Module 9 - Homework 9

Colt Thomas

April 15, 2021

Problem 1

In the fair control experiment, we define the process $X(t)$ as follows:

$$X(t) = \begin{cases} cos(\pi t) & \text{if heads show} \\ 2t & \text{if tails show} \end{cases}$$

Part (A)

Find $E[X(t)]$

For this process, we can find the expected value as follows:

$$E[X(t)] = \sum_{x:p_{X(t)} \geq 0} x p_{X(t)}(x)$$

$$E[X(t)] = cos(\pi t) p_{X(t)}(x) + 2t p_{X(t)}(x)$$

The PDF of a fair coin (note that it is independent of time t) is:

$$p_X(t) = \begin{cases} 1/2 & \text{heads} \\ 1/2 & \text{tails} \end{cases}$$

We end up getting the following expected value for the process:

$$E[X(t)] = cos(\pi t)(1/2) + 2t(1/2)$$

$$\boxed{E[X(t)] = \frac{cos(\pi t)}{2} + t}$$

Part (B)

Find $F_{X(t)}(x)$ for $t = 0.25$,$t = 0.5$, and $t = 1$

For t = 0.25:

$$X(0.25) = \begin{cases} \frac{\sqrt{2}}{2} & \text{if heads show} \\ \frac{1}{2} & \text{if tails show} \end{cases}$$

For t = 0.5

$$X(0.50) = \begin{cases} 0 & \text{if heads show} \\ 1 & \text{if tails show} \end{cases}$$

For t = 1

$$X(1) = \begin{cases} -1 & \text{if heads show} \\ 2 & \text{if tails show} \end{cases}$$

Finding the CDF for any $t_0 \in I$ can be done as follows:

$$F_{X(t_0)} = P(X(t_0) \leq x)$$

Lucky for us, a coin toss is just a Bernoulli random variable. Let $x_1, x_2 \in X$ and $x_1 < x_2$, which are outcomes of $X(t)$:

$$F_{X(t_0)}(x) = (1-p)u(x_1) + pu(x_2 - 1)$$

$$F_{X(t_0)}(x) = \frac{1}{2}u(x) + \frac{1}{2}u(x-1)$$

This can alternatively be represented as:

$$F_{X(t_0)}(x) = \begin{cases} 0 & x < x_1 \\ \frac{1}{2} & x_1 < x < x_2 \\ 1 & x > x_2 \end{cases}$$

And when we plug in t, we get the following CDFs:

$$\boxed{F_{X(0.25)}(x) = \begin{cases} 0 & x < \frac{1}{2} \\ \frac{1}{2} & 1/2 < x < \sqrt{2}/2 \\ 1 & x > \sqrt{2}/2 \end{cases}}$$

$$\boxed{F_{X(0.5)}(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2} & 0 < x < 1 \\ 1 & x > 1 \end{cases}}$$

$$\boxed{F_{X(1)}(x) = \begin{cases} 0 & x < -1 \\ \frac{1}{2} & -1 < x < 2 \\ 1 & x > 2 \end{cases}}$$

Problem 2

Let $X(t), t \in R$ be a continuous-time random process defined below where $A_0, A_1, ..., A_n$ are IID Gaussian normal random variables with mean 0 and varaince 10 and n is a fixed positive integer.

$$X(t) = \sum_{k=0}^{n} A_k t^k$$

Part (A)

Find the mean function $\mu_x(t)$

$$\mu_X(t) = E[X(t)] = E\left[\sum_{k=0}^{n} A_k t^k\right]$$

$$\mu_X(t) = \sum_{k=0}^{n} E[A_k] t^k$$

Remember that the expected value of a Gaussian is just it's mean:

$$\mu_X(t) = \sum_{k=0}^{n} (0) t^k$$

$$\boxed{\mu_X(t) = 0}$$

Part (B)

