Probability of Random Vectors

- **Multiple Random Variables**

  Each outcome of a random experiment may need to be described by a set of $N > 1$ random variables $\{x_1, \cdots, x_N\}$, or in vector form:

  $$X = [x_1, \ \cdots \ , x_N]^T$$

  which is called a *random vector*. In signal processing X is often used to represent a set of N samples of a random signal $x(t)$ (a random process).

- **Joint Distribution Function and Density Function**

  The *joint distribution function* of a random vector X is defined as

  $$\begin{aligned} F_X(u_1, \cdots, u_N) &= P(x_1 < u_1, \cdots, x_N < u_N) \\ &= \int_{-\infty}^{u_1} \cdots \int_{-\infty}^{u_N} p(\xi_1, \cdots, \xi_N) d\xi_1 \cdots d\xi_N \end{aligned}$$

  where $p(\xi_1, \cdots, \xi_N)$ is the *joint density function* of the random vector X.

- **Independent Variables**

  For convenience, let us first consider two of the N variables and rename them as x and y. These two variables are *independent* iff

  $$P(A \cap B) = P(x < u, y < v) = P(x < u)P(y < v) = P(A)P(B)$$

  where events A and B are defined as "$x < u$" and "$y < v$", respectively. This definition is equivalent to

  $$p(x, y) = p(x)p(y)$$

  as this will lead to

  $$\begin{aligned} P(x < u, y < v) &= \int_{-\infty}^{u} \int_{-\infty}^{v} p(\xi, \eta) d\xi d\eta = \int_{-\infty}^{u} \int_{-\infty}^{v} p(\xi)p(\eta) d\xi d\eta \\ &= \int_{-\infty}^{u} p(\xi) d\xi \int_{-\infty}^{v} p(\eta) d\eta = P(x < u)P(y < v) \end{aligned}$$

  Similarly, a set of N variables are independent iff

  $$p(x_1, \cdots, x_N) = p(x_1)\, p(x_2) \cdots p(x_N)$$

- **Mean Vector**

  The *expectation* or *mean* of random variable x_i is defined as

$$\mu_i = E(x_i) \triangleq \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \xi_i \, p(\xi_1, \cdots, \xi_N) \, d\xi_1 \cdots d\xi_N$$

  The *mean vector* of random vector X is defined as

$$M = E(X) \triangleq [E(x_1), \cdots, E(x_N)]^T = [\mu_1, \cdots, \mu_N]^T$$

  which can be interpreted as the center of gravity of an N-dimensional object with $p(x_1, \cdots, x_N)$ being the density function.

- **Covariance Matrix**

  The *variance* of random variable x_i is defined as

$$\begin{aligned} \sigma_i^2 &\triangleq E[(x_i - \mu_i)^2] = E(x_i^2) - \mu_i^2 \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (\xi_i - \mu_i)^2 \, p(\xi_1, \cdots, \xi_N) \, d\xi_1 \cdots d\xi_N \end{aligned}$$

  The *covariance* of x_i and x_j is defined as

$$\begin{aligned} \sigma_{ij}^2 = Cov(x_i, x_j) &\triangleq E[(x_i - \mu_i)(x_j - \mu_j)] = E(x_i x_j) - \mu_i \mu_j \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \xi_i \xi_j \, p(\xi_1, \cdots, \xi_N) \, d\xi_1 \cdots d\xi_N - \mu_i \mu_j \end{aligned}$$

  The *covariance matrix* of a random vector X is defined as

$$\begin{aligned} \Sigma &= E[(X - M)(X - M)^T] = E(XX^T) - MM^T \\ &= \begin{bmatrix} .. & .. & .. \\ .. & \sigma_{ij}^2 & .. \\ .. & .. & .. \end{bmatrix}_{N \times N} \end{aligned}$$

  where

$$\sigma_{ij}^2 = E(x_i x_j) - \mu_i \mu_j$$

  is the covariance of x_i and x_j. When $i = j$, $\sigma_i^2 = E(x_i^2) - \mu_i^2$ is the variance of x_i, which can be interpreted as the amount of information, or energy, contained in the ith component of the signal X. And the total information or energy contained in X is represented by

$$tr\ \Sigma = \sum_{i=1}^{N} \sigma_i^2$$

Σ is symmetric as $\sigma_{ij}^2 = \sigma_{ji}^2$. Moreover, it can be shown that Σ is also *positive definite*, i.e., all its eigenvalues $\{\lambda_1, \cdots, \lambda_N\}$ are greater than zero and we have

$$tr\ \Sigma = \sum_{i=1}^{N} \lambda_i > 0$$

and

$$det\ \Sigma = \prod_{i=1}^{N} \lambda_i > 0$$

Two variables x_i and x_j are *uncorrelated* iff $\sigma_{ij}^2 = 0$, i.e.,

$$E(x_i x_j) = E(x_i)E(x_j) = \mu_i \mu_j$$

If this is true for all $i \neq j$, then X is called *uncorrelated* or *decorrelated* and its covariance matrix Σ becomes a diagonal matrix with only non-zero σ_i^2 $(i = 1, \cdots, N)$ on its diagonal.

If x_i $(i = 1, \cdots, N)$ are independent, $p(x_1, \cdots, x_N) = p(x_1) \cdots p(x_N)$, then it is easy to show that they are also uncorrelated. However, uncorrelated variables are not necessarily independent. (But uncorrelated variables with normal distribution are also independent.)

