INDIAN INSTITUTE OF TECHNOLOGY MADRAS

Department of Chemical Engineering

CH5350 : Applied Time Series Analysis

Solutions to Assignment #1

1

1.1

i) July $\mu = 31.5°$C and sd $\sigma = 4.2°$C

$$
\begin{aligned}
P(25 \le X \le 37) &= \int_{25}^{37} \frac{1}{4.2\sqrt{2\pi}} \exp\left(-0.5\left(\frac{x-31.5}{4.2}\right)^2\right) dx \\
&= 0.844
\end{aligned}
$$

ii) January $\mu = 22.4°$C and sd $\sigma = 3.2°$C

$$
\begin{aligned}
P(25 \le X \le 37) &= \int_{25}^{37} \frac{1}{3.2\sqrt{2\pi}} \exp\left(-0.5\left(\frac{x-22.4}{3.2}\right)^2\right) dx \\
&= 0.208
\end{aligned}
$$

During July

$$
\begin{aligned}
\Pr(T > 25) &= \int_{25}^{\infty} f(x)dx \\
&= 0.9391
\end{aligned}
$$

During January

$$
\begin{aligned}
\Pr(T > 25) &= \int_{25}^{\infty} f(x)dx \\
&= 0.2083
\end{aligned}
$$

The beach runner will run on the beach in January as the probability is at most 0.2 (rounding upto two decimals).

1.2

Gaussian distribution

i) Marks obtained by the students in a class.

ii) Weights of apples in a lot

Poisson distribution

i) Number of accidents occurring on road.

ii) Arrival of persons in queue.

Chi-square distribution

i) Sample variance of normal population.

ii) Power spectral densities of variables with Gaussian distribution.

1.3

The waiting time in a queue when modelled as $\frac{m}{m+1}$ will follow a mixed distribution

$$f(x) = \begin{cases} (1-\rho), & t = 0 \\ \frac{\lambda}{\mu}(\mu - \lambda)\exp(-(\mu - \lambda)x) & \text{elsewhere} \end{cases}$$

In this case $f(x)$ follows a discrete distribution at $t = 0$, where as it follows continuous distribution else where.

2

2.1 Given the density function

$$\text{We know that} \int_{-\infty}^{-\infty}\int_{-\infty}^{-\infty} \left(x^2+y^2\right) dxdy = 1$$

$$K\left(\int_{0}^{2}\int_{0}^{2} \left(x^2+y^2\right) dxdy\right) = 1$$

$$K\int_{0}^{2} \left(2x^2+\frac{8}{3}\right) dx = 1$$

$$K = \frac{3}{32}$$

2.2 Marginal densities

$$f_x(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_{0}^{2} \frac{3}{32}(x^2+y^2)dy = \frac{3}{32}\left(2x^2+\frac{8}{3}\right)$$

$$f_y(y) = \int_{-\infty}^{-\infty} f(x,y)dx = \int_{0}^{2} \frac{3}{32}(x^2+y^2)dx = \frac{3}{32}\left(2y^2+\frac{8}{3}\right)$$

2.3 Probability

$$P\left(0.4 \le X \le 0.8, 0.2 \le Y \le 0.4\right) = \int_{0.4}^{0.8}\int_{0.2}^{0.8} \frac{3}{32}\left(x^2+y^2\right) dxdy = 3.5 * 10^-3$$

2.4 Conditional densities

$$f_x\left(y|X=x\right) = \frac{f\left(x,y\right)}{f(x)} = \frac{\left(x^2+y^2\right)}{2x^2+\frac{8}{3}}$$

$$f_x\left(x|Y=y\right) = \frac{f\left(x,y\right)}{f(y)} = \frac{\left(x^2+y^2\right)}{2y^2+\frac{8}{3}}$$

3

3.1

Given joint cumulative distribution

$$F(x,y) = \frac{1}{6}xy(x+y)$$

The joint pdf of x and y is given by

$$f(x,y) = \frac{\partial^2}{\partial x \partial y}F(x,y)$$
$$f(x,y) = \frac{\partial^2}{\partial x \partial y}\frac{1}{6}(x^2y + xy^2)$$
$$f(x,y) = \frac{\partial}{\partial x}\frac{1}{6}(x^2 + 2xy)$$
$$f(x,y) = \frac{1}{6}(2x + 2y)$$
$$f(x,y) = \frac{1}{3}(x + y)$$

The marginal density in y is given by

$$f_X(x) = \frac{1}{3}\int_0^2 f(x,y)dx$$
$$f_X(x) = \int_0^2 \frac{1}{3}(x+y)dy$$
$$f_X(x) = \frac{2}{3}(y+1)$$

