ECE 275B – Reading & Homework #1 – Due Thurs 2/1/2018

Reading

Kay Chapter 5; Moon Sections 10.5, 10.6, and 12.3.6. Make sure to understand every example. The key ideas are a) *Sufficient Statistics* (especially as a form of data compression and their importance for finding UMVUE's); b) *Exponential Family of distributions* (especially their relationship to finite-dimensional, minimal, sufficient, complete statistics); and c) the *Rao-Blackwell Theorem* as a procedure for finding UMVUE's.

The Exponential Family of Distributions is of key importance. An excellent source of information is the Wikipedia article, Exponential Family. If one does an internet search using the key words "Exponential Family Distribution" many tutorial and summary articles will be found. You should look at these and pick the one(s) that appear most useful and informative for future use.

Also read the excerpt from Steven Pinker's book *How the Mind Works* (Norton, 1997) on the distinction between the frequentist and subjective interpretations of probability.

Homework Time Management

Note that you've been allotted about 3.5 weeks to do the following ***12 problems***. Please begin to think about them and work on them ASAP. If you wait until the last night to do them, you might feel that the homework assignment is too large, even though the number of problems divided by the time allotted to do them is reasonable. *Remember, proper time management is your responsibility.*

Problems

1. Let $\mathbf{x} = (x_1, \cdots, x_n)^T$ be a vector of continuous iid samples. The *order statistics* $\mathbf{x}' = (x'_1, \cdots, x'_n)^T = T(\mathbf{x})$ are defined to be the increasing rearrangement of the elements of $\mathbf{x}$,

$$x'_1 \leq x'_2 \leq \cdots \leq x'_n$$

for $x'_1, \cdots, x'_n$ an appropriate permutation of the elements of $\mathbf{x}$. Because x_i are assumed to be independent realizations of a continuous random variable one can assume that almost surely (i.e., with probability 1),

$$x'_1 < x'_2 < \cdots < x'_n.$$

   (a) Using the definition of sufficiency, prove that T is a sufficient statistics.

   (b) Now prove that T is a sufficient statistic using the Neyman–Fisher Factorization Theorem.

2. For g a statistic (parameter–independent measurable mapping) on a sample space X, for purposes of understanding the statistical properties of sufficiency and minimal sufficiency we treat the statistic g and the partition it induces on the sample space as equivalent.[1]

   As discussed in lecture, the *classical* (Fisherian) *definition of sufficiency* of a statistic T requires that the probability of x conditioned on x being in one of the cosets of the partition induced by T be independent of the unknown parameter vector θ.

   A statistic g (*not* necessarily sufficient) is said to be *necessary* if the partition it induces is coarser than the partition induced by any sufficient statistic. Equivalently, g is necessary iff g is a measurable function of any particular sufficient statistic T, $g = f(T)$, for some g and T dependent measurable function f.

   With the above definitions of sufficiency and minimality, a *minimal sufficient statistic* can then be defined to be a *necessary and sufficient statistic.*

   Consider a statistic which induces a partition (equivalence class) on X such that

   $$x \sim x' \quad \text{iff} \quad \frac{p_\theta(x)}{p_\theta(x')} = g(x, x') = \text{function of } x \text{ and } x' \text{ independent of } \theta.$$

   Prove that this statistic is necessary and sufficient and hence is a minimal sufficient statistic.[2]

3. Consider the pdf's or pmf's listed in problems 5.1, 5.2, 5.3, and 5.14 of Kay and in 10.6-12 of Moon. Place *all* of them into scalar exponential family form and for each one give a sufficient statistic.

4. For a k-parameter exponential family distribution, show that if the elements of the vector of natural statistics[3] together with the function of constant value of one,

   $$\{1, T_1(y), \cdots, T_k(y)\}\ ,$$

   are linearly dependent then the order of the distribution can be reduced below k, showing that the original vector of natural statistics is not minimal. Similarly, if the vector of natural parameters viewed as functions of the parameter vector θ together with the constant function of value one,

   $$\{1, Q_1(\theta), \cdots, Q_k(\theta)\}\ ,$$

   are linearly dependent, again show that the order of the distribution can be reduced.

[1] This corresponds to not distinguishing between statistics on X which induce the same partition of X. Any two such statistics f and g are equivalent, $f \sim g$, as it must be the case that $f(x) = h\,(g(x))$ for all $x \in X$ for some one-to-one mapping $h(\cdot)$.

[2] This proof is essentially done in lecture, so this problem is basically an assignment in directed reading.

[3] Which are necessarily sufficient statistics.

