INDIAN INSTITUTE OF TECHNOLOGY MADRAS
Department of Chemical Engineering
CH5350 Applied Time-Series Analysis

Assignment 1

Due: Tuesday, August 19, 2014 4:00 PM

---

1. [Distributions]

   (a) The morning temperature (between 6:00 AM - 7:00 AM) of Chennai is known to be a random variable with Gaussian distribution. The statistical properties of the temperature during this period vary with the month. During July, the average temperature is $\mu = 31.5$ °C with $\sigma = 4.2$ °C, while in January, the average temperature is $\mu = 22.4$ °C with $\sigma = 3.2$ °C. Then, (i) determine Pr$(25°C \leq T \leq 37°C)$ in each of these months (ii) BeachRunner decides run on the beach in the morning if Pr$(T > 25)$ is at most 0.2 in the morning hours. Will he run on the beach in both months?

   (b) Probability distributions are often used in to describe several real-life phenomena or quantities that arise in estimation. Describe two real situations that can be described by (i) Gaussian (ii) Poisson and (iii) exponential distributions. You may refer to appropriate material on the web for this purpose.

   (c) In the class we learnt that probability **distribution** functions could be either continuous or step-like (discrete RV). Can you think of a phenomenon that is described by a mixed **distribution** function, *i.e.*, partly continuous and partly discrete?

2. [Joint density]

   If two random variables have joint density

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

   Find (i) the value of K (ii) marginal densities of X and Y, (iii) the probability Pr$(0.4 < X < 0.8,\ 0.2 < Y < 0.4)$ (iv) conditional densities $f_Y(y|X=x)$ and $f_X(x|Y=y)$.

3. [Joint distribution]

   (a) The joint *cumulative* distribution function of two continuous random variables is given by

$$F(x,y) = \frac{1}{6}xy(x+y), \qquad 0 \leq x \leq 2,\ 0 \leq y \leq 1$$

   Determine (i) the c.d.f. of y and (ii) the joint p.d.f. of x and y. Plot the three functions using `R`.

   (b) Show that if two random variables with joint Gaussian distribution are *uncorrelated*, they are also independent.

4. [Expectations]

(a) With reference to Q.2, determine the best prediction of Y in two different situations, one when X is unknown and the other when X is within very close vicinity of 0.8.

(b) Given three zero-mean, unit variance, *independent* random variables X_1, X_2 and X_3, compute $E(X_1^3(X_2^2 + 3X_3))$

5. [Correlations in R]

(a) The covariance between two RVs is *estimated* from their samples $x[n]$ and $y[n]$ as

$$\hat{\sigma}_{yx} = \frac{1}{N}\sum_{n=1}^{N}(y[n] - \bar{y})(x[n] - \bar{x})$$

where $\bar{x}$ and $\bar{y}$ are the sample means of X and Y respectively and N is the sample size. Write a **function** in `R` to calculate this **sample covariance matrix** for two random variables, $X \sim \mathcal{N}(1,3)$ and $Y = X^2 + 4X + 2$ (use the `rnorm` routine to generate $N = 1000$ samples of X). Compare the resulting covariance matrix with the theoretical one and the values obtained from `cov` or the `var` command in `R`.

(b) For the partial correlation example in class, $X = 2Z + 3V, Y = Z + W$, compute the **semi-partial correlation** between X and Y given Z, *i.e.*, correlation between X and conditioned Y $(Y|Z)$. Can you explain the resulting answer? Is this the same as partial correlation between X and Y? Also verify your answer by way of estimation in `R`.