Tsinghua-Berkeley Shenzhen Institute
INFERENCE AND INFORMATION
Fall 2017

Problem Set 2

Issued: Monday 9th October, 2017 **Due:** Monday 16th October, 2017

Notations: x, y are random variables. If they are discrete, $\mathcal{X}$ and $\mathcal{Y}$ are corresponding alphabets. $\underline{\mathsf{x}}, \underline{\mathsf{y}}$ are random vectors, and $\underline{f}(\cdot)$ is a vector valued function.

2.1. Mathematical Expectation and Covariance Matrix.

(a) Prove the following properties.

i. $\mathrm{Cov}(\underline{\mathsf{x}}) = \mathbb{E}[\mathrm{Cov}(\underline{\mathsf{x}}|\mathsf{y})] + \mathrm{Cov}(\mathbb{E}[\underline{\mathsf{x}}|\mathsf{y}])$.

ii. $\det[\mathrm{Cov}(\underline{\mathsf{x}})] = 0 \iff \exists \underline{c} \in \mathbb{R}^k, \underline{c} \neq \underline{0}, d \in \mathbb{R}$, such that $\underline{c}^{\mathrm{T}}\underline{\mathsf{x}} = d$.

(b) *Mathematical expectation in estimation.* Similar to the settings in Problem 1.1, we want to estimate y. Instead of estimating y itself, now our goal is estimating a vector $\underline{\mathsf{y}}$ related to y. $\underline{\mathsf{y}}$ is generated as follows: To simplify the statement, assume $\mathcal{Y} = \{1, 2, \cdots, |\mathcal{Y}|\}$ and let $k = |\mathcal{Y}|$. Then $\underline{\mathsf{y}} \triangleq (\mathbb{1}_{\mathsf{y}=1}, \mathbb{1}_{\mathsf{y}=2}, \cdots, \mathbb{1}_{\mathsf{y}=k})^{\mathrm{T}}$, i.e., $\underline{\mathsf{y}}$ becomes the i-th vector of standard basis if $\mathsf{y} = i$. This transformation from y to $\underline{\mathsf{y}}$ is called one-hot encoding.

Now we would use $\underline{\hat{\mathsf{y}}}$ to estimate $\underline{\mathsf{y}}$, and use its MSE to evaluate the goodness of estimate. The MSE is defined similar to the scalar case, except that the scalar quadratic operator is replaced by the square of ℓ_2 norm:

$$\mathrm{MSE}(\underline{\hat{\mathsf{y}}}) \triangleq \mathbb{E}[\|\underline{\mathsf{y}} - \underline{\hat{\mathsf{y}}}\|_2^2].$$

Same to Problem 1.1, the estimator $\underline{\hat{\mathsf{y}}}$ could be chosen from a certain set $\mathcal{A}$.

i. Assume we want to use a point in the real vector space to estimate y, i.e., $\mathcal{A} = \mathbb{R}^k$. Prove that $\underline{P}_{\mathsf{y}}(\cdot)$ is the MMSE estimator:

$$\underline{P}_{\mathsf{y}}(\cdot) = \arg\min_{\underline{\alpha} \in \mathbb{R}^k} \mathbb{E}[\|\underline{\mathsf{y}} - \underline{\alpha}\|_2^2],$$

where $\underline{P}_{\mathsf{y}}(\cdot) \triangleq [P_{\mathsf{y}}(1), P_{\mathsf{y}}(2), \cdots, P_{\mathsf{y}}(k)]^{\mathrm{T}}$.

ii. Now you are allowed to use a multivariant function of x to estimate $\underline{\hat{\mathsf{y}}}$, i.e., $\mathcal{A} = \{\underline{f} : \mathcal{X} \mapsto \mathbb{R}^k\}$. Prove that the MMSE estimator is $\underline{P}_{\mathsf{y}|\mathsf{x}}(\cdot|\mathsf{x})$:

$$\underline{P}_{\mathsf{y}|\mathsf{x}}(\cdot|\mathsf{x}) = \arg\min_{\underline{f} :\, \mathcal{X} \mapsto \mathbb{R}^k} \mathbb{E}[\|\underline{\mathsf{y}} - \underline{f}(\mathsf{x})\|_2^2],$$

where $\underline{P}_{\mathsf{y}|\mathsf{x}}(\cdot|\mathsf{x}) \triangleq [P_{\mathsf{y}|\mathsf{x}}(1|\mathsf{x}), P_{\mathsf{y}|\mathsf{x}}(2|\mathsf{x}), \cdots, P_{\mathsf{y}|\mathsf{x}}(k|\mathsf{x})]^{\mathrm{T}}$.

2.2. Assume $\underline{\mathsf{x}}$ is a d-dimemsional random vectors with $\mathbb{E}[\underline{\mathsf{x}}] = \underline{0}, \mathrm{Cov}(\underline{\mathsf{x}}) = K_{\mathsf{x}}$ and the matrix K_{x} has eigendecomposition $K_{\mathsf{x}} = U^{\mathrm{T}}\Sigma U$ where $U^{\mathrm{T}}U = I, \Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_d)$, $\sigma_1 > \sigma_2 > \cdots > \sigma_d > 0$.