Find the correlation function $R_X(t_1, t_2)$

We can find the autocorrelation as follows:

$$R_X(t_1, t_2) = E[X(t_1)X(t_2)]$$

$$R_X(t_1, t_2) = E\left[\sum_{k=0}^{n} A_k t_1^k \sum_{j=0}^{n} A_k t_2^j\right]$$

$$R_X(t_1, t_2) = E\left[\sum_{k=0}^{n} A_k t_1^k \sum_{j=0}^{n} A_j t_2^j\right]$$

$$R_X(t_1, t_2) = \sum_{k=0}^{n} \sum_{j=0}^{n} E[A_k A_j] t_1^k t_2^j$$

Remember that all A_n are IID

$$R_X(t_1, t_2) = \sum_{k=0}^{n} \sum_{j=0}^{n} E[A_k] E[A_j] t_1^k t_2^j$$

$$R_X(t_1, t_2) = \sum_{k=0}^{n} \sum_{j=0}^{n} (0)(0) t_1^k t_2^j$$

$$\boxed{R_X(t_1, t_2) = 0}$$

Part (C)

Is $X(t)$ a wide-sense stationary process?

A process is WSS if these two conditions are met:

$$\mu_X(t) = E[X(t)] = \mu$$
$$C_X(t_1, t_2) = C_X(t_2 - t_1)$$

The autocovariance for our process is

$$C_X(t_1, t_2) = R_X(t_1, t_2) - \mu_X(t_1)\mu_X(t_2) = 0$$

The second condition does not hold, so our process is not WSS.

Part (D)

Find $P(X(1) < 1)$. Assume $n = 10$.

First find $X(1)$

$$X(1) = \sum_{k=0}^{10} A_k (1)^k$$

Because $X(1)$ is a sum of IID Gaussian RVs,

$$E[X(1)] = E[X_0 + ... + X_{10}] = E[X_1] + ... + E[X_{10}] = 0$$

$$Var(X(1)) = Var(X_0 + ...X_{10}) = Var(X_1) + ... + Var(X_{10}) = 110$$

Which means that:

$$X(1) \sim N(0, 110)$$

$$P(X(1) < 1) = P\left(X(1) < \frac{1-0}{110}\right) = \Phi(1/110) = \boxed{0.5036}$$

Part (E)

Is X(t) a Gaussian process?

A random process $X(t)$ is a Gaussian random process if the samples $X_i = X(t_i)$ for i = 1,2,...,k are jointly Gaussian random variables for all k and all choices of $t_1, t_2, ..., t_k$.

Let's put $X(t)$ into a vector form:

$$X(t) = \sum_{j=0}^{n} A_j t^j$$

$$\begin{bmatrix} X(t_1) \\ ... \\ X(t_k) \end{bmatrix} = \begin{bmatrix} t_1^0 & ... & t_1^n \\ .. & ... & ... \\ t_k^0 & ... & t^n \end{bmatrix} \begin{bmatrix} A_0 \\ ... \\ A_k \end{bmatrix}$$

Notice that regardless of the time index, we still end up with joint Gaussian random variables as indicated by the A_n vector component. $X(t)$ is a Gaussian random process.

Problem 3

Let X(t) be a wide-sense stationary Gaussian random process with $\mu_X(t) = 2$ *and* $R_X(\tau) = 2 + 4sinc(\tau)$

Part (A)

Find $P(1 < X(3) < 2, X(7) < 3)$

Because X(t) is WSS, X(3) and X(7) are Normal RVs. Below are some useful identities to help us solve this problem:

$$E[X(t)] = \mu_X(t)$$
$$Var(X(t)) = C_X(t,t) = R_X(t,t) - \mu_X(t)\mu_X(t)$$
$$P(x_1 < X < x_2, y_1 < Y < y_2) = F_{X,Y}(x_2, y_2) - F_{X,Y}(x_1, y_2) - F_{X,Y}(x_2, y_1) + F_{X,Y}(x_1, y_1)$$
$$F_{XY}(x, -\infty) = F_X(x)$$

