1. The likelihood function, $\boldsymbol{L(\theta) = \prod_{i=1}^{n} f(x_i | \theta)}$ is the product of the n pmf/pdf viewed as a function of a parameter θ.
2. Take the natural log of the likelihood to get the log-likelihood function $\boldsymbol{l(\hat{\theta}) = log(L(\theta))}$ and collect terms involving θ.
3. Find the value of θ for which log-likelihood is maximised. This is typically done by finding $\hat{\theta}$ that solves $\boldsymbol{l'(\hat{\theta}) = \frac{d}{d\theta} log(L(\hat{\theta})) = 0.}$
4. If the estimate $\hat{\theta}$ obtained in step 3 corresponds to a maximum $\boldsymbol{\frac{d^2}{d\theta^2} l(\hat{\theta}) < 0}$, then $\hat{\theta}$ is confirmed as the maximum likelihood estimator (MLE) of θ.

Remarks:

- In large sample sizes, the MLE progressively becomes unbiased, efficient and consistent. This can be proved under mild technical assumptions.
- In small sample sizes there is no such guarantee and the quality of a MLE can vary. Bayesian parameter estimation is an area of statistics that deals with this problem.
- MLE generalizes to multi-parameter distributions. Yet this requires multivariate calculus and the maximization may give more than one answer if l has several peaks (local maxima).