# Method of Maximum Likelihood

Recall that the likelihood

$$L(\theta_1, \dots, \theta_p; y_1, \dots, y_n) = \prod_{i=1}^{n} f_{Y_i}(y_i)$$ for $Y_i$ independent

where $f_{Y_i}$ is the p.d.f. of $Y_i$ for continuous $Y_i$

$f_{Y_i}$ is the p.m.f of $Y_i$ for discrete $Y_i$

The maximum likelihood estimate $\hat{\theta}_1, \dots, \hat{\theta}_p$ are $\theta_1, \dots, \theta_p$ which maximize the likelihood.

The maximum likelihood estimate of $\phi(\theta_1, \dots, \theta_p)$ is $\phi(\hat{\theta}_1, \dots, \hat{\theta}_p)$.

Maximum Likelihood Estimators ~~or~~ (MLE)

have good properties.

We say ~~an e~~ a sequence of estimators is asymptotically unbiased if

$$\lim_{n\to\infty} \text{bias}(T(\underline{Y})) = 0.$$

The efficiency of ~~an asy~~ an asymptotically unbiased sequence of estimators is

$$\lim_{n\to\infty} \text{eff}(T_n) := \lim_{n\to\infty} \frac{CRLB(\phi)}{Var(T(Y))}$$

we know

$$0 \le \text{eff}(T_n) = \frac{CRLB(\phi)}{Var(T_n(\underline{Y}))} \le 1$$

Theorem

Under general conditions, MLEs are asymptotically unbiased, normal, and efficient, where efficient means $\lim_{n\to\infty} \text{eff}(T_n) = 1$.

Because $\ln$ is an increasing function finding the $\hat{\theta}_1, \dots, \hat{\theta}_p$ which maximize

$$L(\theta_1, \dots, \theta_p; y_1, \dots, y_n)$$

is equivalent to finding $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_p$ which maximize

$$\ln(L(\theta_1, \dots, \theta_p; y_1, \dots, y_n)) \quad (*)$$

We will write $\ell(\underline{\theta}; \underline{y})$ for $(*)$

To find the ~~max~~ $\hat{\theta}_1, \dots, \hat{\theta}_p$ we usually solve the equations

$$\frac{\partial \ell}{\partial \theta_i} = 0 \quad i = 1, \dots, p.$$

and check that the solutions are maximize $\ell(\underline{\theta}; \underline{y})$ by using second derivatives. For example, if $p = 1$, check whether $\frac{\partial^2 \ell}{\partial \theta^2} < 0$

Example

Let $Y_i$ be iid Poisson($\lambda$)

$$L(\lambda; \underline{y}) = \prod_{i=1}^{n} \frac{\lambda^{y_i} e^{-\lambda}}{y_i!}$$

$$= \frac{\lambda^{\sum_{i=1}^{n} y_i} e^{-n\lambda}}{\prod_{i=1}^{n} y_i!}$$

$$\ell(\lambda; \underline{y}) = \left(\sum_{i=1}^{n} y_i\right) \ln\lambda - n\lambda - \ln\left(\prod_{i=1}^{n} y_i!\right)$$

$$\frac{\partial \ell}{\partial \lambda} = \frac{\sum_{i=1}^{n} y_i}{\lambda} - n = 0$$

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} y_i}{n} = \bar{y}$$

$$\frac{\partial^2 \ell}{\partial \lambda^2} = -\frac{\sum_{i=1}^{n} y_i}{\lambda^2} < 0$$

so $\hat{\lambda} = \bar{y}$ maximizes the likelihood.

Example (Binomial)

A random sample from Binomial $(n, p)$ distribution with $n, p$ both unknown is

4, 3, 7, 4, 1, 4, 5, 4

$\mu_1 = np$

$\mu_2 = E(Y^2) = E(Y^2) - (E(Y))^2 + (E(Y))^2$

$= Var(Y) + (E(Y))^2$

$= np(1-p) + (np)^2$

Set $np = \frac{1}{8} \sum_{i=1}^{8} y_i = 3.875$

$np(1-p) + (np)^2 = \frac{1}{8} \sum_{i=1}^{8} y_i^2 = 17.85$

$3.875(1-p) + (3.875)^2 = 17.85$

$\tilde{p} = 0.26219$

$\tilde{n}\tilde{p} = 3.875$

$\tilde{n} = \frac{3.875}{0.26219} = 14.78$

take $n = 15$

## Example

Let $Y_i$ be iid. $N(\mu, \sigma^2)$.

$$L(\mu, \sigma^2; \underline{y}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i-\mu)^2}{2\sigma^2}\right)$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right)$$

$$l(\mu, \sigma^2; \underline{y}) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2$$

$$\frac{\partial l}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i-\mu) = 0$$

$$\sum_{i=1}^{n} y_i - n\mu = 0$$

$$\Rightarrow \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y}$$

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(y_i-\mu)^2 = 0$$

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

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2$$

same as method of moments

We won't show $\hat{\mu}, \hat{\sigma}^2$ are maximums.

