Exam: Stochastic Processes

Date and Time: Thursday January 7, 2016, 9:00–13:00.

This entire problem set contains **4 pages**. Please make sure that you have received all pages.

The exam is graded according to your answer as a whole: both quantity and quality count. We value concise arguments showing your command of the topics. Simply answering "Yes." or "No." will not do that!

We recommend that you read through each problem thoroughly before starting to solve it. Should you happen to get stuck at some point, we recommend that you continue and anyway try to solve the rest. You always have the opportunity to sketch or explain how you would have continued if you hadn't got stuck.

It is allowed to use books, lecture notes, your own notes, calculators and computers during the exam. Communication to others during the exam is not allowed—therefore, the use of internet is strictly forbidden.

Problem 1:

Consider the stochastic process $Y(n)$ defined as

$$Y(n) = Z(n) + aZ(n-1), \qquad n = 1, 2, \dots \tag{1}$$

where $a \in \mathbb{R}$ is a known constant and $Z(n)$ is a random process with unknown stationarity properties. Its mean function is $\mathbb{E}[Z(n)] = \mu_Z(n)$, and its autocorrelation function (ACF) is $\mathbb{E}[Z(n_1)Z(n_2)] = R_Z(n_1, n_2)$.

1.1 Find the mean function $\mu_Y(n)$ and the ACF $R_Y(n_1, n_2)$ of the process $Y(n)$.

1.2 In general, what conditions should a stochastic process fulfill to be wide-sense stationary (WSS)?

1.3 If we want $Y(n)$ to be WSS, what properties should the process $Z(n)$ have?

From this point on, assume that the process $Z(n)$ in (1) is defined as $Z(n) \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 2)$.

1.4 Write up the expression of the ACF and power spectral density (PSD) of $Z(n)$. Make a sketch of both functions.

1.5 The process $Y(n)$ can be seen as the output of a linear time-invariant (LTI) system that has $Z(n)$ as input. Find the impulse response $h(n)$ and the transfer function $H(f)$ of this LTI system.

1.6 Compute the ACF and PSD of $Y(n)$ and sketch them.

Figure 1 below shows 3 scatter plots of multiple realizations of the first two samples of $Y(n)$. Each plot has been generated with a different value of the parameter a. More specifically, the scatter plots correspond to realizations generated with the values $a = -1, 0, 1$, although not necessarily in this order.

1.7 Write your best guess about which value of the parameter a has been used for each of the plots, and a short justification of why.

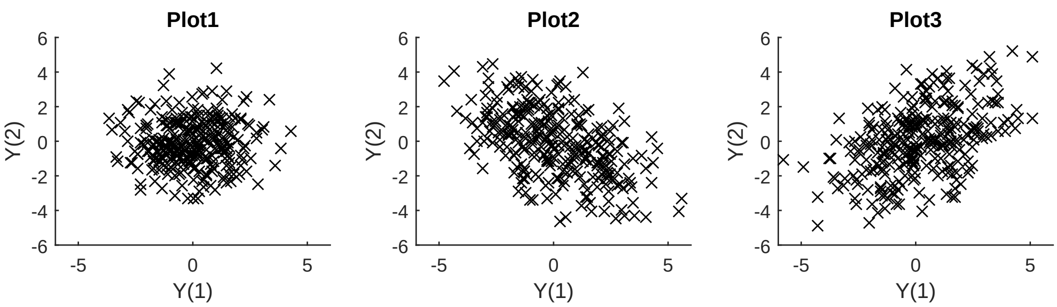

Figure 1

Problem 2:

The digital signal observed at the output of an electronic sensor is modeled as

$$X(n) = Ar^n + W(n), \qquad n = 1, 2, \ldots, N \tag{2}$$

where A is the physical quantity being measured by the sensor, whose value is unknown. The constant r models the response of the sensor, and it is know from the sensor specifications. Unfortunately, some thermal noise $W(n)$ is impairing the outcome, and makes it difficult to read the precise value of the measured quantity A.

