Indian Institute of Technology, Madras

Time series analysis

CH5350

Tutorial 1

Submitted by:
Athul Vijayan
ED11B004

August 19, 2014

1 Solutions

1.1 Problem 1

(a) The probabilty distribution functions for month July and January is defined as a normal distribution with given μ and σ. The PDF of gaussian distribution is defined as:

$$f(T) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(T-\mu)^2}{2\sigma^2}}$$

(i) For month July:

$P_r(25 \leq T \leq 37) = \int\limits_{25}^{37} \frac{1}{4.2\sqrt{2\pi}}e^{-\frac{(T-31.5)^2}{2*4.2^2}}\,\mathrm{d}T$

Which gives

$(P_r(25 \leq T \leq 37))_{July} = 0.843964$

For month January:

$P_r(25 \leq T \leq 37) = \int\limits_{25}^{37} \frac{1}{3.2\sqrt{2\pi}}e^{-\frac{(T-22.4)^2}{2*3.2^2}}\,\mathrm{d}T$

Which gives

$(P_r(25 \leq T \leq 37))_{January} = 0.20825$

(ii) $P_r(T > 25) = \int\limits_{25}^{\infty} \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(T-\mu)^2}{2\sigma^2}}\,\mathrm{d}T$

For month July:

$P_r(T > 25) = \int\limits_{25}^{\infty} \frac{1}{4.2\sqrt{2\pi}}e^{-\frac{(T-31.5)^2}{2*4.2^2}}\,\mathrm{d}T$

Which gives

$(P_r(T > 25))_{July} = 0.939143$

For month January:

$P_r(T > 25) = \int\limits_{25}^{\infty} \frac{1}{3.2\sqrt{2\pi}}e^{-\frac{(T-22.4)^2}{2*3.2^2}}\,\mathrm{d}T$

Which gives

$(P_r(T > 25))_{January} = 0.208252$

(b) Real life quantities that best fit into each of given distribution are

(i) Normal Distribution

- IQ scores
- Heights of males/females
- Body temperatures

(ii) Poisson Distribution

- number of typing errors on a page
- rare diseases (like Leukemia, but not AIDS because it is infectious and so not independent) - especially in legal cases
- The number of deaths per year in a given age group.

(iii) Exponential Distribution

- The time until a radioactive particle decays, or the time between clicks of a geiger counter
- The time it takes before your next telephone call
- The time until default (on payment to company debt holders) in reduced form credit risk modeling

(c)

1.2 Problem 2

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

(i) By axiom of probability $P(\Omega) = 1$ where Ω is the sample space.

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \mathrm{d}x\mathrm{d}y = 1 \tag{1}$$

$$\int_0^2 \int_0^2 K(x^2 + y^2) \mathrm{d}x\mathrm{d}y = 1$$

$$\int_0^2 K[(\frac{x^3}{3} + y^2 x)]_0^2 \mathrm{d}y = 1$$

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

(ii) Marginal density is defined as :

$$P_X(x) = \int_y f(x, y) \mathrm{d}y$$

Which gives us:

$$P_X(x) = \int_0^2 K(x^2 + y^2) \mathrm{d}y \tag{2}$$

$$\implies P_X(x) = K(2x^2 + \frac{8}{3})$$

$$P_Y(y) = \int_0^2 K(x^2 + y^2) \mathrm{d}x$$

$$\implies P_Y(y) = K(2y^2 + \frac{8}{3})$$

(iii)

$$P_r(0.4 < X < 0.8, 0.2 < Y < 0.4) = \int_{0.4}^{0.8} \int_{0.2}^{0.4} f(x, y) \mathrm{d}y\mathrm{d}x \tag{3}$$

$$\implies P_r(0.4 < X < 0.8, 0.2 < Y < 0.4) = 0.037333K$$

(iv) Conditional probability density is defined as:

$$f_Y(y|X=x) = \frac{f_{XY}(x,y)}{f_X(x)} \tag{4}$$

$$\rightarrow f_Y(y|X=x) = \frac{(x^2+y^2)}{(2x^2+\frac{8}{3})}$$

$$and \tag{5}$$

$$\rightarrow f_X(x|Y=y) = \frac{(x^2+y^2)}{(2y^2+\frac{8}{3})}$$

1.3 Problem 3

(a) $F(x,y) = \frac{1}{6}xy(x+y) \quad ; \quad 0 \le x \le 2, 0 \le y \le 1$

(i)

$$F_Y(y) = F_{XY}(x=2,y)$$
$$F_Y(y) = \frac{1}{3}(2+y)$$

(ii) Joint density function can be obtained from joint cumulative distribution by

