Maximum Likelihood Estimation

Motivating Example

Number of years smoking and number of cigarettes smoked

Use maximum likelihood to estimate the mean number of cigarettes smoked

Cigs
9
12
11
3
10
5
7
11
12
11
12
11
10
19
15
8
13
8
10
10

Maximum Likelihood (ML) Estimation

ML identifies the population parameter values that are most consistent with the raw data

A likelihood (or log likelihood) function quantifies the fit of the data to the parameters

ML requires a population distribution for the outcome variable (e.g., multivariate normal)

Probability Density Function

A density function gives the shape of the normal curve

$$L_i = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-.5\frac{(Y_i-\mu)^2}{\sigma^2}\right)$$

L_i (the likelihood) gives the relative probability that Y_i comes from a normal distribution with a particular mean and variance

Simplifying the Likelihood

The likelihood is largely driven by a squared *z*-score to the right of the exponent

$$L_i = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-.5\frac{(Y_i - \mu)^2}{\sigma^2}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{z^2}{2}\right)$$

Small standardized distance = high likelihood (probability) = good fit between the score and the mean

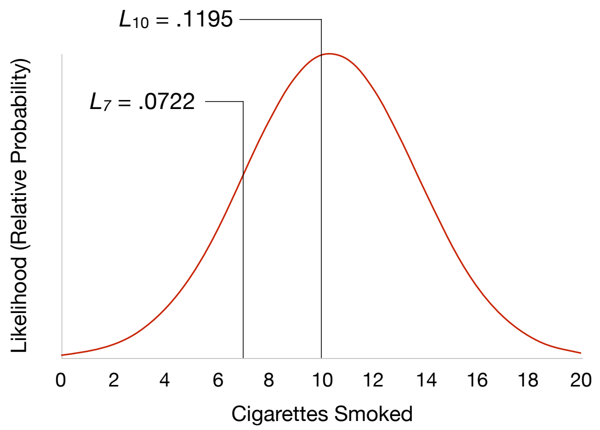
Example

Compute the likelihood of $Y_i = 10$ and $Y_i = 7$ from a normally distributed population with a mean of 10.35 and a standard deviation of 3.32

Substituting the parameters and scores into the density function gives $L_{10} = .1195$ and $L_7 = .0722$

The likelihood values quantify the relative probability of obtaining each score from this population

Graphic



Individual Likelihoods

Smaller deviations between a score and the mean produce higher likelihood values

Higher likelihood values reflect a better fit to the parameter values in question

Cigs	L_i
19	0.0040
15	0.0451
13	0.0874
12	0.1062
12	0.1062
12	0.1062
11	0.1179
11	0.1179
11	0.1179
11	0.1179
10	0.1195
10	0.1195
10	0.1195
10	0.1195
9	0.1106
8	0.0935
8	0.0935
7	0.0722
5	0.0328
3	0.0104

Joint Probability

From probability theory, the joint probability for a set of events is the product of individual probabilities

e.g., The probability of jointly observing two heads is (.50)(.50) = .25

A likelihood is not a probability, but the same rules apply

Sample Likelihood

The sample likelihood is the product of the individual likelihoods

$$L = \prod_{N} \left[\frac{1}{\sqrt{2\pi\sigma^2}} exp \left(-.5 \frac{(Y_i - \mu)^2}{\sigma^2} \right) \right]$$

$\prod_{N}$ is the multiplication operator over all cases

Example

The sample likelihood is the product of 20 individual likelihood contributions

L = (.1106)(.1062)…(.1195)
= 0.0000000000000000000000178

The sample likelihood quantifies the relative probability of obtaining these 20 scores from a normal population with this particular mean and standard deviation

Logarithms

Likelihoods are computationally difficult and introduce precision problems due to rounding error

One rule of logarithms is log[(a)(b)] = log(a) + log(b)

Using logarithms converts a multiplication problem to an addition problem (simpler math)

