Maximum Likelihood Estimation

Dr. Daniel B. Rowe
Professor of Computational Statistics
Department of Mathematical and Statistical Sciences
Marquette University

Outline

MLE General Process

MLEs for Univariate Normal PDF

MLEs for Multivariate Normal PDF

MLEs for Simple Linear Regression

Discussion

Homework

MLE General Process

In Probability when we have a discrete PMF, we describe $P(X=x|\theta)$ as the probability that we observe the specific value x for our random variable X given (assuming that we know) the PMF parameters θ.

When we have a continuous PDF, we describe $f(x|\theta)$ as a continuous function that contains the probability of observing a specific value x for our random variable X between a and b given the PDF parameter θ.
$P(a< x<b|\theta)$

MLE General Process

In Statistics, we observe a sample of random observations $x_1,\ldots,x_n$ from a PMF $P(X=x|\theta)$ or PDF $f(x|\theta)$ and wish to estimate the associated parameters θ.

There is a general technique called Maximum Likelihood Estimation that yields estimators $\hat{\theta}$ for the parameters θ called MLEs.

Estimator = general formula for calculating the MLE.
Estimate = the numerical value using the formula.

MLE General Process

The MLE parameter estimation process is to write down the joint PMF or PDF for the sample of random observations $x_1,\ldots,x_n$, which we will call the likelihood function $L(\theta)$ for the parameters.

$$L(\theta) = f(x_1 | \theta) f(x_2 | \theta) \cdots f(x_n | \theta)$$ ⟵ if independent (joint pdf if not independent)

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

The MLEs for θ are the values that maximize the likelihood

$$\hat{\theta} = \underset{\theta}{ArgMax} L(\theta)$$

works for any PDF but may need to numerically maximize

MLEs for Univariate Normal PDF

As an example, consider the normal PDF for x

$$f(x|\mu,\sigma^2)=\frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$$

iid=independent and identically distributed

if we have $x_1,\ldots,x_n$ that are iid from $f(x|\mu,\sigma^2)$, our likelihood is

$$L(\mu,\sigma^2)=\prod_{i=1}^{n} f(x_i|\mu,\sigma^2)$$

$$L(\mu,\sigma^2)=(2\pi\sigma^2)^{-n/2}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right]$$

MLEs for Univariate Normal PDF

The best way to maximize the likelihood $L(\mu,\sigma^2)$ is to take the natural logarithm, $\ln(L(\mu,\sigma^2))$, which is a monotonic function, and minimizing the monotonic log function yields the same maxima.

$$\ln(L(\mu,\sigma^2)) = -\tfrac{n}{2}\ln(2\pi) - \tfrac{n}{2}\ln(\sigma^2) - \tfrac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

MLEs for Univariate Normal PDF

Taking derivatives of the natural logarithm of the Normal likelihood yields

$$\ln(L(\mu,\sigma^2)) = -\tfrac{n}{2}\ln(2\pi) - \tfrac{n}{2}\ln(\sigma^2) - \tfrac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\frac{\partial}{\partial\mu}\ln(L(\mu,\sigma^2)) = -\tfrac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)(2)(-1)$$

$$\frac{\partial}{\partial\sigma^2}\ln(L(\mu,\sigma^2)) = -\tfrac{n}{2}\frac{1}{\sigma^2} - \tfrac{1}{2}\frac{-1}{(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

MLEs for Univariate Normal PDF

and upon setting equal to zero with min values having hats

$$-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\hat{\mu})(2)(-1)=0$$

$$-\frac{n}{2}\frac{1}{\hat{\sigma}^2}-\frac{1}{2}\frac{-1}{(\hat{\sigma}^2)^2}\sum_{i=1}^{n}(x_i-\hat{\mu})^2=0$$

and solving we get

$$\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}x_i$$

$$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu})^2$$

⟵ Note denominator of n not n-1.

MLEs for Univariate Normal PDF

Let's explore a different approach to maximizing that we will use for the multivariate normal PDF to avoid vector derivatives or gradients and other uncomfortable things.

Looking at the log likelihood, we see that there is only one term that involves μ.

$$\ln(L(\mu,\sigma^2)) = -\tfrac{n}{2}\ln(2\pi) - \tfrac{n}{2}\ln(\sigma^2) - \tfrac{1}{2\sigma^2}\boxed{\sum_{i=1}^{n}(x_i-\mu)^2}$$

So we want to find the value of μ that makes $\sum_{i=1}^{n}(x_i-\mu)^2$ the smallest.