Let $\underline{v} \in \mathbb{R}^d$ be a unit vector, i.e., $\|\underline{v}\| = 1$. Now we can project $\underline{\mathsf{x}}$ on the direction of $\underline{v}$ to get a new random variable $\underline{v}^{\mathrm{T}}\underline{\mathsf{x}}$. We would like to choose a series of $\{\underline{v}_i\}_{i=1}^d$ that maximizes $\mathrm{Var}(\underline{v}_i^{\mathrm{T}}\underline{\mathsf{x}})$:

- $\underline{v}_1 \in \mathbb{R}^d$ is the $\underline{v}$ that maxmizes $\mathrm{Var}(\underline{v}^{\mathrm{T}}\underline{\mathsf{x}})$:

$$\underline{v}_1 \triangleq \operatorname*{arg\,max}_{\underline{v}:\ \|\underline{v}\|=1} \mathrm{Var}(\underline{v}^{\mathrm{T}}\underline{\mathsf{x}})$$

- $\forall k \geq 2$, $\underline{v}_k \in \mathbb{R}^d$ is the $\underline{v}$ that is orthogonal to all $\underline{v}_i (i = 1, 2, \cdots, k-1)$ and maxmizes $\mathrm{Var}(\underline{v}^{\mathrm{T}}\underline{\mathsf{x}})$:

$$\underline{v}_k \triangleq \operatorname*{arg\,max}_{\substack{\underline{v}:\ \|\underline{v}\|=1,\\ \underline{v}\perp\underline{v}_i (i=1,2,\cdots,k-1)}} \mathrm{Var}(\underline{v}^{\mathrm{T}}\underline{\mathsf{x}})$$

Let y_i denote the corresponding projected random variables, i.e., $\mathsf{y}_i \triangleq \underline{v}_i^{\mathrm{T}}\underline{\mathsf{x}} (i = 1, 2, \cdots, d)$. These projections together form a new random vector $\underline{\mathsf{y}} \triangleq (\mathsf{y}_1, \mathsf{y}_2, \cdots, \mathsf{y}_d)^{\mathrm{T}}$.

(a) Find the expressions of $\underline{v}_i (i = 1, 2, \cdots, d)$.

(b) Evaluate $\mathrm{Cov}(\underline{\mathsf{y}})$.

(c) When $\underline{\mathsf{x}}$ is a Gaussian random vector:

  i. Give the probability density function of $\underline{\mathsf{y}}$.

  ii. Prove that random variables $\mathsf{y}_1, \mathsf{y}_2, \cdots, \mathsf{y}_d$ are independent.

2.3. *Joint Gaussian Distribution.* $\underline{\mathsf{x}} = (\mathsf{x}_1, \mathsf{x}_2)^{\mathrm{T}}$ is a Gaussian random vector and $\mathbb{E}[\mathsf{x}_1] = \mathbb{E}[\mathsf{x}_2] = 0, \mathrm{Var}(\mathsf{x}_1) = \mathrm{Var}(\mathsf{x}_2) = \sigma^2$. Let ρ_{x} denote the correlation coefficient between x_1 and x_2: $\rho_{\mathsf{x}} \triangleq \rho(\mathsf{x}_1, \mathsf{x}_2)$. Let $\underline{\mathsf{y}} = (\mathsf{y}_1, \mathsf{y}_2)^{\mathrm{T}} \triangleq A\underline{\mathsf{x}}$, where

$$A = \begin{bmatrix} 1 & -\rho \\ & 1 \end{bmatrix}.$$

$\underline{\mathsf{y}}$ is also a Gaussian random vector, since it is a linear transformation of $\underline{\mathsf{x}}$.

(a) Calculate $K_{\mathsf{x}} \triangleq \mathrm{Cov}(\underline{\mathsf{x}})$ and $K_{\mathsf{y}} \triangleq \mathrm{Cov}(\underline{\mathsf{y}})$.

(b) Prove that $\rho(\mathsf{y}_1, g(\mathsf{y}_2)) = 0$, for all[1] functions $g(\cdot)$. *Hint:* First prove that $\mathsf{y}_1 \perp\!\!\!\perp \mathsf{y}_2$.

(c) Prove that $\mathbb{E}[\mathsf{y}_1^2] \leq \mathbb{E}[(\mathsf{y}_1 - \rho_{\mathsf{x}}\mathsf{y}_2 + g(\mathsf{y}_2))^2]$, i.e., $\mathbb{E}[(\mathsf{x}_1 - \rho\mathsf{x}_2)^2] \leq \mathbb{E}[(\mathsf{x}_1 - g(\mathsf{x}_2))^2], \forall g : \mathbb{R} \to \mathbb{R}$.

Remaks: From what we have known about MMSE estimation in Problem 1.1.(b), this indicates $\mathbb{E}[\mathsf{x}_1|\mathsf{x}_2] = \rho\mathsf{x}_2$. On the other hand, from Problem 1.2, we know that the linear MMSE estimator that uses x_2 to estimate x_1 is $\rho\mathsf{x}_2$. So for the case of joint Gaussian distribution, when using x_2 to estimate x_1, the MMSE estimator coincide with the linear MMSE estimator.

2.4. This is the placeholder for your 2.4. Problem 2.4 is a coding exercise related to 2.3, and you will need MATLAB to finish this task. It is on the way!

Don't worry, the due date of Problem 2.4 will also be *later* than the above ones. I will give a separate assignment for Problem 2.4 on Web Learning.

[1]Strictly speaking, $g(\cdot)$ is required to be measurable.