Maximum likelihood estimation

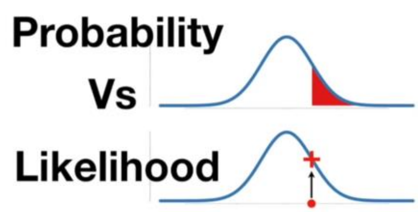
Definition and Purpose

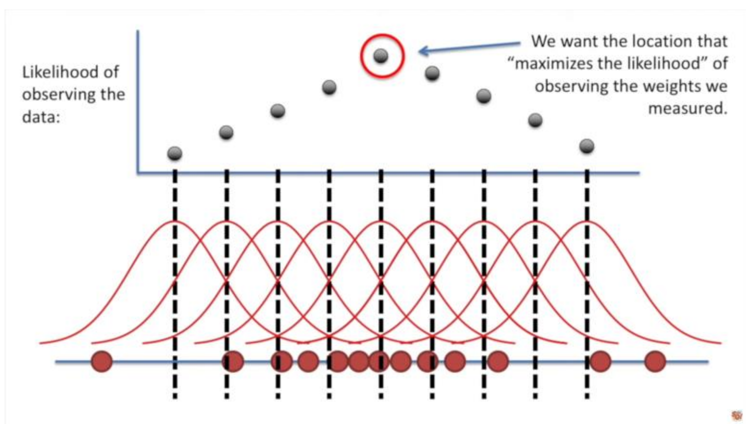
Maximum likelihood estimation (MLE) is a method of **estimating the parameters of an assumed probability distribution, given some observed data.** If the likelihood function is differentiable, then we could use the derivative test for determining maxima value.

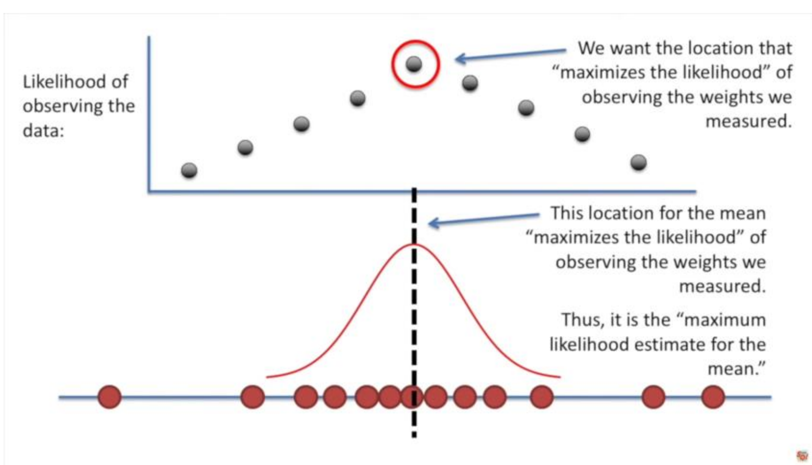
In mathematical notation, the goal of maximum likelihood estimation is to find the value of model parameters that maximize the likelihood function over the parameter space

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \widehat{L_n}(\theta; y)$$

Where $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ is the observed data, Θ is called the parameter space, L_n is the likelihood function.







Examples

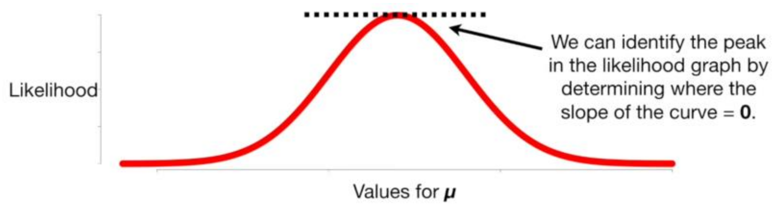
If we assume the population distribution follows $N(\mu = 28, \sigma = 2)$, while we observe $y = 32$, the likelihood of the data from our assumed distribution is

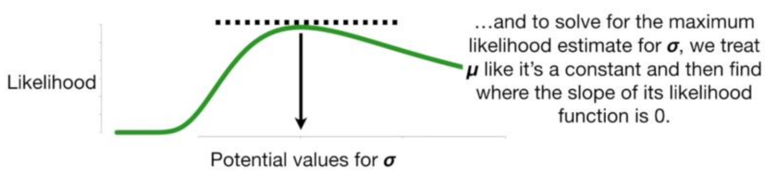
$$L(\mu = 28, \sigma = 2 \mid x = 32) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = 0.03$$

