

Optimization applied 1.5:

Maximum likelihood estimation continued

Marcin Lewandowski

November 28, 2024

Last meeting:

1. We have assumed that the data is generated from Bernoulli:

Last meeting:

1. We have assumed that the data is generated from Bernoulli:

- $Y \sim \text{Bern}(p)$

$$f(y; p) = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{if } y = 0 \end{cases}$$

Last meeting:

1. We have assumed that the data is generated from Bernoulli:

- $Y \sim \text{Bern}(p)$

$$f(y; p) = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{if } y = 0 \end{cases}$$

- The same can be also written as:

$$f(y; p) = p^y (1 - p)^{1-y}$$

Last meeting:

1. We have assumed that the data is generated from Bernoulli:

- $Y \sim \text{Bern}(p)$

$$f(y; p) = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{if } y = 0 \end{cases}$$

- The same can be also written as:

$$f(y; p) = p^y (1 - p)^{1-y}$$

- Examples:

Last meeting:

1. We have assumed that the data is generated from Bernoulli:

- $Y \sim \text{Bern}(p)$

$$f(y; p) = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{if } y = 0 \end{cases}$$

- The same can be also written as:

$$f(y; p) = p^y (1 - p)^{1-y}$$

- Examples:
 - Whether a person has a disease or not

Last meeting:

1. We have assumed that the data is generated from Bernoulli:

- $Y \sim \text{Bern}(p)$

$$f(y; p) = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{if } y = 0 \end{cases}$$

- The same can be also written as:

$$f(y; p) = p^y (1 - p)^{1-y}$$

- Examples:
 - Whether a person has a disease or not
 - Whether a person will default on a loan or not

Last meeting:

1. We have assumed that the data is generated from Bernoulli:

- $Y \sim \text{Bern}(p)$

$$f(y; p) = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{if } y = 0 \end{cases}$$

- The same can be also written as:

$$f(y; p) = p^y (1 - p)^{1-y}$$

- Examples:

- Whether a person has a disease or not
- Whether a person will default on a loan or not
- Whether a person will get to college or not

Last meeting:

1. We have assumed that the data is generated from Bernoulli:

- $Y \sim \text{Bern}(p)$

$$f(y; p) = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{if } y = 0 \end{cases}$$

- The same can be also written as:

$$f(y; p) = p^y (1 - p)^{1-y}$$

- Examples:

- Whether a person has a disease or not
- Whether a person will default on a loan or not
- Whether a person will get to college or not

2. Next: we have assumed that we **observe** a vector of data.

Last meeting:

1. We have assumed that the data is generated from Bernoulli:

- $Y \sim \text{Bern}(p)$

$$f(y; p) = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{if } y = 0 \end{cases}$$

- The same can be also written as:

$$f(y; p) = p^y (1 - p)^{1-y}$$

- Examples:

- Whether a person has a disease or not
- Whether a person will default on a loan or not
- Whether a person will get to college or not

2. Next: we have assumed that we **observe** a vector of data.

3. And want to estimate the underlying parameter p .

Last meeting:

- We've assumed y_i are independent
- The likelihood of observing $\mathbf{y} = \{1, 1, 1, 0, 1, 0\}$ is:

$$\begin{aligned} L(p \mid \mathbf{y}) &= p \cdot p \cdot p \cdot (1 - p) \cdot p \cdot (1 - p) \\ &= p^4(1 - p)^2 \end{aligned}$$

Last meeting:

- We've assumed y_i are independent
- The likelihood of observing $\mathbf{y} = \{1, 1, 1, 0, 1, 0\}$ is:

$$\begin{aligned}L(p \mid \mathbf{y}) &= p \cdot p \cdot p \cdot (1 - p) \cdot p \cdot (1 - p) \\ &= p^4(1 - p)^2\end{aligned}$$

Or:

$$L(p \mid \mathbf{y}) = \prod_{i=1}^6 p^{y_i} (1 - p)^{1 - y_i}$$

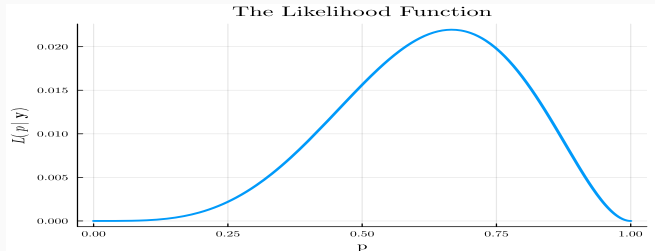
Last meeting:

