Maximum Likelihood estimation

Laboratory 6, DEDP

Table of contents

1 Objective

Experiment with Maximum Likelihood estimation of signal parameters.

2 Maximum Likelihood estimation

2.1 ML estimation in gaussian noise

Given a signal $s_{\Theta}(t)$ which depends on some unknown parameters Θ, with gaussian noise, the ML estimation means finding the parameter values $\hat{\Theta}$ which best match the noisy data:

$$\hat{\Theta} = \arg\min_{\Theta} d(r, s_{\Theta})$$

This can be solved numerically, or, in simple cases, analytically.

Example

Consider a constant signal with unknown value A:

$$s_\Theta(t) = A$$

The signal is affected by AWGN.

We take three samples from the signal, and we obtain:

$$r = [3.46, 3.522, 3.48]$$

If there was no noise, we should have obtained:

$$s = [A, A, A]$$

We look for the best A which makes $[A, A, A]$ closest to $[3.46, 3.522, 3.48]$, i.e. minimizes the distance:

$$d(r,s)^2 = (A - 3.46)^2 + (A - 3.522)^2 + (A - 3.48)^2$$

Solution: derivate with respect to A, make = 0, solve for A.

2.2 Matlab

For ML estimation we can use `x = fminsearch(fun, x0)` function, which searches for the values which minimize a given function.

Example:

```
% Define r
r = [3.46, 3.522, 3.48]

% The function to minimize is the distance between r and s
% We use an anonymous function (lambda function) , defined with @(x)
% We can also make a separate function called distance()
distance = @(A) sum( (r - A).^2 )

% Find A which minimizes the function `distance`
% Use initial search value 0
x0 = [0]
A_ML = fminsearch(distance, x0)
```

3 Exercises

1. Generate a 300-samples long sinusoidal signal $s_\Theta = 4{*}\sin(2\pi f n)$ with frequency $f = 0.02$, and add over it normal noise with distribution $\mathcal{N}(0, \sigma^2 = 2)$. Name the resulting vector `r`. Plot the `r` vector.

2. Estimate the frequency $\hat{f}$ of the signal via Maximum Likelihood estimation, based only on the `r` vector.

   - Write the mathematical expression of the Maximum Likelihood estimation in case of Gaussian noise (`Hint:` based on the Euclidean distance)
   - Generate 1000 candidate frequencies f_k equally spaced from 0 to 0.5
   - Compute the Euclidean distance between `r` and a sine signal with each candidate frequency
   - Maximum Likelihood: choose $\hat{f}_{ML}$ as the candidate frequency which minimizes the Euclidean distance
   - Display $\hat{f}_{ML}$, and plot the resulting sinusoidal along the original
   - Try changing the length of the data. How is the estimation accuracy affected?
   - Try changing the variance of the noise. How is the estimation accuracy affected?

3. Estimate the amplitude A of the signal via Maximum Likelihood Estimation, assuming the frequency is known to be 0.02, based only on the r vector.

   Use a similar approach as in Exercise 2.

4. Use `fminsearch()` function to estimate both `A` and `f` simultaneously.

5. Use `fminsearch()` to fit a linear curve $y = ax + b$ through the following points:

```
x = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10];
y = 0.75*x + 3 + 0.1*randn(1,10);
```

   1. Estimate the unknown parameters are a and b (the true values are 0.75 and 3, respectively).
   2. Generate the vector `y_est` with the estimated parameters.
   3. Plot `y_est` and `y` on the same image.

Not done

3. TO UPDATE: Suppose that for f we know a *prior distribution* $w(f)$, displayed on the whiteboard. Modify the previous example to implement Bayesian estimation.

   - Multiply the computed likelihood function from previous exercise with the prior distribution, for each point. The result is the *posterior* distribution.
   - Maximum A Posteriori: choose $\hat{f}_{MAP}$ as the value which maximizes the posterior distribution

- Minimum Mean Squared Error: : choose $\hat{f}_{MMSE}$ as the average value of the posterior distribution
- Display $\hat{f}_{MAP}$ and $\hat{f}_{MMSE}$, and plot the resulting sinusoidal signals along the original and the ML one

4. *Signal inpainting (recover missing parts of signal).* Randomly replace 20 samples from `data` with 0, to simulate missing data. Rerun exercise 3 and estimate the original signal. Plot the result(s) against the starting data (with the missing samples) to visualize the result.

4 Final questions

1. TBD