Joint Statistical Measures

Consider a joint experimental venture comprising of two random experiments X and Y, whose outcomes are defined on the sample space $\mathbf{S}$ via the Cartesian set product of the sets X and Y, i.e,

$$\mathbf{S} \;=\; \{(x, y) \ni x \in X, y \in Y\} = X \times Y.$$

The joint CDF between two random variables X and Y defined on the sample space $\mathbf{S}$ is given by:

$$F_{XY}(x, y) \;=\; \Pr\{X \leq x, Y \leq y\}.$$

The joint PDF between two random variables X and Y defined on the sample space $\mathbf{S}$ is defined via the second partial derivative:

$$f_{XY}(x, y) \;=\; \frac{\partial^2}{\partial x \partial y}\left(F_{XY}(x, y)\right).$$

If the two random variables X and Y are *statistically independent* then the joint PDF and CDF are separable:

$$F_{XY}(x, y) \;=\; F_X(x)F_Y(y) \;\;\&\;\; f_{XY}(x, y) \;=\; f_X(x)f_Y(y).$$

The *correlation* between two random variables X and Y defined on the sample space $\mathbf{S}$ is given by:

$$r_{xy} \;=\; E\{XY\} \;=\; \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{XY}(x, y) dx dy.$$

The *covariance* between two random variables X and Y is defined via:

$$\sigma_{xy} \;=\; E\{(X - \mu_x)(Y - \mu_y)\} \;=\; \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x - \mu_x)(y - \mu_y) f_{XY}(x, y) dx dy$$

Upon simplification the covariance of the variables X and Y can be related to the correlation via

$$\sigma_{xy} = E\{(X - \mu_x)(Y - \mu_y)\} = r_{xy} - \mu_x\mu_y.$$

The *normalized correlation* or statistical *correlation coefficient* between two random variables X and Y is defined as:

$$\rho_{xy} \;=\; \frac{|\sigma_{xy}|}{\sigma_x\sigma_y},$$

where σ_{xy} is the covariance between the random variables X and Y and σ_x is the standard deviation of the random variable X. The Cauchy-Schwartz (C-S) Inequality for the pair of random variables X and Y is given by:

$$E^2\{XY\} - E\{X^2\}E\{Y^2\} \leq 0.$$

In a similar fashion, the angle between two random variables X and Y is defined as :

$$\theta_{xy} \;=\; \cos^{-1}\left(\frac{|\sigma_{xy}|}{\sigma_x \sigma_y}\right).$$

Two random variables X and Y are said to be statistically uncorrelated if :

$$\rho_{xy} = 0 \quad \equiv \quad \sigma_{xy} = 0 \quad \equiv \quad E\{XY\} = E\{X\}E\{Y\} \quad \equiv \quad \theta_{xy} = 90^\circ$$

i.e., they are statistically perpendicular to each other. Using the C-S inequality it is easy to see that the correlation coefficient has a maximum value of 1, i.e.,

$$\rho_{xy} \;=\; \frac{|\sigma_{xy}|}{\sigma_x \sigma_y} \leq 1.$$

Random variables X and Y are said to be statistically collinear or statistically linearly dependent if :

$$\rho_{xy} = 1 \quad \equiv \quad \sigma_{xy} = \sigma_x \sigma_y \quad \equiv \quad \theta_{XY} = 0^\circ.$$

If two random variables are statistically independent then it can be shown that they are also uncorrelated, i.e.,

$$f_{XY}(x,y) \;=\; f_X(x) f_Y(y) \iff \sigma_{xy} = 0 \iff \rho_{xy} \;=\; 0.$$

In the absence of information on the underlying probability distribution the covariance between the random variables X and Y can be approximated using their corresponding n-point observations as:

$$\sigma_{xy} \;\approx\; \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \frac{1}{n}\sum_{i=1}^{n} X_i\right)\left(Y_i - \frac{1}{n}\sum_{i=1}^{n} Y_i\right).$$