bivariate Regression Model via Maximum Likelihood

$$\frac{\partial l(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{Y}, \boldsymbol{X})}{\partial \beta_0} = \sum_{i=1}^{N} \frac{Y_i - \beta_0 - \beta_1 X_{i1}}{\sigma^2}$$

$$\frac{\partial l(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{Y}, \boldsymbol{X})}{\partial \beta_1} = \sum_{i=1}^{N} \frac{Y_i - \beta_0 - \beta_1 X_{i1}}{\sigma^2} X_{i1}$$

$$\frac{\partial l(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{Y}, \boldsymbol{X})}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \sum_{i=1}^{N} \frac{(Y_i - \beta_0 - \beta_1 X_{i1})^2}{2(\sigma^2)^2}$$

Score function: $s_i(\boldsymbol{\beta}, \sigma^2) = \nabla l(\boldsymbol{\beta}, \sigma^2 | Y_i, X_i) =$

$$\left(\frac{Y_i - \beta_0 - \beta_1 X_{i1}}{\sigma^2}, \frac{Y_i - \beta_0 - \beta_1 X_{i1}}{\sigma^2} X_{i1}, -\frac{1}{2\sigma^2} + \frac{(Y_i - \beta_0 - \beta_1 X_{i1})^2}{2(\sigma^2)^2} \right)$$

$s_N(\boldsymbol{\beta}, \sigma^2) = \sum_{i=1}^{N} s_i(\boldsymbol{\beta}, \sigma^2)$.

Bivariate Regression Model via Maximum Likelihood

$$
\begin{aligned}
0 &= \sum_{i=1}^{N} \frac{Y_i - \beta_0^{*} - \beta_1^{*} X_{i1}}{\sigma^{2,*}} \\
0 &= \sum_{i=1}^{N} \frac{Y_i - \beta_0^{*} - \beta_1^{*} X_{i1}}{\sigma^{2,*}} X_{i1} \\
0 &= -\frac{N}{2\sigma^{2,*}} + \sum_{i=1}^{N} \frac{(Y_i - \beta_0^{*} - \beta_1^{*} X_{i1})^2}{2(\sigma^{2,*})^2}
\end{aligned}
$$

Algebra

Bivariate Regression Model via Maximum Likelihood

$$
\begin{aligned}
\beta_1^* &= \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^N (X_i - \bar{X})^2} \\
\beta_0^* &= \bar{Y} - \beta_1^* \bar{X} \\
\sigma^{2,*} &= \frac{1}{N} \sum_{i=1}^N (Y_i - \beta_0^* - \beta_1^* X_i)^2
\end{aligned}
$$

Hessian Matrix

Define the Hessian for observation i as

$$H_i(\boldsymbol{\beta}, \sigma^2) \quad = \quad \begin{pmatrix} \frac{\partial l(\boldsymbol{\beta},\sigma^2|X_i,Y_i)}{\partial \beta_0 \beta_0} & \frac{\partial l(\boldsymbol{\beta},\sigma^2|X_i,Y_i)}{\partial \beta_0 \beta_1} & \cdots & \frac{\partial l(\boldsymbol{\beta},\sigma^2|X_i,Y_i)}{\partial \beta_0 \sigma^2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial l(\boldsymbol{\beta},\sigma^2|X_i,Y_i)}{\partial \beta_0 \sigma^2} & \frac{\partial l(\boldsymbol{\beta},\sigma^2|X_i,Y_i)}{\partial \beta_1 \sigma^2} & \cdots & \frac{\partial l(\boldsymbol{\beta},\sigma^2|X_i,Y_i)}{\partial \sigma^2 \sigma^2} \end{pmatrix}$$

Bivariate Regression and Maximum Likelihood

Fisher Information matrix

$$\begin{aligned}
I(\boldsymbol{\beta}, \sigma^2) &= E[s_N(\boldsymbol{\beta}, \sigma^2)s_N(\boldsymbol{\beta}, \sigma^2)^{'}] \\
I(\boldsymbol{\beta}, \sigma^2) &= -E[H_i(\boldsymbol{\beta}, \sigma^2)] \text{ (under regularity conditions)} \\
I_{ij}(\boldsymbol{\beta}, \sigma^2) &= -E[\frac{\partial^2 l(\boldsymbol{\beta}, \sigma^2 | Y_i, \boldsymbol{X}_i)}{\partial \beta_i \beta_j} | \boldsymbol{\beta}, \sigma^2] \\
I_{ij}(\boldsymbol{\beta}, \sigma^2) &= - \int \frac{\partial^2 l(\boldsymbol{\beta}, \sigma^2 | Y_i, \boldsymbol{X}_i)}{\partial \beta_i \beta_j} f(Y_i | \boldsymbol{\beta}, \sigma^2, \boldsymbol{X}) d\boldsymbol{Y}
\end{aligned}$$

Observed Information Matrix

$$I_N(\boldsymbol{\beta}, \sigma^2) = -\sum_{i=1}^{N} H_i(\boldsymbol{\beta}, \sigma^2) = -\sum_{i=1}^{N} \nabla^2 \log l(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{X}_i, Y_i)$$

Three Ways to Estimate Variance

$$
\begin{aligned}
\text{Var}(\widehat{\boldsymbol{\beta}, \sigma^2})_{OS} &= \left[\sum_{i=1}^{N} s_i(\boldsymbol{\beta}, \sigma^2) s_i(\boldsymbol{\beta}, \sigma^2)^{'}\right]^{-1} \quad \text{(BHHH estimator)} \\
\text{Var}(\widehat{\boldsymbol{\beta}, \sigma^2})_{OH} &= -\left[\sum_{i=1}^{N} H_i(\boldsymbol{\beta}, \sigma^2)\right]^{-1} \\
\text{Var}(\widehat{\boldsymbol{\beta}, \sigma^2})_{EH} &= -\left[\sum_{i=1}^{N} E[H_i(\boldsymbol{\beta}, \sigma^2)]\right]^{-1}
\end{aligned}
$$

Bivariate Regression Model via Maximum Likelihood

$$
\begin{aligned}
\frac{\partial^2 l(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{Y}, \boldsymbol{X})}{\partial \beta_0 \partial \beta_0} &= -\sum_{i=1}^{N} \frac{1}{\sigma^2} \\
\frac{\partial^2 l(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{Y}, \boldsymbol{X})}{\partial \beta_1 \partial \beta_1} &= -\sum_{i=1}^{N} \frac{X_i^2}{\sigma^2} \\
\frac{\partial^2 l(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{Y}, \boldsymbol{X})}{\partial \sigma^2 \partial \sigma^2} &= \frac{N}{2(\sigma^2)^2} - \sum_{i=1}^{N} \frac{(Y_i - \beta_0 - \beta_1 X_{i1})^2}{(\sigma^2)^4} \\
\frac{\partial^2 l(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{Y}, \boldsymbol{X})}{\partial \beta_0 \partial \beta_1} &= -\sum_{i=1}^{N} \frac{X_i}{\sigma^2} \\
\frac{\partial^2 l(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{Y}, \boldsymbol{X})}{\partial \sigma^2 \partial \beta_1} &= -\sum_{i=1}^{N} \frac{(Y_i - \beta_0 - \beta_1 X_{i1}) X_i}{(\sigma^2)^2}
\end{aligned}
$$

Bivariate Regression Model via Maximum Likelihood

$$-E\left[\frac{\partial^2 l(\boldsymbol{\beta}, \sigma^2|\boldsymbol{Y}, \boldsymbol{X})}{\partial\beta_0\partial\beta_0}\right] = -E\left[-\frac{1}{\sigma^2}\right]$$

Bivariate Regression Model via Maximum Likelihood

$$
\begin{aligned}
-E\left[\frac{\partial^2 l(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{Y}, \boldsymbol{X})}{\partial \beta_0 \partial \beta_0}\right] &= -E\left[-\frac{1}{\sigma^2}\right] \\
&= -\sum_{i=1}^{N} \int_{-\infty}^{\infty} -\frac{1}{\sigma^2} f(Y_i | \boldsymbol{\beta}, \sigma^2, X_i) dY_i
\end{aligned}
$$

Bivariate Regression Model via Maximum Likelihood

$$
\begin{aligned}
-E\left[\frac{\partial^2 l(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{Y}, \boldsymbol{X})}{\partial \beta_0 \partial \beta_0}\right] &= -E\left[-\frac{1}{\sigma^2}\right] \\
&= -\sum_{i=1}^{N} \int_{-\infty}^{\infty} -\frac{1}{\sigma^2} f(Y_i | \boldsymbol{\beta}, \sigma^2, X_i) dY_i \\
&= \sum_{i=1}^{N} \frac{1}{\sigma^2} \int_{-\infty}^{\infty} f(Y_i | \boldsymbol{\beta}, \sigma^2, X_i) dY_i
\end{aligned}
$$

Bivariate Regression Model via Maximum Likelihood

$$
\begin{aligned}
-E\left[\frac{\partial^2 l(\boldsymbol{\beta}, \sigma^2|\boldsymbol{Y}, \boldsymbol{X})}{\partial \beta_0 \partial \beta_0}\right] &= -E\left[-\frac{1}{\sigma^2}\right] \\
&= -\sum_{i=1}^{N} \int_{-\infty}^{\infty} -\frac{1}{\sigma^2} f(Y_i|\boldsymbol{\beta}, \sigma^2, X_i) dY_i \\
&= \sum_{i=1}^{N} \frac{1}{\sigma^2} \int_{-\infty}^{\infty} f(Y_i|\boldsymbol{\beta}, \sigma^2, X_i) dY_i \\
&= \sum_{i=1}^{N} \frac{1}{\sigma^2} \times 1 = \frac{N}{\sigma^2}
\end{aligned}
$$

Bivariate Regression Model via Maximum Likelihood

$$
\begin{aligned}
-E[\frac{\partial^2 l(\boldsymbol{\beta}, \sigma^2|\boldsymbol{Y}, \boldsymbol{X})}{\partial \beta_1 \partial \beta_1}] &= -E[-\sum_{i=1}^{N} \frac{X_i^2}{\sigma^2}] \\
&= \sum_{i=1}^{N} \frac{X_i^2}{\sigma^2}
\end{aligned}
$$

Bivariate Regression Model via Maximum Likelihood

$$
\begin{aligned}
-E[\frac{\partial^2 l(\boldsymbol{\beta}, \sigma^2|\boldsymbol{Y}, \boldsymbol{X})}{\partial \beta_1 \partial \beta_1}] &= -E[-\sum_{i=1}^{N} \frac{X_i^2}{\sigma^2}] \\
&= \sum_{i=1}^{N} \frac{X_i^2}{\sigma^2} \\
-E[\frac{\partial^2 l(\boldsymbol{\beta}, \sigma^2|\boldsymbol{Y}, \boldsymbol{X})}{\partial \beta_0 \partial \beta_1}] &= -E[-\sum_{i=1}^{N} \frac{X_i}{\sigma^2}] \\
&= \sum_{i=1}^{N} \frac{X_i}{\sigma^2}
\end{aligned}
$$

$$-E[\frac{\partial^2 l(\boldsymbol{\beta}, \sigma^2|\boldsymbol{Y}, \boldsymbol{X})}{\partial \sigma^2 \partial \sigma^2}] = E[\frac{N}{2(\sigma^2)^2} - \sum_{i=1}^{N} \frac{(Y_i - \beta_0 - \beta_1 X_{i1})^2}{(\sigma^2)^4}]$$

$$
\begin{aligned}
-E[\frac{\partial^2 l(\boldsymbol{\beta}, \sigma^2|\boldsymbol{Y}, \boldsymbol{X})}{\partial\sigma^2\partial\sigma^2}] &= E[\frac{N}{2(\sigma^2)^2} - \sum_{i=1}^{N}\frac{(Y_i - \beta_0 - \beta_1 X_{i1})^2}{(\sigma^2)^4}] \\
&= -\frac{N}{2(\sigma^2)^2} + \sum_{i=1}^{N} E\left[\frac{(Y_i - \beta_0 - \beta_1 X_{i1})^2}{(\sigma^2)^4}\right]
\end{aligned}
$$

$$
\begin{aligned}
-E[\frac{\partial^2 l(\boldsymbol{\beta}, \sigma^2|\boldsymbol{Y}, \boldsymbol{X})}{\partial \sigma^2 \partial \sigma^2}] &= E[\frac{N}{2(\sigma^2)^2} - \sum_{i=1}^{N} \frac{(Y_i - \beta_0 - \beta_1 X_{i1})^2}{(\sigma^2)^4}] \\
&= -\frac{N}{2(\sigma^2)^2} + \sum_{i=1}^{N} E\left[\frac{(Y_i - \beta_0 - \beta_1 X_{i1})^2}{(\sigma^2)^4}\right] \\
&= -\frac{N}{2(\sigma^2)^2} + \sum_{i=1}^{N} \frac{\sigma^2}{(\sigma^2)^4}
\end{aligned}
$$

$$
\begin{aligned}
-E[\frac{\partial^2 l(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{Y}, \boldsymbol{X})}{\partial \sigma^2 \partial \sigma^2}] &= E[\frac{N}{2(\sigma^2)^2} - \sum_{i=1}^{N} \frac{(Y_i - \beta_0 - \beta_1 X_{i1})^2}{(\sigma^2)^4}] \\
&= -\frac{N}{2(\sigma^2)^2} + \sum_{i=1}^{N} E\left[\frac{(Y_i - \beta_0 - \beta_1 X_{i1})^2}{(\sigma^2)^4}\right] \\
&= -\frac{N}{2(\sigma^2)^2} + \sum_{i=1}^{N} \frac{\sigma^2}{(\sigma^2)^4} \\
&= -\frac{N}{2(\sigma^2)^2} + \frac{2N}{2(\sigma^2)^2}
\end{aligned}
$$