MLEs for Univariate Normal PDF

$$\ln(L(\mu,\sigma^2)) = -\tfrac{n}{2}\ln(2\pi) - \tfrac{n}{2}\ln(\sigma^2) - \tfrac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

We can define $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ then add and subtract as

$$\begin{aligned}
\sum_{i=1}^{n}(x_i-\mu)^2 &= \sum_{i=1}^{n}[(x_i-\bar{x})+(\bar{x}-\mu)]^2 \\
&= \sum_{i=1}^{n}[(x_i-\bar{x})^2 - 2(x_i-\bar{x})(\bar{x}-\mu)+(\bar{x}-\mu)^2] \\
&= \sum_{i=1}^{n}(x_i-\bar{x})^2 - 2\sum_{i=1}^{n}(\bar{x}x_i-\mu x_i-\bar{x}\bar{x}+\bar{x}\mu)+n(\bar{x}-\mu)^2 \\
&= \sum_{i=1}^{n}(x_i-\bar{x})^2 - 2(n\bar{x}^2-n\mu\bar{x}-n\bar{x}^2+n\bar{x}\mu)+n(\bar{x}-\mu)^2 \\
&= \sum_{i=1}^{n}(x_i-\bar{x})^2 + n(\bar{x}-\mu)^2
\end{aligned}$$

and the value of μ being $\hat{\mu} = \bar{x}$ yields the min, therefore MLE!

MLEs for Multivariate Normal PDF

This MLE process also works for multivariate PDFs!

Consider the bivariate normal PDF for x

$$f(x|\mu,\Sigma)=(2\pi)^{-p/2}|\Sigma|^{-1/2}e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)} \qquad x,\mu\in\mathbb{R} \quad \Sigma>0 \quad p=2$$

.

If we have $x_1,\ldots,x_n$ that are iid from $f(x|\mu,\Sigma)$, our likelihood is

$$L(\mu,\Sigma)=\prod_{i=1}^{n}f(x_i|\mu,\Sigma)$$

$$L(\mu,\Sigma)=(2\pi)^{-np/2}|\Sigma|^{-n/2}\exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)'\Sigma^{-1}(x_i-\mu)\right]$$

Don't forget that $f()$ and $L()$ are scalar functions.

MLEs for Multivariate Normal PDF

The best way to maximize the likelihood $L(\mu,\Sigma)$ is to take the natural logarithm, $\ln(L(\mu,\Sigma))$, which is a monotonic function, and minimizing the monotonic log function yields the same maxima.

$$\ln(L(\mu,\Sigma)) = -\tfrac{n}{2}\ln(2\pi) - \tfrac{n}{2}\ln(|\Sigma|) - \tfrac{1}{2}\sum_{i=1}^{n}(x_i-\mu)'\Sigma^{-1}(x_i-\mu)$$

$$L(\mu,\Sigma) = (2\pi)^{-np/2}|\Sigma|^{-n/2}\exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)'\Sigma^{-1}(x_i-\mu)\right]$$

MLEs for Multivariate Normal PDF

Taking derivatives of the natural logarithm of the Normal likelihood yields

$$\ln(L(\mu,\Sigma)) = -\tfrac{n}{2}\ln(2\pi) - \tfrac{n}{2}\ln(|\Sigma|) - \tfrac{1}{2}\sum_{i=1}^{n}(x_i-\mu)'\Sigma^{-1}(x_i-\mu)$$

$$\frac{\partial}{\partial\mu}\ln(L(\mu,\Sigma)) = -\tfrac{1}{2}\sum_{i=1}^{n}\Sigma^{-1}(x_i-\mu)(2)(-1)$$

$$\frac{\partial}{\partial\Sigma}\ln(L(\mu,\Sigma)) = -\tfrac{n}{2}\Sigma^{-1} - \tfrac{1}{2}\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)'\Sigma^{-1}(-1)$$

MLEs for Multivariate Normal PDF

and upon setting equal to zero

$$-\tfrac{1}{2}\hat{\Sigma}^{-1}\sum_{i=1}^{n}(x_i-\hat{\mu})(2)(-1)=0$$

$$-\tfrac{n}{2}\hat{\Sigma}^{-1}-\tfrac{1}{2}\sum_{i=1}^{n}(x_i-\hat{\mu})(x_i-\hat{\mu})'\hat{\Sigma}^{-2}(-1)=0$$

and solving we get

$$\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}x_i$$

$$\underset{p\times p}{\hat{\Sigma}}=\frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu})(x_i-\hat{\mu})'$$

← Note denominator of n not n-1.

MLEs for Univariate Normal PDF

$$\sum_{i=1}^{n}(x_i-\mu)'\Sigma^{-1}(x_i-\mu)=tr\Sigma^{-1}(X-1_n\mu')'(X-1_n\mu')$$

Let's explore a different approach to maximizing that will avoid vector derivatives and other uncomfortable things.

Looking at the log likelihood, we see that there is only one term that involves μ.

$$\ln(L(\mu,\Sigma))=-\tfrac{n}{2}\ln(2\pi)-\tfrac{n}{2}\ln(|\Sigma|)-\tfrac{1}{2}\boxed{\sum_{i=1}^{n}(x_i-\mu)'\Sigma^{-1}(x_i-\mu)}$$

So we want to find the value of μ that makes $\sum_{i=1}^{n}(x_i-\mu)'\Sigma^{-1}(x_i-\mu)$ the smallest.