For our problem,

$$E[X(3)] = \mu_X(3) = 2,$$
$$E[X(7)] = \mu_X(7) = 2,$$

$$Var(X(3)) = C_X(3,3) = R_X(3,3) - \mu_X(3)\mu_X(3)$$
$$Var(X(3)) = R_X(3-3) - \mu_X(3)\mu_X(3)$$
$$Var(X(3)) = 2 + 4sinc(0) - 2 * 2 = 2,$$

$$Var(X(7)) = 2 + 4sinc(0) - 2 * 2 = 2$$

$$\text{Therefore: } X(3), X(7) \sim N(2,2)$$

$$P(1 < X(3) < 2, X(7) < 3) = F_{X,Y}(2,3) - F_{X,Y}(1,3) - F_{X,Y}(2,-\infty) + F_{X,Y}(1,-\infty)$$

$$= F_{X(3)}(2)F_{X(7)}(3) - F_{X(3)}(1)F_{X(7)}(3) - F_{X(3)}(2)F_{X(7)}(-\infty) + F_{X(3)}(1)F_{X(7)}(-\infty)$$
$$= \Phi(\frac{2 - \mu_{X(3)}}{\sigma_{X(3)}})\Phi(\frac{3 - \mu_{X(7)}}{\sigma_{X(7)}}) - \Phi(\frac{1 - \mu_{X(3)}}{\sigma_{X(3)}})\Phi(\frac{3 - \mu_{X(7)}}{\sigma_{X(7)}})$$

$$= \Phi(\frac{2-2}{2})\Phi(\frac{3-2}{2}) - \Phi(\frac{1-2}{2})\Phi(\frac{3-2}{2})$$
$$= \Phi(0)\Phi(\frac{1}{2}) - \Phi(-\frac{1}{2})\Phi(\frac{1}{2})$$

Don't forget to use normcdf in Matlab to calculate these out!

$$\boxed{P(1 < X(3) < 2, X(7) < 3) = (0.5)(0.6915) - (0.3085)(0.6915) = 0.132}$$

Also fun fact: $P(1 < X(3) < 2, X(7) < 3) = P(1 < X(3) < 2)P(X(7) < 3)$. You can verify this in matlab with:

*(normcdf((2-2)/2)-normcdf((1-2)/2)) * (normcdf((3-2)/2))*

Part (B)

Find the joint pdf of X(3) and X(7)

The pdf of a multivariate Gaussian distrubtion can be found as follows:

$$C_X(3,7) = R_X(3-7) - E[X(3)]E[X(7)]$$
$$C_X(3,7) = 2 + 4sinc(-4) - (2)(2) = 4sinc(-4) - 2 = -2$$

Now we have enough info to write out the multivariate Gaussian distribution:

$$f_{X_1,X_2,...X_n}(x_1,x_2,...x_n) = \frac{1}{(2\pi)^{n/2}|C_X|^{1/2}} exp\Big\{-\frac{1}{2}(x-\mu_X)^T C_X^{-1}(x-\mu_X)\Big\}$$

$$f_{X(3),X(7)}(x_1,x_2) = \frac{1}{(2\pi)^{2/2}|-2|^{1/2}} exp\Big\{-\frac{1}{2}(x_1-2)(-2)^{-1}(x_2-2)\Big\}$$

$$\boxed{f_{X(3),X(7)}(x_1,x_2) = \frac{\sqrt{2}}{4\pi} exp\Big\{\frac{x_1x_2}{4} - \frac{x_1}{2} - \frac{x_2}{2} + 1\Big\}}$$

Problem 4

On Blackboard, there is a file called AnnualRainfall.txt that roughly contains the annual summer rainfall in inches from 1895-2002 from a New England state. Let X[n]be the annual summer rainfall total for yearn. We want to determine if the annual summer rainfall totals are increasing. Assume $\mu_X[n] = an + b$. We want to determine if $a > 0$, which would indicate that the mean summer rainfall is increasing.

Part (A)

Estimate a and b by minimizing the least square error

$$J(a,b) = \sum_{n=0}^{N-1} (x[n] - (an+b))^2$$

where N= 108 for our dataset. This can be done by differentiating J with respect to a and with respect to b(and setting each equation to 0) to get the matrix vector equation

$$Ax = y$$

where A is a 2x2 matrix, y is a 2x1 vector and

$$x = \begin{bmatrix} b \\ a \end{bmatrix}$$

Solving this equation will yield the estimates $\hat{a}$ and $\hat{b}$. Give your answers for $\hat{a}$ and $\hat{b}$ and provide a generalized representation for A and y as a function of N.