# Example

$Y_i$ are iid ~~Uniform~~ Uniform $[0, \theta]$

method of moments $\qquad \frac{1}{\theta - 0}$

If $Y \sim \text{Uniform}[a, b]$

$$E(Y) = \frac{a+b}{2}$$

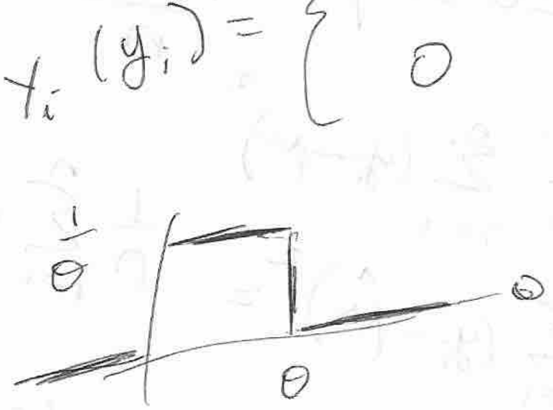
$$\mu_1 = Y_i \sim \text{Uniform}[0, \theta] = \frac{0+\theta}{2} = \frac{\theta}{2}$$

set $\frac{\theta}{2} = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y}$

$$\tilde{\theta} = 2\bar{y}$$

maximum likelihood

$$f_{Y_i}(y_i) = \begin{cases} \frac{1}{\theta} & \text{if } 0 \le y_i \le \theta \\ 0 & \text{else} \end{cases}$$



$$L(\theta; \underline{y}) = \prod_{i=1}^{n} f_{Y_i}(y_i)$$

$$= \begin{cases} \left(\frac{1}{\theta}\right)^n \text{ if } 0 \le y_i \le \theta \text{ for all } i \\ 0 \text{ else} \end{cases}$$

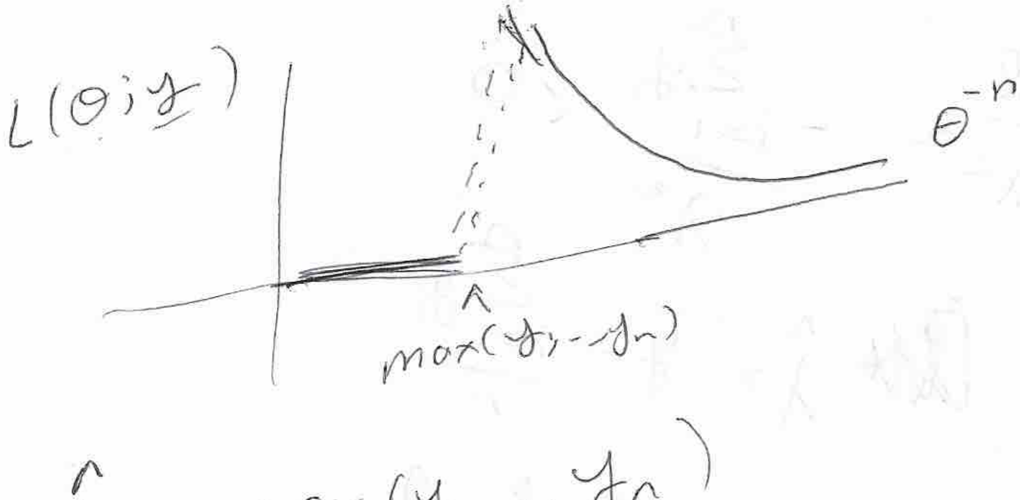
$=$ [illegible]

Try to take $\frac{d}{d\theta}$

$$\frac{d}{d\theta}\left(\frac{1}{\theta}\right)^n = -\frac{1}{\theta^{n+1}} = 0$$

$\theta = \infty$? no solution

$$L(\theta; \underline{y}) = \begin{cases} \left(\frac{1}{\theta}\right)^n \text{ if } 0 \le \max(y_1, \ldots, y_n) \le \theta \\ 0 \text{ else} \end{cases}$$




$$\hat{\theta} = \max(y_1, \ldots, y_n)$$

the method of maximum likelihood can be applied when we have censored data, i.e. partial information.