Log Likelihood

The log likelihood the natural logarithm of a likelihood

$$lnL = \sum_N ln \left[\frac{1}{\sqrt{2\pi\sigma^2}} exp \left(-.5 \frac{(Y_i - \mu)^2}{\sigma^2} \right) \right]$$

Log likelihood values also quantify relative probability, but they do so on a different metric

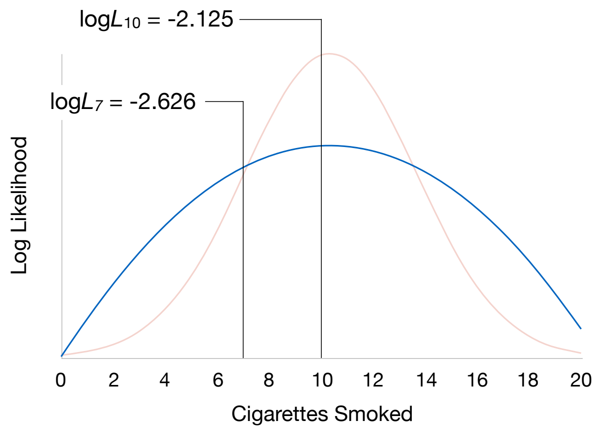
Individual Log Likelihoods

Smaller deviations between a score and the mean produce higher (less negative) log likelihood values

Higher log likelihood values reflect a better fit to the parameter values in question

Cigs	L_i	logL_i
19	0.0040	-5.5117
15	0.0451	-3.0995
13	0.0874	-2.4375
12	0.1062	-2.2426
12	0.1062	-2.2426
12	0.1062	-2.2426
11	0.1179	-2.1383
11	0.1179	-2.1383
11	0.1179	-2.1383
11	0.1179	-2.1383
10	0.1195	-2.1247
10	0.1195	-2.1247
10	0.1195	-2.1247
10	0.1195	-2.1247
9	0.1106	-2.2018
8	0.0935	-2.3695
8	0.0935	-2.3695
7	0.0722	-2.6280
5	0.0328	-3.4169
3	0.0104	-4.5686

Graphic



Log Likelihood

The sample likelihood is the sum of the individual log likelihood contributions

$$lnL = \sum_N ln \left[\frac{1}{\sqrt{2\pi\sigma^2}} exp \left(-.5 \frac{(Y_i - \mu)^2}{\sigma^2} \right) \right]$$

The log likelihood represents the joint probability of the sample data, given the parameters

Example

The sample log likelihood is the sum of 20 individual log likelihood contributions

$$\log L = (-5.117)(-3.0995)\ldots(-4.5686) = -52.3827$$

The sample likelihood quantifies the relative probability of obtaining these 20 scores from a normal population with this particular mean and standard deviation

Interpreting the Log Likelihood

The log likelihood quantifies the fit between the sample data and the population parameters

The log likelihood depends on the sample size, number of variables, number of parameters in the model, missing data, etc.

No absolute criterion for a good or a bad value

Estimation Strategy

The sample log likelihood provides a mechanism for identifying unknown parameter values

Compute the log likelihood for different values of the population mean and find the value that produces the highest log likelihood (best fit to the data)

The variance terms in the log likelihood can be fixed at any value because the mean and variance are uncorrelated

Example: Mean = 8

$\log L = (-7.605) + (-4.341) + \ldots + (-3.253) = -57.3906$

Does increasing the mean from 8 to 9 increase (improve) or decrease (make worse) the log likelihood?

Cigs	**logL_i**
19	-7.605
15	-4.341
13	-3.253
12	-2.845
12	-2.845
12	-2.845
11	-2.527
11	-2.527
11	-2.527
11	-2.527
10	-2.300
10	-2.300
10	-2.300
10	-2.300
9	-2.164
8	-2.119
8	-2.119
7	-2.164
5	-2.527
3	-3.253

Example: Mean = 9

log*L* = (-6.653) + (-3.751) + … + (-3.751) = -54.0354

The log likelihood improved

Does increasing the mean from 9 to 10 increase (improve) or decrease (make worse) the log likelihood?