$$
\begin{aligned}
-E[\frac{\partial^2 l(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{Y}, \boldsymbol{X})}{\partial \sigma^2 \partial \sigma^2}] &= E[\frac{N}{2(\sigma^2)^2} - \sum_{i=1}^{N} \frac{(Y_i - \beta_0 - \beta_1 X_{i1})^2}{(\sigma^2)^4}] \\
&= -\frac{N}{2(\sigma^2)^2} + \sum_{i=1}^{N} E\left[\frac{(Y_i - \beta_0 - \beta_1 X_{i1})^2}{(\sigma^2)^4}\right] \\
&= -\frac{N}{2(\sigma^2)^2} + \sum_{i=1}^{N} \frac{\sigma^2}{(\sigma^2)^4} \\
&= -\frac{N}{2(\sigma^2)^2} + \frac{2N}{2(\sigma^2)^2} \\
&= \frac{N}{2(\sigma^2)^2}
\end{aligned}
$$

$$-E\left[\frac{\partial^2 l(\boldsymbol{\beta},\sigma^2|\boldsymbol{Y},\boldsymbol{X})}{\partial\sigma^2\partial\beta_1}\right] = -E\left[-\sum_{i=1}^{N}\frac{(Y_i-\beta_0-\beta_1X_{i1})X_i}{(\sigma^2)^2}\right]$$

$$
\begin{aligned}
-E\left[\frac{\partial^2 l(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{Y}, \boldsymbol{X})}{\partial \sigma^2 \partial \beta_1}\right] &= -E\left[-\sum_{i=1}^{N} \frac{(Y_i - \beta_0 - \beta_1 X_{i1}) X_i}{(\sigma^2)^2}\right] \\
&= \frac{X_i}{(\sigma^2)^2}\left(\sum_{i=1}^{N} E[Y_i] - \beta_0 - \beta_1 X_{i1}\right)
\end{aligned}
$$

$$
\begin{aligned}
-E\left[\frac{\partial^2 l(\boldsymbol{\beta}, \sigma^2|\boldsymbol{Y}, \boldsymbol{X})}{\partial\sigma^2\partial\beta_1}\right] &= -E\left[-\sum_{i=1}^{N}\frac{(Y_i - \beta_0 - \beta_1 X_{i1})X_i}{(\sigma^2)^2}\right] \\
&= \frac{X_i}{(\sigma^2)^2}\left(\sum_{i=1}^{N} E[Y_i] - \beta_0 - \beta_1 X_{i1}\right) \\
&= \frac{X_i}{(\sigma^2)^2}\left(\sum_{i=1}^{N} \beta_0 + \beta_1 X_{i1} - \beta_0 - \beta_1 X_{i1}\right) = 0
\end{aligned}
$$

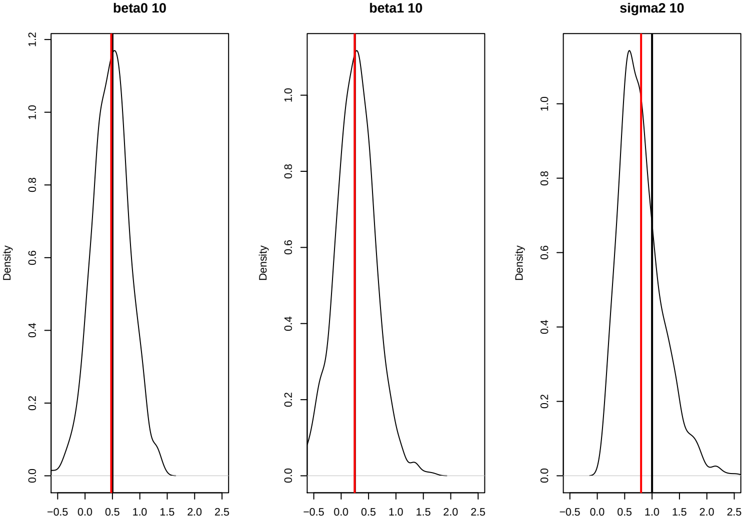
Bivariate Regression Model via Maximum Likelihood

$$
\begin{aligned}
I(\boldsymbol{\beta}, \sigma^2)_N &= \begin{pmatrix} \frac{N}{\sigma^2} & \sum_{i=1}^N \frac{X_i}{\sigma^2} & 0 \\ \sum_{i=1}^N \frac{X_i}{\sigma^2} & \sum_{i=1}^N \frac{X_i^2}{\sigma^2} & 0 \\ 0 & 0 & \frac{N}{2(\sigma^2)^2} \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{\sigma^2}\boldsymbol{X}^{'}\boldsymbol{X} & \boldsymbol{0} \\ \boldsymbol{0} & \frac{N}{2(\sigma^2)^2} \end{pmatrix}
\end{aligned}
$$

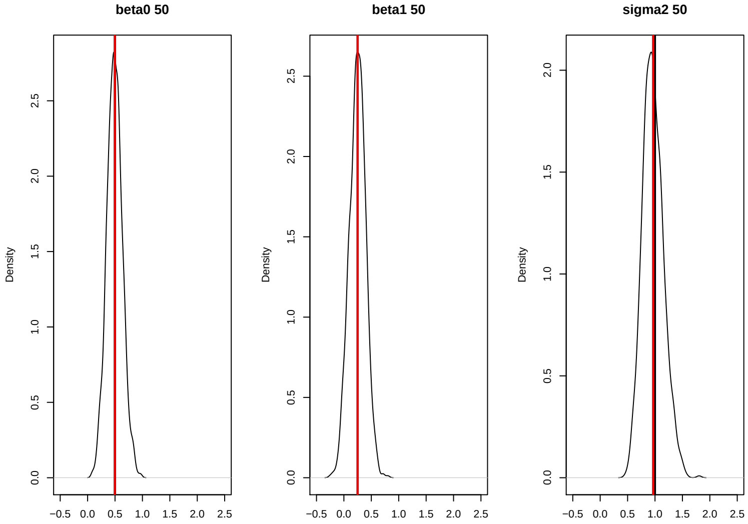
Inverting yields:

$$
\widehat{\text{Var}(\boldsymbol{\beta}, \sigma^2)} = \begin{pmatrix} \sigma^2 \left(\boldsymbol{X}^{'}\boldsymbol{X}\right)^{-1} & 0 \\ 0 & \frac{2(\sigma^2)^2}{N} \end{pmatrix}
$$

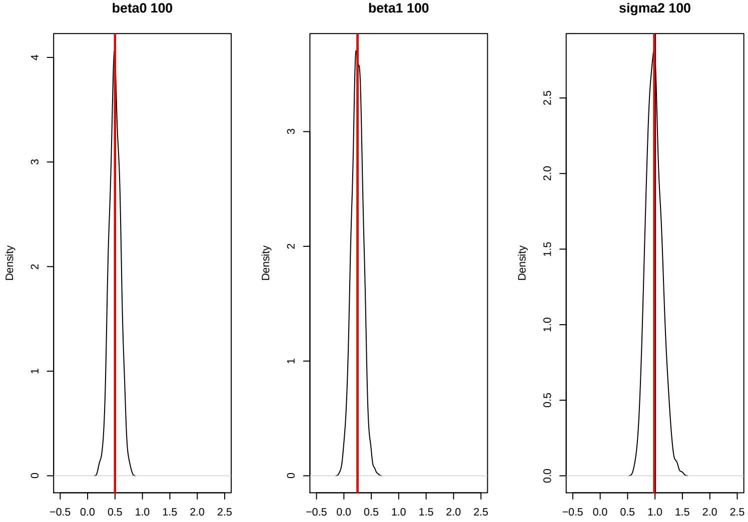
Bivariate Regression with Maximum Likelihood: Numerical Simulation



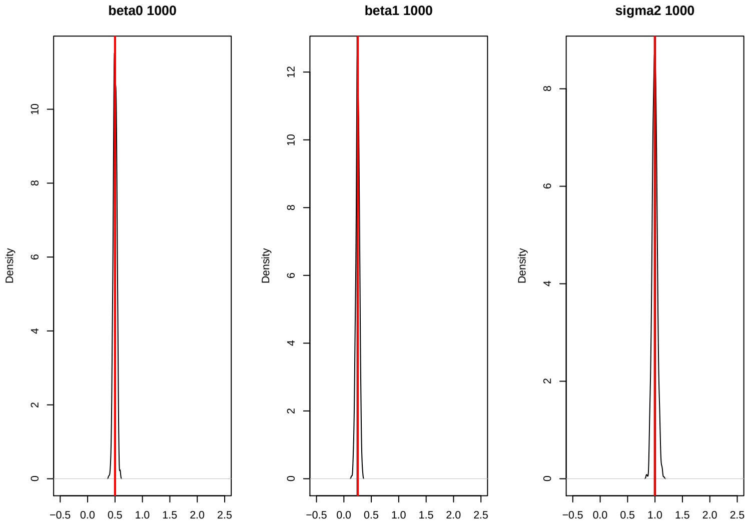
Bivariate Regression with Maximum Likelihood: Numerical Simulation



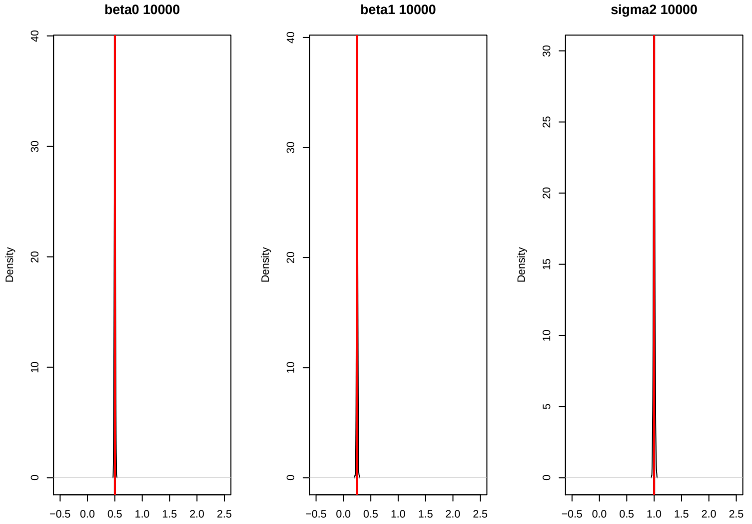
Bivariate Regression with Maximum Likelihood: Numerical Simulation



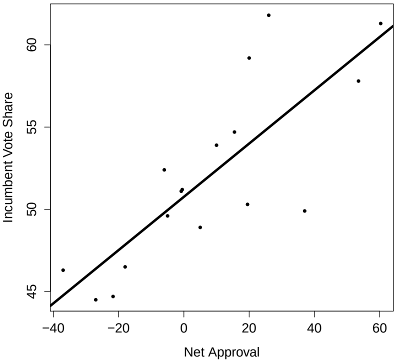
Bivariate Regression with Maximum Likelihood: Numerical Simulation



Bivariate Regression with Maximum Likelihood: Numerical Simulation

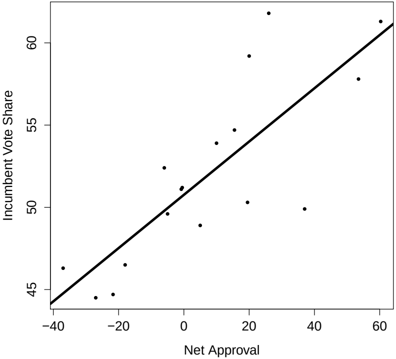


Bivariate Regression via Maximum Likelihood



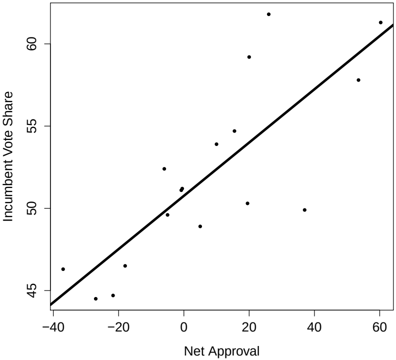
Bivariate Regression via Maximum Likelihood

$$\text{Incumbent}_i = \beta_0^* + \beta_1^* \text{Approval}_i + \epsilon_i$$



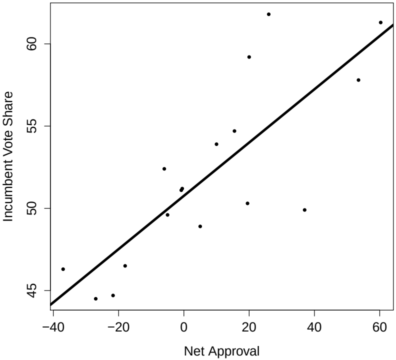
Bivariate Regression via Maximum Likelihood

$$\begin{aligned}\textsf{Incumbent}_i &= \beta_0^* + \beta_1^* \textsf{Approval}_i + \epsilon_i \\ \textsf{Incumbent}_i &= \underbrace{50.76 + 0.16 \times \textsf{Approval}_i}_{\widehat{\textsf{Incumbent}}_i} + \epsilon_i\end{aligned}$$



Bivariate Regression via Maximum Likelihood

$$\begin{aligned}\textsf{Incumbent}_i &= \beta_0^* + \beta_1^* \textsf{Approval}_i + \epsilon_i \\ \textsf{Incumbent}_i &= \underbrace{50.76 + 0.16 \times \textsf{Approval}_i}_{\widehat{\textsf{Incumbent}}_i} + \epsilon_i\end{aligned}$$



Multivariate Regression via Maximum Likelihood

Define:

$$\begin{aligned}
\boldsymbol{X}_i &= (1, X_{i1}, X_{i2}, X_{i3}) \\
\boldsymbol{\beta} &= (\beta_0, \beta_1, \beta_2)
\end{aligned}$$

Then our maximum likelihood estimators are:

$$\begin{aligned}
\boldsymbol{\beta}^* &= \left(\boldsymbol{X}^{'}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{'}\boldsymbol{Y} \\
\sigma^{2,*} &= \frac{1}{N}(Y - \boldsymbol{X}\boldsymbol{\beta})^{'}(Y - \boldsymbol{X}\boldsymbol{\beta}) \\
&= \frac{1}{N}\sum_{i=1}^{N}(Y_i - \beta_0 - \beta_1 X_{i1} - \beta_2 X_{i2} - \beta_3 X_{i3})^2
\end{aligned}$$

