Tsinghua-Berkeley Shenzhen Institute
INFERENCE AND INFORMATION
Fall 2017

Problem Set 1

Issued: Saturday 23rd September, 2017 **Due:** Saturday 30th September, 2017

POLICIES

- **Acknowledgments:** We expect you to make an honest effort to solve the problems individually. As we sometimes reuse problem set questions from previous years, covered by papers and web pages, we expect the students not to copy, refer to, or look at the solutions in preparing their answers (relating to an unauthorized material is considered a violation of the honor principle). Similarly, we expect to not to google directly for answers (though you are free to google for knowledge about the topic). If you do happen to use other material, it must be acknowledged here, with a citation on the submitted solution.
- **Required homework submission format:** You can submit homework either as one single pdf document or as handwritten papers. Written homework needs to be provided during the class in due date, and pdf document needs to be submitted through Tsinghua's Web Learning (`http://learn.tsinghua.edu.cn/`) before the end of due date.

  It is encouraged you LaTeX all your work, and we would provide a LaTeX template for your homework.
- **Collaborators:** In a separate section (before your answers), list the names of all people you collaborated with and for which question(s). If you did the HW entirely on your, please state this. Each student must understand, write, and hand in answers of their own.

Notations: x and y are discrete random variables taking values from finite sets $\mathcal{X}$ and $\mathcal{Y}$ respectively. $f(\cdot)$ and $g(\cdot)$ are functions on these two sets: $f : \mathcal{X} \mapsto \mathbb{R}, g : \mathcal{Y} \mapsto \mathbb{R}$. All other random variables in this problem set could be treated as discrete random variables unless otherwise specified.

1.1. Mathematical Expectation and Variance.

(a) Prove the following properties.

i. $\mathbb{E}[\mathsf{x}|\mathsf{y}] = \mathbb{E}[\mathbb{E}[\mathsf{x}|\mathsf{yz}]|\mathsf{y}]$.

ii. $\mathbb{E}[\mathsf{x}g(\mathsf{y})|\mathsf{y}] = g(\mathsf{y})\,\mathbb{E}[\mathsf{x}|\mathsf{y}]$.

iii. $\mathbb{E}[\mathsf{x}\,\mathbb{E}[\mathsf{x}|\mathsf{y}]] = \mathbb{E}[(\mathbb{E}[\mathsf{x}|\mathsf{y}])^2]$.

iv. $\mathrm{Var}(\mathsf{x}) = \mathbb{E}[\mathrm{Var}(\mathsf{x}|\mathsf{y})] + \mathrm{Var}(\mathbb{E}[\mathsf{x}|\mathsf{y}])$.

(b) *Mathematical expectation and variance in estimation.* Suppose we want to estimate the value of y using an estimator $\hat{\mathsf{y}}$, and using its MSE (Mean Square Error) to evaluate the goodness of estimate, defined as

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

The estimator $\hat{\mathsf{y}}$ could be chosen from a certain set $\mathcal{A}$, then our goal is to find the best estimator in $\mathcal{A}$ which has the least MSE. This best estimator is called the MMSE (Minimum Mean Square Error) estimator.

i. Assume we want to use a real number to estimate y, i.e., $\mathcal{A} = \mathbb{R}$.

α) Prove that $\mathbb{E}[\mathsf{y}]$ is the MMSE estimator:

$$\mathbb{E}[\mathsf{y}] = \operatorname*{arg\,min}_{\alpha \in \mathbb{R}} \mathbb{E}[(\mathsf{y}-\alpha)^2].$$

β) Evaluate this estimator's MSE.

ii. Now you are allowed to use a function of x to estimate y, i.e., $\mathcal{A} = \{f(\cdot) : \mathcal{X} \mapsto \mathbb{R}\}$. Prove that:

α) $\mathbb{E}[\mathsf{y}|\mathsf{x}]$ is the MMSE estimator:

$$\mathbb{E}[\mathsf{y}|\mathsf{x}] = \operatorname*{arg\,min}_{f\colon\, \mathcal{X}\mapsto\mathbb{R}} \mathbb{E}[(\mathsf{y}-f(\mathsf{x}))^2],$$

β) The MSE of estimator $\mathbb{E}[\mathsf{y}|\mathsf{x}]$ is

$$\mathrm{MSE}(\mathbb{E}[\mathsf{y}|\mathsf{x}]) = \mathbb{E}[\mathrm{Var}(\mathsf{y}|\mathsf{x})].$$

iii. Comparing these two estimators. Prove that

$$\mathsf{x} \perp\!\!\!\perp y \implies \mathrm{MSE}(\mathbb{E}[\mathsf{y}]) = \mathrm{MSE}(\mathbb{E}[\mathsf{y}|\mathsf{x}]) \implies \forall f,\ \rho(f(\mathsf{x}),\mathsf{y}) = 0,$$

where $\rho(\cdot,\cdot)$ is the Pearson correlation coefficient. Generally, which estimator in these two would have less MSE than the other?

1.2. The Pearson correlation coefficient $\rho(\mathsf{x},\mathsf{y})$ of two random variables x and y is defined as

$$\rho(\mathsf{x},\mathsf{y}) \triangleq \frac{\mathbb{E}[(\mathsf{x}-\mathbb{E}[\mathsf{x}])(\mathsf{y}-\mathbb{E}[\mathsf{y}])]}{\sqrt{\mathrm{Var}(\mathsf{x})\,\mathrm{Var}(\mathsf{y})}}. \tag{1}$$

(a) *Pearson correlation coefficient as the coefficient of linear regression.* When $\mathrm{Var}(\mathsf{x}) = \mathrm{Var}(\mathsf{y})$, prove that $\rho = a^*$ where a^* is the coefficient in linear regression of x and y:

$$(a^*, b^*) \triangleq \operatorname*{arg\,min}_{(a,b)\in\mathbb{R}^2} \mathbb{E}[(\mathsf{y}-a\mathsf{x}-b)^2].$$

(b) Prove that

$$\mathsf{x} \perp\!\!\!\perp \mathsf{y} \iff \forall f, g,\ \rho(f(\mathsf{x}), g(\mathsf{y})) = 0.$$