Department of Signal Theory and Communications **06/27/07**
UCIIIM

DIGITAL SIGNAL PROCESSING

THEORY
(Time: 70 minutes. Mark: 3/8)

T1.- An order N Erlang probability density function is characterized by the following expression:

$$p(x) = \frac{a^N x^{N-1} \exp(-ax)}{(N-1)!} \quad x > 0 \quad \text{and} \quad a > 0$$

Assume N is known. Taking into account that the mean of this distribution is given by $m = N/a$:

a) Obtain the ML estimator of the mean, $\hat{m}_{ML}$, using K independent observations of the random variable .

b) Calculate the bias of $\hat{m}_{ML}$.

c) Is $\hat{m}_{ML}$ consistent in variance?

(20 min; 1p)

T2.-

a) The histogram, K-NN and Parzen windows with rectangular kernel are non-parametric estimators that admit the following expression for the estimated pdf:

$$\hat{p}(x) = \frac{k}{KV_R}$$

K being the total number of samples and k the number of samples inside a volume V_R. Explain why, in spite of admitting a common expression, these three estimators are different.

b) Check if an 1-NN estimator verifies / does not verify the property

$$\int_{-\infty}^{\infty} \hat{p}(x)\,dx = 1$$

considering the case in which there is only one sample available, $x^{(1)}$.

(20 min; 1 p)

T3.- Consider a regularized Newton's algorithm characterized by the following update rule:

$$\mathbf{w}^{(k+1)} = \mathbf{w}^{(k)} - \left[\varepsilon\mathbf{I} + \mathbf{H}_w\left(\mathbf{w}^{(k)}\right)\right]^{-1}\mathbf{g}_w\left(\mathbf{w}^{(k)}\right) \qquad (1)$$

Similarly to the sequential (stochastic) gradient algorithm, it is also possible to derive a sequential Netwon's algorithm by replacing in (1) the Hessian and the gradient, $\mathbf{H}_\mathbf{w}(w^{(k)})$ and $\mathbf{g}_\mathbf{w}(w^{(k)})$, respectively, by their corresponding approximations calculated exclusively from the k-th pattern of the training set.

For a network which obtains its output as $y^{(k)} = w_0 + \mathbf{w}^T\mathbf{x}^{(k)} = \mathbf{w}_e^T\mathbf{x}_e^{(k)}$, and assuming a square cost function:

a) Show that the update rule of Newton's sequential algorithm is

$$\mathbf{w}_e^{(k+1)} = \mathbf{w}_e^{(k)} + (d^{(k)} - y^{(k)})\left(\varepsilon\mathbf{I} + \mathbf{x}_e^{(k)}\mathbf{x}_e^{(k)T}\right)^{-1}\mathbf{x}_e^{(k)} \qquad (2)$$

b) Using the following inversion Lemma (Woodbury identity):

$$\left(\mathbf{A} + \mathbf{c}\mathbf{c}^T\right)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{c}\mathbf{c}^T\mathbf{A}^{-1}}{1 + \mathbf{c}^T\mathbf{A}^{-1}\mathbf{c}}$$

where **c** is a column vector and **A** is a matrix of compatible dimensions, show that (2) is equivalent to the NLMS algorithm.

c) From a practical point of view, what advantages offer the NLMS algorithm with respect to the direct use of (2)? And with respect to standard (non-normalized) LMS?

(30 min; 1 p)

Department of Signal Theory and Communications **06/27/07**
UCIIIM

DIGITAL SIGNAL PROCESSING

PROBLEMS

(Time: 150 minutes. Mark: 5/8)

Please, do not write in the areas inside wide lines

Nº	
1	
2	

Surname
Name
ID Number Group 95
Signature:

P1.- An analyst designs a linear estimator of a random variable s as a function of an observation x using a set of labeled data, $\{s^{(k)}, x^{(k)}\}$, by replacing in the standard formulation of the minimum mean square error estimator the real statistics, ($E\{s\}, E\{x\}, v_{xx}, v_{sx}, v_{ss}$), by their sample estimations ($\bar{s}, \bar{x}, \bar{v}_{xx}, \bar{v}_{sx}, \bar{v}_{ss}$, respectively), which are considered constant. The analyst ignores that pairs (s,x) follow a distribution $G\left\{\mathbf{0}, \begin{bmatrix} v & \rho v \\ \rho v & v \end{bmatrix}\right\}$.

a) Obtain the estimator designed by the analyst, $\hat{s}_a(x)$, and compare its theoretical mean square error (which would result from its repeated application to new samples) to the one that would be achieved by the linear minimum mean square error estimator (i.e., the theoretically optimum linear estimator), $\hat{s}_t(x)$.

b) Afterwards, the analyst discovers, from physical considerations, that s and x have zero mean; consequently he designs a new estimator, $\hat{s}'_a(x)$, with the previous sample covariance estimations, and zeros means. Which are the differences between $\hat{s}'_a(x)$ and $\hat{s}_a(x)$?

c) Determine the mean square error reduction provided by $\hat{s}'_a(x)$ over $\hat{s}_a(x)$.

d) Would it be possible for the analyst to perceive the advantage calculated in c) if he replaced the real statistics by their sample estimations in the expressions for the mean square errors?

(75 min; 2.5 p)

P2.- Consider two independent and identically distributed random variables, x_0 and x_1, which follow a continuous pdf p(x) with mean m and median M. After observing the value of x_0, and without knowing x_1, one of the two following hypotheses must be chosen:

$$H_0: x_0 > x_1$$
$$H_1: x_1 \geq x_0$$

a) Design the decider with minimum error probability (MAP).

b) Which is the error probability of the decider obtained in a)?

c) Assume now the following costs associated to the decisions:

$$C_{00} = C_{01} = x_1$$
$$C_{10} = C_{11} = x_0$$

(note that these costs are independent of the hypotheses, but they depend on the observations)

Obtain the decider which minimizes the mean cost (for a given x_0).

d) Give an expression for the global mean cost of the decider designed in the previous item.

(75 min; 2.5 p)