Exercises (5)

1. Given is a random vector \( Y = [Y_1, Y_2, \ldots, Y_N]^T \). We want to predict a random variable \( X \) using \( Y \). The random variable \( X \) and the random vector \( Y \) are not independent.

(a) Proof the iterative expectation rule: \( E\{E\{X|Y\}\} = E\{X\} \)

(b) In the following, it is proofed that the optimal predictor \( \hat{X} = A(Y) \) for minimizing the mean square prediction error is given by the conditional mean \( E\{X|Y\} \). Explain all steps (i to ix) of the proof.

Proof:

\[
\begin{align*}
\varepsilon^2 &= E \left\{ (X - \hat{X})^2 \right\} \\
(i) &= E \left\{ (X - A(Y))^2 \right\} \\
(ii) &= E \left\{ (X - E\{X|Y\} + E\{X|Y\} - A(Y))^2 \right\} \\
(iii) &= E \left\{ (X - E\{X|Y\})^2 \right\} + (E\{X|Y\} - A(Y))^2 \\
& \quad + 2 E \left\{ (X - E\{X|Y\})(E\{X|Y\} - A(Y)) \right\} \\
(iv) &= E \left\{ (X - E\{X|Y\})^2 \right\} + (E\{X|Y\} - A(Y))^2 \\
& \quad + 2 E \left\{ (E\{X|Y\} - A(Y)) (E\{X|Y\} - A(Y)) | Y \right\} \\
(v) &= E \left\{ (X - E\{X|Y\})^2 \right\} + (E\{X|Y\} - A(Y))^2 \\
& \quad + 2 E \left\{ (E\{X|Y\} - A(Y)) E \left\{ X - E\{X|Y\} | Y \right\} \right\} \\
(vi) &= E \left\{ (X - E\{X|Y\})^2 \right\} + (E\{X|Y\} - A(Y))^2 \\
& \quad + 2 E \left\{ (E\{X|Y\} - A(Y)) (E\{X|Y\} - E\{X|Y\}) \right\} \\
(vii) &= E \left\{ (X - E\{X|Y\})^2 \right\} + (E\{X|Y\} - A(Y))^2 \\
(viii) \geq E \left\{ (X - E\{X|Y\})^2 \right\} \\
(ix) \quad \Rightarrow \text{optimal predictor } \hat{X} = A(Y) = E\{X|Y\} \\
\end{align*}
\]

(c) Given are two random variables \( X \) and \( Y \) with means \( \mu_X \) and \( \mu_Y \), respectively, and variances \( \sigma^2_X \) and \( \sigma^2_Y \), respectively. The joint distribution of the random variables is Gaussian. The correlation coefficient between the random variables is \( \rho \).

Derive the optimal predictor for predicting \( X \) given \( Y \).
2. Given is a stationary random process $S = \{S_n\}$. We consider linear and affine prediction of a random variable $S_n$ given the $N$ preceding random variables $S_{n-1} = [S_{n-1}, S_{n-2}, \cdots, S_{n-N}]^T$.

Derive all formulas (as given below) as function of the mean $\mu_s$, the variance $\sigma^2_S$, the $N$-th order autocovariance matrix $C_N$ and the autocovariance vector $c_1 = E\{(S_n - \mu_S)(S_{n-1} - \mu_S e_N)\}$, where $e_N$ is a $N$-dimensional vector with all entries equal to 1.

(a) Derive the affine predictor that minimizes the variance of the prediction error.

Derive expressions for the mean and the variance of the resulting prediction error as well as for the mean squared error.

(b) Derive the affine predictor that minimizes the mean squared prediction error.

Derive expressions for the mean and the variance of the resulting prediction error as well as for the mean squared error.

(c) Derive the linear predictor that minimizes the variance of the prediction error.

Derive expressions for the mean and the variance of the resulting prediction error as well as for the mean squared error.

(d) Derive the linear predictor that minimizes the mean squared prediction error.

Derive expressions for the mean and the variance of the resulting prediction error as well as for the mean squared error.

(e) Consider a stationary random process $S = \{S_n\}$ with mean $\mu_S$, variance $\sigma^2_S$ and the first-order correlation coefficient $\rho$.

A random variable $S_n$ shall be predicted by linear prediction using the previous random variable $S_{n-1}$.

Derive the prediction coefficient $h$, the mean of the prediction error $\mu_U$, the variance of the prediction error $\sigma^2_U$, and the mean squared prediction error $\varepsilon^2_U$ for the linear predictor that minimizes the variance $\sigma^2_U$. 

3. In image and video coding, a sample $S_n$ is often predicted by directly using a previous sample $S_{n-1}$, i.e., by $\hat{S}_n = S_{n-1}$.

Consider a stationary process $\mathbf{S} = \{S_n\}$ with the first-order correlation factor $\rho$.

(a) Assume that the process $\mathbf{S}$ has a mean of zero.

For what correlation factors $\rho$ do we observe a prediction gain (the mean squared prediction error is smaller than the second moment of the input)?

For what correlation factors is the loss versus optimal linear prediction smaller than 0.1 dB?