MLEs for Multivariate Normal PDF

It can be shown that

$$\sum_{i=1}^{n} \underbrace{(x_i-\mu)'}_{1\times 2} \underset{2\times 2}{\Sigma^{-1}} \underbrace{(x_i-\mu)}_{2\times 1} = tr \underset{2\times 2}{\Sigma^{-1}} \underbrace{(X-1_n\mu')'}_{2\times n} \underbrace{(X-1_n\mu')}_{n\times 2}$$

$tr()$=trace

$$tr(AB)=tr(BA)$$

if conformable

where $\underset{n\times 2}{X}=(x_1,\ldots,x_n)', \underset{n\times 1}{1_n}=(1,\ldots,1)'$, and $\underset{2\times 1}{\mu}=(\mu_1,\mu_2)'$.

MLEs for Multivariate Normal PDF

tr=trace

We can define $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$ then add and subtract as

$$
\begin{aligned}
(X-1_n\mu')'(X-1_n\mu') &= [(X-1_n\bar{x}')+(1_n\bar{x}'-1_n\mu')]'[(X-1_n\bar{x}')+(1_n\bar{x}'-1_n\mu')] \\
&= \begin{array}{l}(X-1_n\bar{x}')'(X-1_n\bar{x}')+(X-1_n\bar{x}')'(1_n\bar{x}'-1_n\mu') \\ +(1_n\bar{x}'-1_n\mu')'(X-1_n\bar{x}')+(1_n\bar{x}'-1_n\mu')'(1_n\bar{x}'-1_n\mu')\end{array} \\
&= \begin{array}{l}(X-1_n\bar{x}')'(X-1_n\bar{x}')+(1_n\bar{x}'-1_n\mu')'(1_n\bar{x}'-1_n\mu') \\ +(X'-\bar{x}1_n')(1_n\bar{x}'-1_n\mu')+(\bar{x}1_n'-\mu 1_n')(X-1_n\bar{x}')\end{array} \\
&= \begin{array}{l}(X-1_n\bar{x}')'(X-1_n\bar{x}')+(1_n\bar{x}'-1_n\mu')'(1_n\bar{x}'-1_n\mu') \\ +X'1_n\bar{x}'-X'1_n\mu'-\bar{x}1_n'1_n\bar{x}'+\bar{x}1_n'1_n\mu' \\ +\bar{x}1_n'X-\bar{x}1_n'1_n\bar{x}'-\mu 1_n'X+\mu 1_n'1_n\bar{x}'\end{array} \\
&= (X-1_n\bar{x}')'(X-1_n\bar{x}')+(\bar{x}-\mu)(1_n'1_n)(\bar{x}-\mu)'
\end{aligned}
$$

and the value of μ being $\hat{\mu}=\bar{x}$ yields the min, therefore MLE!

