ESTIMATION - DETECTION

TD 1 — Estimation

Exercise 1

We consider n independent identically distributed random variables $X_1, \cdots, X_n$ from a Gamma law $\mathcal{G}(\beta, \theta)$.

1. Maximum likelihood estimation
   (a) Suppose that the shape parameter β is known. Express the likelihood of n observation $(x_1, \cdots, x_n)$ and derive the maximum likelihood estimator $\hat{\theta}_{ML}$ for θ.
   (b) Determine the bias and variance of $\hat{\theta}_{ML}$. Is $\hat{\theta}_{ML}$ unbiased and convergent?
   (c) Analyze the bias and variance of $\hat{\theta}_{ML}$ and interpret the quality of $\hat{\theta}_{ML}$ in view of the Cramer-Rao bound.
   (d) Now suppose that θ is known and β is unknown. Derive the expression determining the maximum likelihood estimator $\hat{\beta}_{ML}$ for β.
2. Method of moments
   (a) Suppose that both θ and β are unknown. Derive estimators for θ and β using the first and second moments of X.
3. Bayesian estimation with inverse Gamma prior $\mathcal{IG}(k, \tau)$ for θ: $\quad \theta \sim \mathcal{IG}(k, \tau)$
   (a) Derive the posterior law, show that it is $\mathcal{IG}(a, b)$ and determine its parameters.
   (b) Derive the MAP estimator $\hat{\theta}_{MAP}$ for θ.
   (c) Show that the expectation of an inverse Gamma random variable $X \sim \mathcal{IG}(k, \tau)$ is given by $\mathbb{E}[X] = \frac{\tau}{k-1}$ and determine the MMSE estimator $\hat{\theta}_{MMSE}$ for θ.

- *Gamma distribution* $\mathcal{G}(k, \theta)$*:* $\qquad k > 0,\ \theta > 0,\ x > 0$
  - *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: $\quad y \sim \mathcal{G}(k, \theta) \implies cy \sim \mathcal{G}(k, c\theta)$
- *Inverse Gamma distribution* $\mathcal{IG}(a, b)$ *:* $\qquad a > 0,\ b > 0,\ x > 0$
  - *density* $f(x; a, b) = \frac{b^a}{\Gamma(a)} x^{-(a+1)} \exp\left(-\frac{b}{x}\right)$
- *log-Normal distribution* $\ln \mathcal{N}(\eta, \nu^2)$ *:* $\qquad \eta \in \mathbb{R},\ \nu^2 > 0$
  - *density* $f(x; \eta, \nu^2) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\ln x - \eta)^2}{2\nu^2}\right)$
- *Rayleigh distribution* $\mathcal{R}(\sigma^2)$*:* $\qquad \sigma > 0,\ x \geq 0$
  - *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* $\nu^2 = \frac{4-\pi}{2}\sigma^2$

Exercise 2

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

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

where $\boldsymbol{y} = (y_1, \dots, y_n)^T$ 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 predications $x_{ij}\beta_j$ (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}.)

   (b) Find the maximum likelihood estimator for σ^2 and interpret the result.

   (c) Suppose that the design variables are given by a known and fixed *nonlinear* transformation $\phi_{ij}(x_{i1}, \dots, x_{iJ})$ (for instance, $\phi_{ij} = a_j x_{ij}^2$). How does the maximum likelihood estimator change? How do you interpret this result?

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 $\boldsymbol{w} = (w_1, \dots, w_n)^T$ 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 $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^T$ 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).

Exercise 3

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 $m = 1, \dots, M$, each of which is associated with a distinct discrete value $\eta_m \in \{\eta_1, \dots, \eta_M\}$. The parameter ν^2 is known, and the problem consists in determining the state of the system from n measurements $(x_1, \dots, x_n)$.

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 $P[\eta = \eta_m] = p_m$, $0 \leq p_m \leq 1$, $\sum_{m=1}^{M} pm = 1$. 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 4

The distribution of the size of files in internet traffic (TCP protocol) can be modeled by a Pareto distribution which has density given by $f(x) = \frac{\alpha x_m^\alpha}{x^{\alpha+1}}, \quad x \geq x_m$ with scale parameter $x_m > 0$ and shape parameter $\alpha > 0$. We want to estimate the parameters of this model from observations $(x_1, \cdots, x_n)$ in order to, for instance, use them in a procedure for detecting abnormal traffic (attacks).

1. Calculate the mean and variance of the model. How do they behave as a function of α?
2. Derive the maximum likelihood estimators $\hat{x}_{m,ML}$ and $\hat{\alpha}_{ML}$.
3. Calculate the Fisher information matrix for the parameter vector $\theta = (x_m, \alpha)$. How do you interpret the off-diagonal terms?
4. Suppose that x_m (in our example, the minimal possible file size) is known and that $\alpha > 1$. Derive the moment based estimator $\hat{\alpha}_M$ for α using the first moment $\mathbb{E}[X]$.
5. In Bayesian statistics, when no (reliable) prior information is available for a parameter θ, one often uses a *non-informative* prior. One possible choice for a non-informative prior for θ is the Jeffreys prior which is proportional to the square root of the Fisher information, $p(\theta) \propto \sqrt{I(\theta)}$.
   (a) Assume that x_m is known and calculate the Jeffreys prior for α for our problem.
   (b) Assume that x_m is known and derive the MAP estimator for α using Jeffreys prior.
6. Assume that x_m is known and that α follows a Gamma law $\mathcal{G}(k, \theta)$. Derive the MAP and the MMSE estimator for α.

