QUANTITATIVE ANALYSIS

Greg Wdowiak

MAXIMUM LIKELIHOOD ESTIMATION of MEAN and VARIANCE as an example of MULTIVARIATE OPTIMIZATION

→ Maximum Likelihood Estimation (MLE) is a statistical technique for estimating the parameters of a probability distribution that generated an observed set of data.

→ In other words: "We see this data here. I'm willing to assume that it comes from a NORMAL DISTRIBUTION. If that's the case, then what are the MEAN and VARIANCE of that distribution?"

→ The ESTIMATION QUESTION is:

$$\boxed{\text{What values of } \mu \text{ and } \gamma \text{ are MOST LIKELY to have generated this data?}}$$

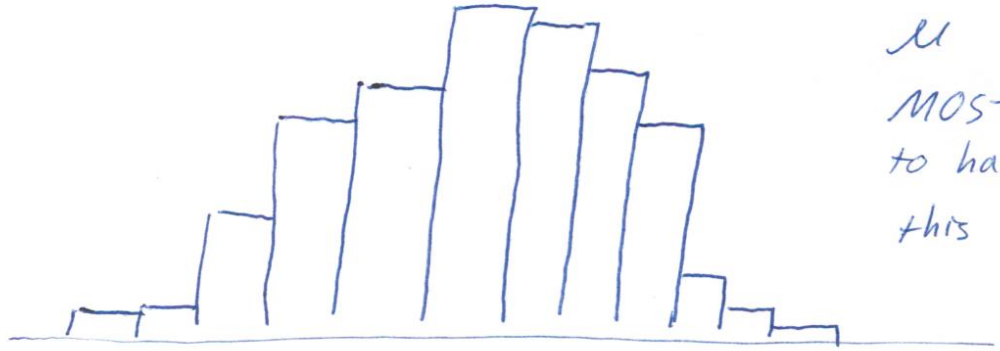
Optimization in two variables: μ and γ

→ The simplest example of MLE is to estimate the mean μ and variance σ of a normal distribution

→ This example is easily extensible to higher level analyses. For example, multivariate regression models, probability models and models with distributions other than normal (e.g, lognormal, logistic)

→ And because this example only has two variables it serves as a good example for a course in quantitative methods.

→ If your data is distributed normally then a histogram of the data may look like this:



"What values of μ and σ are MOST LIKELY to have generated this data?"

→ If we assume that the true, underlying distribution is NORMAL, then we want to <u>maximize</u> the following <u>likelihood function</u> with respect to μ and γ:

$$\max_{\mu,\gamma} \mathcal{L} = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\gamma}} \cdot e^{-\frac{1}{2}\frac{(x_i-\mu)^2}{\gamma}}$$

which is equivalent to maximizing the log of that same function:

$$\max_{\mu,\gamma} \ln \mathcal{L} = \sum_{i=1}^{N} \left\{ \ln\left[(2\pi\gamma)^{-\frac{1}{2}}\right] - \frac{(x_i-\mu)^2}{2\gamma} \right\}$$

$$= -\frac{N}{2}\ln(2\pi\gamma) - \frac{1}{2\gamma}\cdot\sum_{i=1}^{N}(x_i-\mu)^2$$

→ The FIRST-ORDER CONDITIONS are:

$$\frac{\partial \ln \mathcal{L}}{\partial \mu} = \frac{1}{\gamma}\sum_{i=1}^{N}(x_i-\mu) = 0 \quad \text{which implies:} \quad \mu = \frac{1}{N}\sum x_i$$

$$\frac{\partial \ln \mathcal{L}}{\partial \gamma} = \frac{-N}{2\gamma} + \frac{1}{2\gamma^2}\sum_{i=1}^{N}(x_i-\mu)^2 = 0 \quad \text{which implies:}$$

$$\gamma = \frac{1}{N}\sum(x_i-\mu)^2$$

→ The SECOND-ORDER CONDITIONS are

$$\frac{\partial^2 \ln \mathcal{L}}{\partial \mu^2} = \frac{-N}{\gamma} < 0$$

"MEAN's" OWN SECOND PARTIAL IS NEGATIVE

$$\frac{\partial^2 \ln \mathcal{L}}{\partial \gamma^2} = \frac{N}{2\gamma^2} - \frac{1}{\gamma^3}\sum(x_i-\mu)^2$$