We would like to find an estimator of the unknown quantity A. To do so, A is assumed to be uniformly distributed between $-a$ and a, i.e. $A \sim \mathcal{U}(-a, a)$. In addition, the samples of $W(n)$ are modeled as uncorrelated Gaussian noise samples with variance σ_w^2 ($W(n) \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_w^2)$), and are independent of A. We decide to use an estimator of the form:

$$\hat{A} = h_0 + \boldsymbol{h}^{\mathrm{T}}\boldsymbol{X} \tag{3}$$

where $\boldsymbol{h} = [h_1, h_2, \ldots, h_N]^{\mathrm{T}}$ and $\boldsymbol{X} = [X(1), X(2), \ldots, X(N)]^{\mathrm{T}}$.

2.1 Find the expression for the parameters h_0 and $\boldsymbol{h}$ that yield the lowest mean square error (MSE) achievable with an estimator of this form.

2.2 Would it be possible to find a different estimator with lower MSE? If yes, explain shortly how it can be calculated.

2.3 Calculate the estimate $\hat{A}$ when we have only two observations (N = 2). Assume the following values: $X(1) = 0.65$, $X(2) = 0.2$, $r = 1/2$, $a = 2\sqrt{3}$, and $\sigma_w^2 = 0.1$. *Hint:* The inverse of any invertible 2×2 matrix is given by

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

2.4 For the values above, calculate the expected mean squared error of $\hat{A}$.

Next, assume that we have the same observation model as in (2). However, A is this time an unknown, binary variable that can only take values $A = 1$ or $A = 0$. Based on the observations (2), we would like to decide which of the two values is the true one.

2.5 Would you still use the estimator in (3)? If not, what type of tools would you use to solve this problem?

Problem 3:

A shipping company that offers parcel delivery services is interested in predicting the activity and revenues that they experience during the 24 days preceding the winter holidays. This is a particularly busy (and profitable) period for the company and they want to make sure they have enough personnel and transport means to cope with the increased demand.

Looking at data from the previous years, they have experienced that the amount of parcels they process during this period increases linearly with time, and that the weight of each of the parcels appears to be exponentially distributed. Given this, they have developed a stochastic model of the time of arrival and weight of each of the parcels during the 24 day period. They model them as a two-dimensional Poisson point process with intensity function

$$\varrho(x, y) = \begin{cases} 3x\exp(-y/2), & 0 \leq x \leq 24, y \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

The dimension x indicates the time of arrival (in days) of each parcel, while the dimension y represents the weight (in kg) of the parcel.

3.1 Make a qualitative sketch of how a realization of this Poisson point process may look like. Indicate in which region of the 2D plane you expect the density of points to be largest.

3.2 The total number of parcels processed in the period between two arbitrary days x_1 and x_2, $0 \leq x_i \leq 24$, can be expressed as the region count of the process over an appropriate region B. Make a sketch of this region and calculate the intensity measure over the region and the probability mass function of its region count.

3.3 If no parcels are received during the first day, does this fact influence the probability that some parcel will be received during the second day? Why? (Hint: think of how region counts in a Poisson point process are statistically related).

During this period of the year, the company has a simplified billing scheme, in which the amount billed for each parcel depends on the parcel's weight. Parcel's weighing less than 2 kg are billed for 200 DKK each, while parcels exceeding that weight are billed for 400 DKK. With this, the total amount billed by the company during the 24 days can be expressed as

$$A = \sum_{(x,y)\in \boldsymbol{X}} f(x, y)$$

where $\boldsymbol{X}$ denotes the point process, and

$$f(x, y) = \begin{cases} 200, & 0 \leq y \leq 2, x \geq 0 \\ 400, & y \geq 2, x \geq 0 \end{cases}.$$

3.4 Use Campbell's theorem to calculate the expected value of the total billing amount A. How much of the expected billing is due to the parcels heavier than 2 kg?