Let's take the partial derivatives $\frac{dJ}{da}$ and $\frac{dJ}{db}$, and then solve for 0:

$$\frac{dJ}{da} = \sum_{n=0}^{N-1} 2n(an - x[n] + b),$$

$$0 = \sum_{n=0}^{N-1} 2n(an - x[n] + b)$$

$$0 = \sum_{n=0}^{N-1} (an^2 - nx[n] + bn)$$

$$0 = a\sum_{n=0}^{N-1} n^2 - \sum_{n=0}^{N-1} nx[n] + b\sum_{n=0}^{N-1} n$$

$$a\sum_{n=0}^{N-1} n^2 + b\sum_{n=0}^{N-1} n = \sum_{n=0}^{N-1} nx[n]$$

$$\frac{dJ}{db} = \sum_{n=0}^{N-1} 2(b - x[n] + an),$$

$$0 = \sum_{n=0}^{N-1} 2(b - x[n] + an)$$

$$0 = b\sum_{n=0}^{N-1} 1 - \sum_{n=0}^{N-1} x[n] + a\sum_{n=0}^{N-1} n$$

$$b\sum_{n=0}^{N-1} 1 + a\sum_{n=0}^{N-1} n = \sum_{n=0}^{N-1} x[n]$$

We can take these results and come up with a generalized $Ax = y$:

$$\begin{bmatrix} A_{1,1}(N) & A_{1,2}(N) \\ A_{2,1}(N) & A_{2,2}(N) \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

Where we have:

$$A_{1,1}(N) = \sum_{n=0}^{N-1} 1$$

$$A_{1,2}(N) = \sum_{n=0}^{N-1} n$$

$$A_{2,1}(N) = \sum_{n=0}^{N-1} n$$

$$A_{2,2}(N) = \sum_{n=0}^{N-1} n^2$$

$$y_1 = \sum_{n=0}^{N-1} x[n]$$

$$y_2 = \sum_{n=0}^{N-1} nx[n]$$

With N=108 and the results from the rainfall data, we get

$$\begin{bmatrix} 108 & 5778 \\ 5778 & 414090 \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 1.0464e3 \\ 5.7773e4 \end{bmatrix}$$

$$\begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} \frac{215}{5886} & \frac{-1}{1962} \\ \frac{-1}{1962} & \frac{1}{104967} \end{bmatrix} \begin{bmatrix} 1.0464e3 \\ 5.7773e4 \end{bmatrix}$$

$$\boxed{\begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 8.776 \\ 17.058e-3 \end{bmatrix}}$$

(b) Plot your estimate for $\mu_X[n]$, given by $\hat{\mu}_X[n]$, along with the rainfall data on the same graph. Note that n= 0 represents the year 1895 and n= 107 represents the year 2002.

Below is a plot of the results in Figure . $\hat{\mu}_X[n]$ should show the general trend of the data