Then,

$$\boldsymbol{\beta}^{*}, \sigma^{2,*} \quad \rightarrow^{D} \quad \mathsf{MVN}((\boldsymbol{\beta}^{*}, \sigma^{2,*}), I(\boldsymbol{\beta}^{*}, \sigma^{2,*})^{-1})$$

Where

$$\mathsf{Var}(\widehat{\boldsymbol{\beta}^{*}, \sigma^{2,*}})) \quad = \quad \begin{pmatrix} \sigma^{2}(\boldsymbol{X}^{'}\boldsymbol{X})^{-1} & 0 \\ 0 & \frac{2(\sigma^{2})^{2}}{N} \end{pmatrix}$$

Analytic vs Computational

Analytic optimization:

- Calculus + pencil and paper
- Search for closed form solutions
- Big guarantees, but solutions can be hard (impossible) to obtain

Computational Optimization:

- Iterative procedure
- Search for approximate solution
- Fewer guarantees, but you'll obtain solutions

Overall strategy:
One-parameter Newton Raphson $\rightsquigarrow$ Multi-parameter Newton Raphson $\rightsquigarrow$ BFGS (Wednesday)

Newton-Raphson Method

Iterative procedure to find a root

Newton-Raphson Method

Iterative procedure to find a root
Often solving for x when $f(x) = 0$ is hard $\rightsquigarrow$ complicated function

Newton-Raphson Method

Iterative procedure to find a root

Often solving for x when $f(x) = 0$ is hard $\rightsquigarrow$ complicated function

Solving for x when $f(x)$ is linear $\rightsquigarrow$ easy

Newton-Raphson Method

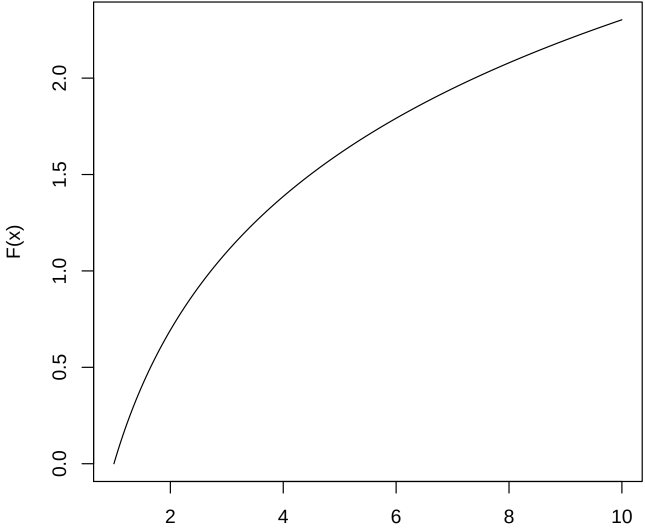
Iterative procedure to find a root

Often solving for x when $f(x) = 0$ is hard $\rightsquigarrow$ complicated function

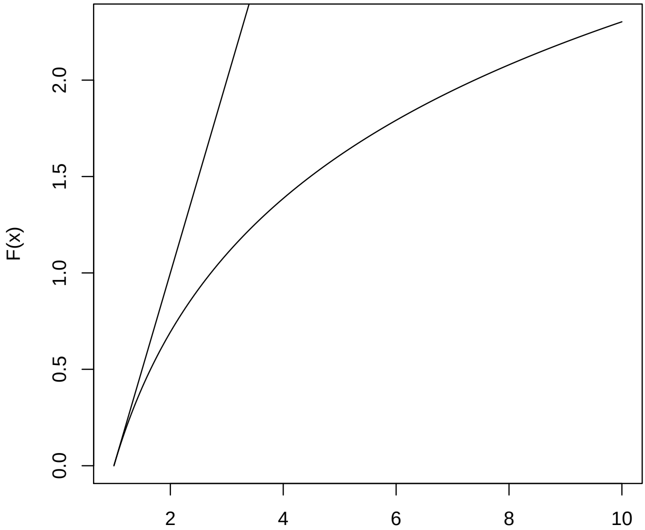
Solving for x when $f(x)$ is linear $\rightsquigarrow$ easy

