Maximum likelihood estimation

Recap: ways of fitting linear regression models

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \cdots + \beta_k X_{i,k} + \varepsilon_i \qquad \varepsilon_i \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$$

We observe data $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$, where $X_i = (1, X_{i,1}, \ldots, X_{i,k})^T$. We want to estimate

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$$

Possible methods

- Minimize SSE
- Projection ($\Leftrightarrow$ to minimizing SSE)
- Maximize likelihood

$$Y_i \sim N(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik},\ \sigma_\varepsilon^2)$$

$$L(\beta_0, \ldots, \beta_k, \sigma_\varepsilon^2 \mid (X_1, Y_1), \ldots, (X_n, Y_n))$$

$$\overset{\text{(independence)}}{=} \prod_{i=1}^{n} \underbrace{f(Y_i \mid \beta_0, \ldots, \beta_k, \sigma_\varepsilon^2, X_i)}_{\frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left\{-\frac{1}{2\sigma_\varepsilon^2}(Y_i - \beta_0 - \beta_1 X_{i1} - \cdots - \beta_k X_{ik})^2\right\}}$$

$$= (2\pi)^{-\frac{n}{2}} \sigma_\varepsilon^{-n} \exp\left\{-\frac{1}{2\sigma_\varepsilon^2} \underbrace{\sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_{i1} - \cdots - \beta_k X_{ik})^2}_{\text{SSE!}}\right\}$$

maximizing likelihood $\iff$ minimize SSE
(for normal data)

Summary: three ways of fitting linear regression models

linear regression:

$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \cdots + \hat{\beta}_k X_{ik}$

$\Rightarrow Y_i - \hat{Y}_i$ makes sense

- Minimize SSE, via derivatives of $$\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_{i,1} - \cdots - \beta_k X_{i,k})^2$$

not appropriate for logistic regression b/c we don't have the same idea of a residual

- Minimize $||Y - \widehat{Y}||$ (equivalent to minimizing SSE)
- Maximize likelihood (for *normal* data, equivalent to minimizing SSE)

could be appropriate for logistic regression, but need to use Bernoulli instead of Normal

> Which of these three methods, if any, is appropriate for fitting a logistic regression model? Do any changes need to be made for the logistic regression setting?

Idea: pick parameter value for which the observed data has the highest "probability"

Step back: likelihoods and estimation

Let $Y \sim Bernoulli(p)$ be a Bernoulli random variable, with $p \in [0, 1]$. We observe 5 independent samples from this distribution:

$$Y_1 = 1,\ Y_2 = 1,\ Y_3 = 0,\ Y_4 = 0,\ Y_5 = 1$$

> The true value of p is unknown, so two friends propose different guesses for the value of p: 0.3 and 0.7. Which do you think is a "better" guess?

Sample proportion = 0.6 (closer to 0.7)

$P(\text{data} \mid p = 0.3) = (0.3)(0.3)(1-0.3)(1-0.3)(0.3)$

$= 0.3^3\, 0.7^2 = 0.013$

$P(\text{data} \mid p = 0.7) = 0.7^3\, 0.3^2 = 0.031$

Likelihood

Definition: Let $\mathbf{Y} = (Y_1, \ldots, Y_n)$ be a sample of n observations, and let $f(\mathbf{y}|\theta)$ denote the joint pdf or pmf of $\mathbf{Y}$, with parameter(s) θ. The *likelihood function* is

$$\underbrace{L(\theta|\mathbf{Y})}_{\text{function of } \theta \text{, given } Y} = \underbrace{f(\mathbf{Y}|\theta)}_{\text{function of } Y \text{, given } \theta}$$

"probability" of the observed data

- $L(\theta|Y)$: condition on the observed data. Want to know how "probability" (joint distribution) of Y changes as a function of θ
- Since $f(y|\theta) \geq 0$, $L(\theta|Y) \geq 0 \quad \forall \theta$