The cumulative distribution function in y is given by

$$F_Y(y) = \int_0^y f_Y(y)dx$$
$$F_Y(y) = \int_0^y \frac{2}{3}(y+1)dy$$
$$F_Y(y) = \frac{2}{3}(\frac{y^2}{2} + y)$$

The plot of joint cumulative distribution is shown in Figure 1. The plot of joint density function is shown in Figure 2. The plot of cumulative distribution in Y is shown in Figure

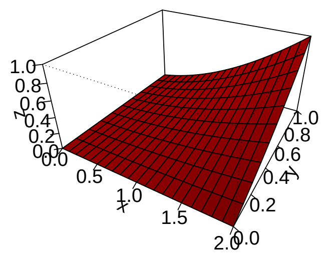

Figure 1: Joint cumulative distribution

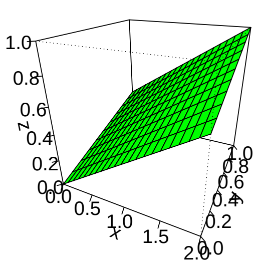

Figure 2: Joint density function

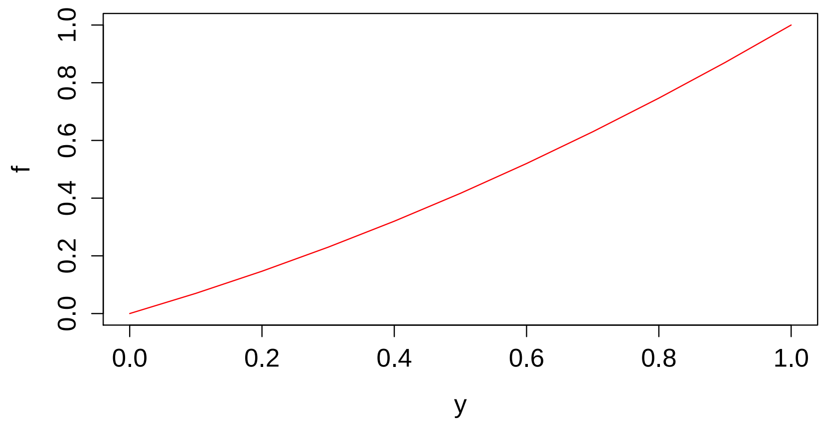

Figure 3: Cumulative distribution

3.

3.2

The joint Gaussian distribution of two random variables $\mathbf{X} = [x_1\ x_2]$ is

$$f(\mathbf{X}) = \frac{1}{(2\pi)^{2/2}|\Sigma_{\mathbf{X}}|^{-1/2}} \exp(-0.5 * (\mathbf{X} - \mu)^T)(\Sigma_{\mathbf{X}}^{-1})(\mathbf{X} - \mu)$$

where $\Sigma_X = \begin{bmatrix} \sigma_{x_1}^2 & \sigma_{x_1 x_2} \\ \sigma_{x_2 x_1} & \sigma_{x_2}^2 \end{bmatrix}$
Since, x_1 and x_2 are uncorrelated $\sigma_{x_1 x_2} = \sigma_{x_2 x_1} = 0$.

$$f(x_1, x_2) = \frac{1}{(2\pi)^{2/2}\sigma_{x_1}\sigma_{x_2}} \exp\left(-0.5\left(\frac{(x_1 - \mu_{x_1})^2}{\sigma_{x_1}^2} + \frac{(x_2 - \mu_{x_2})^2}{\sigma_{x_2}^2}\right)\right)$$

$$= \frac{1}{(2\pi)^{1/2}\sigma_{x_1}} \exp\left(-0.5\left(\frac{(x_1 - \mu_{x_1})^2}{\sigma_{x_1}^2}\right)\right) \frac{1}{(2\pi)^{1/2}\sigma_{x_2}} \exp\left(-0.5\left(\frac{(x_2 - \mu_{x_2})^2}{\sigma_{x_2}^2}\right)\right)$$

$$\implies f(x_1, x_2) = f(x_1) \times f(x_2)$$

Hence, x_1 and x_2 are independent. Hence, x_1 and x_2 are independent.