- We've assumed y_i are independent
- The likelihood of observing $\mathbf{y} = \{1, 1, 1, 0, 1, 0\}$ is:

$$\begin{aligned} L(p \mid \mathbf{y}) &= p \cdot p \cdot p \cdot (1 - p) \cdot p \cdot (1 - p) \\ &= p^4(1 - p)^2 \end{aligned}$$

Or:

$$L(p \mid \mathbf{y}) = \prod_{i=1}^6 p^{y_i} (1 - p)^{1 - y_i}$$



MLE: Log likelihood function (Bernoulli)

$$L(p \mid \mathbf{y}) = \prod_{i=1}^n p^{y_i} \cdot (1 - p)^{1-y_i}$$

Necessary condition for maxima:

MLE: Log likelihood function (Bernoulli)

$$L(p \mid \mathbf{y}) = \prod_{i=1}^n p^{y_i} \cdot (1 - p)^{1-y_i}$$

$$\log(L(p \mid \mathbf{y})) = \sum_{i=1}^n \log(p^{y_i}) + \sum_{i=1}^n \log((1 - p)^{1-y_i})$$

Necessary condition for maxima:

MLE: Log likelihood function (Bernoulli)

$$\begin{aligned} L(p \mid \mathbf{y}) &= \prod_{i=1}^n p^{y_i} \cdot (1-p)^{1-y_i} \\ \log(L(p \mid \mathbf{y})) &= \sum_{i=1}^n \log(p^{y_i}) + \sum_{i=1}^n \log((1-p)^{1-y_i}) \\ &= \sum_{i=1}^n y_i \log(p) + \sum_{i=1}^n (1-y_i) \log(1-p) \end{aligned}$$

Necessary condition for maxima:

MLE: Log likelihood function (Bernoulli)

$$\begin{aligned}L(p \mid \mathbf{y}) &= \prod_{i=1}^n p^{y_i} \cdot (1-p)^{1-y_i} \\ \log(L(p \mid \mathbf{y})) &= \sum_{i=1}^n \log(p^{y_i}) + \sum_{i=1}^n \log((1-p)^{1-y_i}) \\ &= \sum_{i=1}^n y_i \log(p) + \sum_{i=1}^n (1-y_i) \log(1-p) \\ &= \log(p) \sum_{i=1}^n y_i + \log(1-p) \sum_{i=1}^n (1-y_i)\end{aligned}$$

Necessary condition for maxima:

MLE: Log likelihood function (Bernoulli)

$$\begin{aligned}L(p \mid \mathbf{y}) &= \prod_{i=1}^n p^{y_i} \cdot (1-p)^{1-y_i} \\ \log(L(p \mid \mathbf{y})) &= \sum_{i=1}^n \log(p^{y_i}) + \sum_{i=1}^n \log((1-p)^{1-y_i}) \\ &= \sum_{i=1}^n y_i \log(p) + \sum_{i=1}^n (1-y_i) \log(1-p) \\ &= \log(p) \sum_{i=1}^n y_i + \log(1-p) \sum_{i=1}^n (1-y_i)\end{aligned}$$

Necessary condition for maxima:

$$\frac{\partial \log(L(p \mid \mathbf{y}))}{\partial p} = 0$$

MLE: Log likelihood function (Bernoulli)

$$L(p | \mathbf{y}) = \prod_{i=1}^n p^{y_i} \cdot (1 - p)^{1-y_i}$$

$$\begin{aligned}\log(L(p | \mathbf{y})) &= \sum_{i=1}^n \log(p^{y_i}) + \sum_{i=1}^n \log((1 - p)^{1-y_i}) \\ &= \sum_{i=1}^n y_i \log(p) + \sum_{i=1}^n (1 - y_i) \log(1 - p) \\ &= \log(p) \sum_{i=1}^n y_i + \log(1 - p) \sum_{i=1}^n (1 - y_i)\end{aligned}$$

Necessary condition for maxima:

$$\begin{aligned}\frac{\partial \log(L(p | \mathbf{y}))}{\partial p} &= 0 \\ \frac{\sum_{i=1}^n y_i}{p} - \frac{\sum_{i=1}^n (1 - y_i)}{1 - p} &= 0\end{aligned}$$

But what if we want to model the probability:

How to model *probability* of a binary outcome, $Y \in \{0, 1\}$.

Recall that Y can be:

- Whether a person has a disease or not
- Whether a person will default on a loan or not
- Whether a person will get to college or not

But what if we want to model the probability:

How to model *probability* of a binary outcome, $Y \in \{0, 1\}$.