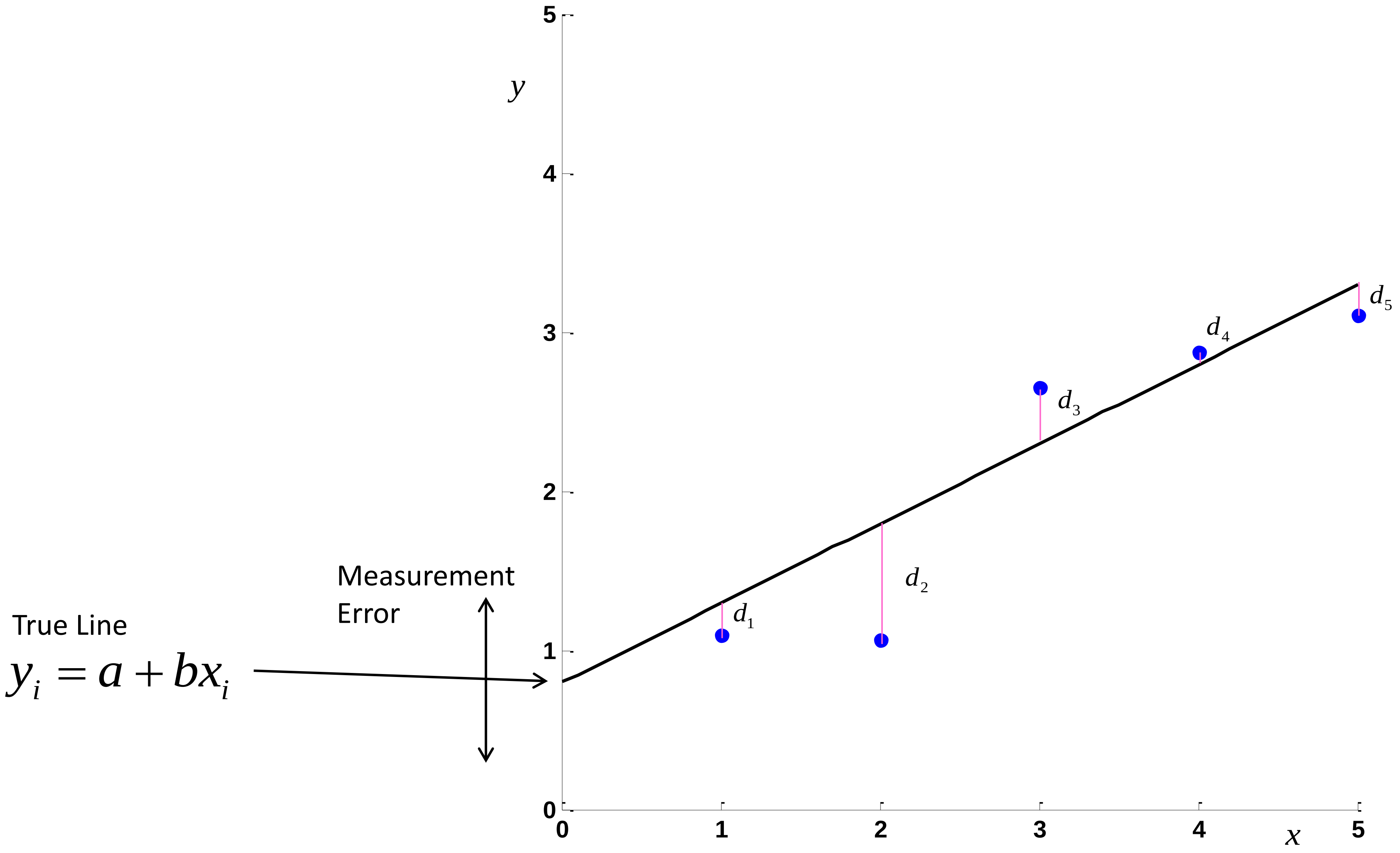
MLEs for Simple Linear Regression

This technique, can be generalized to linear regression.

Let $y_i = a + bx_i + \varepsilon_i$,

where $\varepsilon_i \sim N(0,\sigma^2)$

are independent.



MLEs for Simple Linear Regression

This technique, can be generalized to linear regression.

Let $y_i = a + bx_i + \varepsilon_i$, where $\varepsilon_i \sim N(0,\sigma^2)$ are independent.

Then, the joint PDF or likelihood is

$$f(y_1,...,y_n \mid a,b,\sigma^2) = f(y_1 \mid a,b,\sigma^2)...f(y_n \mid a,b,\sigma^2)$$

$$f(y_1,...,y_n \mid a,b,\sigma^2) = \frac{\exp[-(y_1 - a - bx_1)^2 / 2\sigma^2]}{(2\pi\sigma^2)^{1/2}} ... \frac{\exp[-(y_n - a - bx_n)^2 / 2\sigma^2]}{(2\pi\sigma^2)^{1/2}}$$

$$L(a,b,\sigma^2) = \frac{\exp[-(y_1 - a - bx_1)^2 / 2\sigma^2]}{(2\pi\sigma^2)^{1/2}} ... \frac{\exp[-(y_n - a - bx_n)^2 / 2\sigma^2]}{(2\pi\sigma^2)^{1/2}}$$

MLEs for Simple Linear Regression

This technique, can be generalized to linear regression.

Let $y_i = a + bx_i + \varepsilon_i$, where $\varepsilon_i \sim N(0,\sigma^2)$ are independent.

Then, the likelihood is

$$L(a,b,\sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - a - bx_i)^2 \right]$$

and the log likelihood is

$$LL(a,b,\sigma^2) = \underbrace{-\frac{n}{2}\log(2\pi)}_{\text{no } a \text{ or } b} \underbrace{-\frac{n}{2}\log(\sigma^2)}_{\text{no } a \text{ or } b} \underbrace{-\frac{1}{2\sigma^2}\boxed{\sum_{i=1}^{n} (y_i - a - bx_i)^2}}_{a \text{ or } b} .$$

MLEs for Simple Linear Regression

$L(a,b,\sigma^2)$ is again called the likelihood function.

What we want to do is find the values of (a,b,σ^2)

that maximize $L(a,b,\sigma^2)$. The values (a,b) that maximize

$L(a,b,\sigma^2)$ are the values $(\hat{a},\hat{b})$ that minimize $\sum_{i=1}^{n}(y_i-\hat{a}-\hat{b}x_i)^2$.

The value of σ^2 that maximizes $L(a,b,\sigma^2)$ is

$$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{a}-\hat{b}x_i)^2 \ .$$

⟵ Note denominator of n not n-1.

$$d_i = y_i-\hat{a}-\hat{b}x_i$$

minimize $\sum_{i=1}^{n} d_i^2$

MLEs for Simple Linear Regression

Differentiate $LL(a,b,\sigma^2)$ wrt a, b, and σ^2, then set = 0

$$LL(a,b,\sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - a - bx_i)^2$$

$$\left.\frac{\partial LL(a,b,\sigma^2)}{\partial a}\right|_{\hat{a},\hat{b},\hat{\sigma}^2} = -\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{n}2(y_i - \hat{a} - \hat{b}x_i)(-1) = 0$$

$$\left.\frac{\partial LL(a,b,\sigma^2)}{\partial b}\right|_{\hat{a},\hat{b},\hat{\sigma}^2} = -\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{n}2(y_i - \hat{a} - \hat{b}x_i)(-x_i) = 0$$

$$\left.\frac{\partial LL(a,b,\sigma^2)}{\partial \sigma^2}\right|_{\hat{a},\hat{b},\hat{\sigma}^2} = -\frac{n}{2}\frac{1}{\hat{\sigma}^2} - \frac{-1}{2(\hat{\sigma}^2)^2}\sum_{i=1}^{n}(y_i - \hat{a} - \hat{b}x_i)^2 = 0$$