Approximate with tangent line, iteratively update

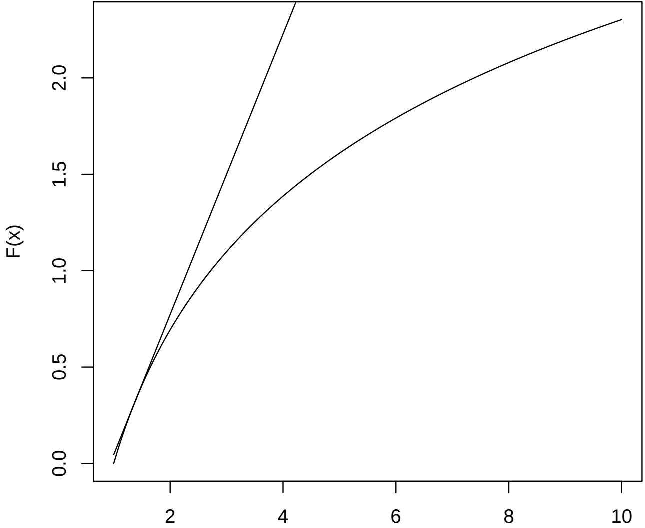
Tangent Line



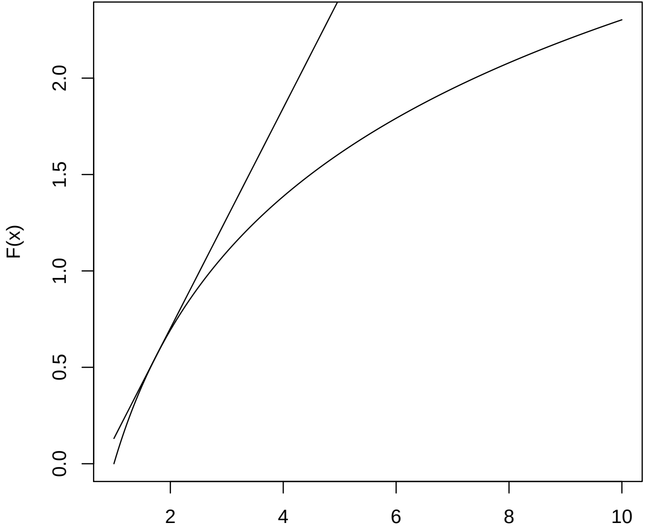
Tangent Line



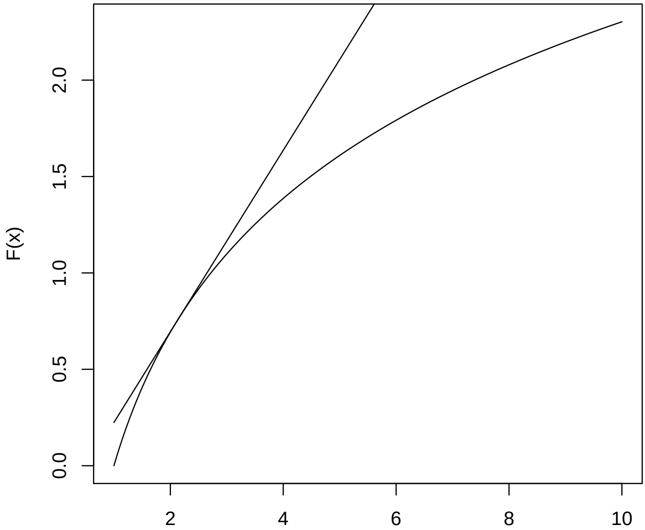
Tangent Line



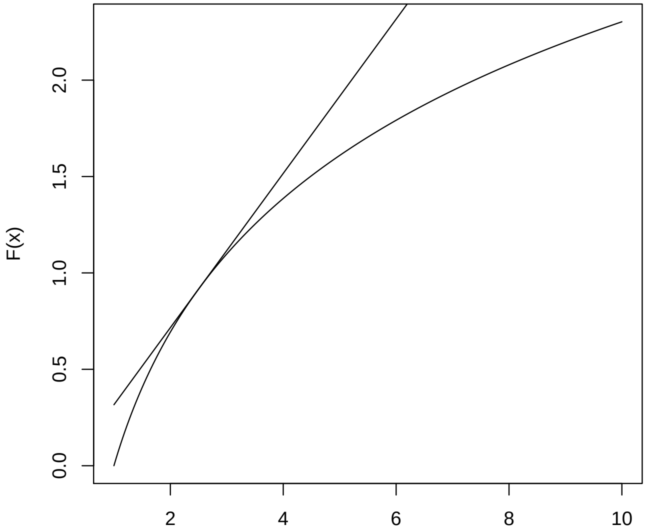
Tangent Line



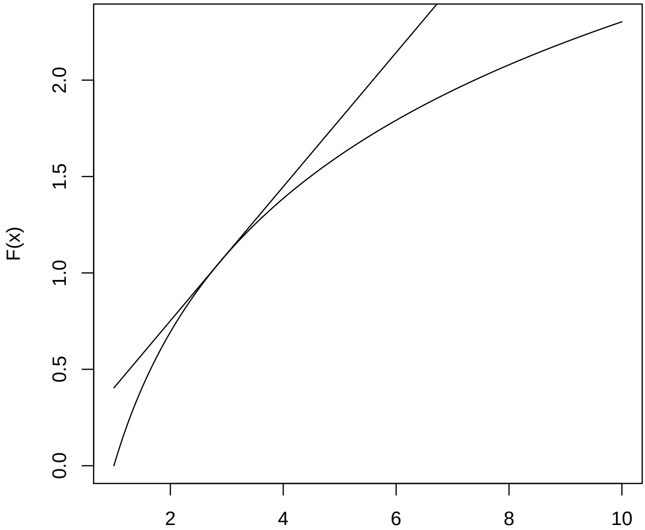
Tangent Line



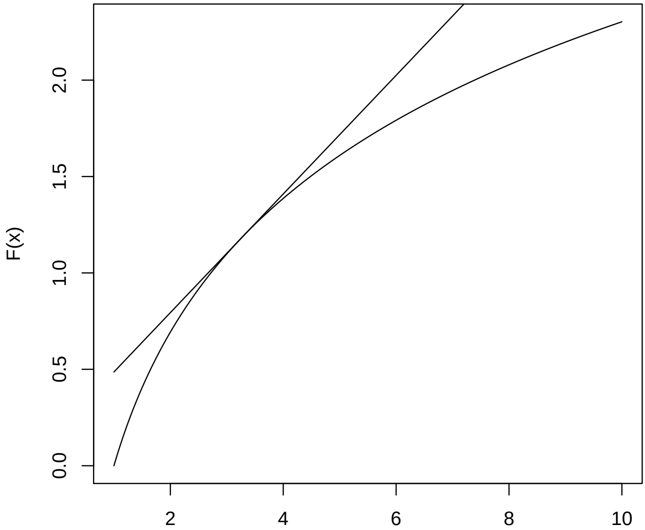
Tangent Line



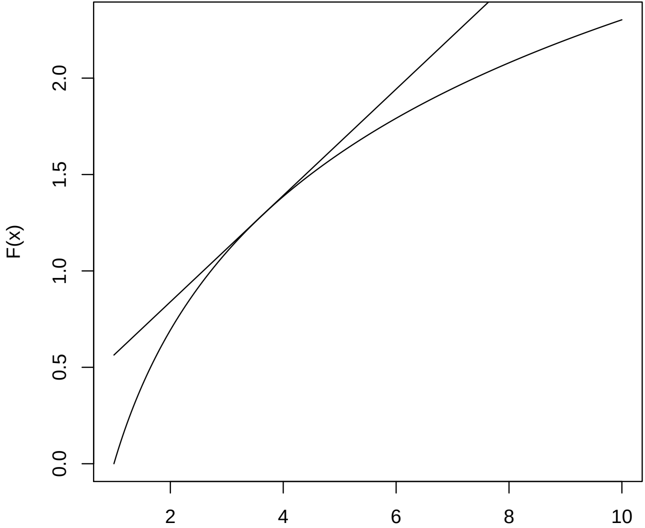
Tangent Line



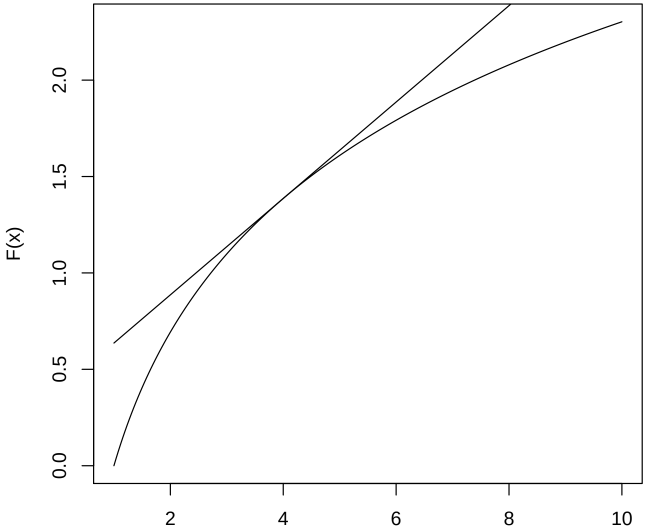
Tangent Line



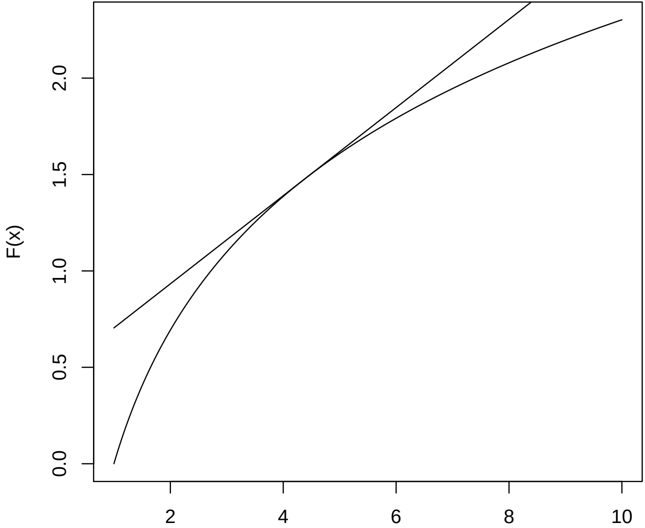
Tangent Line



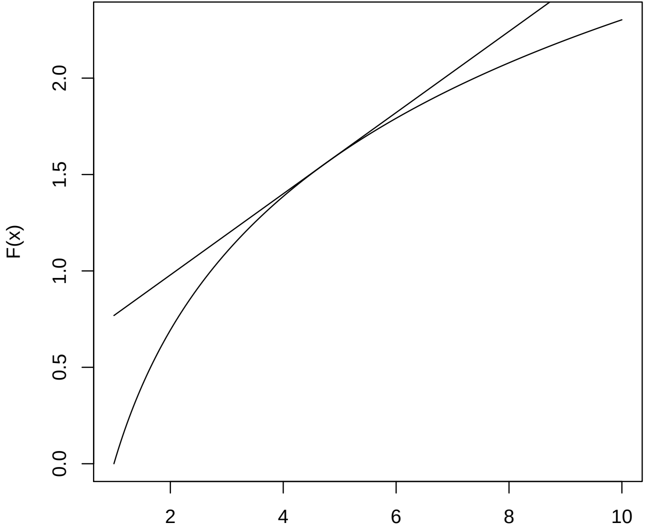
Tangent Line



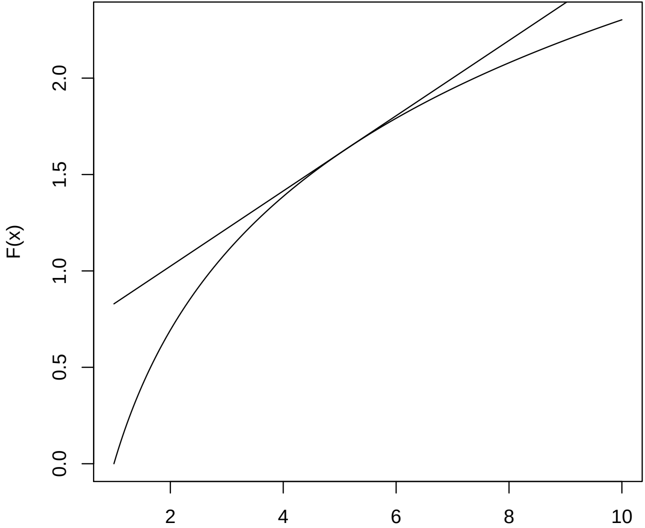
Tangent Line



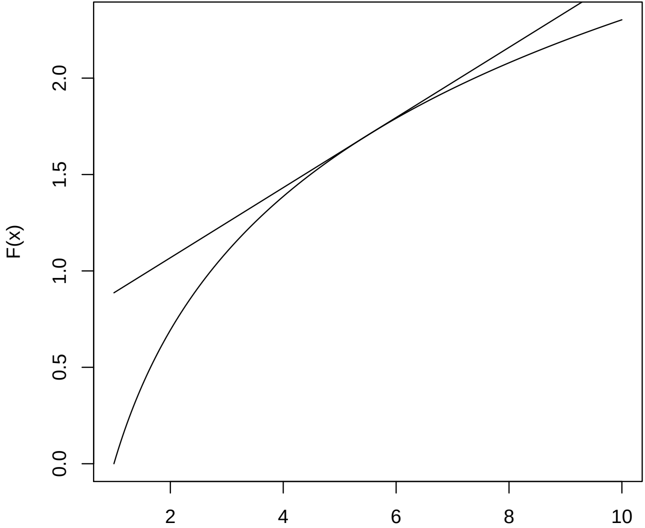
Tangent Line



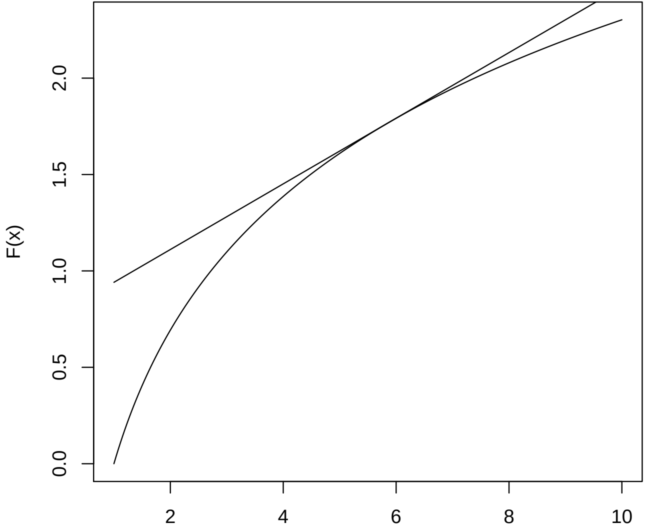
Tangent Line



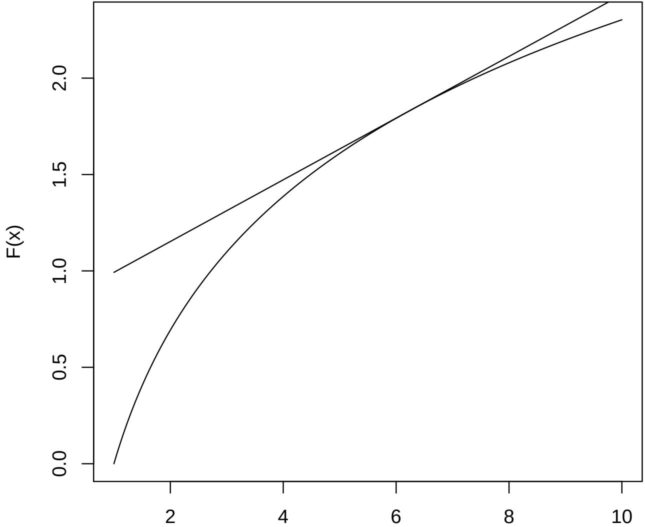
Tangent Line



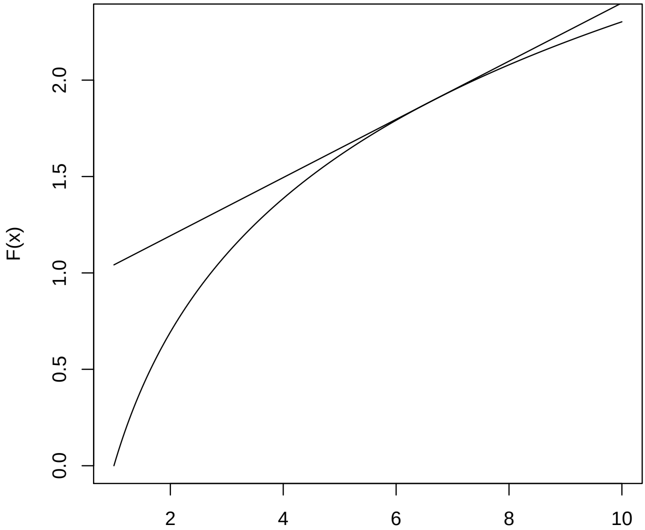
Tangent Line



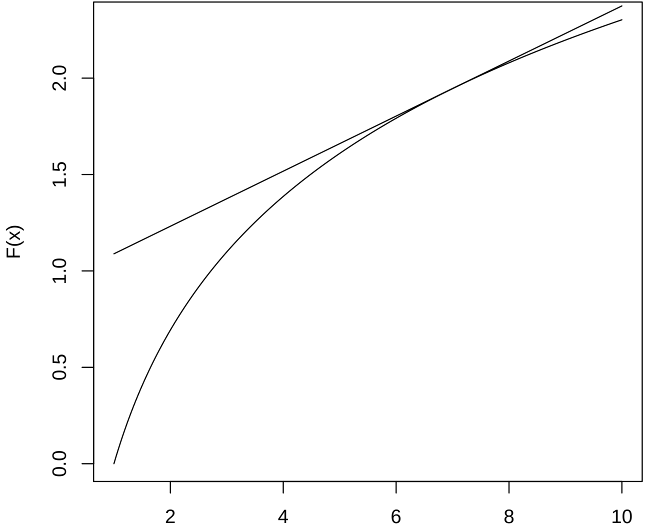
Tangent Line



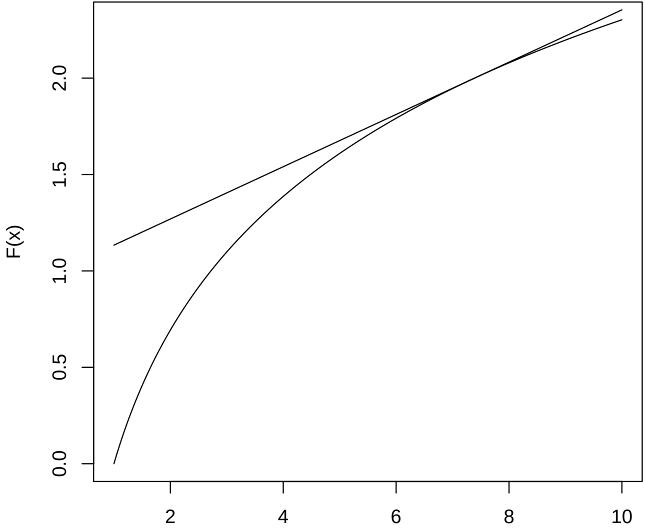
Tangent Line



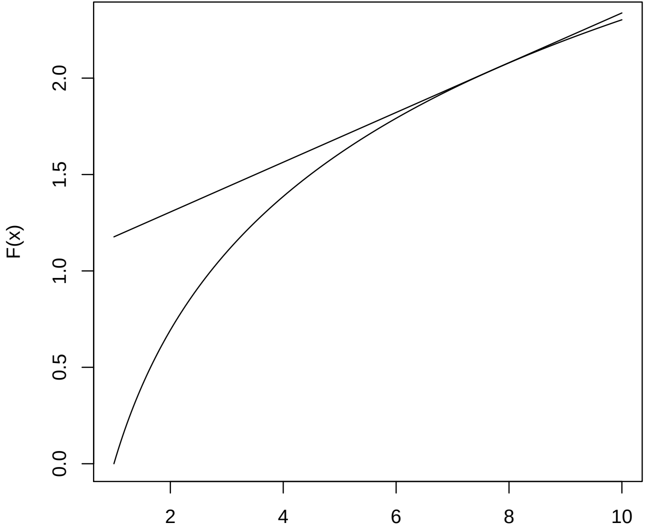
Tangent Line



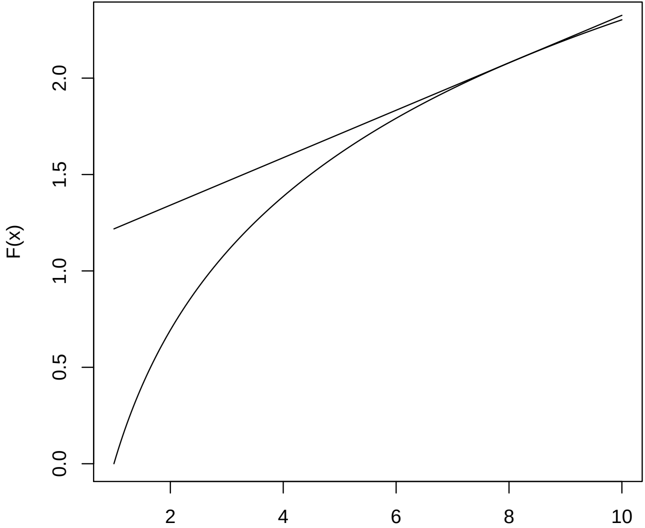
Tangent Line



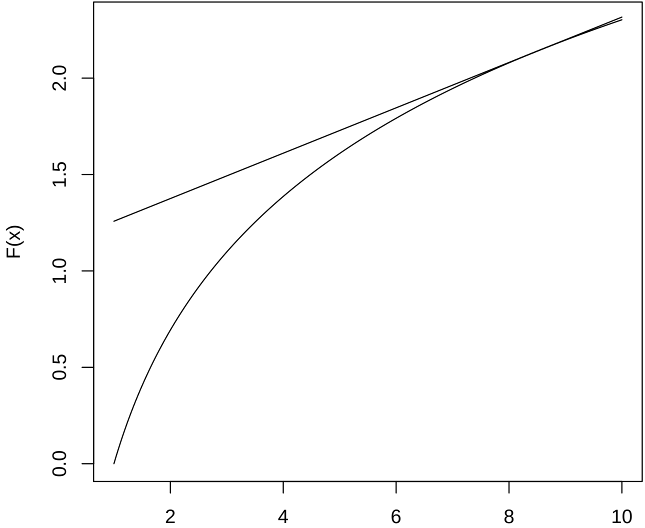
Tangent Line



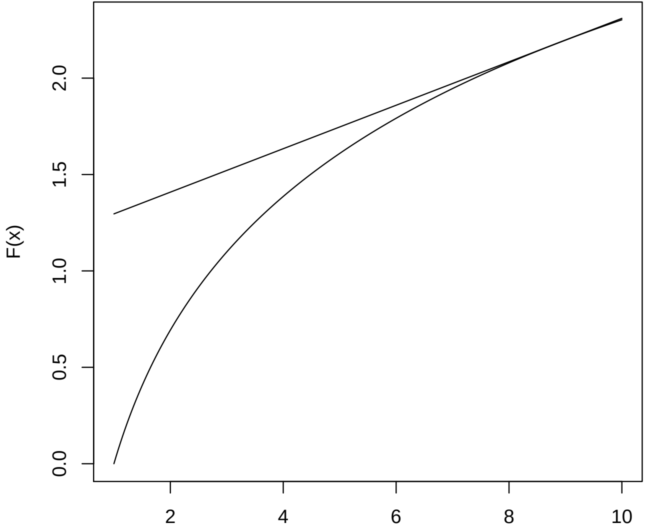
Tangent Line



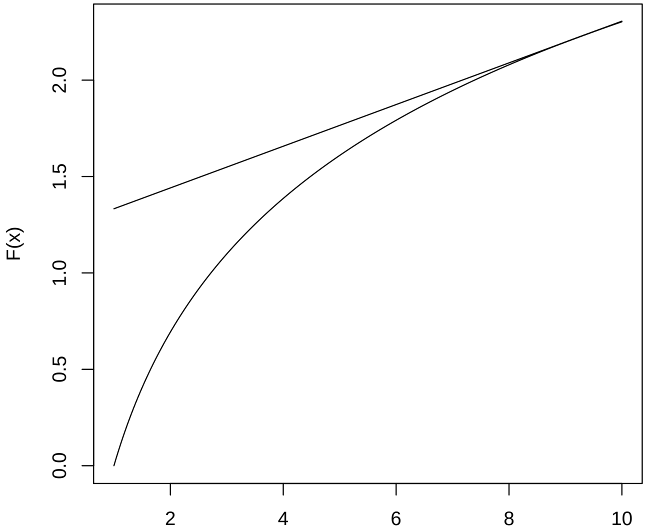
Tangent Line



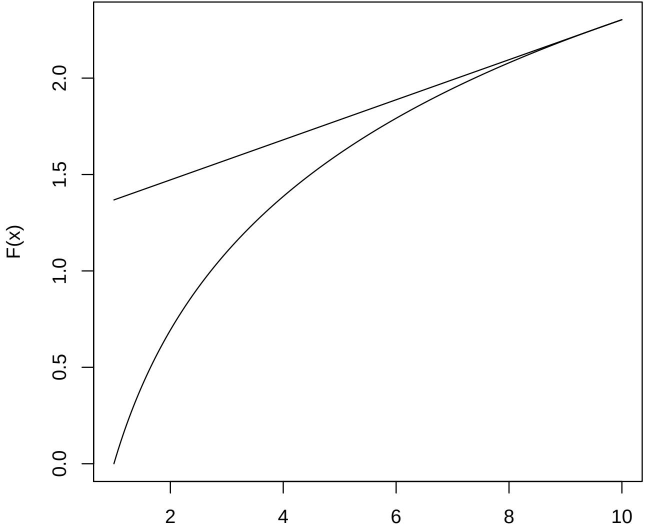
Tangent Line



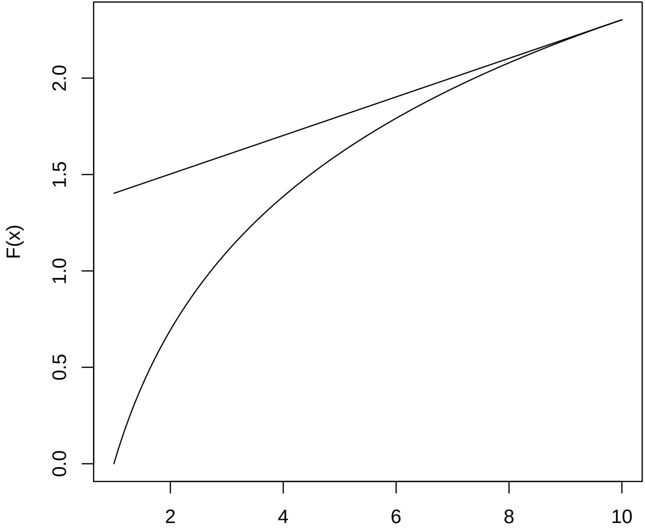
Tangent Line



Tangent Line



Tangent Line



Tangent Line

Formula for Tangent line at x_0:

Tangent Line

Formula for Tangent line at x_0:

$$g(x) = f^{'}(x_0)(x - x_0) + f(x_0)$$

Tangent Line

Formula for Tangent line at x_0:

$$g(x) = f^{'}(x_0)(x - x_0) + f(x_0)$$

