(See [6] for a good introduction to ML and Bayesian estimation.)

Maximum likelihood estimation (MLE) is a way to estimate model parameters based on parameterized probability density functions.

1. Construct a conditional model $P(x|\theta)$ which gives the probability density function (PDF) of observables x in terms of model parameters θ.

2. Interpret the PDF as a function $L(\theta) = P(x_0|\theta)$, setting the observables to a particular observed outcome x_0, and find the θ_0 that maximizes this function.

This gives a θ_0 that *best explains* the data, as the probability of observing x_0 is highest for the parameter vector θ_0.

A simple example of MLE is linear least squares estimation with uniform noise assumption.

From [6] p. 91, the Kalman filter arrises in a Bayesian framework from calculating updates of conditional probabilities, taking into account the next observation step. In the case where PDFs of parameter priors and noise sources are gaussian, there is a efficient update mechanism to compute parameter representation of these PDFs (mean and covariance matrix). In general the KF gives the best possible linear estimator in a MSE sense.

In [6], p. 71-77 the differences between ML, MMSE and MAP estimators is explained using the concept of risk[7] which combines cost with probability. An estimator minimises risk. Different cost functions give rise to different estimators.

In [8] a brief explanation is given about how one arrives at the Kalman filter relations from a recursive Bayesian setting.