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

(b) If two random variables with joint Normal distribution and they are uncorrelated, they are statistically independent.
Proof:
Let $\mathbf{x} = [X, Y]$ is a random vector with random variables X and Y.
$\mu = [\mu_x, \mu_y]$ be the expectations of X and Y.
and $\mathbf{\Sigma} = [Cov(X_i, X_j)]$ be the covariance matrix.
Normal distribution is given by

$$f_{\mathbf{x}}(x,y) = \frac{1}{\sqrt{(2\pi)^2|\Sigma|}} exp(-\frac{1}{2}(\mathbf{x}-\mu)^T \mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)) \tag{6}$$

$$\tag{7}$$

With two variables, we have:
$\mu = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}$
$\mathbf{\Sigma} = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}$
Expanding the matrix gives the joint normal distribution as :

$$f_{\mathbf{x}}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right) \tag{8}$$

Since X and Y are uncorrelated, $\rho = 0$

$$f_{\mathbf{x}}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} exp\left(-\frac{1}{2}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right]\right)$$

$$f_{\mathbf{x}}(x,y) = \left(\frac{1}{\sqrt{2\pi}\sigma_x} exp\left(-\frac{1}{2}\left[\frac{(x-\mu_x)^2}{\sigma_x^2}\right]\right)\right)\left(\frac{1}{\sqrt{2\pi}\sigma_y} exp\left(-\frac{1}{2}\left[\frac{(y-\mu_y)^2}{\sigma_y^2}\right]\right)\right)$$

$$\implies f_{\mathbf{x}}(x,y) = f_X(x)f_Y(y)$$

That proves that uncorrelated variables with joint normal distribution are independent.

1.4 Problem 4

(a) best prediction of Y is the expectation of Y; $E(Y|X)$

When the x is unknown:

$$E(Y) = \int_y y f_Y(y)\mathrm{d}y$$

$$E(Y) = \int_y y(\frac{1}{4}(\frac{3}{4}y^2 + 1))\mathrm{d}y$$

$$E(Y) = 1.25 \tag{9}$$

And at $x = 0.8$, the conditional probability is:

$$E(Y|X) = \int_y y f_Y(y|X=x)\mathrm{d}y$$

$$E(Y|X) = \int_y y(\frac{(x^2+y^2)}{2x^2+\frac{8}{3}})\mathrm{d}y$$

$$E(Y|X) = \left[\frac{(x^2y^2/2 + y^4/4)}{2x^2+\frac{8}{3}}\right]_0^2$$

$$E(Y|X) = \left[\frac{(x^2+2)}{x^2+\frac{4}{3}}\right]$$

and at $x = 0.8$

$E(Y|X = 0.8) = 0.3243243$

(b) For two independent random variables,

$$E(XY) = E(X)E(Y)$$

Expanding the given expression

$$\begin{aligned} E(X_1^3(X_2^2+3X_3)) &= E(X_1^3X_2^2 + 3X_1^3X_3) \\ &= E(X_1^3X_2^2) + 3E(X_1^3X_3) \\ &= E(X_1^3)E(X_2^2) + 3E(X_1^3)E(X_3) \end{aligned}$$

Since unit variance and zero expectation

$$E(X_1^3(X_2^2+3X_3)) = E(X_1^3)$$

1.5 Problem 5

(a) The following function returns the covariance matrix of a random vector:

```
# A general function to produce covariance matrix of a n dimensional
    random vector
covariance <- function(X) {
   muVec <- colMeans(X); # find mean of each random vector
   N <- 1:nrow(x);
   CovMatrix <- matrix(0, ncol(X), ncol(X)); #initialize to zero
   for (i in 1:ncol(X)) { # loop through pair of every random vectors
      for (j in 1:ncol(X)) {
         for (k in N) {
            CovMatrix[i, j] = CovMatrix[i, j] + (X[k, i] -
                muVec[i])*(X[k, j]-muVec[j]); #keep adding
         }
         CovMatrix[i, j] = CovMatrix[i,j]/length(N); #divide by size
      }
   }
   return(CovMatrix); #return the result
}
```

And the given problem can be solved using the above function:

```
# create dataset
# sqrt since rnorm has argument standard deviation and given is
    variance

X <- rnorm(1000, 1, sqrt(3));
Y <- X^2 + 4*X +2;
x <- cbind(X, Y);

# covariance from my function
myCov = covariance(x);
# covariance from R
rCov = cov(x);

View(myCov);
View(rCov);
```

The covariance matrix from the above function is:
$$\begin{pmatrix} 2.921442 & 17.86062 \\ 17.860625 & 125.20517 \end{pmatrix}$$
And the covariance using R is:
$$\begin{pmatrix} 2.924366 & 17.8785 \\ 17.878503 & 125.3305 \end{pmatrix}$$

(b)