Cigs	logL_i
19	-6.653
15	-3.751
13	-2.845
12	-2.527
12	-2.527
12	-2.527
11	-2.300
11	-2.300
11	-2.300
11	-2.300
10	-2.164
10	-2.164
10	-2.164
10	-2.164
9	-2.119
8	-2.164
8	-2.164
7	-2.300
5	-2.845
3	-3.751

Example: Mean = 10

log*L* = (-5.792) + (-3.253) + … + (-4.341) = -52.4938

The log likelihood improved

Does increasing the mean from 10 to 11 increase (improve) or decrease (make worse) the log likelihood?

Cigs	logL_i
19	-5.792
15	-3.253
13	-2.527
12	-2.300
12	-2.300
12	-2.300
11	-2.164
11	-2.164
11	-2.164
11	-2.164
10	-2.119
10	-2.119
10	-2.119
10	-2.119
9	-2.164
8	-2.300
8	-2.300
7	-2.527
5	-3.253
3	-4.341

Example: Mean = 11

log*L* = (-5.021) + (-2.845) + … + (-5.021) = -52.7658

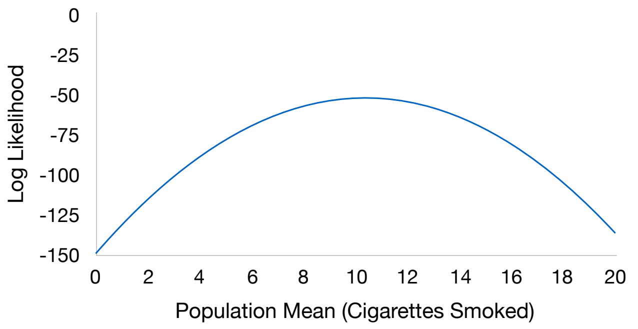
The log likelihood got worse

No need to audition more values, the population mean falls between 9 and 11

Cigs	logL_i
19	-5.021
15	-2.845
13	-2.300
12	-2.164
12	-2.164
12	-2.164
11	-2.119
11	-2.119
11	-2.119
11	-2.119
10	-2.164
10	-2.164
10	-2.164
10	-2.164
9	-2.300
8	-2.527
8	-2.527
7	-2.845
5	-3.751
3	-5.021

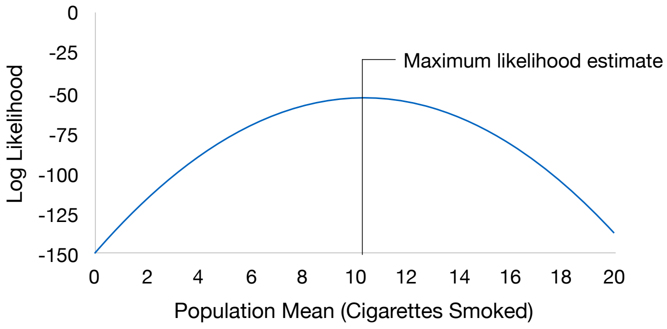
Log Likelihood Function

The sample log likelihood (model-data fit) expressed as a function of the population mean

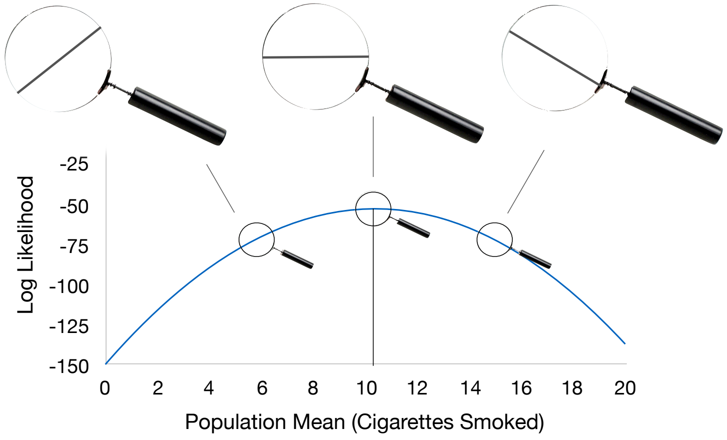


Identifying the Maximum Likelihood Estimate

The population mean most likely to have produced this sample of 20 scores is the value that maximizes the function