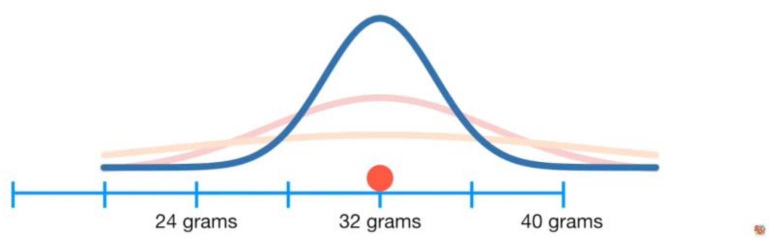
If we assume the population distribution as $N(\mu = 30, \sigma = 2)$, the likelihood function would be

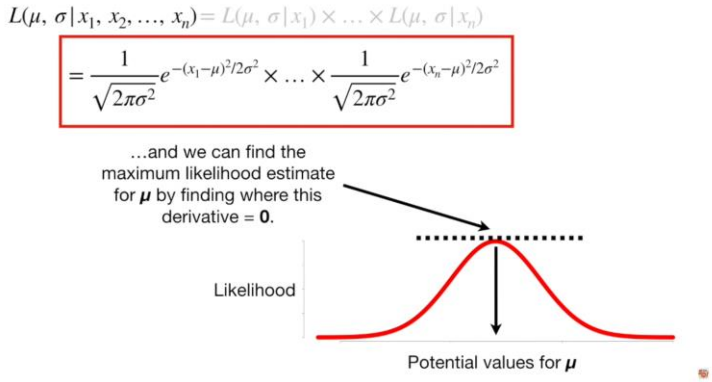
$$L(\mu = 30, \sigma = 2 | x = 32) = 0.12$$

We could treat the μ as the parameter and plot the parameter space









When we have multiple observations, the likelihood of those data is the join probability of the likelihood of all individuals, which is

$$
\begin{aligned}
L(\mu, \sigma | x_1, x_2, \ldots, x_n) &= L(\mu, \sigma | x_1) * \ldots * L(\mu, \sigma | x_n) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_1-\mu)^2}{2\sigma^2}} * \ldots * \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_n-\mu)^2}{2\sigma^2}}
\end{aligned}
$$

For simplify the calculation, we take the log transformation on each likelihood function (since it is monotonous, which would not change the position of peak), so the equation becomes

$$
\begin{aligned}
L(\mu, \sigma | x_1, x_2, \ldots, x_n) &= \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_1-\mu)^2}{2\sigma^2}} * \ldots * \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_n-\mu)^2}{2\sigma^2}}\right) \\
&= \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_1-\mu)^2}{2\sigma^2}}\right) + \cdots + \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_n-\mu)^2}{2\sigma^2}}\right) \\
&= -\frac{1}{2}(\ln(2\pi) + \ln\sigma^2) - \frac{(x_1-\mu)^2}{2\sigma^2} + \cdots \pm \frac{1}{2}(\ln(2\pi) + \ln\sigma^2) \\
&\quad - \frac{(x_n-\mu)^2}{2\sigma^2} \\
&= -\frac{n}{2}\ln(2\pi) - n\ln(\sigma) - \frac{(x_1-\mu)^2}{2\sigma^2} - \cdots - -\frac{(x_n-\mu)^2}{2\sigma^2}
\end{aligned}
$$

And then, take the partial derivative of μ, we get

$$\frac{\partial L}{\partial \mu} = 0 - 0 + \frac{(x_1-\mu)}{\sigma^2} + \cdots + \frac{(x_n-\mu)}{\sigma^2} = \frac{1}{\sigma^2}[(x_1 + \cdots + x_n) - n\mu]$$

Let $\frac{\partial L}{\partial \mu}$ equal to 0, we get $\mu = \frac{(x_1+\cdots+x_n)}{n}$

take the partial derivative of σ, we get

$$\frac{\partial L}{\partial \sigma} = 0 - \frac{n}{\sigma} + \frac{(x_1-\mu)^2}{\sigma^3} + \cdots + \frac{(x_n-\mu)^2}{\sigma^3} = \frac{-n}{\sigma} + \frac{1}{\sigma^3}[(x_1-\mu)^2 + \cdots + (x_n-\mu)^2]$$

Let $\frac{\partial L}{\partial \sigma}$ equal to 0, we get $\sigma = \sqrt{\frac{(x_1-\mu)^2+\cdots+(x_n-\mu)^2}{n}}$

…and the likelihood function and the
log of the likelihood function both peak
at the same values for $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$.