Tangent Line

Formula for Tangent line at x_0:

$$g(x) = f^{'}(x_0)(x - x_0) + f(x_0)$$

Tangent Line

Formula for Tangent line at x_0:

$$g(x) = f^{'}(x_0)(x - x_0) + f(x_0)$$

Newton-Raphson Method

Suppose we have some initial guess x_0. We're going to approximate $f^{'}(x)$ with the tangent line to generate a new guess

Newton-Raphson Method

Suppose we have some initial guess x_0. We're going to approximate $f^{'}(x)$ with the tangent line to generate a new guess

$$g(x) \quad = \quad f^{''}(x_0)(x - x_0) + f^{'}(x_0)$$

Newton-Raphson Method

Suppose we have some initial guess x_0. We're going to approximate $f^{'}(x)$ with the tangent line to generate a new guess

$$\begin{aligned} g(x) &= f^{''}(x_0)(x - x_0) + f^{'}(x_0) \\ 0 &= f^{''}(x_0)(x_1 - x_0) + f^{'}(x_0) \end{aligned}$$

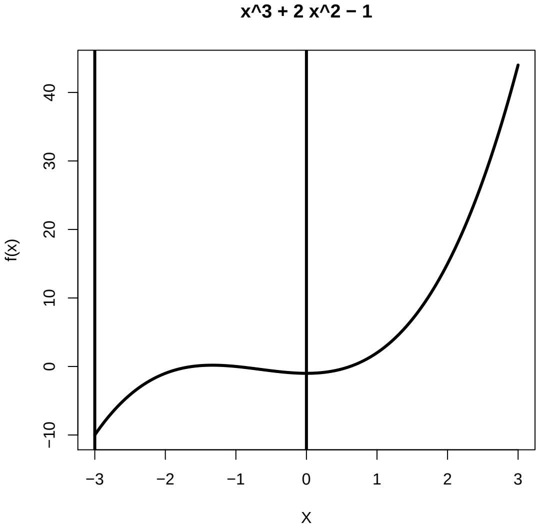
Newton-Raphson Method

Suppose we have some initial guess x_0. We're going to approximate $f^{'}(x)$ with the tangent line to generate a new guess

$$\begin{aligned}
g(x) &= f^{''}(x_0)(x - x_0) + f^{'}(x_0) \\
0 &= f^{''}(x_0)(x_1 - x_0) + f^{'}(x_0) \\
x_1 &= x_0 - \frac{f^{'}(x_0)}{f^{''}(x_0)}
\end{aligned}$$

Example Function

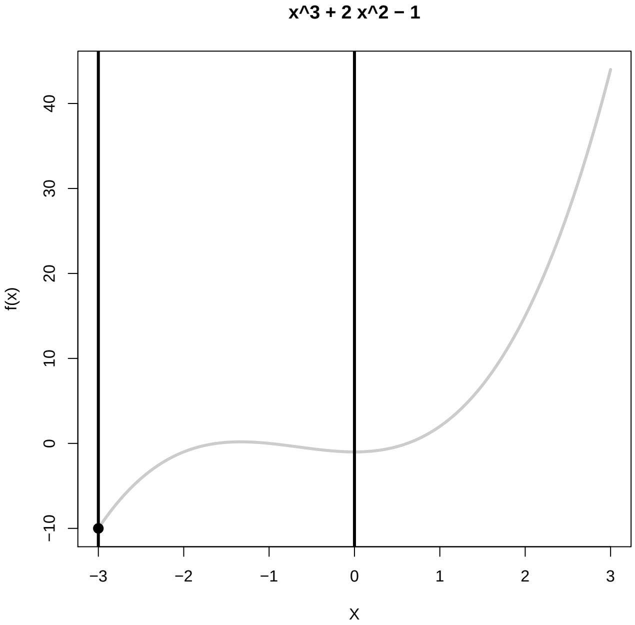
$f(x) = x^3 + 2x^2 - 1$ find x that maximizes $f(x)$ with $x \in [-3, 0]$

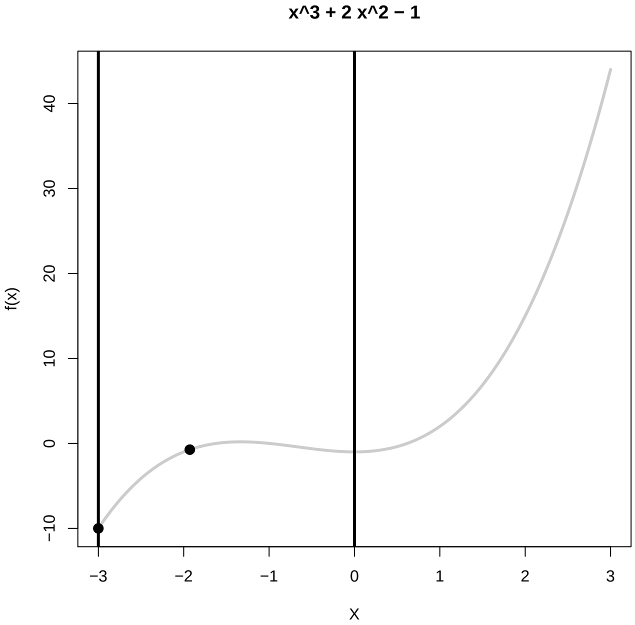


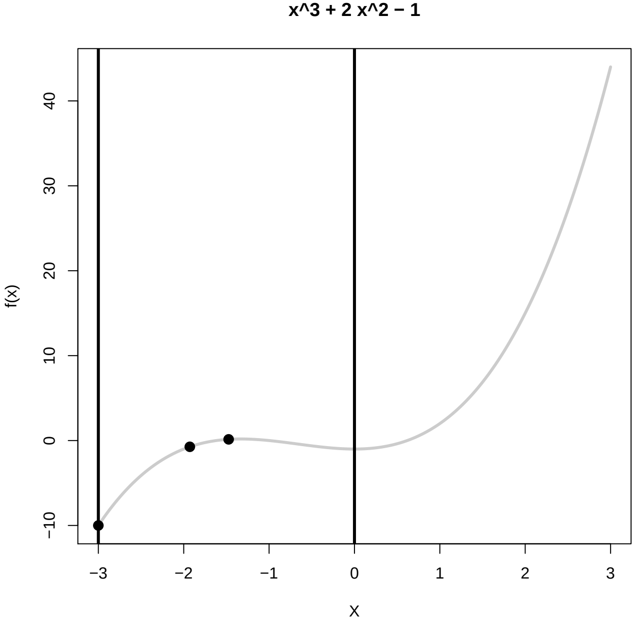
$$
\begin{aligned}
f^{'}(x) &= 3x^2 + 4x \\
f^{''}(x) &= 6x + 4
\end{aligned}
$$

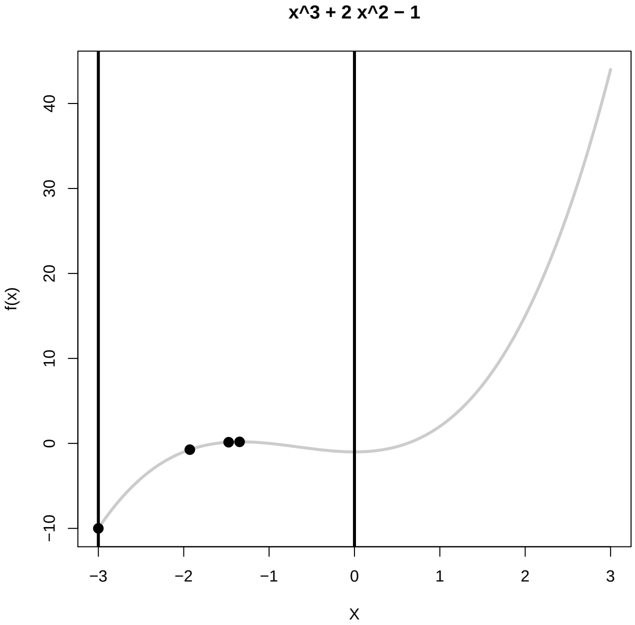
Suppose we have guess x_t then the next step is:

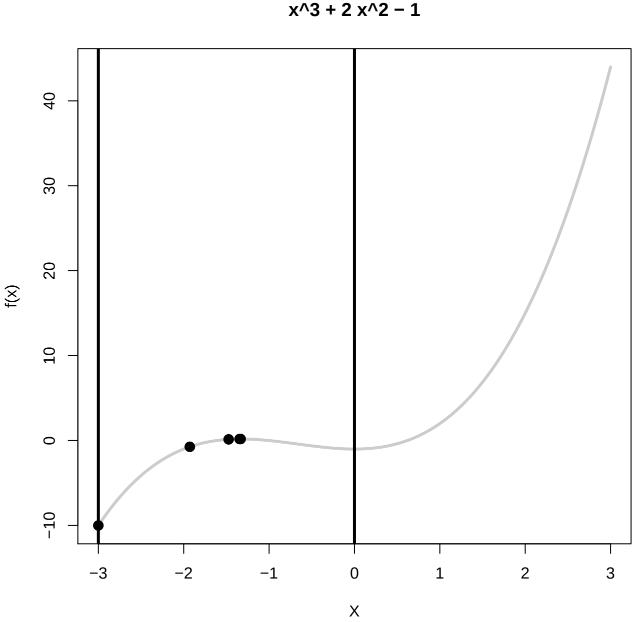
$$
x_{t+1} = x_t - \frac{3x_t^2 + 4x_t}{6x_t + 4}
$$