Special case: $Y_1, \ldots, Y_n$ iid w/ pdf or pmf f

$$L(\theta|Y) = \prod_{i=1}^{n} f(Y_i|\theta)$$

Example: Bernoulli data

Let $Y_1, \ldots, Y_n \overset{iid}{\sim}$ Bernoulli(p) $\qquad f(y|p) = p^y (1-p)^{1-y}$

$$L(p|Y) = \prod_{i=1}^{n} f(Y_i|p) = \prod_{i=1}^{n} p^{Y_i} (1-p)^{1-Y_i}$$

$$= p^{\sum_{i=1}^{n} Y_i} (1-p)^{n - \sum_{i=1}^{n} Y_i}$$

<u>Example</u>: $\quad Y = (1,1,0,0,1)$

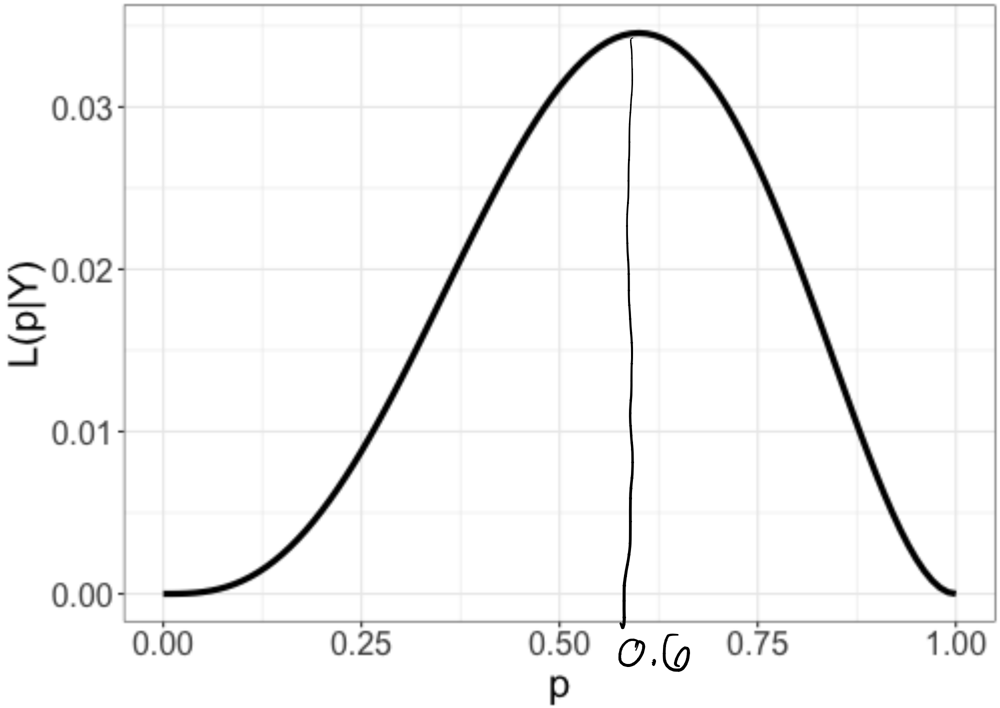
$$L(p|Y) = p^3 (1-p)^2$$

Example: Bernoulli data

$Y_1, \ldots, Y_5 \overset{iid}{\sim} Bernoulli(p)$, with observed data

$$Y_1 = 1,\ Y_2 = 1,\ Y_3 = 0,\ Y_4 = 0,\ Y_5 = 1$$

$L(p|\mathbf{Y}) = p^3(1-p)^2$



Maximum likelihood estimator

Definition: Let $\mathbf{Y} = (Y_1, \ldots, Y_n)$ be a sample of n observations. The *maximum likelihood estimator* (MLE) is

$$\hat{\theta} = \operatorname{argmax}_\theta L(\theta|\mathbf{Y})$$

$\operatorname{argmax}_\theta$ means "value of θ that maximizes..."

