Problem Set I.

Quasi maximum likelihood estimation and Sieve model selection in linear models.

Advanced Statistical Methods
January 23, 2017

Student: Arshak Minasyan

Problem 1.1. If a homogeneous noise is assumed, that is, if $\Sigma = \sigma^2 I_n$ and $\epsilon \sim \mathcal{N}(0, \sigma^2 I_n)$, then the formulae for the MLEs $\tilde{\theta}$, $\tilde{\mathbf{f}}$ slightly simplify. In particular, the variance σ^2 cancels and the resulting estimate is the ordinary least squares (oLSE). Derive the formulae for $\tilde{\theta}$ and $\mathbf{f}$ directly from the log-likelihood $L(\theta)$ for honogenuous noise.

$$\tilde{\theta} = (\Psi\Psi^T)^{-1}\Psi^T\mathbf{Y} = \mathcal{S}\mathbf{Y} \tag{1}$$

with $\mathcal{S} = (\Psi\Psi^T)^{-1}\Psi^T$. Also,

$$\tilde{\mathbf{f}} = \Psi^T(\Psi\Psi^T)^{-1}\Psi^T = \Pi\mathbf{Y} \tag{2}$$

with $\Pi = \Psi^T(\Psi\Psi^T)^{-1}\Psi^T$.

Solution. The log-likehood has the following form

$$\tilde{\theta} = \arg\max_{\theta} L(\theta) = -\frac{\|\mathbf{Y} - \Psi^T\theta\|^2}{2\sigma^2} + R, \tag{3}$$

where R is a remainder and does not depend on θ, hence the maximum of this quadratic form achieved when

$$\tilde{\theta} = (\Psi\Psi^T)^{-1}\Psi\mathbf{Y}. \tag{4}$$

Hence,

$$\tilde{\mathbf{f}} = \Psi^T\tilde{\theta} = \Psi^T(\Psi\Psi^T)^{-1}\Psi^T = \Pi\mathbf{Y} \tag{5}$$

■

Problem 1.2. Consider univariate polynomial regression of degree $p1$. This means that $\mathbf{f}$ is a polynomial function of degree $p1$ observed at the points X_i with errors ϵ_i that are assumed to be i.i.d. normal. The function $\mathbf{f}$ can be represented as

$$f(x) = \theta_1^* + \theta_2^* x + \cdots + \theta_p^* x^{p-1} \tag{6}$$

using the basis functions $\psi_j(x) = x^{j-1}$ for $j = 0, \ldots, p-1$. At the same time, for any point x_0, this function can also be written as

$$f(x) = u_1^* + u_2^*(x - x_0) + \cdots + u_p^*(x - x_0)^{p-1} \tag{7}$$

using the basis functions $\breve{\psi}_j = (x - x_0)^{j-1}$.

- Write the matrices Ψ and $\Psi\Psi^T$ and similarly $\breve{\Psi}$ and $\breve{\Psi}\breve{\Psi}^T$.
- Describe the linear transformation A such that $u = A\theta$ for $p = 1$.
- Describe the transformation A such that $u = A\theta$ for $p > 1$.

Solution.

- Each jth row of the matrix $\Psi(\breve{\Psi})$ consists of the following entries $x_1^j, \ldots, x_n^j((x_1 - x_0)^j, \ldots (x_n - x_0)^j)$. The product $\Psi\Psi^T$ has the following form

$$\begin{pmatrix} n & \sum x_i & \sum x_i^2 & \ldots \\ \sum x_i & \sum x_i^2 & \sum x_i^3 & \ldots \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 & \ldots \\ \ldots & \ldots & \ddots & \ldots \\ \sum x_i^{p-1} & \sum x_i^p & \ldots & x_i^{2p-2} \end{pmatrix} \tag{8}$$

  For $\breve{\Psi}\breve{\Psi}^T$ the matrix has the same form with the only difference $(x_i - x_0)$ instead of x_i.

- We will describe the matrix A is general case skipping the second sub-problem.

$$\begin{pmatrix} u_1 \\ \vdots \\ u_p \end{pmatrix} = \mathbf{u} = \begin{pmatrix} 1 & x_0 & \ldots & x_0^{p-1} \\ x_0 & 1 & 2x_0 & \ldots \\ \ldots & \ldots & \ddots & \ldots \\ \ldots & \ldots & \ldots & 1 \end{pmatrix} \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_p \end{pmatrix} = A\theta. \tag{9}$$

■

Problem 1.3. Prove the result of Theorem 1.1.2.