"VARIANCE's" OWN SECOND PARTIAL IS NEGATIVE

$$= \frac{N}{2\gamma^2} - \frac{N}{\gamma^2} = \frac{N}{\gamma^2}\left(\frac{1}{2}-1\right) = \frac{-N}{2\gamma^2} < 0$$

BY the 1st Order Condition

AND THE PRODUCT OF THE OWN SECOND PARTIALS IS GREATER THAN THE PRODUCT OF THE CROSS PARTIALS (i.e. the square of the cross partial by Young's Theorem)

$$\frac{\partial^2 \ln \mathcal{L}}{\partial \mu \partial \gamma} = \frac{-1}{\gamma^2}\sum(x_i-\mu)$$

$$= \frac{-1}{\gamma^2}\left[\left(\sum x_i\right) - N\mu\right] = 0$$

← BY THE FIRST ORDER CONDITION

$$H = \begin{bmatrix} \frac{\partial^2 \ln \mathcal{L}}{\partial \mu^2} & \frac{\partial^2 \ln \mathcal{L}}{\partial \mu \partial \gamma} \\ \frac{\partial^2 \ln \mathcal{L}}{\partial \mu \partial \gamma} & \frac{\partial^2 \ln \mathcal{L}}{\partial \gamma^2} \end{bmatrix} = \begin{bmatrix} -\frac{N}{\gamma} & 0 \\ 0 & \frac{-N}{2\gamma^2} \end{bmatrix}$$

☆ HESSIAN MATRIX ☆

→ The DETERMINANT of THE HESSIAN $|H|$ is the product of the own second partials minus the square of the cross partial

$$|H| = \left(\frac{-N}{\gamma}\right)\left(\frac{-N}{2\gamma^2}\right) - 0 \cdot 0$$

$$= \frac{N^2}{2\gamma^3} > 0$$

SATISFACTION OF THIS SECOND ORDER CONDITION ENSURES THAT WE ARE AT A MAXIMUM (as opposed to a SADDLE ~~POINT~~ POINT)

→ The <u>INFORMATION</u> <u>MATRIX</u> is the negative of the INVERSE of the Hessian Matrix

$$I = -1 \cdot H^{-1} = \begin{bmatrix} \gamma/N & 0 \\ 0 & \frac{2\gamma^2}{N} \end{bmatrix}$$

The "information" contained ~~in~~ in this matrix are our measures of the precision with which we estimated the mean + variance.

Std. error of our estimate of the mean $\sqrt{\frac{\gamma}{N}}$

Variance of our estimate of the variance $\frac{2\gamma^2}{N}$

→ The true mean of x μ is unknown to us. What MLE provides is an ESTIMATE of μ. So it would be more accurate to write:

$$\hat{\mu} = \frac{1}{N} \sum x_i \qquad \underline{\text{MEAN of } X}$$

→ Similarly, we only obtain an estimate of the variance of x so it would be more accurate:

$$\hat{\gamma} = \frac{1}{N} \sum (x_i - \hat{\mu})^2 \qquad \underline{\text{VARIANCE of } X}$$

→ Our estimate of the standard error of X is

$$\hat{\sigma} = \sqrt{\hat{\gamma}} = \sqrt{\frac{1}{N} \sum (x_i - \hat{\mu})^2}$$

→ <u>BUT</u> <u>REMEMBER</u>: We measure the mean of x and the variance of x with error.