4 Expectations

4.1 Best predictions

Given

$$f(x,y) = \begin{cases} \frac{3}{32}(x^2+y^2) & 0<x<2, 0<y<2 \\ 0 & \text{elsewhere} \end{cases}$$

The best prediction of Y without knowledge of X is expectation value of Y *i.e*

$$\begin{aligned} \hat{Y} &= E(Y) \\ &= \int_{-\infty}^{\infty} y f(y) dy \\ &= \int_0^2 \frac{3y}{32}\left(\frac{8}{3}+2y^2\right) \\ &= \frac{5}{4} \end{aligned}$$

The best prediction of y with knowledge of $x = 0.8$ is conditional expectation value of y given x *i.e.*

$$\begin{aligned} \hat{Y} &= E(Y|X=x) \\ &= \int_{-\infty}^{\infty} y f(Y|X=x) \\ &= \int_0^2 \frac{0.64y+y^3}{2(0.64)+(8/3)} \\ &= 1.34 \end{aligned}$$

4.2 Compute $E(X_1^3(X_2^2-4X_3))$

Given X_1, X_2 and X_3 are independent random variables with zero mean and unit variance

$$E(X_1^3(X_2^2-4X_3)) = E(X_1^3)E(X_2^2) - 4E(X_1^3)E(X_3)$$

Since, $E(X_1) = E(X_2) = E(X_3) = 0$

$$E(X_1^3(X_2^2-4X_3)) = E(X_1^3)E(X_2^2)$$

Since, $E(X_1^2) = E(X_2^2) = E(X_3^2) = 1$

$$E(X_1^3(X_2^2 - 4X_3)) = E(X_1^3)$$

which is third moment of X_1 called as skewness.
If x_1 follows Gaussian distribution the value of $E(x_1^3) = 0$

5 Correlations in R

5.1 Estimate covariance matrix in R

The theoretical covariance is computed as

$$\begin{aligned}
\sigma_{xy} &= E((x - \mu_x)(y - \mu_y)) \\
&= E((x - 1)(x^2 + 4x + 2 - 10) \\
&= E(x^3) + 3E(x^2) - 12E(x) + 8
\end{aligned}$$

Since X follow Gaussian distribution, the third central moment will be zero $i.e$

$$\begin{aligned}
E((x_1 - \mu_x)^3) &= 0 \\
E(x_1^3) &= 10
\end{aligned}$$

σ_{xy} is computed to be

$$\sigma_{xy} = 18$$

Similarly σ_{xx} and σ_{yy} is computed to be

$$\sigma_{xx} = 3 \qquad \text{and} \qquad \sigma_{yy} = 140$$

The theoretical covariance matrix is computed as

$$\sum_{xx} = \begin{bmatrix} 3 & 18 \\ 18 & 140 \end{bmatrix}$$

The function to estimate covariance matrix in R is shown below

```
covar <- function(x,y) # defining function name
{
mat=matrix(rep(0,4),nrow=2,ncol=2); # define matrix with all zeros
N=length(x); # calculate size of vector x
p1=0;p2=0;p3=0;
for (i in 1:N){  # initialise for loop
p1=p1+(((y[i]-mean(y))*(x[i]-mean(x)))/N);
p2=p2+(((x[i]-mean(x))*(x[i]-mean(x)))/N);
p3=p3+(((y[i]-mean(y))*(y[i]-mean(y)))/N);
}
mat[1,1]=p2;
mat[1,2]=p1;
mat[2,1]=p1;
mat[2,2]=p3;
return(mat) #returns the value of mat
}
```

The covariance matrix calculated using *cov* command in R is

$$\hat{\sum}_{xx} = \begin{bmatrix} 3.144 & 18.907 \\ 18.907 & 133.45 \end{bmatrix}$$

The covarince matrix computed using the user defined function is

$$\hat{\sum}_{xx} = \begin{bmatrix} 3.141 & 18.904 \\ 18.904 & 133.42 \end{bmatrix}$$

The estimated matrix using inbuilt function as well as user defined function in R are matched with theoretical ones calculated.

5.2 Partial correlation

Given

$$X = 2Z + 3V$$
$$Y = Z + W$$

In class, we derived the partial correlation $\rho_{xv|z} = 0$. The semi partial correlation $i.e$ correlation between x and $y|z$ is derived as follows Assume

$$Y^* = Y - \hat{Y}|Z$$

The best prediction of Y given Z is

$$\begin{aligned}\hat{Y}|Z &= E(Y|Z) \\ &= Z\end{aligned}$$

Then

$$Y^* = W$$

The semi partial correlation between X and Y^* is obtained as zero because W is uncorrelated with Z and V.

In this case both partial correlation and semi partial correlation is zero since, only one variable Z is affecting X and Y. If both X and Y are causing by a series of confounding variables then the values will be different. R code for partial and semi partial correlations is given below

```r
library(ppcor)
v = rnorm(1000) # defining random variables
w = rnorm(1000)
z = rnorm(1000)
x = 2*z+3*v
y = z+w
pcorrelation = pcor.test(x,y,z) # Computing partial correlation
spcorrelation = spcor.test(x,y,z) # Computing semi partial correlation
```

The autocorrelation obtained is 0.0461 and semi partial correlation is 0.0397. The estimated values of PCF, SPCF are in closer agreement with the theoretical ones.