Statistical Signal Processing. Homework set #1
due November 11, 2022

1. Suppose that a random variable Z has the following Rayleigh density function

$$f(z|\alpha) = \begin{cases} \frac{z}{\alpha^2} exp\{\frac{-z^2}{2\alpha^2}\}, & \text{if } \; z > 0 \\ 0, & \text{if } \; z \leq 0 \end{cases}$$

where $\alpha > 0$.

a) Given N statistically independent measurements of z, find the maximum likelihood estimate of α.

b) Find the Cramer-Rao lower bound on the variance of the unbiased estimator of parameter α.

2. Let $y(1), \ldots, y(N)$ be i.i.d. observations from a location-shifted exponential distribution with density function

$$f(y|\alpha, \lambda) = \frac{1}{\lambda} exp\left(\frac{-(y-\alpha)}{\lambda}\right), \quad y \geq \alpha, \quad \alpha \in \mathbb{R}, \quad \lambda > 0.$$

Find the MLE of (α,λ).

3. Let $y(1), \ldots, y(N)$ be i.i.d. observations from a zero-inflated Poisson distribution with density function given below. Find the MLE of θ.

$$f(y|\theta, \lambda) = \begin{cases} \theta + (1-\theta)e^{-\lambda}, & y = 0 \\ (1-\theta)\frac{e^{-\lambda}\lambda^y}{y!}, & y = 1, 2, 3, \ldots \end{cases}$$

4. Let Y_1, Y_2, Y_3 be three independent and identically distributed Exponential$(1/\mu)$ random variables. Let

$$S = \frac{1}{4}(2Y_1 + Y_2 + Y_3), \quad T = \frac{1}{3}(Y_1 + Y_2 + Y_3).$$

(a) Check that T is the maximum likelihood estimator of μ.

(b) Show that both S and T are unbiased estimators of μ.

(c) Give a reason for why estimator T should be preferred over S.

(d) Show that in fact T is a minimum-variance unbiased estimator (MVUE) of μ.

- The probability density function of $Y \sim$ Exponential(λ) is

$$f(y|\lambda) = \lambda e^{-\lambda y}, \quad y > 0, \quad \lambda > 0$$

and $\mathbb{E}(Y) = \frac{1}{\lambda}$, $\text{var}(Y) = \frac{1}{\lambda^2}$.

5. Suppose that given 256 measurements $y(n)$ $(n = 0, ..., 255)$ we need to estimate the amplitude $\theta = [\alpha_0\ \alpha_1\ \alpha_2]^T$ of a sinusoidal in additive Gaussian noise with $\sigma^2 = 0.81$:

$$y(n) = \alpha_0 + \alpha_1 cos(2\pi f_0 n) + \alpha_2 sin(2\pi f_0 n) + v(n),$$

where $f_0 = \frac{1}{16}$ is known. Furthermore, 10% of the original observations are randomly replaced by outliers that have the maximum signal value (8.0). Desired value of α_0 is 0. Pick desired values of $\alpha_1 \in [2.0, 3.0]$ and $\alpha_2 \in [0.5, 1.0]$ randomly.

Estimate α_0, α_1 and α_2 using M-estimation and so-called Andrew's sine ψ-function given as follows:

$$\psi(\tilde{y}) = \begin{cases} sin(\tilde{y}/a), & |\tilde{y}| \leq a\pi \\ 0, & |\tilde{y}| > a\pi \end{cases}$$

and the value of the tuning parameter a is set such that samples yielding measurement residuals larger that 3σ are rejected completely (have no influence). Use the IRLS method described in the lecture notes.

In your solution show plots of the desired signal, the noisy signal, the weighting function, the estimated signal using both M-estimation and Maximum likelihood (LS in this case) estimation, as well as the desired values and estimated values of the parameters using M-estimation and LS-estimation. Enclose your matlab code as well.

6. Expectation-Maximization Algorithm The EM algorithm simultaneously segments and fits data generated from multiple parametric models. We consider the measured data $y(n)$ generated by the two linear models:

$$\begin{aligned} y(n) &= a_1 x(n) + b_1 + v_1(n) \\ y(n) &= a_2 x(n) + b_2 + v_2(n) \end{aligned} \tag{1}$$

where a_1, b_1 and a_2, b_2 are the model parameters. The noise terms $v_1(n)$ and $v_2(n)$ are assumed to be Gaussian and zero mean. For generating the data for the Matlab numerical simulation, choose a_1 randomly from interval $[-0.5, -0.1]$, a_2 from $[0.5, 2]$, use $b_1 = 1$ and $b_2 = -0.7$. The explaining variable n runs from $n = 0$ up with increment of 0.1. Generate the data corresponding to the two models, considering $N = 64$ data points produced by each model. Add the noise to the generated signals and estimate the model parameters from the noisy data by using the EM Algorithm. The noise variance for the two models is known $\sigma_1^2 = \sigma_2^2 = \sigma^2 = 0.1$.

Requirements:

a) Write the analytical derivation for the EM estimation (E-step, M-step).

b) By using Matlab, simulate the data fitting of the two linear models. Plot the data points together with the estimated models (the two lines in (x, y) plane) after every iteration.

c) Enclose your Matlab code as well

Hints:

- EM Algorithm does not assign directly each data point to one of the models. EM assigns to each data point, a probability of belonging to one of the models (likelihood).
- E-step: Assume random model parameters in the beginning. Calculate the likelihood of each data point belonging to each model. For this, you consider the residual error of each point n for each model k, $r_k(n) = a_k x(n) + b_k - y(n), \quad k = 1, 2$. The likelihood of each data point is $w_k(n) = P(a_k, b_k | r_k(n))$.
- M-step: Take the likelihood of each data point belonging to each model and re-estimate the model parameters using Weighted Least Squares. For this, you need to build a weighted quadratic error function. (The weight for the squared error of each point is the likelihood of that point).