Solution. Since the matrix $\Psi\Psi^T$ is symmetric then the eigen decomposition gives us the following result

$$\Psi\Psi^T U = U\Lambda, \tag{10}$$

where Q is an orthogonal matrix, hence we can denote $\tilde{\Psi} = U\Psi$ and get that $\tilde{\Psi}\tilde{\Psi}^T = \Lambda$. Define $Z = \tilde{\Psi}\mathbf{Y}$ then

$$Z = \tilde{\Psi}\mathbf{Y} = \tilde{\Psi}\tilde{\Psi}^T\mathbf{u} + \tilde{\Psi}\epsilon = \Lambda\mathbf{u} + \tilde{\Psi}\epsilon = \Lambda\mathbf{u} + \Lambda^{1/2}\xi, \tag{11}$$

where $\xi = \Lambda^{-1/2}\tilde{\Psi}\epsilon$ and

$$\mathbb{E}\xi\xi^T = \Lambda^{-1/2}\tilde{\Psi}\sigma^2 I_n \tilde{\Psi}^T \Lambda^{-1/2} = \sigma^2 I_p. \tag{12}$$

which leads to the end of the proof of the first part. Second part follows directly from the fact that the linear combination of Gaussian random variables is still also Gaussian. ■

Problem 1.4. Consider the case of an orthogonal design with $\Psi^T = I_p$. Specify the projector Π of Lemma 1.2.1. for this situations, particularly its decomposition from (vi).

Solution. The projector matrix that we have now is $\Pi = \Psi\Psi^T$. It obviously a projector, since it is symmetric and $\Pi^2 = \Pi$. Considering to the (vi) property from Lemma 1.2.1. we have the same representation

$$\Pi = U\Lambda_p U^T, \tag{13}$$

with the only difference that $\Pi = \sum_{i=1}^{p} \psi_i \psi_i^T$, which implies that the first columns of matrix U, since the diagonal matrix Λ_p has p 1s and $n-p$ 0s. ■

Problem 1.5. Let δ be the stochastic component of $\tilde{\theta}$ built for the misspecified linear model $\mathbf{Y} = \Psi^T\theta^* + \epsilon$ with $\operatorname{Var}\epsilon = \Sigma$. Let also the true noise variance is Σ_0. Then $\operatorname{Var}(\tilde{\theta}) = W^2$ with

$$W^2 = \left(\Psi\Sigma^{-1}\Psi^T\right)^{-1}\Psi\Sigma^{-1}\Sigma_0\Sigma^{-1}\Psi^T\left(\Psi\Sigma^{-1}\Psi^T\right)^{-1}. \tag{14}$$

Solution. Too see this recall that the estimator $\tilde{\theta}$ has the following form

$$\tilde{\theta} = \mathcal{S}\mathbf{Y}, \tag{15}$$

where $\mathcal{S} = \left(\Psi\Sigma^{-1}\Psi^T\right)^{-1}\Psi\Sigma^{-1}$ under the parametric assumption (PA) that $\operatorname{Var}\epsilon = \Sigma$. Hence,

$$\operatorname{Var}(\tilde{\theta}) = \mathcal{S}\operatorname{Var}(\epsilon)\mathcal{S}^T = \mathcal{S}\operatorname{Var}(\mathbf{Y})\mathcal{S}^T = W^2, \tag{16}$$

which completes the proof. ■

Problem 1.6. State the result of Theorems 1.3.2 and 1.3.3 for the MLE $\tilde{\theta}$ built in the model $\mathbf{Y} = \Psi^T\theta^* + \epsilon$ with $\operatorname{Var}\epsilon = \Sigma$.

Solution. Again, recall that under the PA $\operatorname{Var}\epsilon = \Sigma$ we have $\mathcal{S} = \left(\Psi\Sigma^{-1}\Psi^T\right)^{-1}\Psi\Sigma^{-1}$ and the result of Theorem 1.3.2. continues to apply with the the new $\mathcal{S}$.

