Tsinghua-Berkeley Shenzhen Institute
INFORMATION THEORY AND STATISTICAL LEARNING
Fall 2020

Problem Set 5

Issued: Monday 30th November, 2020 **Due:** Monday 14th December, 2020

Notations: We use $\mathsf{x}, \mathsf{y}, \mathsf{w}$ and $\underline{\mathsf{x}}, \underline{\mathsf{y}}, \underline{\mathsf{w}}$ to denote random variables and random vectors.

5.1. *Cramer-Rao inequality with a bias term.* Let $\mathsf{y} \sim f(y;x)$ and let $\hat{x}(\mathsf{y})$ be an estimator for x. Let $b(x) = \mathbb{E}[\hat{x}(\mathsf{y})] - x$ be the bias of the estimator. Show that

$$\mathbb{E}\left[(\hat{x}(\mathsf{y}) - x)^2\right] \geq \frac{[1 + b'(x)]^2}{J_{\mathsf{y}}(x)} + b^2(x)$$

5.2. (a) Let

$$p_{\mathsf{y}}(y;x) = \begin{cases} x & \text{if } 0 \leq y \leq 1/x \\ 0 & \text{otherwise} \end{cases}$$

for $x > 0$. Show that there exist no unbiased estimators $\hat{x}(\mathsf{y})$ for x. (Note that because only $x > 0$ are possible values, an unbiased estimator need only be unbiased for $x > 0$ rather than all x.)

(b) Suppose instead that

$$p_{\mathsf{y}}(y;x) = \begin{cases} 1/x & \text{if } 0 \leq y \leq x \\ 0 & \text{otherwise} \end{cases}$$

for $x > 0$. Does a minimum-variance unbiased estimator for x based on y exist? If your answer is yes, determine $\hat{x}_{\mathrm{MVU}}(\mathsf{y})$. If your answer is no, explain.

5.3. Suppose, for $i = 1, 2$

$$\mathsf{y}_i = x + \mathsf{w}_i$$

where x is an unknown but non-zero constant, w_1 and w_2 are statistically independent, zero-mean Gaussian random variables with

$$\mathrm{var}(\mathsf{w}_1) = 1$$
$$\mathrm{var}(\mathsf{w}_2) = \begin{cases} 1 & x > 0 \\ 2 & x < 0 \end{cases}.$$

(a) Calculate the Cramér-Rao bound for unbiased estimators of x based on observation of

$$\underline{\mathsf{y}} = \begin{bmatrix} \mathsf{y}_1 \\ \mathsf{y}_2 \end{bmatrix}.$$

(b) Show that a minimum variance unbiased estimator $\hat{x}_{\mathrm{MVU}}(\underline{\mathsf{y}})$ does not exist.

Hint: Consider the estimators

$$\hat{x}_1(\underline{\mathsf{y}}) = \frac{1}{2}\mathsf{y}_1 + \frac{1}{2}\mathsf{y}_2,$$
$$\hat{x}_2(\underline{\mathsf{y}}) = \frac{2}{3}\mathsf{y}_1 + \frac{1}{3}\mathsf{y}_2.$$

5.4. Let $\underline{\mathsf{y}} = [\mathsf{y}_1\ \mathsf{y}_2]^{\mathrm{T}}$ be a vector random variable whose components are i.i.d. Bernoulli random variables with parameter x, $0 < x < 1$, i.e., $\mathbb{P}(\mathsf{y}_i = 1) = x, i = 1, 2$.

(a) Show that $t(\underline{y}) = y_1 + y_2$ is a sufficient statistic.

(b) Let $\hat{x}(\underline{y}) = y_1$ be an estimator of the parameter x from the observation $\underline{y}$. Find $\mathrm{MSE}_{\hat{x}}(x)$, the mean-square error of this estimator.

(c) Let $\hat{x}'(t) = \mathbb{E}[\hat{x}(\underline{\mathsf{y}})|\mathsf{t} = t]$ be an estimator of the parameter x that uses the sufficient statistic t instead of the observations $\underline{y}$.

i. Show that $\hat{x}'(t)$ is a valid estimator, i.e., it does not depend on x.

ii. Show that $\mathrm{MSE}_{\hat{x}'}(x) = \gamma\, \mathrm{MSE}_{\hat{x}}(x)$ and find the constant γ.

(d) We now consider a generalization of this problem. Let $\underline{\mathsf{y}}$ be a random variable generated by a distribution $p_{\underline{\mathsf{y}}}(\cdot; x)$ and $\underline{t}(\underline{y})$ be a sufficient statistic. Let $\hat{x}(\underline{y})$ be an estimator of the parameter x based on the observation $\underline{y}$. We define an alternate estimator $\hat{x}'(\underline{t}) = \mathbb{E}[\hat{x}(\underline{\mathsf{y}})|\underline{\mathsf{t}} = \underline{t}]$.

i. Show that $\hat{x}'(\underline{t})$ is a valid estimator, i.e., it does not depend on x.

ii. Show that for any cost function $C(x, \hat{x})$ that is convex in $\hat{x}$, the following inequality holds:

$$\mathbb{E}[C(x, \hat{x}'(\underline{\mathsf{t}}))] \leq \mathbb{E}[C(x, \hat{x}(\underline{\mathsf{y}}))].$$

5.5. For a non-bayesian case $p_{\mathsf{y}}(y; x)$, we do a binary hypothesis testing where $x \in \{H_0, H_1\}$. Please prove that $t(y) = \frac{p_{\mathsf{y}}(y;H_1)}{p_{\mathsf{y}}(y;H_0)}$ is a complete sufficient statistics.

5.6. In class we developed the EM algorithm for maximum likelihood estimation (EM-ML). That is, we gave an iterative procedure to compute

$$\hat{x}_{ML}(y) = \arg\max_a p_{\mathsf{y}}(y; a).$$

and showed that the likelihood was non-decreasing with each iteration.
Please develop the EM-MAP algorithm for MAP estimation:

$$\hat{x}_{MAP}(y) = \arg\max_a p_{\mathsf{x}|\mathsf{y}}(a|y)$$

where the complete data z is an arbitrary random vector. (Please follow the procedures in the lecture note)