MLEs for Simple Linear Regression

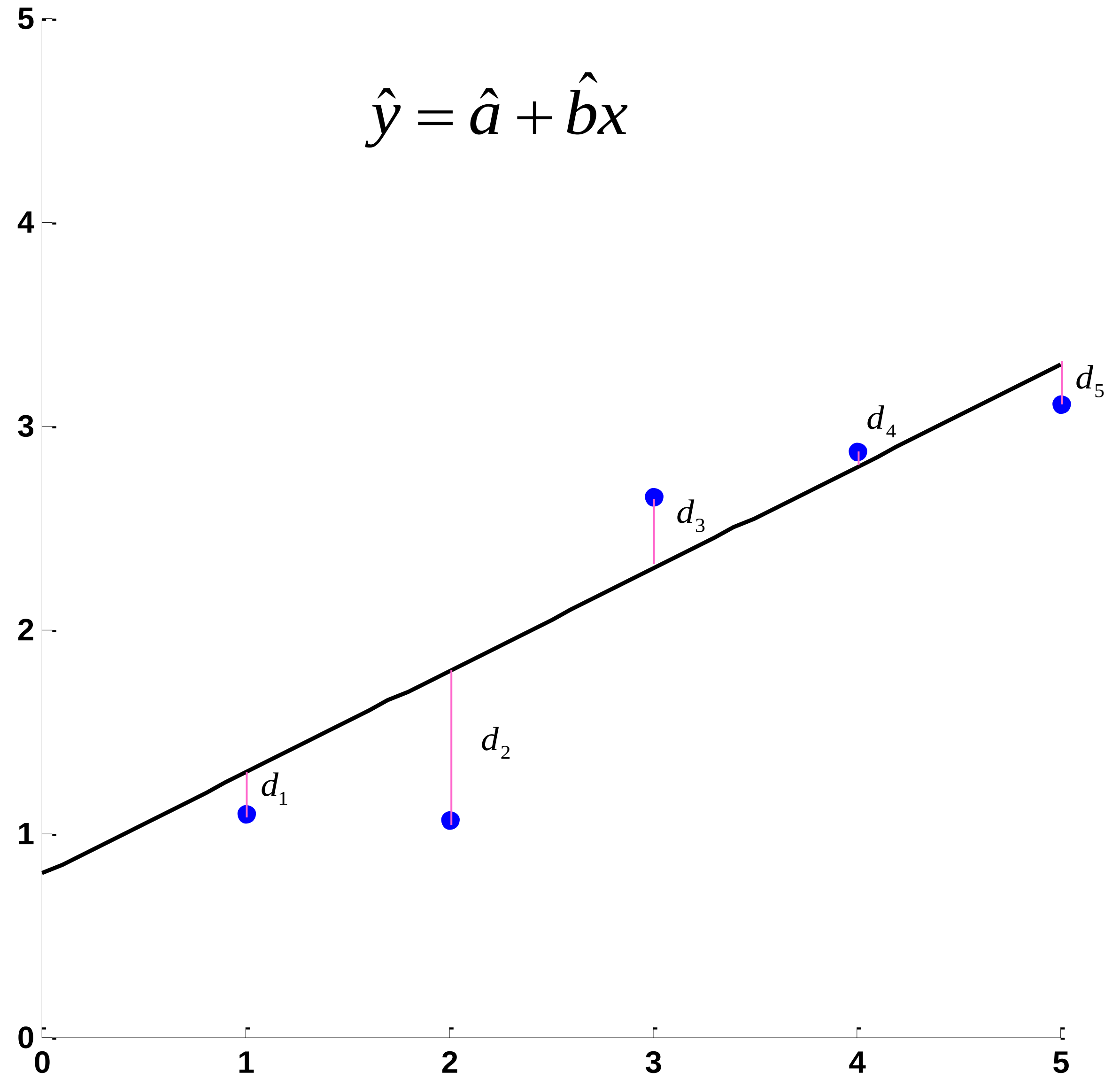
Solving for the estimated parameters yields

$$\hat{b} = \frac{n(\sum_{i=1}^{n} x_i y_i) - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n(\sum_{i=1}^{n} x_i^2) - (\sum_{i=1}^{n} x_i)^2}$$

$$\hat{a} = \frac{(\sum_{i=1}^{n} y_i)(\sum_{i=1}^{n} x_i^2) - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} x_i y_i)}{n(\sum_{i=1}^{n} x_i^2) - (\sum_{i=1}^{n} x_i)^2}$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{a} - \hat{b}x_i)^2$$



MLEs for Simple Linear Regression

The regression model $y_i = a + bx_i + \varepsilon_i$ where $\varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)$, $i=1,\ldots,n$,

that we presented, can be equivalently written as

measured data, design matrix, regression coefficients, measurement error

$$y = X\beta + \varepsilon$$

where

$$\underset{n\times 1}{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \underset{n\times 2}{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \underset{2\times 1}{\beta} = \begin{pmatrix} a \\ b \end{pmatrix}, \underset{n\times 1}{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix},$$