→ The values of the own second partials (or more precisely the negative of the inverse of those values) tell us how precisely we have measured the mean of x and the variance of x.

→ From the own second partials we get the STANDARD ERROR of the ESTIMATE of the MEAN (as opposed to the STANDARD ERROR of X)

std error of estimate of mean $\sqrt{\dfrac{\hat{\sigma}}{N}}$ ← increasing fn of estimated variance ← decreasing fn of number of observations

→ We also obtain the variance of our estimate of the variance:

$$\frac{2\sigma^2}{N}$$

← INCREASING FN of ESTIMATED VARIANCE

← DECREASING FN of NUMBER of OBSERVATIONS

→ SO the OWN SECOND PARTIALS provide measures of HOW PRECISELY we have estimated the mean and variance.

→ The remaining question is:

"Why do the measures of precision come from the own second partials?"

→ The answer can be seen by comparing the following functions

$$f(x) = -x^2 \quad \text{and} \quad g(x) = -2x^2$$

WIDER and LESS PRECISE

$f(x) = -x^2$

$f'(x) = -2x$

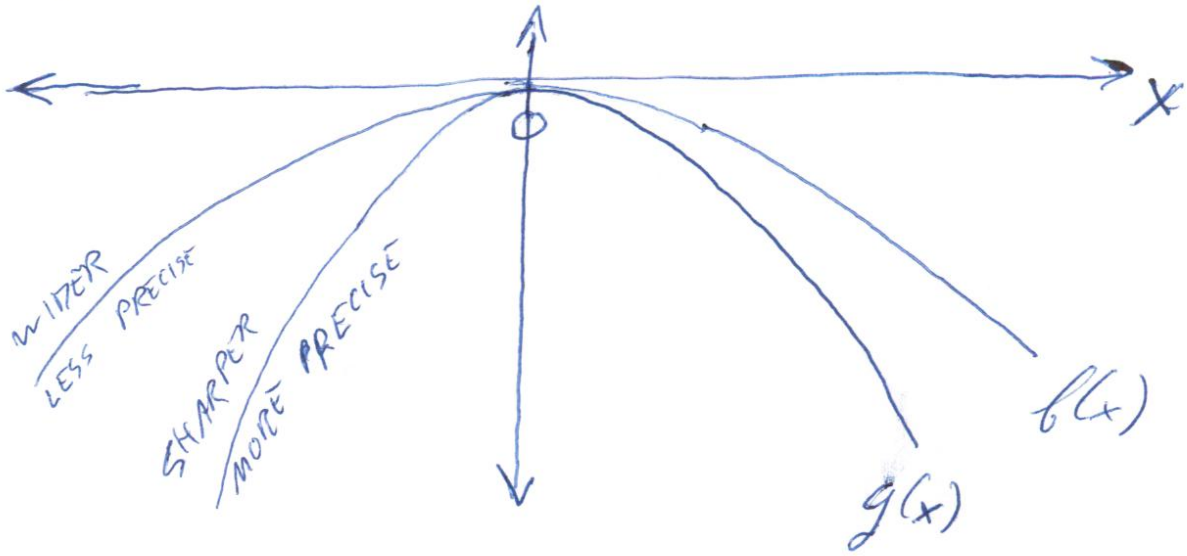
$f''(x) = -2$

$g(x) = -2x^2$

$g'(x) = -4x$

$g''(x) = -4$

SHARPER therefore MORE PRECISE



→ Both $f(x)$ and $g(x)$ are maximized at $x=0$, but $g(x)$ comes to a sharper peak, thus representing a more precise estimate

→ Comparing the negative of the inverses allows us to compare what would be the measure of precision

Pretend that this is the variance of the estimate

$$\frac{-1}{f''(x=0)} > \frac{-1}{g''(x=0)}$$

$$\frac{1}{2} > \frac{1}{4}$$

The $g(x)$ function would yield lower variance of the estimate.

It is a more precise measure.