5. Let y and x be jointly random vectors. For notational convenience, define the quantities $\langle y \rangle \triangleq \mathrm{E}\{y\}$ and $\overline{y} \triangleq \mathrm{E}\{y|\, x\}$, and note that $\langle \overline{y} \rangle = \langle y \rangle$.[4] Prove that the covariance matrices for $\overline{y}$ and y satisfy the relationship

$$\mathrm{Cov}\{y\} = \mathrm{Cov}\{\mathrm{E}\{y|\, x\}\} + \mathrm{E}\{\mathrm{Cov}\{y|\, x\}\} = \mathrm{Cov}\{\overline{y}\} + \mathrm{E}\{\mathrm{Cov}\{y|\, x\}\}\,,$$

which, in particular, implies the *important fact* that

$$\mathrm{Cov}\{y\} \geq \mathrm{Cov}\{\overline{y}\}\ .$$

This fact shows that the conditional mean random variable $\overline{y} = \mathrm{E}\{y \mid x\}$ is more tightly distributed around its mean than y. Thus the processing of a *measurement* x to form the conditional mean $\overline{y}$ will generally result in less uncertainty about an *unknown* random variable y, *provided* that x and y are *not* independent. Later, we will see that the conditional mean provides an optimal reduction of the uncertainty in y in the matrix mean-squared error sense.

6. Let the real scalar random variable y be uniformly distributed over the closed interval $[0,\, \theta]$, for unknown θ, $0 < \theta < \infty$. Recall that this pdf does not satisfy the regularity condition allowing for the determination of the Cramér–Rao lower bound.

   (a) Given iid measurement samples, $y_1, \cdots, y_m$ drawn from this distribution, find a sufficient statistic for θ and prove that it is complete.

   (b) Find the MVUE of θ.

   (c) Find the mse of the MVUE and compare the to the mse of the BLUE (use your result from last quarter).

7. Kay 5.13

8. Moon 10.5.8

9. Kay 5.15

10. Suppose that the m–dimensional random vector $\mathbf{y}$ obeys $\mathbf{y} \sim N(\mathbf{A}\boldsymbol{\theta}, \mathbf{R})$ for known matrices $\mathbf{A}$ and $\mathbf{R}$ and unknown k–dimensional parameter vector $\boldsymbol{\theta}$. Assume that $\mathbf{R}$ is full rank and that $\mathbf{A}$ is full column rank.

    (a) Show that $T(\mathbf{y}) = \mathbf{A}^T\mathbf{R}^{-1}\,\mathbf{y}$ is a minimal complete sufficient statistic for $\boldsymbol{\theta}$. Give the dimension of $T(\mathbf{y})$ and explain why it is no larger, and in general is smaller, than the dimension of the "raw" data $\mathbf{y}$.

    (b) Use the RBLS theorem to find the UMVU of $\boldsymbol{\theta}$.

[4] I.e., $\overline{y}$ is an unbiased estimator of y.

11. Note that scalar measurements

$$y[\ell] = \mathbf{r}_\ell \, \boldsymbol{\theta} + v[\ell], \quad \ell = 1, 2, \cdots ,$$

for an independent noise sequence $v[\ell] \sim N(0, \sigma_\ell)$ and $\boldsymbol{\theta}$ n–dimensional, can be written as

$$\mathbf{y}_k = \mathbf{A}_k \, \boldsymbol{\theta} + \mathbf{v}_k$$

where

$$\mathbf{y}_k = \begin{pmatrix} \mathbf{y}_{k-1} \\ y[k] \end{pmatrix}, \quad \mathbf{v}_k = \begin{pmatrix} \mathbf{v}_{k-1} \\ v[k] \end{pmatrix}, \quad \mathbf{A}_k = \begin{pmatrix} \mathbf{A}_{k-1} \\ \mathbf{r}_k \end{pmatrix}.$$

Note that the column dimensions of these quantities grow linearly with the number of measurements. Assume in the sequel that $k \geq n$ and that the corresponding $\mathbf{A}_k$ has full column rank.

Show that the minimal sufficient statistic derived in the previous problem can be updated recursively as

$$T(\mathbf{y}_k) = T(\mathbf{y}_{k-1}) + \mathbf{b}_k \, y[k]$$

and give the functional form of $\mathbf{b}_k$. Also give the dimensions of $T(\mathbf{y}_k)$ and $\mathbf{b}_k$, noting that the minimal sufficient statistic $T(\mathbf{y}_k)$ and the "gain" $\mathbf{b}_k$ are *constant* in dimension and *do not* increase in size as k increases.

12. Kay 7.7