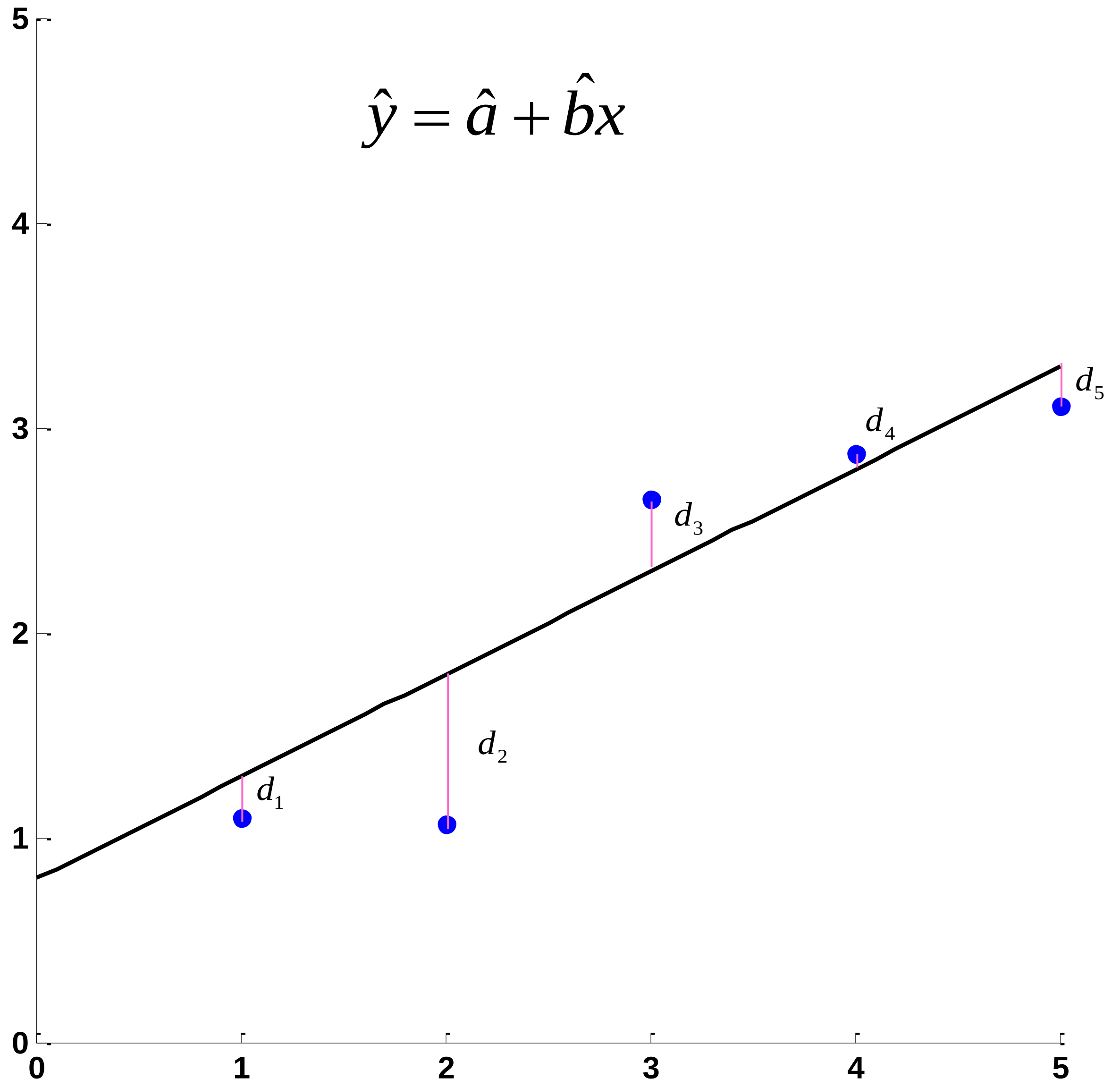
and $\varepsilon \sim N(0,\sigma^2 I_n)$. I_n is an n-dimensional identity matrix.

MLEs for Simple Linear Regression

The regression model is $y = X\beta + \varepsilon$ where $\varepsilon \sim N(0,\sigma^2 I_n)$.

$$\underset{n\times 1}{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}} = \underset{n\times 2}{\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}} \underset{2\times 1}{\begin{pmatrix} a \\ b \end{pmatrix}} + \underset{n\times 1}{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}}$$

$y_i = a + bx_i + \varepsilon_i$ for $i=1,\ldots,n$.



MLEs for Simple Linear Regression

With $\underset{n\times 1}{y} = X\beta + \underset{n\times 1}{\varepsilon}$ and $\varepsilon \sim N(0,\sigma^2 I_n)$

The likelihood is

$$f(y_1,\ldots,y_n \mid a,b,\sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right]$$

compare to

$$f(y_1,\ldots,y_n \mid a,b,\sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - a - bx_i)^2\right]$$

and the log likelihood is

$$LL(a,b,\sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\boxed{(y - X\beta)'(y - X\beta)}\ .$$

MLEs for Simple Linear Regression

$L(a,b,\sigma^2)$ is again called the likelihood function.