Since $\mathbb{E}\epsilon^T\epsilon$ does not depend on θ then the proof is identical to the one provided in script with only difference that $\mathcal{S} = \left(\Psi\Sigma^{-1}\Psi^T\right)^{-1}\Psi\Sigma^{-1}$. ■

Problem 1.7. Consider the regression model

$$Y_i = \theta_1^*\psi_1(X_i) + \cdots + \theta_p^*\psi_p(X_i) + \epsilon_i \tag{17}$$

with independent heterogeneous errors $\operatorname{Var}\epsilon_i = \sigma_i^2$. Consider the MLE $\tilde{\theta}$ and the LSE $\tilde{\theta}_{\mathrm{LSE}} = (\Psi\Psi^T)^{-1}\Psi\mathbf{Y}$ and the corresponding to homogeneous errors.

- Compute $\tilde{\theta}$;
- Show that $\mathbb{E}\tilde{\theta} = \mathbb{E}\tilde{\theta}_{\mathrm{LSE}} = \theta^*$;
- Compute the variance $\operatorname{Var}\tilde{\theta}$ and the variance $\operatorname{Var}\tilde{\theta}_{\mathrm{LSE}}$;
- Show that $\operatorname{Var}\tilde{\theta}_{\mathrm{LSE}} \geq \operatorname{Var}\tilde{\theta}$
- Check that $\operatorname{Var}\tilde{\theta}_{\mathrm{LSE}} = \operatorname{Var}\tilde{\theta}$ iff all the σ_i are equal to each other.

Solution.

- Assuming Var $\epsilon = \Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_n)$ then we have that

$$\tilde{\theta} = \mathcal{S}\mathbf{Y} = \left(\Psi\Sigma^{-1}\Psi^T\right)^{-1}\Psi\Sigma^{-1}\mathbf{Y} \tag{18}$$

- Since $\mathbf{f}^* = \Psi^T\theta^*$, where $\mathbb{E}\mathbf{Y} = \mathbf{f}^*$ then ot os easy to see that $\mathbb{E}\tilde{\theta} = \mathbb{E}\tilde{\theta}_{\mathrm{LSE}} = \theta^*$.
- It is straightforwards to check that

$$\operatorname{Var}\tilde{\theta}_{\mathrm{LSE}} = (\Psi\Psi^T)^{-1}\Psi\operatorname{Var}(\mathbf{Y})\Psi^T(\Psi\Psi^T)^{-1} = (\Psi\Sigma^{-1}\Psi^T)^{-1}, \tag{19}$$

  where the latter is obtained using the assumption about homogenous noise. Similarly,

$$\operatorname{Var}\tilde{\theta} = \operatorname{Var}(\mathcal{S}\mathbf{Y}) = \mathcal{S}\operatorname{Var}(\mathbf{Y})\mathcal{S}^T = \left(\Psi\Sigma^{-1}\Psi^T\right)^{-1}\Psi\Sigma^{-1}\underbrace{\operatorname{Var}(\mathbf{Y})}_{\Sigma}\Sigma^{-1}\Psi^T\left(\Psi\Sigma^{-1}\Psi^T\right)^{-1}.$$

  Hence, we derived at a closed form expression for $\operatorname{Var}\tilde{\theta}$:

$$\operatorname{Var}\tilde{\theta} = \left(\Psi\Sigma^{-1}\Psi^T\right)^{-1}\Psi\Sigma^{-1}\Psi^T\left(\Psi\Sigma^{-1}\Psi^T\right)^{-1} \tag{20}$$

- The result immediately follows from Gauss-Markov theorem.
- It obviously holds for the case when $\sigma_1 = \cdots = \sigma_n = \sigma$, while in the opposite direction is not that trivial.

$\blacksquare$

Problem 1.8. Consider the nonparametric model

$$Y_i = f(X_i) + \epsilon_i, \qquad \operatorname{Var}(\epsilon_i) = \sigma_i^2 \tag{21}$$

and the parametric approximation (17). Consider the qMLE (LSE) $\tilde{\mathbf{f}}_{\mathrm{LSE}} = \Pi\mathbf{Y}$ for $\Pi = \Psi^T(\Psi\Psi^T)^{-1}\Psi$.