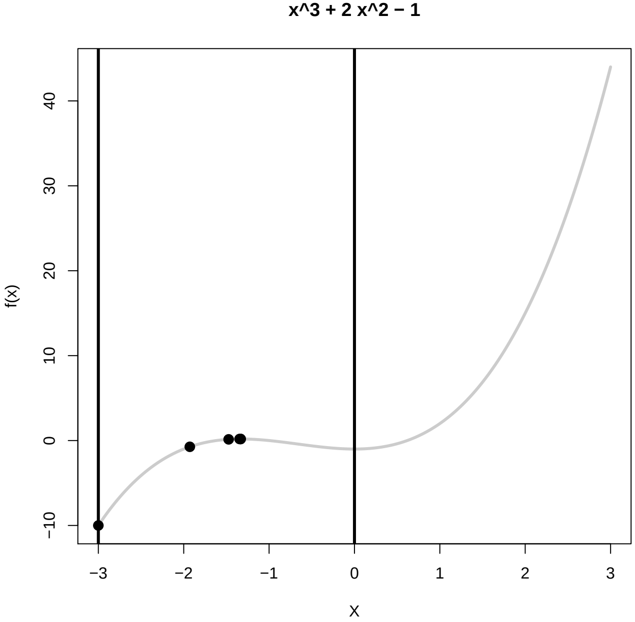

x^3 + 2 x^2 − 1
f(x)
−10
0
10
20
30
40
−3
−2
−1
0
1
2
3
X


x^3 + 2 x^2 − 1
f(x)
40
30
20
10
0
−10
−3
−2
−1
0
1
2
3
X

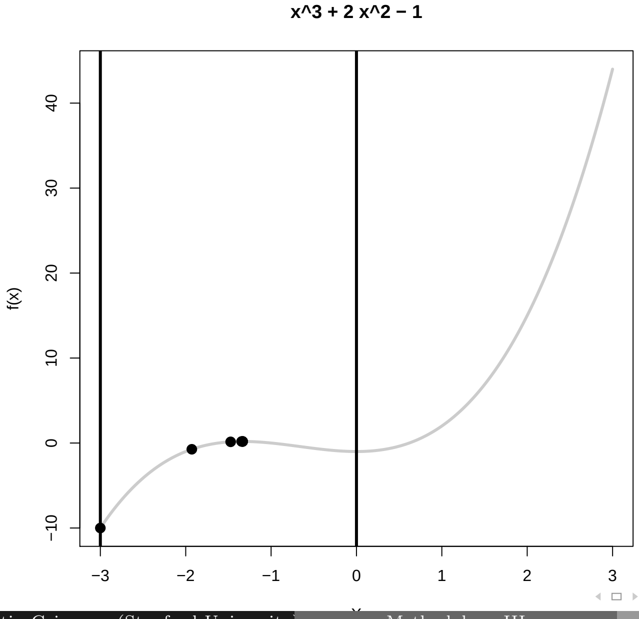

x^3 + 2 x^2 − 1
f(x)
−10
0
10
20
30
40
−3
−2
−1
0
1
2
3
X


x^3 + 2 x^2 − 1
f(x)
−10
0
10
20
30
40
−3
−2
−1
0
1
2
3
X

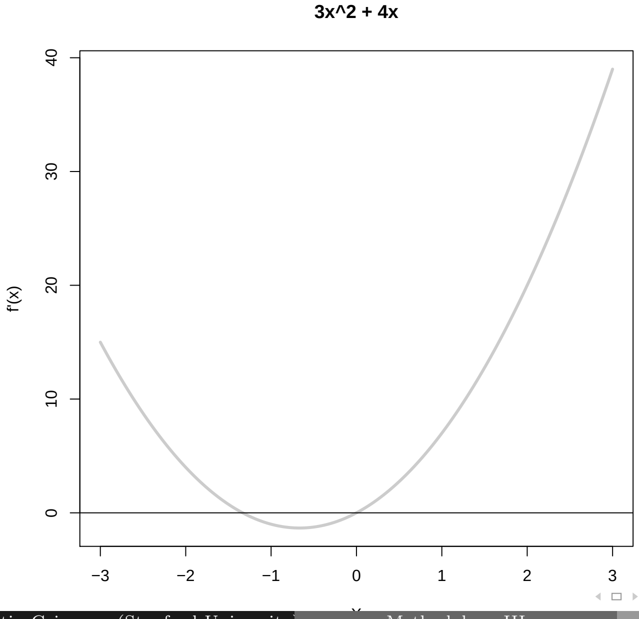

x^3 + 2 x^2 − 1
f(x)
40
30
20
10
0
−10
−3
−2
−1
0
1
2
3
X


x^3 + 2 x^2 − 1
f(x)
X
−10
0
10
20
30
40
−3
−2
−1
0
1
2
3

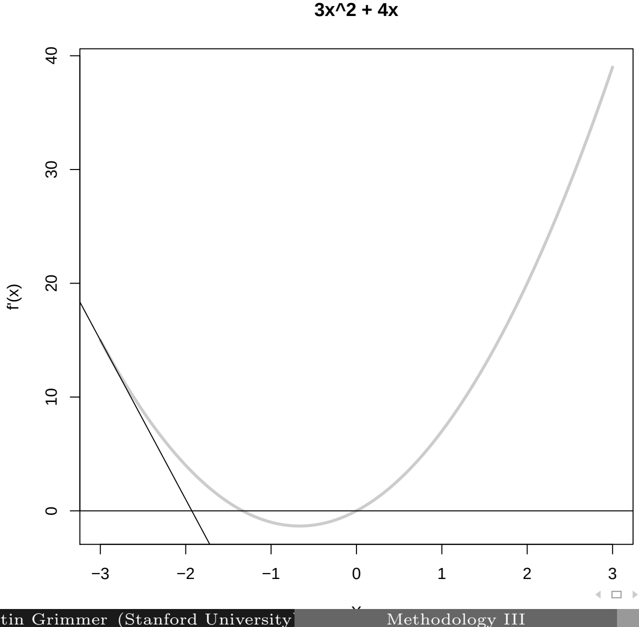
$x^* = -1.3333$



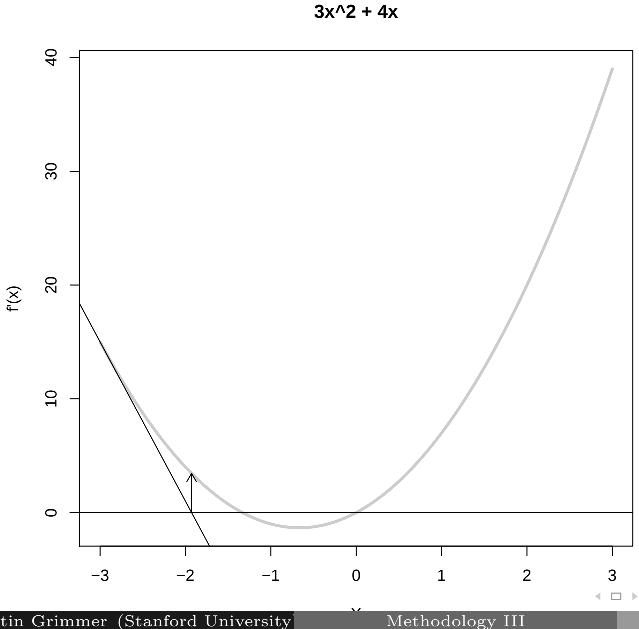
What is Happening with the Roots



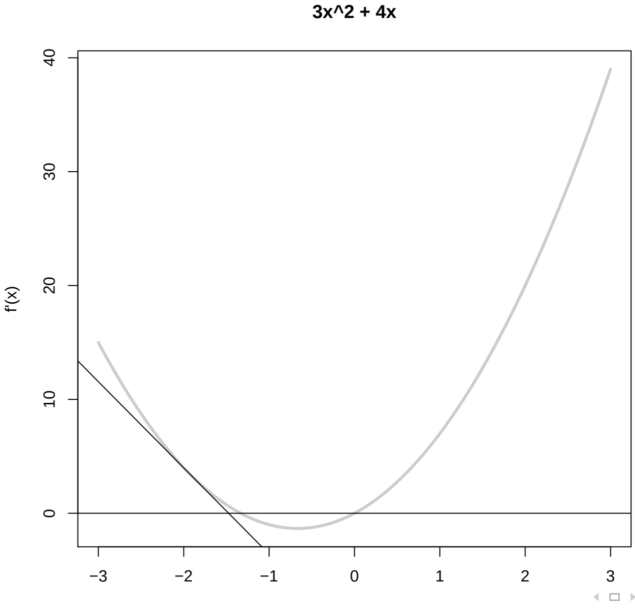
What is Happening with the Roots



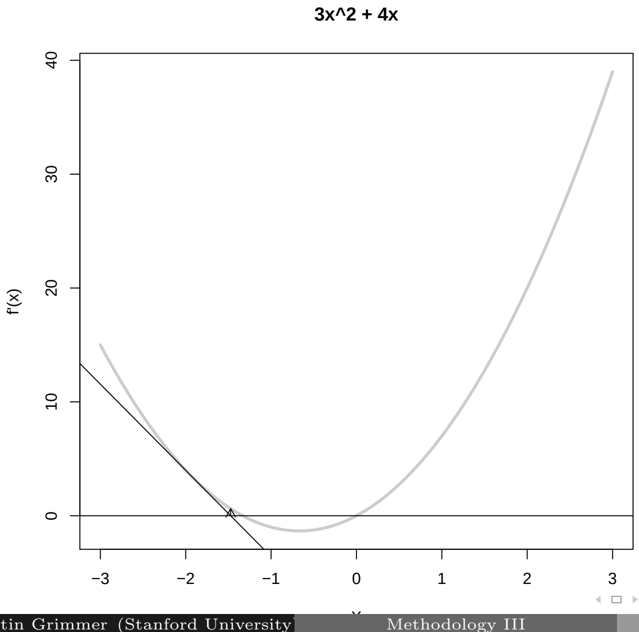
What is Happening with the Roots



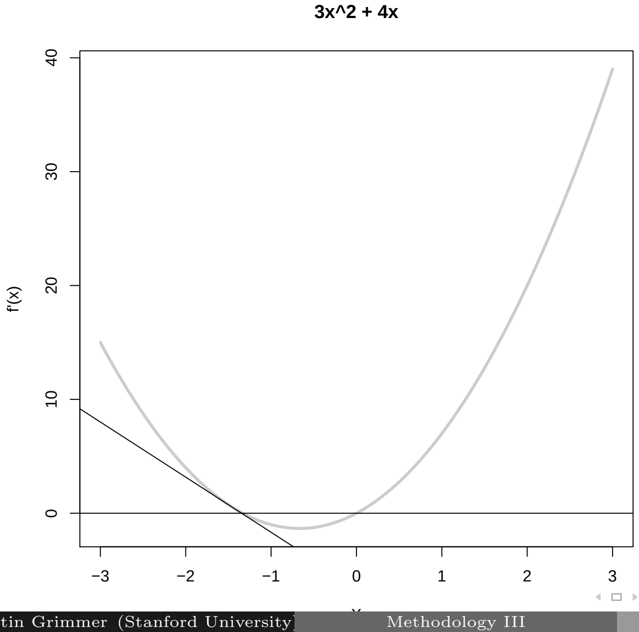
What is Happening with the Roots



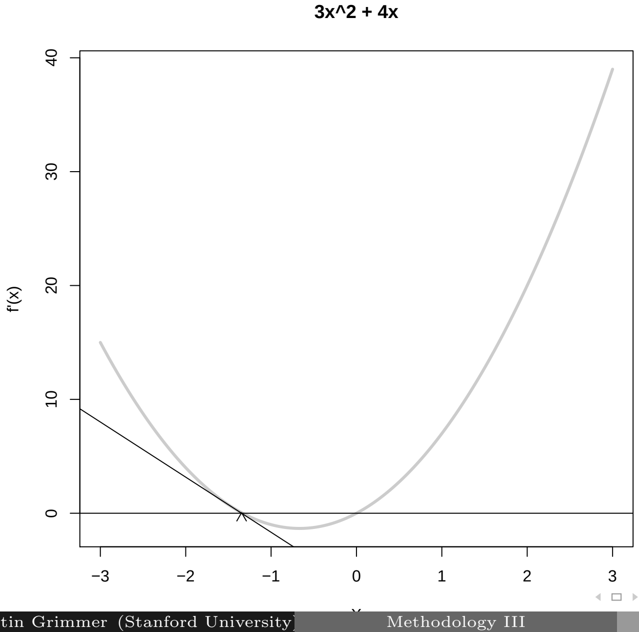
What is Happening with the Roots



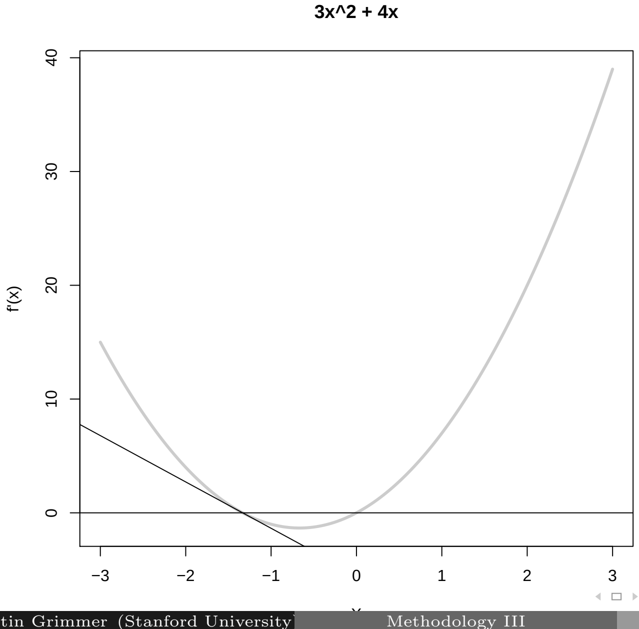
What is Happening with the Roots



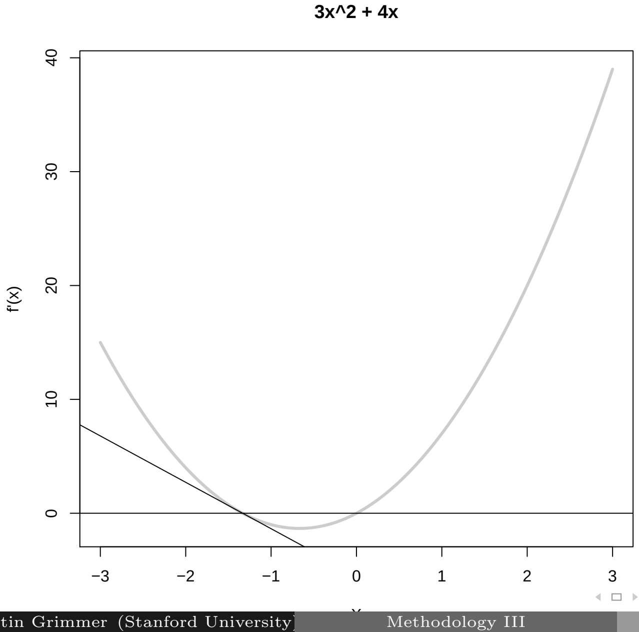
What is Happening with the Roots



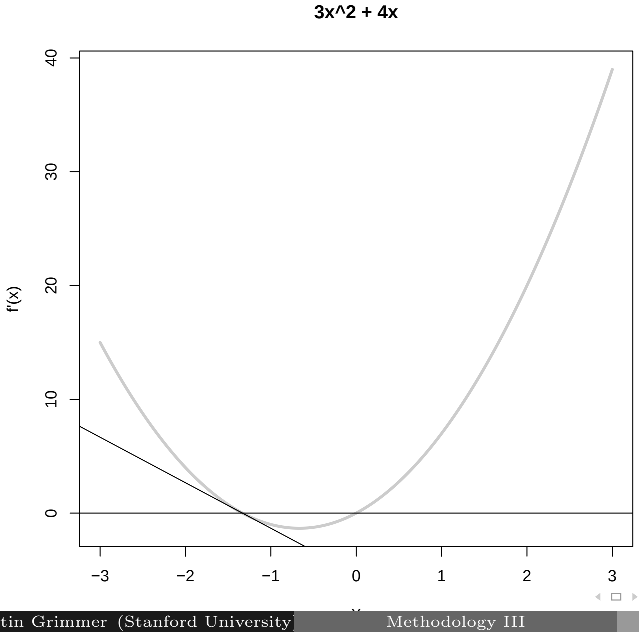
What is Happening with the Roots



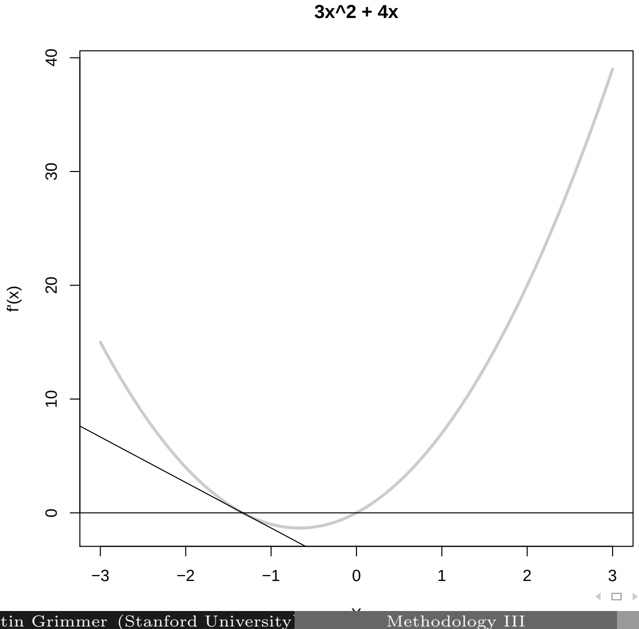
What is Happening with the Roots



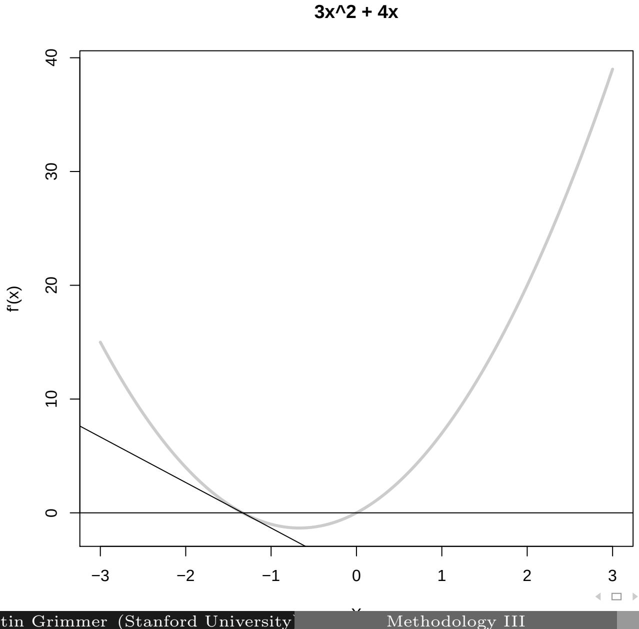
What is Happening with the Roots



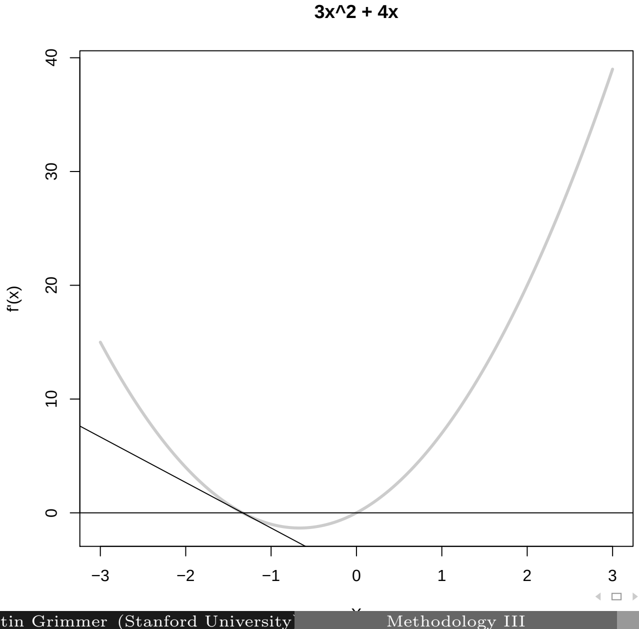
What is Happening with the Roots

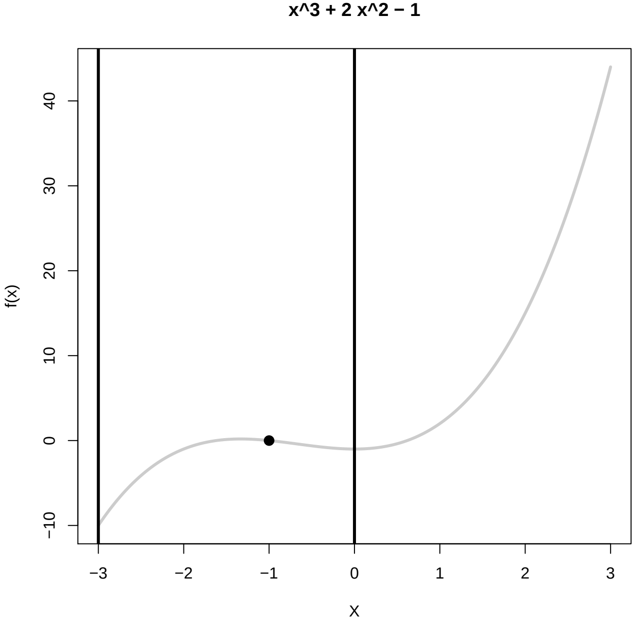


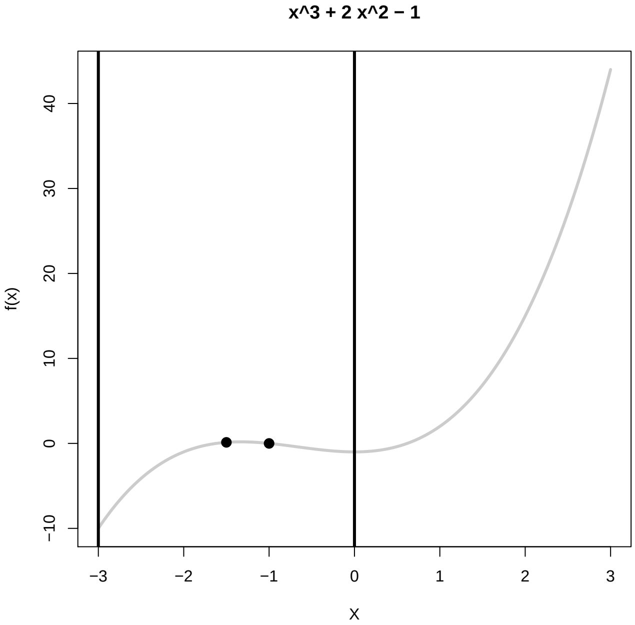
What is Happening with the Roots

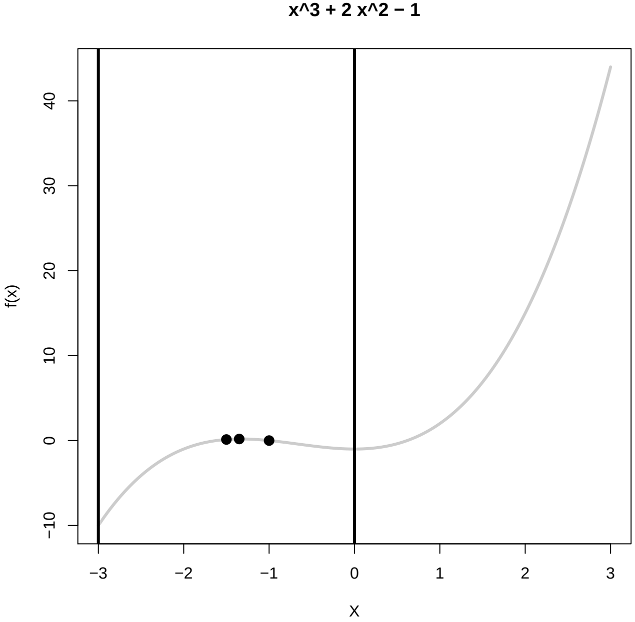


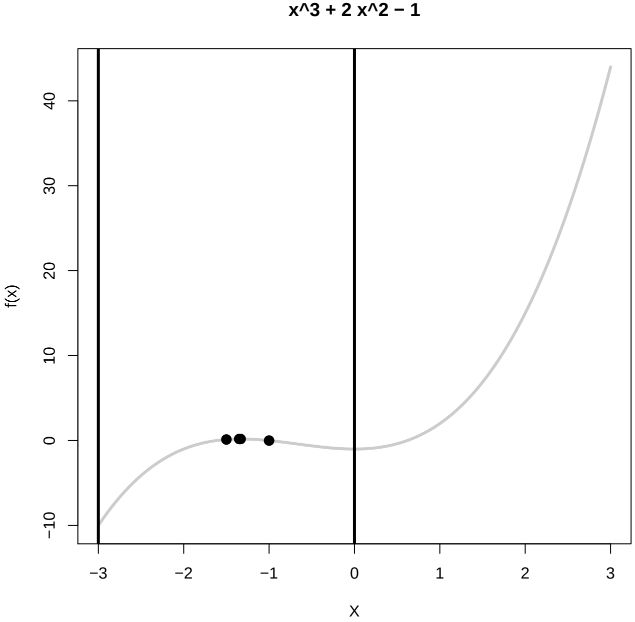
What is Happening with the Roots

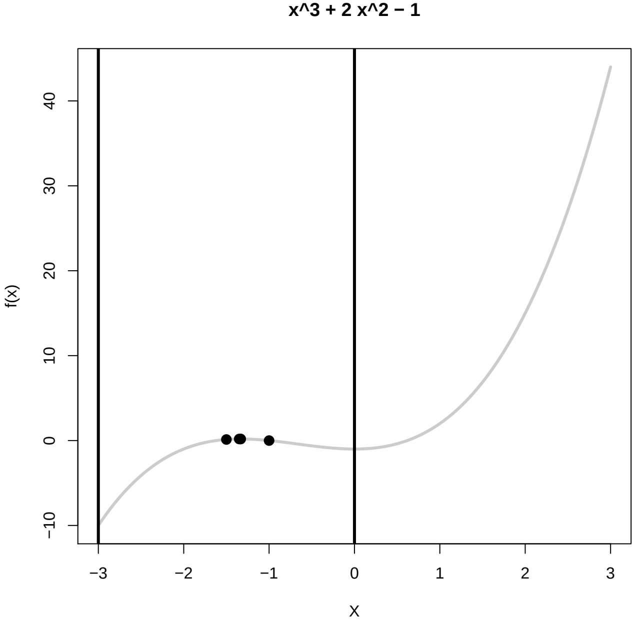


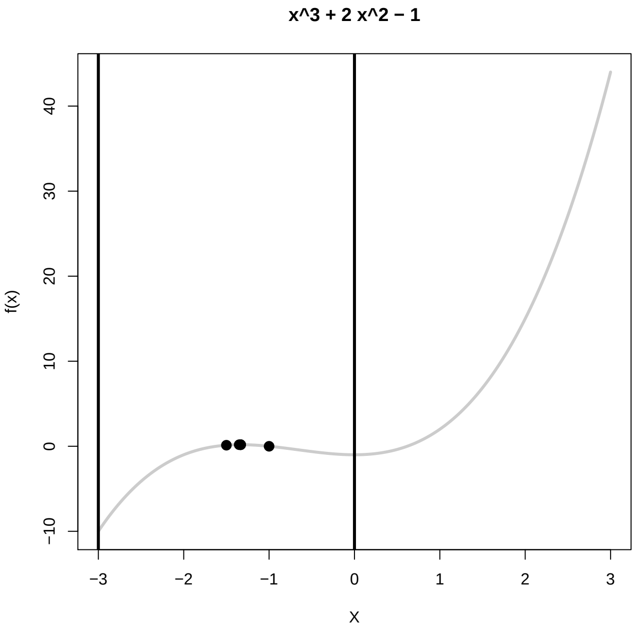

x^3 + 2 x^2 − 1
f(x)
−10
0
10
20
30
40
−3
−2
−1
0
1
2
3
X

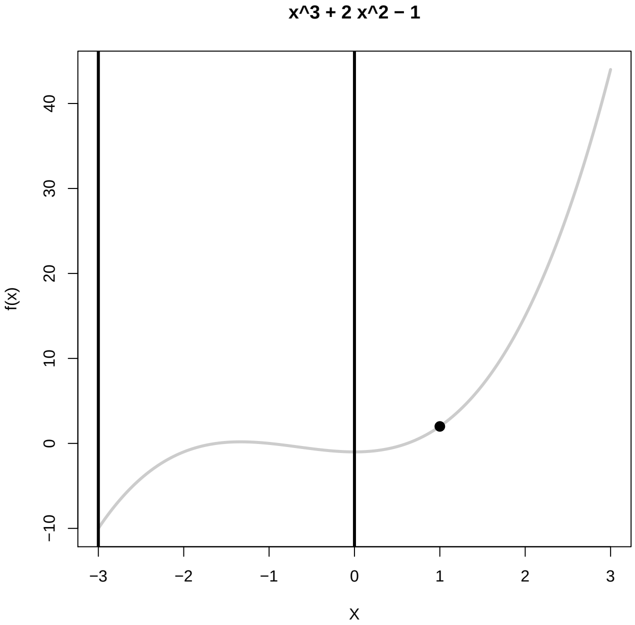

x^3 + 2 x^2 − 1
f(x)
−10
0
10
20
30
40
−3
−2
−1
0
1
2
3
X

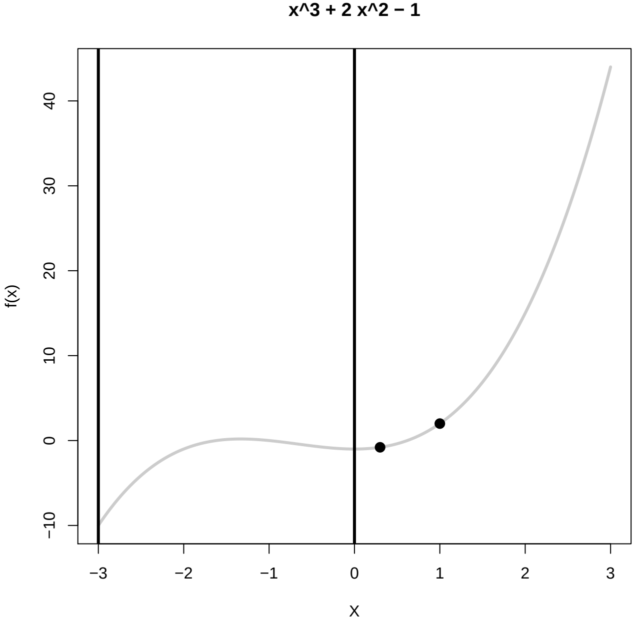

x^3 + 2 x^2 − 1
f(x)
X
−10
0
10
20
30
40
−3
−2
−1
0
1
2
3

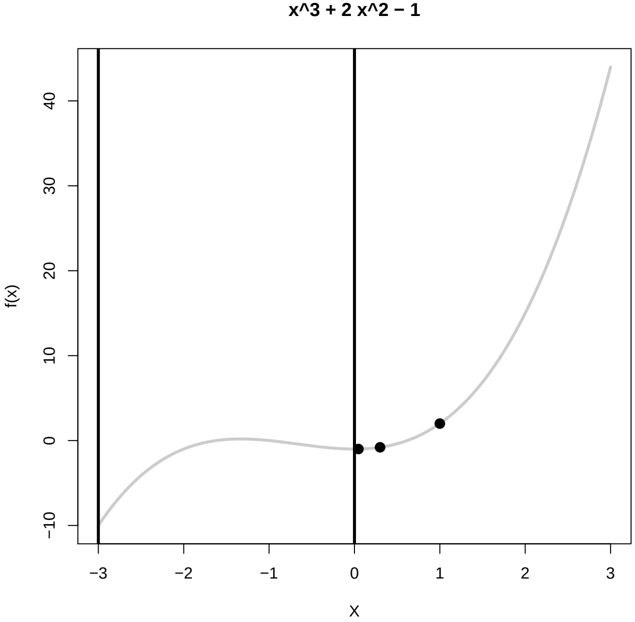

x^3 + 2 x^2 − 1
f(x)
X
−10
0
10
20
30
40
−3
−2
−1
0
1
2
3

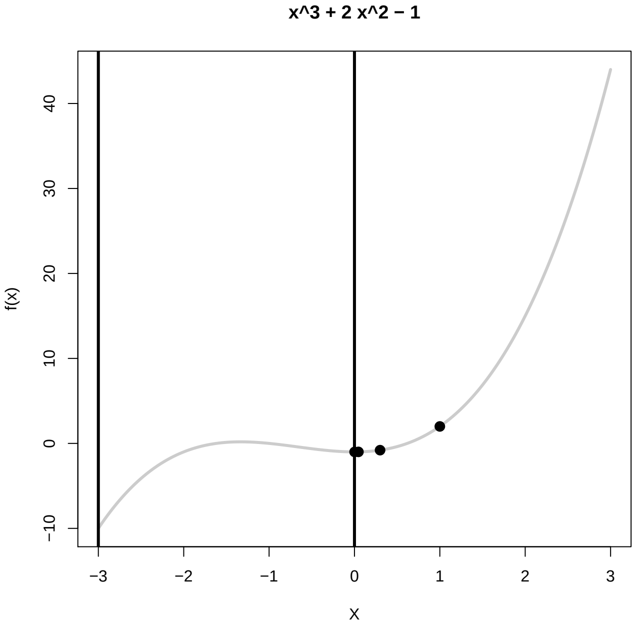

x^3 + 2 x^2 − 1
f(x)
−10
0
10
20
30
40
−3
−2
−1
0
1
2
3
X

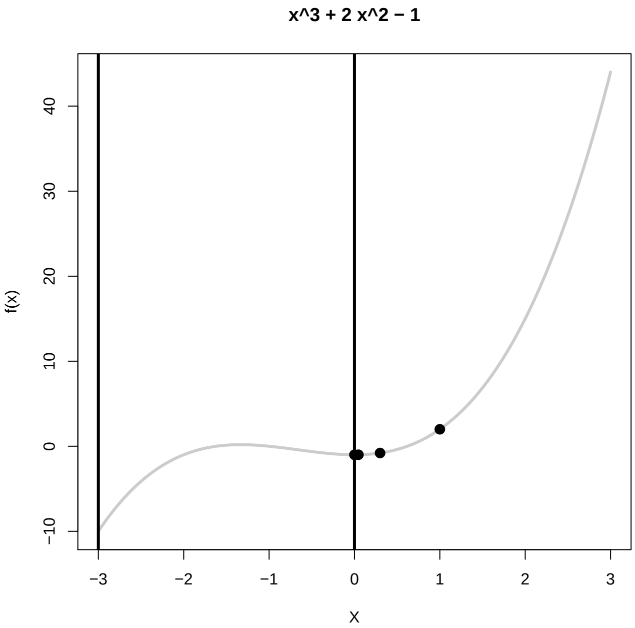

