ESTIMATION - DETECTION

TD 1 — Estimation

Exercise 1

We consider n independent identically distributed random variables $X_1, \ldots, X_n$ with law given by the Rayleigh distribution, $X \sim \mathcal{R}(\sigma^2)$.

1. Maximum likelihood estimation
   (a) Express the likelihood of n observations $(x_1, \cdots, x_n)$ and derive the maximum likelihood estimator $\hat{\sigma}^2_{ML}$ for σ^2.
   (b) Determine the bias and variance of $\hat{\sigma}^2_{ML}$. Is $\hat{\sigma}^2_{ML}$ unbiased and convergent?
   Note: use the result that if $Y_k \overset{i.i.d.}{\sim} \mathcal{R}(\sigma^2)$*, then* $Z = \sum_{k=1}^{K} Y_k^2 \sim \mathcal{G}(K, 2\sigma^2)$.
   (c) Determine the Cramér-Rao bound for σ^2. Is $\hat{\sigma}^2_{ML}$ efficient?
2. Method of moments
   (a) Derive an estimator for σ^2 using the second moment of X, denoted by $\hat{\sigma}^2_{m2}$. Determine if it is biased, convergent, and efficient.
   (b) Derive an estimator for σ^2 using the first moment of X, denoted by $\hat{\sigma}^2_{m1}$ and determine its bias. Which of the two estimators $\hat{\sigma}^2_{m1}$ and $\hat{\sigma}^2_{m2}$ is preferable?
   (c) From $\hat{\sigma}^2_{m1}$, derive an alternative estimator $\tilde{\sigma}^2_{m1}$ based on the first moment of X that is unbiased. Which of the two estimators $\hat{\sigma}^2_{m1}$ and $\tilde{\sigma}^2_{m1}$ has smaller variance, and which one is overall preferable?
3. Bayesian estimation

   Suppose the parameter σ^2 is known to follow an Inverse Gamma distribution $\mathcal{IG}(a, b)$ with parameters $a = \alpha > 0$ (shape) and $b = \beta/2 > 0$ (scale).

   (a) Determine, up to multiplicative constants, the posterior and the log-posterior law for σ^2. Show that the posterior is $\mathcal{IG}(a', b')$ and determine its parameters.
   (b) Calculate the MAP estimator for σ^2, denoted by $\hat{\sigma}^2_{MAP}$.
   (c) Calculate the MMSE estimator for σ^2, denoted by $\hat{\sigma}^2_{MMSE}$.

Exercise 2

Suppose that the output of a physical system can be well modeled by random variables X_i) that are independently and identically distributed according to a log-Normal distribution $\ln \mathcal{N}(\eta, \nu^2)$. The system can take on M discrete states $m = 1, \ldots, M$, each of which is associated with a distinct discrete value $\eta_m \in \{\eta_1, \ldots, \eta_M\}$. The parameter ν^2 is known, and the problem consists in determining the state of the system from n measurements $(x_1, \ldots, x_n)$.

1. Derive the maximum likelihood estimator for η_m (and hence the state m of the system). Interpret the result in view of the result that would be obtained if the output of the physical system was modeled by a Gaussian distribution $\mathcal{N}(\eta, \nu^2)$.

2. Suppose that we have prior information in the form $P[\eta = \eta_m] = p_m$, $0 \leq p_m \leq 1$, $\sum_{m=1}^{M} p_m = 1$. Derive the MAP estimator for m. How do you interpret the result with respect to the expression for the maximum likelihood estimator in 1. ?

3. Can the MMSE estimator be useful for this problem? Why / why not?

Exercise 3

Least squares (LS) methods can be used for estimating the parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_J)^T$ of a *linear model*

$$y_i = \sum_{j=1}^{J} x_{ij}\beta_j + \varepsilon_i, \tag{1}$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)^T$ are the observations, x_{ij} are fixed design variables and ε_i are scalar random variables that account for the discrepancies between the observations and the predictions $x_{ij}\beta_j$ (errors, noise). The LS methods aims at minimizing the squared difference between the observations and the fitted model, $\sum_{i=1}^{n}(y_i - \sum_{j=1}^{J} x_{ij}\beta_j)^2$.

1. Ordinary least squares method and maximum likelihood estimation

   (a) Suppose that the errors ε_i are independent and identically distributed with Normal law $\varepsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$. Show that the solution $\hat{\boldsymbol{\beta}}_{OLS} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}(\boldsymbol{X}^T\boldsymbol{y})$ to the *ordinary least squares* (OLS) problem $\hat{\boldsymbol{\beta}}_{OLS} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n}(y_i - \sum_{j=1}^{J} x_{ij}\beta_j)^2$ is identical to the maximum likelihood estimate $\hat{\boldsymbol{\beta}}_{ML}$. ($\boldsymbol{X}$ is the design matrix with elements x_{ij}.)