- Derive the bias-variance decomposition for the quadratic losses $\|\tilde{\mathbf{f}}_{\mathrm{LSE}} - \mathbf{f}^*\|^2$ and of the risk $\mathbb{E}\|\tilde{\mathbf{f}}_{\mathrm{LSE}} - \mathbf{f}^*\|^2$.
- Compute the variance term of $\tilde{\mathbf{f}}_{\mathrm{LSE}}$ and $\tilde{\mathbf{f}} = \Psi^T\tilde{\theta}$ for the MLE $\tilde{\theta}$.

Solution.

-

$$\|\tilde{\mathbf{f}}_{\mathrm{LSE}} - \mathbf{f}^*\|^2 = \|\Pi f + \Pi\epsilon - f\|^2 = \|(I-\Pi)f\|^2 + \|\Pi\epsilon\|^2 + \underbrace{2\epsilon^T\Pi(I-\Pi)f}_{=0} = \|(I-\Pi)f\|^2 + \|\Pi\epsilon\|^2$$

$$\mathbb{E}\|\tilde{\mathbf{f}}_{\mathrm{LSE}} - \mathbf{f}^*\|^2 = \|\Pi f + \Pi\epsilon - f\|^2 = \|(I-\Pi)f\|^2 + \mathbb{E}\|\Pi\epsilon\|^2 = \|(I-\Pi)f\|^2 + \operatorname{tr}\Pi\Sigma.$$

- $$\mathrm{Var}\ \tilde{\mathbf{f}}_{\mathrm{LSE}} = \mathrm{Var}\ (\Pi \mathbf{Y}) = \Pi \Sigma \Pi^T, \tag{22}$$

  where $\Pi = \Psi^T(\Psi\Psi^T)^{-1}\Psi$.

  As for the $\tilde{\mathbf{f}}$ we have similar expression

  $$\mathrm{Var}\ \tilde{\mathbf{f}} = \mathrm{Var}\ (\Pi \mathbf{Y}) = \Pi \Sigma \Pi^T, \tag{23}$$

  where $\Pi = \Psi^T(\Psi\Sigma^{-1}\Psi^T)^{-1}\Psi\Sigma^{-1}$.

$\blacksquare$

Problem 1.9. Consider the projection estimator $\tilde{\mathbf{f}}_m = \Pi_m \mathbf{Y}$ for the model (17) with $\Pi_m = \Psi_m^T(\Psi_m\Psi_m^T)^{-1}\Psi_m$. For two different values $m' > m$:

- Show that $\Pi_{m',m} = \Pi_{m'} - \Pi_m$ is a projector in $\mathbb{R}^n$. Describe its image in the orthogonal case when $\Psi\Psi^T$ is a diagonal matrix.
- Check the identities

  $$\|\mathbf{Y} - \Pi_m \mathbf{Y}\|^2 - \|\mathbf{Y} - \Pi_{m'}\mathbf{Y}\|^2 = \|\tilde{\mathbf{f}}_{m'} - \tilde{\mathbf{f}}_m\|^2 = \|\Pi_{m',m}\mathbf{f} + \Pi_{m',m}\epsilon\|^2 \tag{24}$$

  $$\|\tilde{\mathbf{f}}_{m'} - \mathbf{f}\|^2 - \|\tilde{\mathbf{f}}_m - \mathbf{f}\|^2 = -\|\Pi_{m',m}\mathbf{f}\|^2 + \|\Pi_{m',m}\epsilon\|^2. \tag{25}$$

- Compute $\mathbb{E}\|\tilde{\mathbf{f}}_{m'} - \tilde{\mathbf{f}}_m\|^2$ and $\mathbb{E}\left[\|\tilde{\mathbf{f}}_{m'} - \mathbf{f}\|^2 - \|\tilde{\mathbf{f}}_m - \mathbf{f}\|^2\right]$

Solution.

- Let us examine the term $\Pi_{m',m} = \Pi_{m'} - \Pi_m$:

  $\Pi_{m',m}^2 = (\Pi_{m'} - \Pi_m)(\Pi_{m'} - \Pi_m) = \Pi_{m'}^2 - 2\Pi_m(\Pi_{m'} - \Pi_m) + \Pi_m^2 = \Pi_{m'} - \Pi_m = \Pi_{m',m}$,

  hence $\Pi_{m',m} = \Pi_{m'} - \Pi_m$ is indeed a projection. Here we have used the fact that $\Pi_m(\Pi_{m'} - \Pi_m) = 0$ for $m' > m$.