Exercise 5

Magnetic resonance imaging (MRI) results in complex-valued images. These are usually analyzed through their magnitude. Suppose that the recorded data are corrupted by background noise $\underline{X} = X_R + \mathrm{i}X_I$. The real and imaginary parts of $\underline{X}$ are independent and distributed according to a centered Gaussian distribution, $X_R \sim \mathcal{N}(0, \sigma^2)$ and $X_I \sim \mathcal{N}(0, \sigma^2)$, respectively. It follows that the magnitude of the noise, $X = \sqrt{X_R^2 + X_I^2}$, is distributed according to a Rayleigh distribution, $X \sim \mathcal{R}(\sigma^2)$. We want to estimate the noise level σ^2 from a sample $(X_1, \cdots, X_n)$ of n pixels of an image of the background.

1. Express the likelihood of the n pixels $(x_1, \cdots, x_n)$. Derive the maximum likelihood estimator for σ^2, denoted by $\hat{\sigma}^2_{ML}$.
2. Determine whether $\hat{\sigma}^2_{ML}$ is unbiased / convergent / efficient or not.
   (hint: If $x_i \overset{i.i.d.}{\sim} \mathcal{R}(\sigma^2)$, then $\sum_{i=1}^{n} x_i^2 \sim \mathcal{G}(n, 2\sigma^2)$.)
3. Method of moments
   (a) Derive an estimator for σ^2 using the second moment of X, denoted by $\hat{\sigma}^2_{m2}$, and study its bias, convergence, and efficiency.
   (b) Derive an estimator for σ^2 using the first moment of X, denoted by $\hat{\sigma}^2_{m1}$ and study 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 $\tilde{\sigma}^2_{m1}$ that is unbiased. Which of the two estimators $\hat{\sigma}^2_{m1}$ and $\tilde{\sigma}^2_{m1}$ has smaller variance? Which one is overall preferable?
4. 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) Derive the posterior and log-posterior law for σ^2. Show that the posterior is $\mathcal{IG}(a', b')$ and determine its parameters.
   (b) Derive the MAP estimator for σ^2, denoted by $\hat{\sigma}^2_{MAP}$.
   (c) Derive the MMSE estimator for σ^2, denoted by $\hat{\sigma}^2_{MMSE}$.
   (d) Which of the two estimators $\hat{\sigma}^2_{MAP}$ and $\hat{\sigma}^2_{MMSE}$ has smaller variance / smaller bias?
5. Compare the behavior of $\hat{\sigma}^2_{ML}$, $\hat{\sigma}^2_{MAP}$ and $\hat{\sigma}^2_{MMSE}$ as $n \to \infty$.

Exercise 6

The Poisson distribution $P(\lambda)$ with rate parameter $\lambda > 0$ is widely used for modeling and predicting the number of failures occurring in a time interval. Its probability mass function is given by

$$P(X = x) = \frac{\lambda^x}{x!} \exp(-\lambda)$$

and its mean and variance are both given by the rate parameter λ. Suppose that in n unit time intervals we have observed the number of failures $(x_1, \cdots, x_n)$.

1. Derive the maximum likelihood estimator $\hat{\lambda}_{ML}$ for the rate parameter λ. Show that it is unbiased, convergent, and efficient.
2. Suppose now that we know that the rate parameter λ follows a Gamma law

   $$p(\lambda) = \mathcal{G}(\alpha, \beta).$$

   Derive the MMSE estimator $\hat{\lambda}_{MMSE}$ and the MAP estimator $\hat{\lambda}_{MAP}$ for λ.
3. Calculate the bias and variance of $\hat{\lambda}_{ML}$, $\hat{\lambda}_{MMSE}$ and $\hat{\lambda}_{MAP}$ and compare them as the sample size grows large ($n \longrightarrow \infty$).
4. Calculate the MSE of $\hat{\lambda}_{MMSE}$ and $\hat{\lambda}_{MAP}$. Which of the two estimators would you prefer, and why?

Exercise 7

Suppose that binary information $\xi \in \{0, 1\}$ is submitted twice over some transmission channel. The channel adds noise, which we assume to be zero mean and Gaussian with variance σ^2. The received message hence reads $z = (z_1, z_2)$ where $z_k = \xi + \varepsilon_k,\ k = 1, 2$ and $\varepsilon_k \sim \mathcal{N}(0, \sigma^2)$. The problem consists in recovering the symbol ξ that has been transmitted from the received signal $z = (z_1, z_2)$.

1. Derive the maximum likelihood estimator for ξ.
2. Suppose that we have prior information on ξ in the form $P[\xi = 0] = p$ and $P[\xi = 1] = 1 - p$. Derive the MAP estimator for ξ. 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?