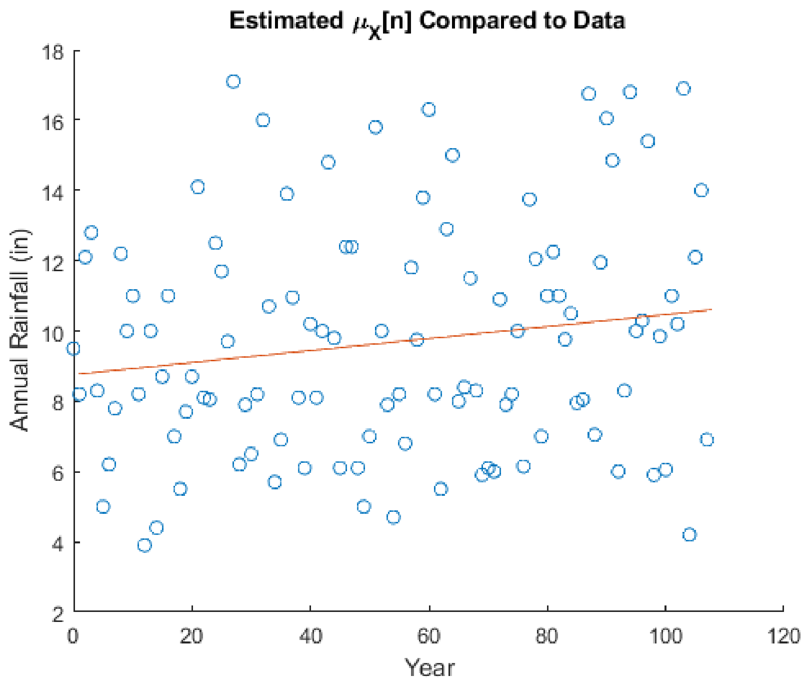

Figure 1: This is a plot of our $\hat{\mu}_X[n]$ that shows a general increasing trend compared to the data.

Part (C)

Plot the least square error sequence, also known as the fitting error, which is given by

$$e[n] = x[n] - (an + b)$$

Also find the estimated sample variance $\sigma^2 = \frac{1}{N}\sum_{n=0}^{N-1} e^2[n]$

Below is a plot of the fitting error in Figure , with the sample variance included.

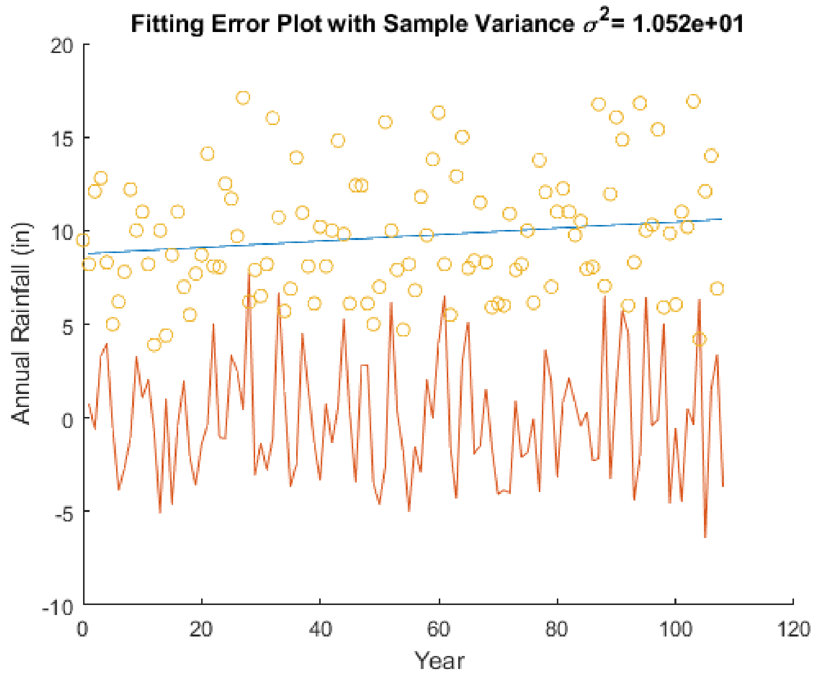

Figure 2: This plot shows us the error in our calculated fitted line, compared to the data. The further from zero the line goes, the more error. Variance is in the title of the plot.

Problem 5

One can show that the variance of $\hat{a}$ from the previous problem, is given by

$$Var(\hat{a}) = \frac{\hat{\sigma}^2}{\sum_{n=0}^{N-1} n^2 - \frac{1}{N}(\sum_{n=0}^{N-1} n)^2}$$

If we suspect that the mean summer rainfall is a constant, then we could model $\hat{a}$ as a Gaussian with zero mean so that $\hat{a} \sim N(0, Var(\hat{a}))$. Determine the required value of N so that the probability that $\hat{a} > 0.0171$ is less than 10^{-6}.Use the value of $\hat{\sigma}^2$ that you found in the previous problem. You may find the following summation formulas useful

$$\sum_{n=1}^{N} n = \frac{N(N+1)}{2}; \sum_{n=1}^{N} n^2 = \frac{N(N+1)(2N+1)}{6}$$