Example: $Bernoulli(p)$

$Y_1, \dots, Y_n \overset{iid}{\sim} \text{Bernoulli}(p)$

$L(p|Y) = p^{\sum_i Y_i} (1-p)^{n - \sum_i Y_i}$

Now maximize!

① Take log to make life easier (log is monotone increasing, so if $\hat{p}$ maximizes $\log L(p|Y)$ then $\hat{p}$ maximizes $L(p|Y)$)

$$\ell(p|Y) = \log L(p|Y) = \left(\sum_i Y_i\right) \log p + \left(n - \sum_i Y_i\right) \log(1-p)$$

② Differentiate wrt parameter of interest:

$$\frac{\partial}{\partial p} \ell(p|Y) = \frac{\sum_i Y_i}{p} - \frac{\left(n - \sum_i Y_i\right)}{1-p}$$

(3) set $= 0$ & solve

$$\frac{\partial}{\partial p} \ell(p|Y) = \frac{\sum_i Y_i}{p} - \frac{(n - \sum_i Y_i)}{1-p} \overset{set}{=} 0$$

$$\Rightarrow \quad \frac{\sum_i Y_i}{p} = \frac{(n - \sum_i Y_i)}{1-p}$$

$$\Rightarrow \quad \frac{1-p}{p} = \frac{n - \sum_i Y_i}{\sum_i Y_i}$$

$$\Rightarrow \quad \frac{1}{p} = \frac{n}{\sum_i Y_i} \quad \Rightarrow \quad p = \frac{\sum_i Y_i}{n}$$ (sample proportion)

So $\frac{\partial}{\partial p} \ell(p|Y)$ has a max. or a min. at $p = \frac{\sum_i Y_i}{n} = \bar{Y}$

(4) Check max. or min:

$$\frac{\partial^2}{\partial p^2} \ell(p|Y)\Big|_{p=\bar{Y}} = -\frac{\sum_i Y_i}{p^2} - \frac{(n - \sum_i Y_i)}{(1-p^2)}\Bigg|_{p=\bar{Y}} < 0$$

$\Rightarrow$ maximum!

(5) check bondary points: $p=0$ and $p=1$

$$\ell(p \mid Y) = \begin{cases} n \log(1-p) & p=0 \\ n \log p & p=1 \end{cases}$$

If either $p=0$ or $p=1$ has highest likelihood,

$p = \dfrac{\sum_i Y_i}{n}$ anyway

So $\boxed{\hat{p} = \dfrac{\sum_i Y_i}{n} = \bar{Y}}$

Example: $N(\theta, 1)$

$\theta \in (-\infty, \infty)$

$Y_1, \dots, Y_n \overset{iid}{\sim} N(\theta, 1)$ $\qquad f(y \mid \theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-\theta)^2}$

$$L(\theta \mid Y) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(Y_i - \theta)^2\right\}$$

$$= (2\pi)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_i (Y_i - \theta)^2\right\}$$

$$\Rightarrow \ell(\theta \mid Y) = \cancel{-\frac{n}{2}\log(2\pi)} - \frac{1}{2}\sum_{i=1}^{n}(Y_i - \theta)^2$$

$$\frac{\partial}{\partial \theta}\ell(\theta \mid Y) = -\frac{1}{2}\sum_{i=1}^{n}\frac{\partial}{\partial \theta}\left((Y_i - \theta)^2\right)$$

$$= -\frac{1}{2}\sum_{i=1}^{n} 2(Y_i - \theta)(-1) = \sum_{i=1}^{n}(Y_i - \theta)$$

$$\Rightarrow \sum_{i=1}^{n} Y_i = \sum_{i=1}^{n}\theta = n\theta \qquad (\overset{set}{=} 0)$$

$$\Rightarrow \theta = \frac{\sum_i Y_i}{n} = \overline{Y}$$

Example: $Uniform(0, \theta)$