(b) Now assume that the mean of $\mathbf{S}$ is unequal to 0.

Is the loss versus optimal linear prediction greater than, less than, or equal to the loss for a process with zero-mean but the same variance $\sigma_S^2$ and correlation factor $\rho$? Does it depend on $\mu_S$, $\sigma_S^2$, and $\rho$?

4. Consider prediction in images. Assume that an image can be considered as a realization of a stationary 2-d process with mean $\mu_S$ and variance $\sigma_S^2$.

We want to linearly predict a current sample based on up to three (already coded) neighbouring samples: the sample left of the current sample, the sample above the current sample, and the sample to the top-left of the current sample.

The correlation factor between two horizontally adjacent samples is $\rho_H$, the correlation factor between two vertically adjacent samples is $\rho_V$, and the correlation factor between two diagonally adjacent samples is $\rho_D$ (same in both directions).

The goal is to design linear predictors that minimize the mean squared prediction error. The mean $\mu_S$ is subtracted before doing the prediction.

(a) Assume that $\rho_H > \rho_V$.

Compare optimal linear prediction using only the horizontally adjacent sample and optimal linear prediction using both the horizontally and the vertically adjacent sample.

Under which circumstances is the prediction using both samples better than the prediction using only the horizontally adjacent sample?

(b) Consider the special case $\rho_H = \rho_V = \rho$ and $\rho_D = \rho^2$.

Derive the prediction gain $g = \sigma_S^2/\epsilon^2$ for the optimal vertical predictors using

- the sample to the left
- the sample to the left and the sample above
- the sample to the left, the sample above, and the sample to the top-left

What are the prediction gains in dB for $\rho = 0.95$?
5. Given is a stationary AR(2) process

\[ S_n = Z_n + \alpha_1 \cdot S_{n-1} + \alpha_2 \cdot S_{n-2} \]

where \( \{Z_n\} \) represents zero-mean white noise.

The AR parameters are \( \alpha_1 = 0.7 \) and \( \alpha_2 = 0.2 \).

(a) Determine the correlation factors \( \rho_1 \) and \( \rho_2 \), where \( \rho_1 \) is the correlation factor between adjacent samples \( S_n \) and \( S_{n-1} \), and \( \rho_2 \) is the correlation factor between samples \( S_n \) and \( S_{n-2} \) that are two sampling intervals apart.

(b) Derive the optimal linear predictor (minimizing the MSE) using the 2 previous samples.

Determine the prediction gain in dB.

(c) Derive the optimal linear predictor (minimizing the MSE) using only the directly preceding sample.

What is the prediction gain in dB?

What is the loss relative to an optimal prediction using the last two samples?

(d) Can the linear predictor using the directly preceding sample, given by

\[ U_n = S_n - \rho_1 \cdot S_{n-1}. \]

be improved by adding a second prediction stage

\[ V_n = U_n - h \cdot U_{n-1}. \]

What is the optimal linear predictor for the second prediction stage?

What is the prediction gain achieved by the second prediction stage?

How big is the loss versus optimal linear prediction using the last two samples?

6. Consider a stationary process \( S = \{S_n\} \) with zero-mean, variance \( \sigma^2_S \), the \( N \)-th order autocovariance matrix \( C_N \), and the autocovariance vector \( c_1 = E[S_nS_{n-1}] \) with \( S_{n-1} = [S_{n-1} \cdots S_{n-N}]^T \).

The random variables \( S_n \) are not directly accessible. Only a noisy measurement \( X_n = S_n + Z_n \) can be observed, where \( Z = \{Z_n\} \) is white noise with zero mean and the variance \( \sigma^2_Z \).

(a) Derive the linear predictor that yields the minimum MSE for predicting \( X_n \) given the \( N \) previous samples \( X_{n-1} = [X_{n-1} \cdots X_{n-N}]^T \).

(b) Consider the special case \( N = 1 \) with the first-order correlation coefficient \( \rho \).

Let \( \rho \) be unequal to zero and \( \sigma^2_S > 0 \). For which variance ratios \( \gamma = \sigma^2_Z / \sigma^2_S \) does the prediction reduce the MSE (relative to using the samples \( X_n \) directly)?

What is the loss in prediction gain (in dB) relative to a prediction using noise-free data if \( \gamma = 0.1 \) and \( \rho = 0.9 \)?
7. Consider a zero-mean Gauss-Markov process with the correlation factor \( \rho = 0.9 \).

The Gauss-Markov source is coded using DPCM at high rates. The quantizer is an entropy-constrained Lloyd quantizer with optimal entropy coding.

(a) Neglect the quantization and derive the optimal linear predictor (minimizing the MSE) using the previous sample. Determine the prediction gain.

(b) Use the predictor derived in (7a) inside the DPCM loop. Assume that the prediction error has a Gaussian distribution. What is the approximate coding gain compared to ECSQ without prediction at the rates \( R_1 = 1 \) bit per sample, \( R_2 = 2 \) bit per sample, \( R_3 = 3 \) bit per sample, \( R_4 = 4 \) bit per sample, and \( R_5 = 8 \) bit per sample?