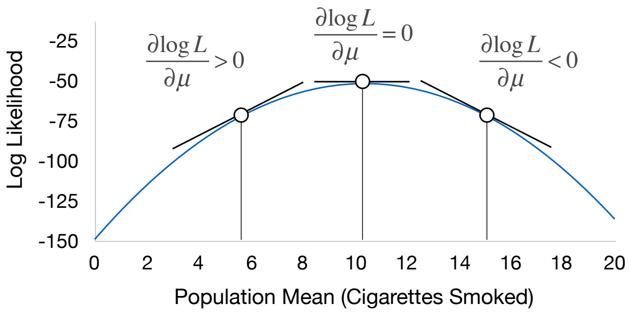


Slide of the Function



First Derivatives

The first (partial) derivative is the slope of the function at a particular population parameter on the horizontal axis



Estimation Via Derivatives

Apply differential calculus rules to the log likelihood function to obtain the first derivative of the function (i.e., the slope, given some value for the parameter)

Set the derivative (slope) formula equal to zero

Solve for the unknown parameter value

Maximum Likelihood Estimate of the Mean

Differentiating the log likelihood with respect to the mean gives its first (partial) derivative

$$\frac{\partial lnL}{\partial \mu} = \frac{-N\mu + \sum Y_i}{\sigma^2}$$

Setting the slope to zero and solving yields a closed-form ML solution for the mean

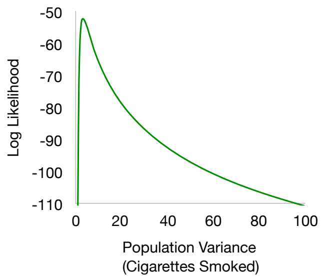
$$\frac{-N\mu + \sum Y_i}{\sigma^2} = 0$$

$$\mu = \frac{\sum Y_i}{N}$$

The Variance

We can use the maximum likelihood estimate of the mean to estimate the variance

We can write or graph a function that shows the change in the log likelihood as a function of the population variance



Maximum Likelihood Estimate of the Variance

Differentiating the log likelihood with respect to the variance gives its first derivative

$$\frac{\partial lnL}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{\sum (Y_i - \mu)^2}{2\left(\sigma^2\right)^2}$$

Setting the slope to zero and solving gives the biased expression that corresponds to the population variance

$$-\frac{N}{2\sigma^2} + \frac{\sum (Y_i - \mu)^2}{2\left(\sigma^2\right)^2} = 0$$

$$\sigma^2 = \frac{\sum (Y_i - \mu)^2}{N}$$

Curvature of the Function

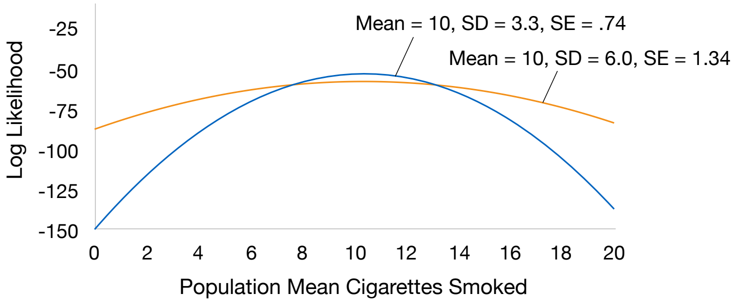
The curvature of the log likelihood function captures the precision (standard error) of the ML estimate

Peaked functions imply greater precision (and lower standard errors) because a given change in the parameter produces a larger change in the log likelihood

Second derivatives quantify curvature and are the building blocks for ML standard errors

Curvature and Precision

Two functions with the same maximum but different levels of precision (different standard errors)