x^3 + 2 x^2 − 1
f(x)
−10
0
10
20
30
40
−3
−2
−1
0
1
2
3
X

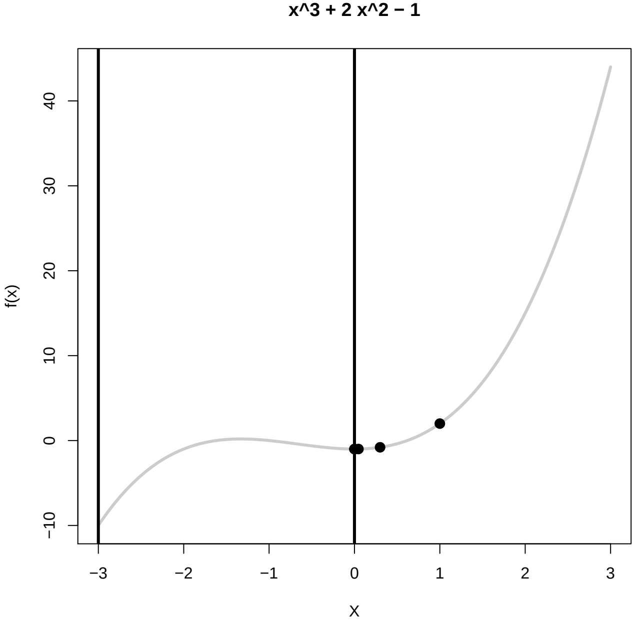

x^3 + 2 x^2 − 1
f(x)
−10
0
10
20
30
40
−3
−2
−1
0
1
2
3
X


x^3 + 2 x^2 − 1
f(x)
X
−10
0
10
20
30
40
−3
−2
−1
0
1
2
3


x^3 + 2 x^2 − 1
f(x)
X
−3
−2
−1
0
1
2
3
−10
0
10
20
30
40


x^3 + 2 x^2 − 1
f(x)
X
−10
0
10
20
30
40
−3
−2
−1
0
1
2
3


x^3 + 2 x^2 − 1
f(x)
−10
0
10
20
30
40
−3
−2
−1
0
1
2
3
X


x^3 + 2 x^2 − 1
f(x)
X
−10
0
10
20
30
40
−3
−2
−1
0
1
2
3

Multivariate Newton Raphson

Suppose $f : \Re^n \rightarrow \Re$. Suppose we have guess $\boldsymbol{x}_t$.

Multivariate Newton Raphson

Suppose $f : \Re^n \rightarrow \Re$. Suppose we have guess $\boldsymbol{x}_t$. Then our update is:

Multivariate Newton Raphson

Suppose $f : \Re^n \rightarrow \Re$. Suppose we have guess $\boldsymbol{x}_t$. Then our update is:

$$\boldsymbol{x}_{t+1} = \boldsymbol{x}_t - \underbrace{\mathbf{H}(f)(\boldsymbol{x}_t)^{-1}}_{\text{Hessian}} \underbrace{\nabla f(\boldsymbol{x}_t)}_{\text{gradient}}$$

Multivariate Newton Raphson

Suppose $f : \Re^n \rightarrow \Re$. Suppose we have guess $\boldsymbol{x}_t$. Then our update is:

$$\boldsymbol{x}_{t+1} \quad = \quad \boldsymbol{x}_t - \underbrace{\mathbf{H}(f)(\boldsymbol{x}_t)^{-1}}_{\text{Hessian}} \underbrace{\nabla f(\boldsymbol{x}_t)}_{\text{gradient}}$$

Derivation (intuition):

Multivariate Newton Raphson

Suppose $f : \Re^n \rightarrow \Re$. Suppose we have guess $\boldsymbol{x}_t$. Then our update is:

$$\boldsymbol{x}_{t+1} = \boldsymbol{x}_t - \underbrace{\mathbf{H}(f)(\boldsymbol{x}_t)^{-1}}_{\text{Hessian}} \underbrace{\nabla f(\boldsymbol{x}_t)}_{\text{gradient}}$$

Derivation (intuition): Approximate function with tangent plane.

Multivariate Newton Raphson

Suppose $f : \Re^n \rightarrow \Re$. Suppose we have guess $\boldsymbol{x}_t$. Then our update is:

$$\boldsymbol{x}_{t+1} = \boldsymbol{x}_t - \underbrace{\mathbf{H}(f)(\boldsymbol{x}_t)^{-1}}_{\text{Hessian}} \underbrace{\nabla f(\boldsymbol{x}_t)}_{\text{gradient}}$$

Derivation (intuition): Approximate function with tangent plane. Find value of x_{t+1} that makes the plane equal to zero. Update again.

Multivariate Newton Raphson: Multivariate Linear Regression Model