2. Weighted and generalized least squares methods and maximum likelihood estimation

   (a) *Weighted least squares* (WLS) can be used if different confidence is attributed to the observations y_i, $\hat{\boldsymbol{\beta}}_{WLS} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} w_i(y_i - \sum_{j=1}^{J} x_{ij}\beta_j)^2$, where the weights $\boldsymbol{w} = (w_1, \ldots, w_n)^T$ reflect the confidence in the observation $\boldsymbol{y}$. Suppose that the errors ε_i are independent but distributed according to $\varepsilon_i \sim \mathcal{N}(0, \sigma_i^2)$ and determine the weights w_i in the WLS solution $\hat{\boldsymbol{\beta}}_{WLS} = (\boldsymbol{X}^T\boldsymbol{w}\boldsymbol{X})^{-1}(\boldsymbol{X}^T\boldsymbol{w}y)$ such that the WLS estimte is identical to the maximum likelihood estimate, $\hat{\boldsymbol{\beta}}_{WLS} = \hat{\boldsymbol{\beta}}_{ML}$.

   (b) Finally, *generalized least squares* (GLS) can be used if the errors $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^T$ are zero mean but dependent with known covariance $\boldsymbol{\Sigma} = \mathbb{E}\left[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\right]$. Show that in the case of Gaussian errors ε_i, the maximum likelihood estimate $\hat{\boldsymbol{\beta}}_{ML}$ is identical to the GLS estimate $\hat{\boldsymbol{\beta}}_{GLS} = (\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}(\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y})$.

3. Bayesian estimation. Suppose that $\beta \in \mathbb{R}$ ($J = 1$), i.e. $y_i = x_i\beta + \varepsilon_i$, and that a Normal prior $\beta \sim \mathcal{N}(\tilde{\beta}, \nu^2)$ is assigned to β.

   (a) Let $\varepsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$ and derive the MAP and MMSE estimate for β.

   (b) Let $\varepsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma_i^2)$ and derive the MAP and MMSE estimate for β.

   (c) Compare the MAP and MMSE estimates in both cases to the corresponding LS estimates. How do you interpret the result?

4. Suppose that $x = [1, 2, 3, 4, 5]$ and $y = [1.3, 0.5, 2.4, 6.5, 5.1]$, $\tilde{\beta} = 1$ and $\nu^2 = 0.1$, and that the variances of the independent Gaussian noise ε_i are given by $\boldsymbol{\sigma}^2 = [1, 2, 3, 4, 5]$.

   (a) Calculate $\hat{\beta}$ using the results from 1.(a) and 3.(a) (using $\sigma^2 = 3$ where needed).

   (b) Calculate $\hat{\beta}$ using the estimators derived in 2.(a) and 3.(b).

- *Rayleigh distribution* $\mathcal{R}(\sigma^2)$*:* $\sigma > 0,\ x \geq 0$
  - *density* $f(x;\sigma^2) = \frac{x}{\sigma^2}\exp\left(-\frac{x^2}{2\sigma^2}\right)$
  - *mean* $\mu = \sigma\sqrt{\frac{\pi}{2}}$
  - *variance* $\nu^2 = \frac{4-\pi}{2}\sigma^2$
- *Gamma distribution* $\mathcal{G}(k,\theta)$*:* $k > 0,\ \theta > 0,\ x > 0$
  - *density* $f(x;k,\theta) = \frac{1}{\Gamma(k)\theta^k}x^{k-1}\exp\left(-\frac{x}{\theta}\right)$
  - *mean* $\mu = k\theta$
  - *variance* $\nu^2 = k\theta^2$

    note: $y \sim \mathcal{G}(k,\theta) \implies cy \sim \mathcal{G}(k,c\theta)$
- *Inverse Gamma distribution* $\mathcal{IG}(a,b)$ *:* $a > 0,\ b > 0,\ x > 0$
  - *density* $f(x;a,b) = \frac{b^a}{\Gamma(a)}x^{-(a+1)}\exp\left(-\frac{b}{x}\right)$
- *log-Normal distribution* $\ln\mathcal{N}(\eta,\nu^2)$ *:* $\eta \in \mathbb{R},\ \nu^2 > 0$
  - *density* $f(x;\eta,\nu^2) = \frac{1}{x\sqrt{2\pi\nu^2}}\exp\left(-\frac{(\ln x-\eta)^2}{2\nu^2}\right)$