Recall that Y can be:

- Whether a person has a disease or not
- Whether a person will default on a loan or not
- Whether a person will get to college or not

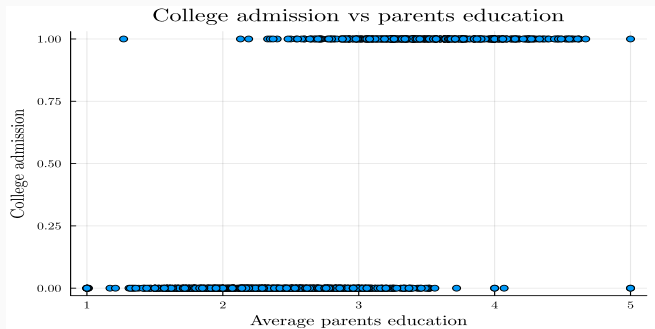
Let's assume that $P(Y = y_i | x)$ depends on a single predictor, x .

Making it more concrete:

Suppose you were asked to model the relationship between the probability of getting to college and the (average) parents education.

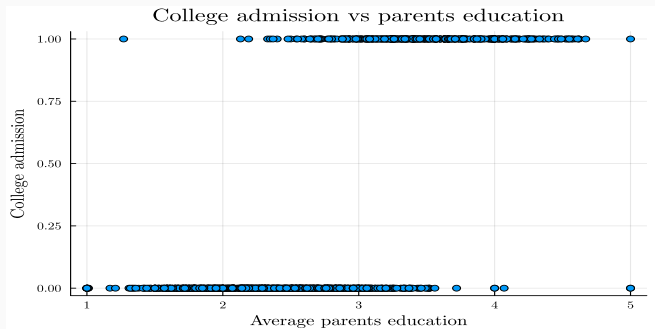
Making it more concrete:

Suppose you were asked to model the relationship between the probability of getting to college and the (average) parents education.



Making it more concrete:

Suppose you were asked to model the relationship between the probability of getting to college and the (average) parents education.



In this case:

- Y is whether a person gets to college or not
- x is the average parents education

The Logit Model:

- Let's assume that $P(Y = y_i | x)$ depends on a single predictor, x .

The Logit Model:

- Let's assume that $P(Y = y_i | x)$ depends on a single predictor, x .
- Thus p is some function of x !

The Logit Model:

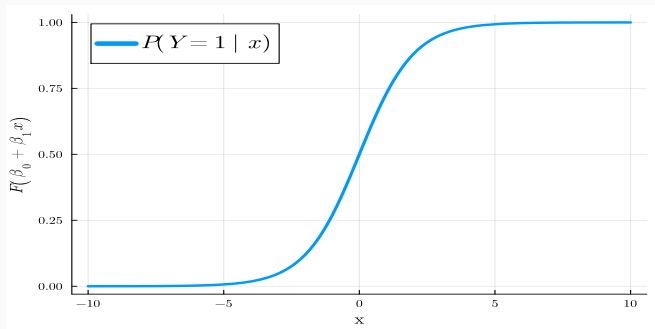
- Let's assume that $P(Y = y_i | x)$ depends on a single predictor, x .
- Thus p is some function of x !
- $P(Y = y_i | x_i) = F(\beta_0 + \beta_1 x)$

The Logit Model:

- Let's assume that $P(Y = y_i | x)$ depends on a single predictor, x .
- Thus p is some function of x !
- $P(Y = y_i | x_i) = F(\beta_0 + \beta_1 x)$

The Logit Model:

- Let's assume that $P(Y = y_i | x)$ depends on a single predictor, x .
- Thus p is some function of x !
- $P(Y = y_i | x_i) = F(\beta_0 + \beta_1 x)$



The Logit Model:

$$\begin{aligned}P(y = 1 \mid x) &= F(\beta_0 + \beta_1 x) \\&= \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \\&= \frac{1}{\frac{1}{e^{\beta_0 + \beta_1 x}} + 1}\end{aligned}$$

The Logit Model:

$$\begin{aligned}P(y = 1 \mid x) &= F(\beta_0 + \beta_1 x) \\&= \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \\&= \frac{1}{\frac{1}{e^{\beta_0 + \beta_1 x}} + 1}\end{aligned}$$

$$\begin{aligned}P(y = 1 \mid x) &= 1 - F(\beta_0 + \beta_1 x) \\&= 1 - \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \\&= \frac{1}{1 + e^{\beta_0 + \beta_1 x}}\end{aligned}$$

The Logit Model:

Suppose we want to know the probability of $Y = y_i$, where $y_i \in \{0, 1\}$.

$$P(Y = y_i | x_i) = P(Y = y_i | x)^{y_i} P(Y = 1 - y_i | x)^{1-y_i}$$

The Logit Model:

Suppose we want to know the probability of $Y = y_i$, where $y_i \in \{0, 1\}$.

$$P(Y = y_i | x_i) = P(Y = y_i | x)^{y_i} P(Y = 1 - y_i | x)^{1 - y_i}$$

Recall that in the Bernoulli case this was simply:

$$P(Y = y_i) = p^{y_i} (1 - p)^{1 - y_i}$$

The Logit Model:

Suppose we want to know the probability of $Y = y_i$, where $y_i \in \{0, 1\}$.

$$P(Y = y_i | x_i) = P(Y = y_i | x)^{y_i} P(Y = 1 - y_i | x)^{1-y_i}$$

Recall that in the Bernoulli case this was simply:

$$P(Y = y_i) = p^{y_i} (1 - p)^{1-y_i}$$

In logit model:

$$P(Y = y_i | x_i) = \left(\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right)^{y_i} \left(\frac{1}{1 + e^{\beta_0 + \beta_1 x_i}} \right)^{1-y_i}$$

The Logit Model:

Suppose we want to know the probability of $Y = y_i$, where $y_i \in \{0, 1\}$.

$$P(Y = y_i | x_i) = P(Y = y_i | x)^{y_i} P(Y = 1 - y_i | x)^{1-y_i}$$

Recall that in the Bernoulli case this was simply:

$$P(Y = y_i) = p^{y_i} (1 - p)^{1-y_i}$$

In logit model:

$$P(Y = y_i | x_i) = \left(\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right)^{y_i} \left(\frac{1}{1 + e^{\beta_0 + \beta_1 x_i}} \right)^{1-y_i}$$

The likelihood function is:

$$L(\beta_0, \beta_1 | \mathbf{y}, \mathbf{x}) = \prod_{i=1}^N \left(\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right)^{y_i} \left(\frac{1}{1 + e^{\beta_0 + \beta_1 x_i}} \right)^{1-y_i}$$

The likelihood function:

$$L(\beta_0, \beta_1 \mid \mathbf{y}, \mathbf{x}) = \prod_{i=1}^N \left(\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right)^{y_i} \left(\frac{1}{1 + e^{\beta_0 + \beta_1 x_i}} \right)^{1-y_i}$$

The likelihood function:

$$L(\beta_0, \beta_1 \mid \mathbf{y}, \mathbf{x}) = \prod_{i=1}^N \left(\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right)^{y_i} \left(\frac{1}{1 + e^{\beta_0 + \beta_1 x_i}} \right)^{1-y_i}$$

For example, suppose we have 3 observations:

- $y = [1, 0, 1]$
- $x = [5, 2, 1]$

The likelihood function:

$$L(\beta_0, \beta_1 \mid \mathbf{y}, \mathbf{x}) = \prod_{i=1}^N \left(\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right)^{y_i} \left(\frac{1}{1 + e^{\beta_0 + \beta_1 x_i}} \right)^{1-y_i}$$

For example, suppose we have 3 observations:

- $y = [1, 0, 1]$
- $x = [5, 2, 1]$

$$L(\beta_0, \beta_1 \mid \mathbf{y}, \mathbf{x}) = \left(\frac{e^{\beta_0 + 5\beta_1}}{1 + e^{\beta_0 + 5\beta_1}} \right) \cdot 1 \cdot 1 \left(\frac{1}{1 + e^{\beta_0 + 2\beta_1}} \right) \cdot 1 \cdot \left(\frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}} \right)$$

The log likelihood function:

Given:

$$L(\beta_0, \beta_1 \mid \mathbf{y}, \mathbf{x}) = \prod_{i=1}^N \left(\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right)^{y_i} \left(\frac{1}{1 + e^{\beta_0 + \beta_1 x_i}} \right)^{1-y_i}$$

The log likelihood function is:

$$\log(L(\beta_0, \beta_1 \mid \mathbf{y}, \mathbf{x})) = \sum_{i=1}^N \left\{ y_i \log \left(\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right) + (1 - y_i) \log \left(\frac{1}{1 + e^{\beta_0 + \beta_1 x_i}} \right) \right\}$$

The log likelihood function

Note that if needed one can split the log likelihood function into two parts:

$$\begin{aligned}\log(L(\beta_0, \beta_1 \mid \mathbf{y}, \mathbf{x})) &= \sum_{i=1}^N y_i \underbrace{\log\left(\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}\right)}_{\text{The first part}} \\ &+ \sum_{i=1}^N (1 - y_i) \underbrace{\log\left(\frac{1}{1 + e^{\beta_0 + \beta_1 x_i}}\right)}_{\text{The second part}}\end{aligned}$$