Regression Example

Number of years smoking and number of cigarettes smoked

Use maximum likelihood to estimate the the regression of cigarettes smoked on years smoking

Years	Cigs
7	9
8	12
1	11
4	3
6	10
8	5
8	7
10	11
15	12
5	11
9	12
11	11
14	10
13	19
12	15
11	8
10	13
10	8
7	10
11	10

Log Likelihood

A predicted value and residual variance replace the mean and variance in the individual log likelihood expression

$$lnL = \sum_{N} ln \left[\frac{1}{\sqrt{2\pi\sigma_e^2}} exp \left(-.5 \frac{(Y_i - [\beta_0 + \beta_1 X_i])^2}{\sigma_e^2} \right) \right]$$

The log likelihood gives the relative probability of different *Y* values, the predictors are treated as known constants

Estimating Multiple Parameters

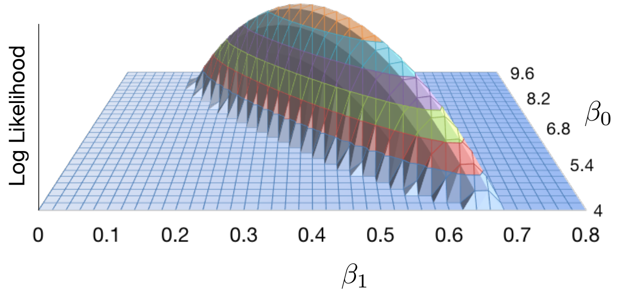
The population mean is expressed as a function of an intercept and slope, so estimation involves a search for two parameters

Estimation auditions different combinations of parameter values and computes the log likelihood for each

The goal is to identify the combination of parameter values that maximizes the log likelihood

Log Likelihood Surface

The log likelihood varies across combinations of intercept and slope values



Multivariate Normal Density Function

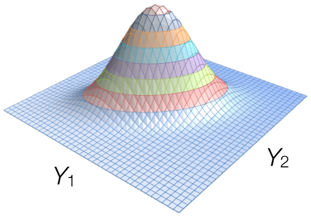
Density function for the multivariate normal distribution

$$lnL = \sum_N ln\left[\frac{1}{\sqrt{2\pi}^{k/2} \mid \boldsymbol{\Sigma} \mid^{.5}} exp\left(-\frac{\left(\mathbf{Y}_i - \boldsymbol{\mu}\right)' \boldsymbol{\Sigma}^{-1} \left(\mathbf{Y}_i - \boldsymbol{\mu}\right)}{2}\right)\right]$$

L_i fromgives the relative probability that the set of scores in **Y** came from a multivariate normal distribution with a particular mean vector and covariance matrix

Multivariate Normal Density Function

L_i gives the relative probability that the set of scores in **Y** came from a multivariate normal distribution with a particular mean vector and covariance matrix



$$L_i = \frac{1}{\sqrt{2\pi}^{k/2} \mid \boldsymbol{\Sigma} \mid^{.5}} exp\left(-\frac{\left(\mathbf{Y}_i - \boldsymbol{\mu}\right)' \boldsymbol{\Sigma}^{-1} \left(\mathbf{Y}_i - \boldsymbol{\mu}\right)}{2}\right)$$

Simplifying the Likelihood

The likelihood is still driven by a squared z score (called Mahalanobis distance) that captures the sum of standardized deviation scores

$$lnL = \sum_N ln\left[\frac{1}{\sqrt{2\pi}^{k/2} \mid \boldsymbol{\Sigma} \mid^{.5}} exp\left(-\frac{\left(\mathbf{Y}_i - \boldsymbol{\mu}\right)' \boldsymbol{\Sigma}^{-1} \left(\mathbf{Y}_i - \boldsymbol{\mu}\right)}{2}\right)\right]$$

$$= \sum_N ln\left[\frac{1}{\sqrt{2\pi}^{k/2} \mid \boldsymbol{\Sigma} \mid^{.5}} exp\left(-\frac{z_i^2}{2}\right)\right]$$