What we want to do is find the values of (β,σ^2)

that maximize $L(\beta,\sigma^2)$. The value of β that maximizes

$L(\beta,\sigma^2)$ is the value $\hat{\beta}$ that minimizes $(y - X\beta)'(y - X\beta)$.

The value of σ^2 that maximizes $L(\beta,\sigma^2)$ is

$$d_i = y_i - \hat{a} - \hat{b}x_i$$

$$\hat{\sigma}^2 = \frac{1}{n}(y - X\hat{\beta})'(y - X\hat{\beta})$$ ⟵ Note denominator of n not n-1.

$$\underset{\text{wrt } \beta}{\text{minimize}}\ (y - X\beta)'(y - X\beta)$$

We need to find $\hat{\beta}$.

MLEs for Simple Linear Regression

We don't need to take the derivative of $L(\beta,\sigma^2)$

wrt β (although we could). We can write with algebra

$$(y-X\beta)'(y-X\beta)=(y-X\hat{\beta})'(y-X\hat{\beta})+(\beta-\hat{\beta})'(X'X)(\beta-\hat{\beta})$$

add and subtract $X\hat{\beta}$ does not depend on β

invertible

where $\hat{\beta}=(X'X)^{-1}X'y$. It can be seen that $\beta=\hat{\beta}$

maximizes $LL(\beta,\sigma^2)$ because it makes $(y - X\beta)'(y - X\beta)$ smallest

$$LL(\beta,\sigma^2)=-\frac{n}{2}\log(2\pi)-\frac{n}{2}\log(\sigma^2)-\frac{1}{2\sigma^2}\left[(y-X\hat{\beta})'(y-X\hat{\beta})+(\beta-\hat{\beta})'(X'X)(\beta-\hat{\beta})\right]$$

MLEs for Simple Linear Regression

More generally, we can have a multiple regression model

$$\underset{n\times 1}{y} = X\beta + \varepsilon \quad \text{where } \underset{n\times 1}{\varepsilon} \sim N(0,\sigma^2 I_n) \quad \text{and}$$

measured data
design matrix
regression coefficients
measurement error

$$\underset{n\times 1}{y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}}, \quad \underset{n\times(q+1)}{X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1q} \\ 1 & x_{21} & & x_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nq} \end{pmatrix}}, \quad \underset{(q+1)\times 1}{\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_q \end{pmatrix}}, \quad \underset{n\times 1}{\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}}.$$

MLEs for Simple Linear Regression

The MLEs are the same,

$$\underset{(q+1)\times 1}{\hat{\beta}} = (X'X)^{-1}X'y \text{ and } \underset{1\times 1}{\hat{\sigma}^2} = \frac{1}{n}(y - X\hat{\beta})'(y - X\hat{\beta}).$$ ⟵ Note denominator of n not n-1.

In addition,

$$\underset{(q+1)\times 1}{\hat{\beta}} \mid \beta, \sigma^2, X \sim N\left(\underset{(q+1)\times 1}{\beta}, \sigma^2 \underset{(q+1)\times(q+1)}{(X'X)^{-1}}\right) \text{ and } n\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-q-1) \ .$$

$$\underbrace{(y - X\beta)'(y - X\beta)}_{\sigma^2\chi^2(n)} = \underbrace{(y - X\hat{\beta})'(y - X\hat{\beta})}_{\sigma^2\chi^2(n-q-1)} + \underbrace{(\beta - \hat{\beta})'(X'X)(\beta - \hat{\beta})}_{\sigma^2\chi^2(q+1)}$$

could + & - $X\hat{\beta}$

This means we should use a denominator of n-q-1 for unbiased estimator of σ^2.