Let's work backwards and symbolically find $\hat{a} > 0.0171$:

$$P(\hat{a} > 0.0171) = 1 - \hat{a} < 0.0171$$
$$P(\hat{a} > 0.0171) = 1 - \Phi(\frac{0.0171 - 0}{\sqrt{Var(\hat{a})}})$$

Convert to the target inequality:

$$10^{-6} > 1 - \Phi(\frac{0.0171 - 0}{\sqrt{Var(\hat{a})}})$$
$$\Phi(\frac{0.0171}{\sqrt{Var(\hat{a})}}) > 999.999 \times 10^{-3}$$

I did some trial and error to find out that $\Phi(4.7) \approx 999.999 \times 10^{-3}$. We can solve for that value and find the N required to hit that level of precision:

$$4.7 = \frac{0.0171}{\sqrt{Var(\hat{a})}}$$

$$Var(\hat{a}) = 6.459 \times 10^{-3}$$

$$6.459 \times 10^{-3} = \frac{\hat{\sigma}^2}{\sum_{n=0}^{N-1} n^2 - \frac{1}{N}(\sum_{n=0}^{N-1} n)^2}$$
$$\sum_{n=0}^{N-1} n^2 - \frac{1}{N}(\sum_{n=0}^{N-1} n)^2 = \frac{10.52}{6.459 \times 10^{-3}}$$
$$\frac{N(N+1)(2N+1)}{6} - \frac{1}{N}(\frac{N(N+1)}{2})^2 = 1.629 \times 10^3$$
$$\frac{N(N+1)(2N+1)}{6} - \frac{N(N+1)^2}{4} = 1.629 \times 10^3$$

$$\frac{1}{6}N(N+1)(2N+1) - \frac{1}{4}N(N+1)^2 = 1.629 \times 10^3$$

$$\frac{1}{3}N^3 + \frac{1}{2}N^2 + \frac{1}{6}N - \frac{1}{4}N^3 - \frac{1}{2}N^2 - \frac{1}{4}N = 1.629 \times 10^3$$

$$\frac{1}{12}N^3 - \frac{1}{12}N = 1.629 \times 10^3$$

Solving for N will get us $N = 26.9$, so we can round up to $\boxed{N = 30}$

Matlab Code

```
%% %%%%%%%%%%%%%%%
% Problem 4
%%%%%%%%%%%%%%%%%

%% Part A verification

% import your data
txt = readtable('AnnualRainFall.txt');
N = 108;
n = [1:108];
x = table2array(txt(:,2));

% calculate a and b
n = [0:N-1];
A_1_1 = 108;
A_1_2 = sum(n);
A_2_1 = sum(n);
A_2_2 = sum(n.^2);
y_1 = sum(x);
y_2 = sum(x'.*n);


% set up the matrix equation Ax = y
A = [A_1_1,A_1_2;A_2_1,A_2_2];
Y = [y_1;y_2];
X = A^(-1)*Y;

% estimates for a and b from part (a)
b = X(1);
a = X(2);
% What I calculated
% a = 17.0633e-3;
% b = 8.776;

%% Part B - Plotting mu_X[n] and comparing to the data

y = [0:107];
mu_X = a*n+b; % estimated average

figure(1)
scatter(y,x);
hold on;
plot(mu_X);
```

```
title("Estimated \mu_X[n] Compared to Data")
xlabel("Year")
ylabel("Annual Rainfall (in)")
%% Part C - least squares error sequence/fitting error
fittingError = x'-(a*n+b);
actual = mu_X + fittingError;
sampleVariance = sum(fittingError.^2)./N;

figure(2)

title(['Fitting Error Plot with Sample Variance \sigma^2=
    ',num2str(sampleVariance,'%.3e')])
xlabel("Year")
ylabel("Annual Rainfall (in)")
hold on
plot(mu_X);
plot(fittingError);
scatter(y,x);
```