We may know the data lies in a certain range, but not their actual value's

We may be given n specific values and know that m other values ~~are~~ are bigger than z.

$$L(\underline{\theta}; \underline{y}) = \prod_{i=1}^{n} f_i(y_i; \underline{\theta}) \times P(Y > z)^{m}$$

where $f_i$ are p.d.f.s or p.m.f.s

or if we know m other values lie between x and z

$$L(\underline{\theta}; \underline{x}) = \prod_{i=1}^{n} f_i(y_i; \underline{\theta}) \; P(x \leq y \leq z)^{m}$$

~~nor~~ claim

Example

Claims in 1000's of pounds ~~on~~ per on a particular policy have p.d.f.

$$f(y) = 2cye^{-cy^2}, \quad y > 0.$$

Seven of the last claims are 1.85, 2.37, 2.71, 3.22, 3.26, 1.05, 3.38, while ~~the~~ three other claims are known to be bigger than 6000.

$$L(c; \underline{y}) = \prod_{i=1}^{7} f_{Y_i}(y_i) \times P(Y > 6)^3$$

$$P(Y > 6) = \int_6^\infty 2cye^{-cy^2}\,dy$$

$$= -e^{-cy^2}\Big|_{y=6}^{y=\infty}$$

$$= e^{-36c}$$

Note that $\sum_{i=1}^{7} y_i^2 = 49.91$

$$\cancel{L(c;\underline{y}) = \prod_{i=1}^{7} 2cy_i e}$$

$$L(c;\underline{y}) = \prod_{i=1}^{7} 2cy_i e^{-cy_i^2} \times e^{-108c}$$

$$= 2^7\left(\prod_{i=1}^{7} y_i\right) c^7 e^{-c\sum_{i=1}^{7} y_i^2} e^{-108c}$$

$$= 2^7\left(\prod_{i=1}^{7} y_i\right) c^7 e^{-49.91c} e^{-108c}$$

$$= 2^7\left(\prod_{i=1}^{7} y_i\right) c^7 e^{-157.91c}$$

$$l(c;\underline{y}) = \ln\left(2^7\prod_{i=1}^{7} y_i\right) + 7\ln c - 157.91c$$

$$\frac{\partial l}{\partial c} = \frac{7}{c} - 157.91$$

$$\Rightarrow \hat{c} = \frac{7}{157.91} = 0.0443$$

$$\frac{\partial^2 l}{\partial c^2} = -\frac{7}{c^2} < 0$$

so $\hat{c}$ maximizes $l(c;\underline{y})$

There may be other situations for which we have no information about what number of claims take some particular value. This is an example of <u>truncated data</u>.

The technique we use is to condition the distribution of $Y$ on the event that only the values which we can see occur. ~~Theo~~ Then we proceed as before.

<u>Example</u>

The number of claims $N$ per policy in a year on a pet insurance policy has distribution

$$P(N=0)=5\theta, \quad P(N=1)=3\theta, \quad P(N=2)=\theta,$$
$$P(N\geq 3)=1-9\theta$$

For a particular year, there wer
60 policies with 1 claim
24 policies with 2 claims
16 policies with at least 3 claims.

We don't know how many had 0 claims.

$$P(N=1 \mid N>0) = \frac{3\theta}{3\theta+\theta+1-9\theta} = \frac{3\theta}{1-5\theta}$$

$$P(N=2 \mid N>0) = \frac{\theta}{1-5\theta}$$

$$P(N\geq 3 \mid N>0) = \frac{1-9\theta}{1-5\theta}$$

The likelihood is

$$L(\theta;\underline{y}) = \left(\frac{3\theta}{1-5\theta}\right)^{60}\left(\frac{\theta}{1-5\theta}\right)^{24}\left(\frac{1-9\theta}{1-5\theta}\right)^{16}$$

$$\ell(\theta;\underline{y}) = 60\ln\theta + 24\ln\theta + 16\ln(1-9\theta) - 100\ln(1-5\theta) + \text{constant}$$

$$= 84\ln\theta + 16\ln(1-9\theta) - 100\ln(1-5\theta)$$

>l

$$\frac{\partial l}{\partial \theta} = \frac{84}{\theta} - \frac{144}{1-9\theta} + \frac{500}{1-5\theta}$$

$$= \frac{84(1-9\theta)(1-5\theta) - 144\theta(1-5\theta) + 500\theta(1-9\theta)}{\theta(1-9\theta)(1-5\theta)} = 0$$

$$= 84 - 820\theta = 0$$

$$\hat{\theta} = \frac{84}{820} = \cancel{0.120}\ 0.102$$

You can check $\frac{d^2 l}{d\theta^2} < 0$ at $\hat{\theta}$