independent

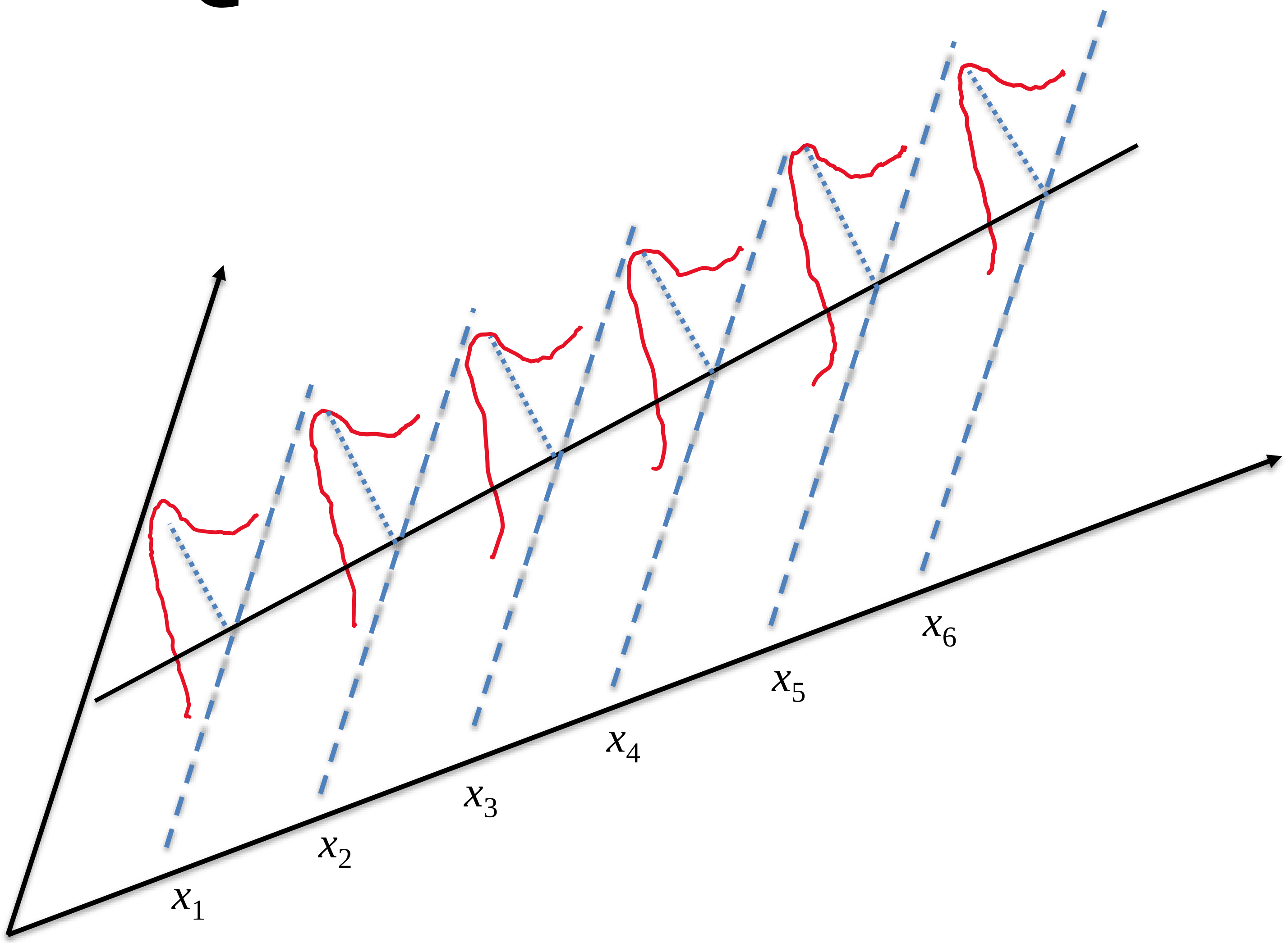
Note:

$$E(\hat{\beta} \mid \beta, \sigma^2, X) = \beta$$

$$\text{cov}(\hat{\beta} \mid \beta, \sigma^2, X) = \sigma^2 (X'X)^{-1}$$

Discussion

Questions?



Homework 7

1. Assume that you have a random sample $x_1,\ldots,x_n$ from each of the below PDFs. Derive with pencil and paper the MLEs of each of the below PDFs then generate n=10^4 observations from each PDF and compute your MLEs. Pick your own true parameter values and compare.

a) $f(x|\theta)=\theta^x(1-\theta)^{1-x} \quad x=0,1 \quad 0<\theta<1$

b) $f(x|\theta)=\dfrac{\theta^x e^{-\theta}}{x!} \quad x=0,1,2,.... \quad 0\le\theta<\infty \qquad f(0|\theta=1)\equiv 1$

c) $f(x|\theta)=\theta x^{\theta-1} \quad 0<x<1 \quad 0<\theta<\infty$

d) $f(x|\theta)=\dfrac{1}{\theta}e^{-x/\theta} \quad 0<x<\infty \quad 0<\theta<\infty$

Homework 7

2**.With pencil and paper show that

$$\sum_{i=1}^{n}(x_i-\mu)'\Sigma^{-1}(x_i-\mu)=tr\Sigma^{-1}(X-1_n\mu')'(X-1_n\mu')$$

3*.With pencil and paper show that

$$(y-X\beta)'(y-X\beta)=(y-X\hat{\beta})'(y-X\hat{\beta})+(\beta-\hat{\beta})'(X'X)(\beta-\hat{\beta})$$

** For students that have had 6010 and 6020.

* For students in MSSC 5790.

Homework 7

4***.With pencil and paper complete the square to show that

$$(y - X\beta)'\underset{n\times n}{\Phi}(y - X\beta) = (\beta - \hat{\beta})'(X'\Phi X)(\beta - \hat{\beta}) + Stuff$$

MLE

$$\hat{\beta} = (X'\Phi X)^{-1} X'\Phi y$$

weighted least squares

which arises from

$$f(y_1,\ldots,y_n \mid a,b,\sigma^2,\underset{n\times n}{\Phi}) = (2\pi\sigma^2)^{-\frac{n}{2}} \mid \underset{n\times n}{\Phi} \mid^{-1/2} \exp\left[-\frac{1}{2\sigma^2}(y - X\beta)'\underset{n\times n}{\Phi}^{-1}(y - X\beta)\right]$$

*** For students that want to show off.