- **Autocorrelation Matrix**

The *autocorrelation* matrix of X is defined as

$$R \triangleq E(XX^T) = \begin{bmatrix} .. & .. & .. \\ .. & r_{ij} & .. \\ .. & .. & .. \end{bmatrix}_{N \times N}$$

where

$$r_{ij} \triangleq E(x_i x_j) = \sigma_{ij}^2 + \mu_i \mu_j$$

Obviously R is symmetric and we have

$$\Sigma = R - MM^T$$

When $M = 0$, we have $\Sigma = R$.

Two variable x_i and x_j are *orthogonal* iff $r_{ij} = 0$. Zero mean random variables which are uncorrelated are also orthogonal.

- **Mean and Covariance under Unitary Transforms**

  A unitary (orthogonal) transform of X is defined as

$$\left\{ \begin{array}{c} Y = A^T X \\ X = AY \end{array} \right.$$

  where A is a unitary (orthogonal) matrix

$$A^{*T} = A^{-1}$$

  and Y is another random vector.

  The mean vector M_Y and the covariance matrix Σ_Y of Y are related to the M_X and Σ_X of X as shown below:

$$M_Y = E(Y) = E(A^T X) = A^T E(X) = A^T M_X$$

$$\begin{aligned} \Sigma_Y &= E(YY^T) - M_Y M_Y^T = E(A^T X X^T A) - A^T M_X M_X^T A \\ &= A^T E(XX^T) A - A^T M_X M_X^T A = A^T [E(XX^T) - M_X M_X^T] A \\ &= A^T \Sigma_X A \end{aligned}$$

  Unitary transform does not change the trace of Σ:

$$\begin{aligned} tr\ \Sigma_Y &= tr\ [E(YY^T) - M_Y M_Y^T] = E[tr\ (YY^T)] - tr\ (M_Y M_Y^T) \\ &= E(Y^T Y) - M_Y^T M_Y = E(X^T A A^T X) - M_X^T A A^T M_X \\ &= E(X^T X) - M_X^T M_X = tr\ \Sigma_X \end{aligned}$$

  which means the total amount of energy or information contained in X is not changed after a unitary transform $Y = A^T X$ (although its distribution among the N components is changed).

- **Normal Distribution**

  The density function of a normally distributed random vector X is:

$$p(x_1, \cdots, x_N) = N(X, M, \Sigma) = \frac{1}{(2\pi)^{N/2} |\Sigma|^{1/2}} exp[-\frac{1}{2}(X-M)^T\Sigma^{-1}(X-M)]$$

  where M and Σ are the mean vector and covariance matrix of X, respectively. When $N = 1$, Σ and M become σ and μ, respectively, and the density function becomes single variable normal distribution.

  To find the shape of a normal distribution, consider the iso-value hyper surface in the N-dimensional space determined by equation

$$N(X, M, \Sigma) = c_0$$

  where c_0 is a constant. This equation can be written as

$$(X - M)^T\Sigma^{-1}(X - M) = c_1$$

  where c_1 is another constant related to c_0, M and Σ. For $N = 2$ variables x and y, we have

$$\begin{aligned} (X - M)^T\Sigma^{-1}(X - M) &= [x - \mu_x, y - \mu_y] \begin{bmatrix} a & b/2 \\ b/2 & c \end{bmatrix} \begin{bmatrix} x - \mu_x \\ y - \mu_y \end{bmatrix} \\ &= a(x - \mu_x)^2 + b(x - \mu_x)(y - \mu_y) + c(y - \mu_y)^2 \\ &= c_1 \end{aligned}$$

  Here we have assumed

$$\begin{bmatrix} a & b/2 \\ b/2 & c \end{bmatrix} = \Sigma^{-1}$$

  The above quadratic equation represents an ellipse (instead of any other quadratic curve) centered at $M = [\mu_1, \mu_2]^T$, because Σ^{-1}, as well as Σ, is positive definite:

$$\left|\Sigma^{-1}\right| = ac - b^2/4 > 0$$

  When $N > 2$, the equation $N(X, M, \Sigma) = c_0$ represents a hyper ellipsoid in the N-dimensional space. The center and spatial distribution of this ellipsoid are determined by M and Σ, respectively.

In particular, when $X = [x_1, \cdots, x_N]^T$ is decorrelated, i.e., $\sigma_{ij} = 0 \;\; for\ all \;\; i \neq j$, Σ becomes a diagonal matrix

$$\Sigma = diag[\sigma_1^2, \cdots, \sigma_N^2] = \left[\begin{array}{cccc} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \sigma_N^2 \end{array} \right]$$

and equation $N(X, M, \Sigma) = c_0$ can be written as

$$(X - M)^T \Sigma^{-1} (X - M) = \sum_{i=1}^{N} \frac{(x_i - \mu_i)^2}{\sigma_i^2} = c_1$$

which represents a standard ellipsoid with all its axes parallel to those of the coordinate system.

- **Estimation of M and Σ**

  When $p(x_1, \cdots, x_N)$ is not known, M and Σ cannot be found by their definitions. However, they can be estimated if a large number of outcomes $(X_j, \;\; j = 1, \cdots, K)$ of the random experiment in question can be observed.

  The mean vector M can be estimated as

$$\hat{M} = \frac{1}{K} \sum_{j=1}^{K} X_j$$

  i.e., the ith element of M is estimated as

$$\hat{\mu}_i = \frac{1}{K} \sum_{j=1}^{K} x_i^{(j)}$$

  where $x_i^{(j)}$ is the ith element of X_j.

  The autocorrelation R can be estimated as

$$\hat{R} = \frac{1}{K} \sum_{j=1}^{K} X_j X_j^T$$

  And the covariance matrix Σ can be estimated as

$$\hat{\Sigma} = \frac{1}{K} \sum_{j=1}^{K} X_j X_j^T - \hat{M}\hat{M}^T = \hat{R} - \hat{M}\hat{M}^T$$