Suppose that we have guess: $\left(\boldsymbol{\beta}_t, \sigma^2_t\right)$

$$
\begin{aligned}
(\boldsymbol{\beta}_{t+1}, \sigma^2_{t+1} &= \left(\boldsymbol{\beta}_t, \sigma^2_t\right) \underbrace{-\mathbf{H}(f)((\left(\boldsymbol{\beta}_t, \sigma^2_t\right))^{-1}}_{\text{Information}} \underbrace{\nabla f((\left(\boldsymbol{\beta}_t, \sigma^2_t\right)}_{\text{Score}} \\
\left(\boldsymbol{\beta}_{t+1}, \sigma^2_{t+1}\right) &= \left(\boldsymbol{\beta}_t, \sigma^2_t\right) + I_N(\boldsymbol{\beta}_t, \sigma^2_t)^{-1} \left(\sum_{i=1}^{N} s(Y_i | X_i, \boldsymbol{\beta}_t, \sigma^2_t) \right)
\end{aligned}
$$

So we have:

$$
\left(\boldsymbol{\beta}_{t+1}, \sigma^2_{t+1}\right) = \left(\boldsymbol{\beta}_t, \sigma^2_t\right)
$$

$$
+ \begin{pmatrix} \sigma^2 \left(\boldsymbol{X}^{'} \boldsymbol{X}\right)^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{2(\sigma^2)^2}{N} \end{pmatrix} \left(\sum_{i=1}^{N} \left(\frac{\left(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}_t\right)^{'} \boldsymbol{X}}{\sigma^2}, -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^{'} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) \right) \right)
$$

Multivariate Newton Raphson

Multivariate Newton Raphson

- Expensive to calculate (requires inverting Hessian)

Multivariate Newton Raphson

- Expensive to calculate (requires inverting Hessian)
- Very sensitive to starting points

Multivariate Newton Raphson

- Expensive to calculate (requires inverting Hessian)
- Very sensitive to starting points
- Ideally: method that exploits Newton-like structure, but is cheaper and more robust

Multivariate Newton Raphson

- Expensive to calculate (requires inverting Hessian)
- Very sensitive to starting points
- Ideally: method that exploits Newton-like structure, but is cheaper and more robust

`BFGS`: Quasi-Newton method

Multivariate Newton Raphson

- Expensive to calculate (requires inverting Hessian)
- Very sensitive to starting points
- Ideally: method that exploits Newton-like structure, but is cheaper and more robust

`BFGS`: Quasi-Newton method

`R` code

1) Multivariate regression
2) Information $\rightsquigarrow$ learning about uncertainty from model
3) Numerical optimization

Probit/Logit + BFGS next!!