- Opening the brackets yields

  $$\begin{aligned} \|\mathbf{Y} - \Pi_m\mathbf{Y}\|^2 - \|\mathbf{Y} - \Pi_{m'}\mathbf{Y}\|^2 = \|(I - \Pi_m)\mathbf{Y}\|^2 - \|(I - \Pi_{m'})\mathbf{Y}\|^2 = \\ \mathbf{Y}^T(I - \Pi_m)\mathbf{Y} - \mathbf{Y}^T(I - \Pi_{m'})\mathbf{Y} = \mathbf{Y}^T\Pi_{m',m}\mathbf{Y} = \|\Pi_{m',m}\mathbf{Y}\|^2 = \\ \|\tilde{\mathbf{f}}_{m'} - \tilde{\mathbf{f}}_m\|^2 = \|\Pi_{m',m}\mathbf{f} + \Pi_{m',m}\epsilon\|^2 \end{aligned}$$

  Using the fact that $\|\tilde{\mathbf{f}}_m - \mathbf{f}\|^2 = \|(I - \Pi_m)\mathbf{f}\|^2 + \|\Pi_m\epsilon\|^2$ one gets

  $$\begin{aligned} \|\tilde{\mathbf{f}}_{m'} - \mathbf{f}\|^2 - \|\tilde{\mathbf{f}}_m - \mathbf{f}\|^2 = \|(I - \Pi_{m'})\mathbf{f}\|^2 + \|\Pi_{m'}\epsilon\|^2 - \|(I - \Pi_m)\mathbf{f}\|^2 - \|\Pi_m\epsilon\|^2 = \\ -\|\Pi_{m',m}\mathbf{f}\|^2 + \|\Pi_{m',m}\epsilon\|^2, \end{aligned}$$

  where the latter equation follows from the previous shown identity.

- The first term

$$\mathbb{E}\|\tilde{\mathbf{f}}_{m'} - \tilde{\mathbf{f}}_m\|^2 = \mathbb{E}\|\Pi_{m',m}\mathbf{Y}\|^2 = \text{tr } \mathbb{E}\Pi_{m',m}\mathbf{Y}\mathbf{Y}^T\Pi_{m',m} = \text{tr } \Pi_{m',m}\Sigma, \tag{26}$$

where we have used the facts that $\Pi_{m',m}$ is a projector and Var $\mathbf{Y} =$ Var $\epsilon = \Sigma$. As for the second term we have

$$\mathbb{E}\left[\|\tilde{\mathbf{f}}_{m'} - \mathbf{f}\|^2 - \|\tilde{\mathbf{f}}_m - \mathbf{f}\|^2\right] = \mathbb{E} - \|\Pi_{m',m}\mathbf{f}\|^2 + \|\Pi_{m',m}\epsilon\|^2 = \tag{27}$$

$$-\|\Pi_{m',m}\mathbf{f}\|^2 + \mathbb{E}\text{tr}\Pi\epsilon\epsilon^T\Pi^T = -\|\Pi_{m',m}\mathbf{f}\|^2 + \text{tr}\Pi\Sigma, \tag{28}$$

where $\mathbb{E}\epsilon\epsilon^T = \Sigma$.

■

Problem 1.10. For the quantities $\triangle^+(k,x)$ and $\triangle^-(k,x)$ defined as follows

$$\zeta^+(k, \triangle^+(k,x); x) = 2k, \qquad \zeta^-(k, \triangle^-(k,x); x) = 2k \tag{29}$$

and $x \geq 1$

- Show that $\triangle^+(k,x) < k$ and $\triangle^-(k,x) > k$.
- Check that Lemma 4.2.4 implies

$$\triangle^+(k,x) \geq \zeta^-(k,x) - 2k^{1/2}z_1(x) \tag{30}$$

$$\triangle^-(k,x) \leq \zeta^+(k,x) + 2k^{1/2}z_1(x) \tag{31}$$

Solution. ■